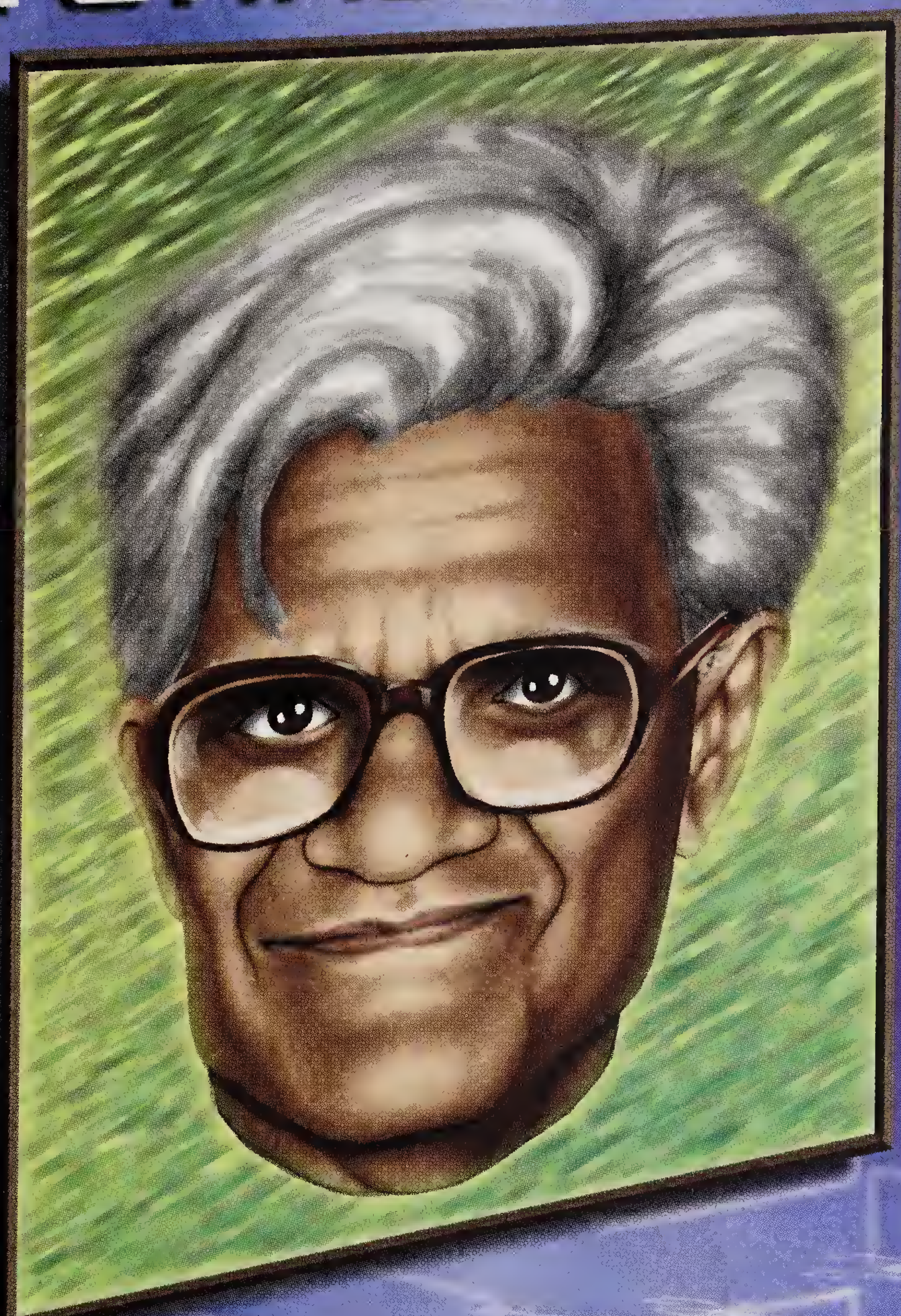
PRAMANA
Journal of Physics

RAYCHAUDHURI EQUATION



AT THE CROSSROADS

EDITORS
NARESH DADHICH
PANKAJ JOSHI
PROBIR ROY

Special issue on

Raychaudhuri equation at the crossroads

Special Issue on

Raychaudhuri equation at the crossroads

Guest Editors

Naresh Dadhich
Pankaj Joshi
Probir Roy

Published by
Indian Academy of Sciences
Bangalore

Reprinted from *Pramana – J. Phys.*
Vol. 69, No. 1, July 2007

ISSN No. 0304-4289

Edited by Naresh Dadhich, Pankaj Joshi and Probir Roy and printed for the Indian Academy of Sciences by Tholasi Prints, Bangalore

Contents

Preface

In 1953 something extraordinary happened at the Indian Association for the Cultivation of Science in Calcutta. Amal Kumar Raychaudhuri, twenty-seven years of age and employed ungainfully as a scientific assistant at the Experimental X-ray Section, made a startling theoretical discovery in General Relativity (GR). Without assuming any symmetry constraint on the underlying spacetime, he derived an equation which showed the unavoidable occurrence of spacetime singularities in GR under quite general conditions. Nearly a decade later, by global topological arguments utilizing the causal structure of spacetime and Einstein's equations, this result was given a complete mathematical generalization and proved rigorously in terms of a set of precisely enunciated theorems, now very well-known as singularity theorems, by Hawking, Penrose and Geroch. The equation of Raychaudhuri was the critical starting point for these theorems which held under more general conditions of which Raychaudhuri's conditions were a subset. Its import and significance were immediately recognized as was evident from the fact that Prof. Charles Misner could obtain a grant from NSF for an year's visit of AKR to the University of Maryland in 1964.

The Raychaudhuri equation has three aspects which are logically sequential. First and foremost, it is a geometric statement on the congruence of non-spacelike paths, including geodesics, in an arbitrary spacetime manifold. Second, on introducing the Principle of Equivalence, it becomes a statement on the congruence of the trajectories of material particles and photons in an arbitrary gravitational field. Finally, the use of Einstein's equations and of the energy conditions leads to the result that in a generally nonflat spacetime manifold there exist trajectories which are necessarily incomplete in the sense that they and their neighbouring trajectories inevitably focus into singularities at finite comoving times. The equation describes how trajectories behave during the course of their dynamical evolution, i.e. how they expand, reconverge, get distorted under shearing effects of gravitational fields and rotate under the influence of the energy density and matter fields present. The scope of the Raychaudhuri equation is very wide since it is a geometric statement on the evolution of paths in a general (not necessarily spacetime) manifold. For gravitational dynamics, it encompasses all spacetime singularities from the cosmological big bang to black holes and naked singularities that could arise in astrophysics from collapsing stars.

More than fifty years have passed since this powerful equation was written down. In this intervening half-century, it has influenced different types of research, not only in classical GR as well as in still incomplete theories of quantum gravity, but also in string theory and even in hydrodynamics. There is every likelihood that research involving the Raychaudhuri equation will take new directions in future. Just to illustrate this point, let us mention that in the currently fashionable Loop Quantum Cosmology, this equation is needed in a new avatar in the possible avoidance of the cosmological big bang singularity. Standing at the crossroads, we feel this to be an appropriate juncture to view this equation in perspective. To this end, we have invited essays from several experts, working in different areas, whose current research not only derives inspiration from this equation, but in fact makes use of it in some way or the other. We feel that, as a celebration of the golden jubilee of the birth of this amazing equation, this is the best tribute that we can offer to the memory of its deceased creator.

We thank all our authors for magnificently responding to our request for their contributions. We are grateful to the editorial board of *Pramana* for agreeing to publish this volume and we especially thank Prof. Rohini Godbole for facilitating that. The article by Professor Jürgen Ehlers is being produced from the *Proceedings of the International Conference on Einstein's Legacy in the New Millennium* published in *Int. J. Mod. Phys.* **D15** (2006). We thank both the author and the publisher, World Scientific, for their permission to reproduce it. Also reproduced is Prof. Raychaudhuri's reminiscence from the *Proceedings of the Symposium on the Fortieth Anniversary of the Raychaudhuri Equation* held at IUCAA in December 1995.

Naresh Dadhich
Inter-University Centre for Astronomy and Astrophysics, Pune 411 007

Pankaj Joshi
Tata Institute of Fundamental Research, Mumbai 400 005

Probir Roy
Tata Institute of Fundamental Research, Mumbai 400 005

Spring 2007

(*Guest Editors*)

PRAMANA — journal of physics

Vol. 69, No. 1
July 2007
pp. 3–5

A little reminiscence*

A K RAYCHAUDHURI

I feel rather exhilerated by the honour that the organizers of ICGC-III have bestowed on me by arranging a particular session in connection with my works. However, I must confess that I am embarassed to speak at this meeting. Firstly, I believe that a characteristic beauty of science is its objectivity but my presence as a speaker apparently brings in a subjective element. Secondly, in a conference like the present one, the discussions are generally on the pressing problems of the present and possible developments that are likely to take place in the foreseeable future. Frankly, I am an old, out-of-date person who is a misfit to take part in such deliberations. In the circumstances, acceding to your request, all I can do is to travel down the memory lane and tell you something about how I came across the equation that bears my name.

It was early fifties when words like the black hole, static limit or geodesic incompleteness had not entered the scientific vocabulary and quite different types of peculiarities were clubbed under the name 'singularity'. My first interest was in the so-called Schwarzschild singularity where some metric tensor components vanished or blew up. Nevertheless, the metric determinant remained well behaved and the signature condition was nowhere violated. This was an indication that this so-called singularity was a peculiarity of the particular coordinate system, nevertheless, there were some awkward questions to be answered:

(a) What does this peculiarity which has no analogue in Newtonian gravitation signify?
(b) Is this singularity physically attainable?

Bergmann in his book 'Introduction to the theory of relativity' referred to an unpublished work of Robertson on the penetrability of the 'Schwarzschild singularity' and also Einstein's model of a rotating cluster in equilibrium. On the basis of his model, Einstein conjectured that matter could not be compactified to an extent sufficient for the appearance of the singularity. Bergmann's concluding remark that the Schwarzschild singularity is only 'partly singular' intrigued me.

---

*Reprinted from *Proceedings of the Symposium on the Fortieth Anniversary of the Raychaudhuri Equation* (IUCAA, 1995).

Bergmann was apparently unaware of the work of Oppenheimer and Snyder on the collapse of a spherical dust distribution and so was I. Thus I took up the study of a collapsing homogeneous dust distribution, which could be imagined to be a portion of a contracting Friedmann universe, in an outside empty space. The problem was solved using a comoving coordinate system and provided a couter-example to Einstein's conjecture. In view of the earlier work of Oppenheimer and Snyder, the result was not new but my paper correctly considered the boundary conditions which were overlooked by Oppenheimer and Snyder. This work led me to the ultimate occurrence of the collapse singularity – the singularity with which one is familiar with as the big bang origin of the universe.

But the then prevailing ideas about the big bang were confusing. Einstein speculated that this singularity signalled a failure of the general theory of relativity for high concentrations and intense fields and might be removed in a unified field theory where there will be no separation between field and matter. Others thought that the origin of the singularity lay in the assumptions of homogeneity and isotropy and in a more realistic picture, the singularity will disappear. However, Einstein did not succeed in building up such a unified field theory and Tolman and Omer investigated a non-homogeneous model only to find that the singularity persisted.

True, the steady state theory did away with the singularity but a breakdown of the long cherished conservation principle of energy did not find favour with the average physicist.

In this background came Gödel's famous paper in 1949 – the Einstein issue of the *Reviews of Modern Physics.* To be sure, I did not understand many parts of the paper but it was remarkable that there was no singularity. No doubt there were undesirable features – the closed timelike lines, absence of expansion and a large cosmological term augmenting gravitational attraction. Nevertheless, there seemed reasons to hope that one could so change the parameters, that an expansion will be there, the closed time like lines will disappear and what seemed most important to me, the singularity will not appear. With this naive hope I devoted considerable time trying to discover such a solution. Little did I realise that what I was doing was like searching in a dark night for a black cat which probably did not exist.

During this investigations, at same stage, a neat result of Einstein and Pauli influenced me. In the process of proving the non-existence of an everywhere regular solution representing a monopole, they have shown that the time–time component of the Ricci tensor can be elegantly expressed as a divergence. Just out of curiosity, I tried to figure out what this component will be in more general (i.e., non-static) case, spherically when rotation (*á la* Gödel) is introduced. Somewhat to my surprise, I found that the expression was not only fairly simple but one could read the equation physically, namely that the gravitation and shear (i.e. anisotropy of the velocity field) augment collapse while centrifugal repulsion opposes it.

That was the elementary form of the equation. However, I had obtained this by a rather clumsy method so that it might appear that the equation was coordinate dependent. Again I had assumed the velocity field to be geodesic and had consequently missed an important term. All these defects were removed by later researchers. Further, while I had restricted myself to material fluids whose velocity vectors are unit timelike, the case of null propagation vectors was also taken up by later workers. Anyway, in spite of all these developments the relativity community has generously continued to refer to the equation in my name.

To conclude, I may tell you that story of the publication of my paper – this will perhaps not be very boring. The result was first communicated to the Editor, *Physical Review* as a letter and was received by them on April 21, 1953. I stated the equation without proof and set out some of the consequences. The referee's report was dated May 27, 1953 and here is an excerpt from that:

"In spite of considerable efforts on my part, I did not understand this paper. The author interprets two equations ... and I have no idea whence these two equations come from. I have looked up the paper by Gödel ... and also Gödel's article in the Einstein volume of the Library of Living Philosophers... and I cannot find any relation similar to these two. ...Somehow I have the feeling that I may be terribly obtuse... If the author would be kind enough to enlighten me concerning the derivation of the equation... I would welcome an opportunity to re-read the paper and to advise concerning its publishability. At present, I feel unable to recommend its publication."

Following this I wrote out a full paper giving the derivation and also made a conjecture that there exist solutions in which the velocity brings about a bounce from a collapsing to an expanding phase. Using the values of universal energy density, Hubble constant and galactic spins which seemed plausible at that time, I tried to argue that the conditions at the bounce agreed with the conditions required in the theory of nucleosynthesis by Gamow *et al.* In view of the astrophysical slant I sent the paper to the *Astrophysical Journal*. However, the paper was rejected as the referee considered the astrophysical considerations of dubious value.

After this I cut off the astrophysical part and sent the truncated paper to the *Physical Review*, naming the paper as 'Relativistic Cosmology I': the number I was put in as I hoped that I shall be able to prove the existence of a rotating non-singular model and also to display some astrophysical consequences in a later communication. However, these hopes were never realised. Anyway the paper was received by the *Physical Review* on December 28, 1953. An acknowledgement duly came from the Editors but then there was a long period of silence when all my enquiries could elicit no reply. Then after about fourteen months, in February 1955, came the letter of acceptance. It carried an interesting remark:

"After much prodding we were finally able to recover your manuscript from the referee. We endeavour to choose as referee those colleagues who accept this task conscientiously. We regret that in this case, there was an extensive delay". So at last the paper appeared in May 15, 1955 issue – a little over two years after the first note was sent.

PRAMANA
— journal of
physics

Vol. 69, No. 1
July 2007
pp. 7–14

A K Raychaudhuri and his equation*

J EHLERS
Max-Planck-Institute for Gravitational Physics, Albert-Einstein-Insitute,
Am Mühlenberg 1, D-14476 Golm, Germany
E-mail: juergen.ehlers@aei.mpg.de

Abstract. Amal Kumar Raychaudhuri died on June 18, 2005. This essay follows the lecture which I gave in honour of this great Indian scientist and teacher on December 26, 2005 in Puri, India.

Keywords. Raychaudhuri equation; cosmology; gravitational collapse.

PACS Nos 01.60; 98.80

1. A K Raychaudhuri's scientific career

AKR, as Raychaudhuri came to be called later by his colleagues and students, was born on September 14, 1923 in Barisal, in what is now Bangladesh. He studied at Presidency College, Kolkata, and obtained his B.Sc. and M.Sc. degrees in 1942 and 1944, respectively.

From 1945 to 1949 AKR worked as Research Assistant at the Indian Association for the Cultivation of Sciences (IACS) in Jadavpur, Kolkata. There, he had to work in experimental physics, which reflected neither his wishes nor his talent – he rather wanted to do research in mathematical physics, and he pursued this interest by studying general relativity on his own.

At that time general relativity was considered a difficult and useless subject, admitting no interaction between theory and experiment. So not only in India, but practically everywhere only few theoreticians pursued general relativity, among them were Pascual Jordan and Otto Heckmann, who from 1949 on were my teachers at the University of Hamburg.

In 1949, AKR got a temporary position as Lecturer of Physics at Asutosh College, where he wrote his third publication 'Arbitrary Concentrations of Matter and the Schwarzschild Singularity' [1]. In this *Physical Review* article he constructed an

*Invited plenary talk delivered at the International Conference on *Einstein's Legacy in the New Millennium*, December 15–22, 2005, Puri, India.

exact model of a collapsing, spherical dust cloud surrounded by its vacuum gravitational field. Apparently unaware of the 1939 Oppenheimer–Snyder paper 'On Continued Gravitational Contraction', he solved the junction conditions required at the surface of the 'star' and showed that 'there is no theoretical limit to the degree of concentration, and the Schwarzschild singularity ... occurs only in some coordinate systems'. He contrasted his result with Einstein's model of circularly orbiting particles which, according to Einstein, establishes beyond reasonable doubt that the Schwarzschild singularity is physically unattainable, and he remarked that he had been led to this investigation by his doubts – no small achievement for a young self-made relativist.

In 1952 AKR returned to IACS, where he got a permanent position as Research Officer. There he had to work on solids, which resulted in three publications. He did not, however, neglect his main interest, turning his attention to structure formation in the universe and the question whether the Big Bang singularity of the cosmological models by Friedmann and Lemaitre will be present also in more general, nonsymmetric models, or can be circumvented. In trying to find an answer to this fundamental question, AKR found in 1953 what came to be called the Raychaudhuri equation, which will be discussed here. He submitted his important result to the *Physical Review* in December 1953, and it was published in 1955 [2]. Its importance and implications were however then not recognized by his Indian colleagues.

At this time exact solutions of Einstein's equations were discussed extensively at the Hamburg Seminar on General Relativity and Cosmology directed by Pascual Jordan. These efforts were closely related to the long-standing interest in cosmology which Otto Heckmann, Director of the Observatory in Bergedorf near Hamburg, pursued. The young members of the Jordan-Seminar (among them W Kundt, E Schucking, R K Sachs, M Truemper and I Ozsvath) were eager not just to review known solutions, but to find new ones, and to understand their intrinsic properties using newly developed mathematical methods like differential forms and moving frames. Good fortune had it that Engelbert Schucking happened to read AKR's article in the *Physical Review* and recognized its importance; he coined the name Raychaudhuri equation [3,4]. The Hamburg group applied and extended AKR's work, of which AKR learned through a letter, presumably from the young Indian student Bramachary, who was a member of the Jordan Seminar around 1956. AKR was very delighted about this recognition, which helped his standing in India a lot. It encouraged him to submit his work as a PhD thesis, and again, it was fortunate for AKR that John A Wheeler was chosen as his external examiner. Wheeler immediately recognized the significance of the work and saw to it that the PhD was granted, in 1960, 'with honours'. Promptly, in 1961, AKR was appointed professor of physics at Presidency College, where he served until his retirement in 1988.

2. The Raychaudhuri equation

Around 1955, Raychaudhuri was interested mainly in two problems:

1. Is the Big Bang singularity due to the strong symmetry of the Friedmann–Lemaitre–Robertson–Walker models due to the strong symmetry assumptions which had been made when setting up these models?
2. Is a cosmological constant needed to solve the time-scale problem? Based on his discovery of different types of Cepheid variables, the astronomer Walter Baade had revised the extragalactic distance scale, and thereby ameliorated the time-scale problem. But was that sufficient?

AKR found his equation while searching for answers to these questions. I will now give a derivation of the equation.

Consider an arbitrary motion of a continuous medium in a general space-time $(M, g_{\alpha\beta})$. Let the world lines of the fluid elements be generated by the 4-velocity field U. Then the spatial tensors $\omega_{\alpha\beta} = \omega_{[\alpha\beta]}$, $\delta_{\alpha\beta} = \delta_{(\alpha\beta)}$, $\Theta = U^{\alpha}_{;\alpha}$, $h_{\alpha\beta} = g_{\alpha\beta} + U_{\alpha}U_{\beta}$ which appear in the decomposition

$$U_{\alpha\beta} = \omega_{\alpha\beta} + \sigma_{\alpha\beta} + \frac{1}{3}\Theta h_{\alpha\beta} - \dot{U}_{\alpha}U_{\beta} \tag{1}$$

represent, respectively, the rate of rotation, shear, expansion, the spatial metric and the acceleration associated with the fluid motion. Insertion of eq. (1) into the equation

$$U_{\alpha;|\beta;\gamma|} = \frac{1}{2}U_{\delta}R^{\delta}_{\alpha\beta\gamma}$$

which defines the Riemann curvature tensor, gives the kinematical identity

$$\dot{\Theta} + \frac{1}{3}\Theta^2 = \dot{U}^{\alpha}_{;\alpha} + 2(\omega^2 - \sigma^2) - R_{\alpha\beta}U^{\alpha}U^{\beta}, \tag{2}$$

where the overdot indicates covariant differentiation along the world lines, e.g. $\dot{\Theta} = \Theta_{;\alpha}U^{\alpha}$. If eq. (2) is combined with Einstein's field equation

$$R_{\alpha\beta} = \kappa\left(T_{\alpha\beta} - \frac{1}{2}Tg_{\alpha\beta}\right) + \Lambda g_{\alpha\beta} \tag{3}$$

and a length scale l is introduced via the definition $\dot{l}/l = \frac{1}{3}\Theta$, the (slightly generalized) Raychaudhuri equation

$$3\frac{\ddot{l}}{l} = \dot{U}^{\alpha}_{;\alpha} + 2(\omega^2 - \sigma^2) - \kappa(T_{\alpha\beta}U^{\alpha}U^{\beta} + \frac{1}{2}T) + \Lambda \tag{4}$$

results. It is a covariant form of the (time, time) component, in a frame adapted to U, of the field equation (3). Equation (4) shows: If the acceleration term is non-positive and, together with the shear and matter terms, dominates the rotation and Λ-terms, the mean motion is decelerating, otherwise accelerating.

In the original version of 1955, Raychaudhuri assumed pressureless matter of density ρ; then eq. (4) simplifies to

$$3\frac{\ddot{l}}{l} = -\frac{\kappa}{2}\rho - \sigma^2 + \Lambda + 2\omega^2. \tag{5}$$

(The same relation holds in Newtonian theory if Poisson's equation is generalized to $\Delta\Phi = 4\pi\rho - \Lambda$.) Raychaudhuri drew several conclusions from his eq. (5):

(a) In the static case, eq. (5) reduces to the relation $\kappa/2\rho = \Lambda$ of Einstein's 'cylindrical' universe. AKR emphasises that in his approach that relation follows 'without any assumption regarding symmetry or conditions obtaining in distant parts of the universe', and he states (without proof) that 'the only static nonspinning universe is the Einstein universe'. This correct statement shows that spatial closure follows from staticity and local properties; it need not be assumed. (I remark in passing that an analogous characterization holds for Gödel's universe: It is the only dust solution of eq. (3) without shear, having covariantly constant rotation rate [5].)
(b) If $\omega = 0$ on a world line and $\Lambda = 0$, and if $\dot{l}(t_0) > 0$ for some t_0, then there exists a density singularity in the past of t_0, and the 'age' is bounded by $(l/\dot{l})_0$. This is the first singularity theorem without symmetry assumptions. The more general eq. (5) has been used, e.g., by Hawking–Ellis to argue for a hot Big Bang [6].
(c) If $\omega = 0$ on a world line, then the time between a state $(\rho_{0,}(\dot{l}/l)_0)$ and a prior state with $l < l_0$, is maximal if $\sigma = 0$.
(d) If a singularity happened on a world line, then the 'age' measured from the singularity increases with Λ and ω and decreases with increasing ρ and shear. 'Thus the longer time scales are due to an increased freedom in the choice of Λ rather than to anisotropy'.

AKR's 1955 paper ends with the remark that in a dust model with vanishing rotation and shear the 3-spaces orthogonal to the fluid world lines have constant curvature. (In fact, the space-time is then of the Robertson–Walker type [5].)

The paper reviewed here is truly a landmark in cosmology. It showed how results are in principle observable and local variables can be obtained without imposition of isometries. It can be considered a forerunner of a programme begun by J Kristian and R K Sachs and continued by G Ellis and his coworkers. Moreover, it paved the way to the Penrose–Hawking singularity theorems.

I should like to recall that R K Sachs in 1960 obtained equations similar to eq. (5) for congruences of light-like geodesics. They first played a part in gravitational radiation theory in connection with peeling theorems, and later and at present are employed in gravitational lensing, especially in locating caustics of optical wavefronts.

In his later work, AKR constructed several exact solutions, among them the one illustrating spherical, but inhomogeneous collapse in an expanding universe and space-times containing charged matter and electromagnetic fields. His monograph 'Theoretical Cosmology' [7], published in 1979 gives a comprehensive survey of the state-of-the-art at that time. The text distinguishes itself by the care with which assumptions and conclusions are stated and critically interpreted, and by the attention given to observations and alternative theoretical explanations thereof. The preface begins: 'The book aims to convey to the reader some of the excitement while not hiding the confusion that remains.' From 1990 onwards AKR became interested in singularity-free space-times and attempted to find properties that might characterize them.

AKR was the president of the Indian Association of General Relativity and Gravitation from 1978 to 1980 and Fellow of the Indian Academy of Sciences since 1982.

Finally I would like to mention that Raychadhuri was not only a pathbreaking researcher, but also an outstanding teacher who created in many young minds enthusiasm for physics, especially general relativity and relativistic cosmology. He continued the line of great Indian physicists such as Saha, Raman, Bose, Bhabha and Chandrasekhar.

List of Publications of Professor A K Raychaudhury

A. Papers

1. Volkoff's massive spheres; *Phys. Rev.* **84**, 166 (1951)
2. Radiation sphere in Einstein universe; *Bull. Calcutta Math. Soc.* **44**, 31 (1952)
3. Condensations in expanding cosmological models; *Phys. Rev.* **86**, 90 (1952)
4. Reine Strahlungsfelder mit Zentralsymmetrie in der allgemeinem Relativitätstheorie; *Z. Physik.* **135**, 225 (1953)
5. Arbitrary concentration of matter and the Schwarzchild singularity; *Phys. Rev.* **89**, 417 (1953)
6. Relativistic cosmology I; *Phys. Rev.* **98**, 1123 (1955)
7. Note on the network approximation in metals; *Proc. Phys. Soc. (London)* **A68**, 439 (1955)
8. Perturbed cosmological models; *Z. Astrophysik* **37**, 103 (1955)
9. Relativistic and Newtonian cosmology; *Z. Astrophysik* **43**, 161 (1957)
10. Singular state in relativistic cosmology; *Phys. Rev.* **106**, 172 (1957)
11. Electronic energy bands in model three dimensional lattices, *Z. Physik* **148**, 435 (1957)
12. An anisotropic cosmological solution in general relativity; *Proc. Phys. Soc. (London)* **A72**, 263 (1958)
13. The isoelectronic series of semiconducting compounds with zinc-blende structure; *Proc. Nat. Inst. Sci. (India)* **25**, 201 (1959)
14. Static electromagnetic fields in general relativity; *Ann. Phys. (NY)* **11**, 501 (1960)
15. A general deduction of two important relations in relativistic cosmology; *Z. Astrophysik* **51**, 88 (1961)
16. Stationary cylindrically symmetric clusters of particles in general relativity (with M M Som); *Proc. Cambridge Philos. Soc.* **58**, 338 (1962)
17. Cosmological evolution with creation of matter (with S. Banerji); *Z. Astrophysik* **58**, 187 (1964)
18. Gravitational collapse in a cosmological background; *Proc. Phys. Soc. (London)* **88**, 545 (1966)
19. Static distribution of charged dust in general relativity (with U K Deo); *Proc. Roy. Soc. (London)* **A303**, 97 (1968)
20. Stationary electromagnetic fields in general relativity (with B K Dutta); *J. Math. Phys.* **9**, 1715 (1968)

21. Cylindrically symmetric charged dust distribution in rigid rotation in general relativity (with M M Som); *Proc. R. Soc. (London)* **A304**, 81 (1968)
22. Charged dust distribution in general relativity (with U K De); *J. Phys.* **A3**, 263 (1970)
23. Tachyons and gravitation; *J. Math. Phys.* **15**, 856 (1974)
24. Null fields in spaces of high symmetry (with A K Datta); *J. Math. Phys.* **15**, 1277 (1974)
25. Spherically symmetric charged dust distribution in general relativity; *Ann. Inst. Henri Poincare* **22**, 229 (1975)
26. Nature of the singularity in some Brans-Dicke universes; *Prog. Theor. Phys.* **53**, 1360 (1975)
27. Einstein–Cartan cosmologies with a magnetic field; *Phys. Rev.* **D12**, 952 (1975)
28. Einstein–Cartan spheres (with S Banerji), *Phys. Rev.* **D16**, 281 (1977)
29. Perfect fluid cosmology with geodesic world lines (with S R Maity); *Phys. Rev.* **D18**, 3595 (1978)
30. Charged dust distribution in Brans–Dicke theory (with N Bandyopadhyay); *Phys. Rev.* **D18**, 2756 (1978)
31. Field of a charged particle in Brans–Dicke theory (with N Bandyopadhyay); *Prog. Theor. Phys.* **59**, 414 (1978)
32. Ferraro's theorem in non-flat space-time; *Mon. Not. R. Astron. Soc.* **189**, 39 (1979)
33. Conformal flatness and the Schwarzchild interior solution (with S R Maity); *J. Math. Phys.* **20**, 245 (1979)
34. Homogenoeous space-times of Godel type (with S Guhathakurta); *Phys. Rev.* **D22**, 802 (1980)
35. Viscous fluid interpretation of electromagnetic fields in general relativity (with S K Saha); *J. Math. Phys.* **22**, 2237 (1981)
36. Rotating charged dust distributions in general relativity; *J. Phys.* **A15**, 831 (1982)
37. Viscous fluid interpretation of electromagnetic fields in general relativity II (with S K Saha); *J. Math. Phys.* **23**, 2554 (1982)
38. Temperature dependent gravitational constant and black hole physics (with B Bagchi); *Phys. Lett.* **B124**, 168 (1983)
39. Dual interpretation of electromagnetic fields in general relativity (with S K Saha); *Gen. Relativ. Gravit.* **15**, 611 (1983)
40. Is the universe near the state of maximal expansion? (with G Mukherjee); *Mon. Not. R. Astron. Soc.* **209**, 353 (1984)
41. Physical approach to cosmological homogeneity (with B Modak); *Phys. Rev.* **D31**, 1807 (1985)
42. Domain walls of finite thickness in general relativity (with G Mukherjee); *Phys. Rev. Lett.* **59**, 1504 (1987)
43. Inflation with arbitrary initial conditions; *Class. Quantum. Gravit.* **5**, 225 (1988)

44. Cosmic strings in general relativity; *Phys. Rev.* **D41**, 3041 (1990)
45. Theorem for non-rotating singularity free universes; *Phys. Rev. Lett.* **80**, 645 (1998)
46. N K Dadhich and A K Raychaudhuri, *Mod. Phys. Lett.* **14**, 2135 (1999)
47. A new theorem in relativistic cosmology; *Mod. Phys. Lett.* **15**, 391 (2000)
48. Singularity free cosmological solutions with non-rotating perfect fluids; *Gen. Relativ. Gravit.* **36**, 343 (2004)

B. Books

1. *Theoretical Cosmology* (Clarendon Press, Oxford; Oxford University Press, 1979)
2. *Classical Mechanics – A Course of Lectures* (Oxford University Press, India, 1983)
3. *Classical Theory of Electromagnetic Fields* (Oxford University Press, India, 1990)
4. *General Relativity, Astrophysics and Cosmology* (with S Banerji and A Banerjee) (Springer-Verlag, 1992)
5. *Uchchataro Gatibidya* (Higher Mechanics – in Bengali) (Paschim Banga Rajya Pustak Parishad, Kolkata)
6. Some outstanding problems in cosmology; in *Gravitation and Relativistic Astrophysics*: edited by Prasanna *et al* (World Scientific, 1984)

Acknowledgements

I am grateful to Naresh Dadhich for inviting me to make this contribution to the Puri conference, and also for providing me with a copy of his article 'A Lighthouse Falls', a memorial on Raychaudhuri, prior to publication. The article 'Solitary Pursuit' in *The Telegraph* by Parthasarathi Majumdar was helpful, too. My thanks go to Anita Ehlers and the excellent www-connections offered by IUCAA, through which I – though a computer illiterate – could establish fast contact with my former fellow students, notably Engelbert Schucking.

References

[1] A K Raychaudhuri, Arbitrary concentration of matter and the Schwarzschild singularity, *Phys. Rev.* **89**, 417 (1953)
[2] A K Raychaudhuri, Relativistic cosmology, *Phys. Rev.* **98**, 1123 (1955)
[3] There they say on p. 448: 'An important theorem for arbitrary models with incoherent matter has been discovered by A K Raychaudhuri in 1955. It rests on the Raychaudhuri equation for the scalar of expansion'

[4] O Heckmann and E Schucking, in *Gravitation: An introduction to current research* (New York, 1962)
[5] J Ehlers, Contributions to the relativistic mechanics of continuous media, *Gen. Relativ. Gravit.* **25**, 1225 (1993), Translation of the 1961 German original
[6] S W Hawking and G F R Ellis, *The large scale structure of spacetime* (Cambridge, 1973)
[7] A K Raychaudhuri, *Theoretical cosmology* (Oxford University Press, 1979)

PRAMANA
— journal of
physics

© Indian Academy of Sciences

Vol. 69, No. 1
July 2007
pp. 15–22

On the Raychaudhuri equation

GEORGE F R ELLIS
Mathematics Department, University of Cape Town, Rondebosch, Cape Town, South Africa
E-mail: Ellis@MATHS.uct.ac.za

Abstract. The Raychaudhuri equation is central to the understanding of gravitational attraction in astrophysics and cosmology, and in particular underlies the famous singularity theorems of general relativity theory. This paper reviews the derivation of the equation, and its significance in cosmology.

Keywords. General relativity; singularity; cosmology.

PACS Nos 04.20.Dw; 47.75.+f; 95.30.Sf; 98.80.-k; 98.80.Jk

1. Introduction

Amalkumar Raychaudhuri's remarkable paper [1] for the first time gave a general derivation of the fundamental equation of gravitational attraction for pressure-free matter, showing the repulsive nature of a positive cosmological constant, and underlying the basic singularity theorem (see below). He used special coordinates in the process, but wrote the result in a transparent covariant way. What was remarkable was how he derived this equation on his own, after reading writings of K Gödel on the ideas of shear and vorticity in cosmology (he defines the shear (eq. (8) in [1]) without fully explaining its meaning, apparently being unaware of Heckmann's writing on the topic, e.g. [2]). When combined with the other fundamental equation of gravity, the energy density conservation equation, it can be integrated to give the Friedmann equation (when the shear and vorticity vanish) and various generalisations of that equation (when the shear and vorticity have simple behaviour, e.g. in the case of Bianchi I models).

This equation was extended to arbitrary matter in an important paper by Ehlers [3], which included acceleration of the matter world lines and arbitrary stress tensors. This established the key result that the active gravitational mass density of a continuous medium is $\mu_{\rm grav} \equiv \mu + 3p/c^2$. Particular versions of this equation had been obtained earlier, at centers of symmetry by Tolman and Synge (see Raychaudhuri's footnote 12 of ref. [1]) and in the case of static space-time by Whittaker [4]. It was obtained for cosmology by Eddington [5] who used it to prove the instability of the Einstein static universe – a key result. It also underlies the estimates of ages in cosmological models obeying the energy conditions.

Its generalisation to the case of null geodesics (the null Raychaudhuri equation) plays a key role in geometrical optics in a curved space-time, as explored by Sachs, Ehlers, Penrose and others (see the summary in [6]). The combination of the time-like and null versions of the equation then played a key role in many singularity theorems – simple ones applicable in the case of Friedmann universes and inhomogeneous or anisotropic pressure-free irrotational models [3] and finally the famous Hawking–Penrose theorems showing the existence of singularities under rather generic conditions [6,7]. When completed with the full set of 1+3 covariant equations, it plays a central role in the investigations of the growth of inhomogeneities in perturbed cosmological models. These further developments will not be considered here. Rather, I focus on the direct application to time-like curves in cosmological and astrophysical contexts.

2. The equation

In fluid flows and cosmology, there is a preferred 4-velocity vector field $u^a : u^a u_a = -1$ that represents the average motion of matter [3,8]. Let τ be the proper time along these world lines: $u^a = \mathrm{d}x^a/\mathrm{d}\tau$. The time derivative of any tensor $T^{ab\cdot\cdot}_{\cdot\cdot cd}$ along the fluid flow lines is

$$\dot{T}^{ab\cdots}_{\cdots cd} = T^{ab\cdots}_{\cdots cd;e}u^e.$$

A particular application of time differentiation is the derivative of the 4-velocity itself in its own direction. This determines the acceleration vector

$$\dot{u}^a \;=\; u^a{}_{;b}u^b \;\Rightarrow\; \dot{u}^a u_a = 0, \tag{1}$$

which vanishes if and only if the flow lines are geodesics. It is convenient to define the representative length $\ell(x^i)$ such that comoving volumes scale like ℓ^3. The expansion $\Theta = u^a_{;a}$ gives the rate of change of volume d^3V:

$$\Theta = 3\dot{\ell}/\ell = (\mathrm{d}^3V)\dot{}/(\mathrm{d}^3V). \tag{2}$$

The shear and vorticity magnitudes are ω^2 and σ^2. Spatial gradients orthogonal to the preferred world lines are determined by the derivative operator $\nabla_a f := (g^b_a + u_a u^b)f_{,b}$.

The Einstein field equations (EFE) are

$$R_{ij} - \frac{1}{2}Rg_{ij} + \Lambda g_{ij} = \kappa T_{ij}, \tag{3}$$

where Λ is the cosmological constant and κ is the gravitational constant. The Raychaudhuri equation, giving the evolution of Θ along the fluid flow lines [1,3,8], is obtained by contracting the equivalent form

$$R_{ij} = \kappa\left\{T_{ij} - \frac{1}{2}Tg_{ij}\right\} + \Lambda g_{ij}, \tag{4}$$

with $(u^a u^b)$ and using the Ricci identity for u^a. Thus one has

Raychaudhuri equation

$$\dot{\Theta} + \frac{1}{3}\Theta^2 + 2(\sigma^2 - \omega^2) - \dot{u}^a{}_{;a} + \frac{1}{2}\kappa(\mu + 3p/c^2) - \Lambda = 0. \tag{5}$$

This is the generic form of the Raychaudhuri equation (eq. (11) in [1] is for the case $p = 0$). It is the fundamental equation of gravitational attraction. To see its implications, we rewrite it in the form

$$3\frac{\ddot{\ell}}{\ell} = -2(\sigma^2 - \omega^2) + \dot{u}^a{}_{;a} - \frac{1}{2}\kappa(\mu + 3p/c^2) + \Lambda \tag{6}$$

which follows from the definition of the scale factor ℓ. This equation for the curvature $\ddot{\ell}$ of the curve $\ell(\tau)$ directly shows that shear, energy density and pressure tend to make matter collapse, as they tend to make the $\ell(\tau)$ curve bend down; while vorticity and a positive cosmological constant tend to make matter expand, as they tend to make the $\ell(\tau)$ curve bend up. The acceleration term is of indefinite sign. The equation shows that $\mu_{\text{grav}} = \mu + 3p/c^2$ is the active gravitational mass density of a fluid. Hence any increase in the internal energy or the pressure increases its active gravitational mass density. The corresponding equation in Newtonian theory is the same except that $(\mu + 3p/c^2) \to \rho$ (the active gravitational mass density is just the mass density) [8].

3. Applications

3.1 *Static star models*

In the case of a static star, $\Theta = \omega = \sigma = 0$ and we can neglect the cosmological constant on this scale. So the equation reduces to $\dot{u}^a{}_{;a} = \frac{1}{2}\kappa(\mu + 3p/c^2)$, where the acceleration is determined from the pressure gradient by the momentum conservation equation. In the case of a perfect fluid,

$$(\mu + p/c^2)\dot{u}^a + \nabla^a p = 0.$$

So we obtain from (5)

$$((\nabla^a p)/(\mu + p/c^2))_{;a} = -\frac{1}{2}\kappa(\mu + 3p/c^2),$$

which is the basic balance equation between gravitational attraction and hydrostatic pressure for a static star. In the corresponding Newtonian equations, $\mu + 3p/c^2 \to \rho$ and $\mu + p/c^2 \to \rho$. It is these differences which cause gravitational collapse to be so much less severe in Newtonian theory than in general relativity.

3.2 *Static FLRW universe models*

In the case of a static Friedman–Lemaître–Robertsom–Walker (FLRW) universe, $\Theta = \omega = \sigma = 0 = \dot{u}^a$. So eq. (5) becomes

$$\frac{1}{2}\kappa(\mu + 3p/c^2) = \Lambda. \tag{7}$$

Thus *static universes with ordinary matter are only possible if* $\Lambda > 0$ [9] (cf. discussion after eq. (14) in [1]). Then, given the equation of state $p = p(\mu)$ and the cosmological constant, there is a unique radius ℓ_s for the static solution, at which the gravitational attraction caused by matter and the repulsion caused by the cosmological constant balance.

However, *this universe is unstable* [5] because if we change ℓ to a larger value, $\ell > \ell_s$, μ decreases but Λ stays constant. So $\ddot{\ell} > 0$ and the universe expands to infinity. Similarly, $\ell < \ell_s \Rightarrow \ddot{\ell} < 0$ and the universe collapses. This instability [10] leads us to believe that *the universe should either be expanding or contracting, but not static.* Indeed, the failure to perceive this in the 1920s can be regarded as one of the major lost opportunities in the history of cosmology (all the major figures in cosmology at that time believed that universe was static).

In the Newtonian case, the corresponding static model satisfies $\kappa\rho/2 = \Lambda$. The same qualitative results hold then as in general relativity theory (i.e. $\Lambda > 0$ and the model is unstable).

3.3 *Friedmann–Lemaître models*

In the FLRW case, the Raychaudhuri equation (6) becomes (cf. eq. (12) in [1]), for the case $p = 0$

$$3\ddot{\ell}/\ell = -\frac{1}{2}\kappa(\mu + 3p/c^2) + \Lambda. \tag{8}$$

Now the energy conservation equation for a perfect fluid is

$$\dot{\mu} + (\mu + p/c^2)\Theta = 0 \tag{9}$$

and this implies that $(\ell^2\mu)^{\cdot} = -\ell\dot{\ell}(\mu+3p/c^2)$. Thus, provided $\dot{\ell} \neq 0$, we can multiply (8) by $\ell\dot{\ell}$ and integrate to find

$$3(\dot{\ell})^2 - \kappa\mu\ell^2 - \Lambda\ell^2 = \text{const.} \tag{10}$$

This is just the *Friedmann equation* which governs the time evolution of FLRW universe models.

4. The basic singularity theorem

The fundamental singularity theorem follows immediately from the Raychaudhuri equation [1,3,8].

Theorem. *Irrotational geodesic singularities. If* $\Lambda \leq 0$, $\mu + 3p/c^2 \geq 0$ *and* $\mu + p/c^2 > 0$ *in a fluid flow for which* $\dot{u} = 0$, $\omega = 0$ *and* $H_0 > 0$ *at some time* s_0, *then a space-time singularity, where either* $\ell(\tau) \to 0$ *or* $\sigma \to \infty$, *occurs at a finite proper time* $\tau_0 \leq 1/H_0$ *before* s_0.

Proof. Consider the curve of ℓ against proper time τ. If $\ddot{\ell} = 0$, then $\ell \to 0$ a time $1/H_0$ ago. However, with the given conditions, $\ddot{\ell} < 0$. So, following the curve $\ell(\tau)$

back in the past, it must drop below this straight line and reach arbitrarily small positive values of ℓ at a time less than $1/H_0$ ago (unless some other space-time singularity intervenes before $\ell \to 0$, which can happen only if the shear diverges first).

In the exceptional case where the shear diverges first, a conformal space-time singularity will occur where $\ell \neq 0$. In the general case where $\ell \to 0$, the matter world lines converge together a finite time ago in the past and a space-time singularity will develop if $\mu + p/c^2 > 0$, for then as the universe contracts, the density and pressure increase indefinitely, implying the space-time curvature does so also. For ordinary matter this will additionally imply that $T \to \infty$, that is, the universe originates at a hot Big Bang. Furthermore, an age problem also becomes possible; for if we observe structures in the universe, such as stars, globular clusters, or galaxies, that are older than $1/H_0$, there is a contradiction with the assumptions of the theorem (for the universe must be older than its contents!). Furthermore, the presence of shear will shorten the lifetime of the universe (for a given value of the Hubble constant today) (cf. discussion after eq. (16) in [1]).

Similarly the argument implies that the universe must experience a very quick evolution through its hot early phase. For example at the time of decoupling the scale function ℓ_d is small: $\ell_d/\ell_0 \approx 1/1000$, which implies the Hubble parameter H_d at that time is $H_d > 1000H_0$ (the straight line estimate $\ddot{\ell} = 0 \Rightarrow H \propto 1/\ell \Rightarrow H_d \approx 1000H_0$; however the high densities will cause a considerable steepening of the $\ell(\tau)$ curve by those times, leading to the inequality). Similarly at the time of nucleosynthesis $\ell/\ell_0 \approx 10^{-8}$, showing that $H > 10^8 H_0$.

Application:

This result applies in particular to an expanding Friedmann–Lemaître universe, where $\ell \to 0$ and a hot Big Bang must occur (the shear is zero in this case, so a conformal singularity cannot occur). The proof makes it clear that *an increase in pressure does not resist the occurrence of the singularity*, but rather decreases the age of the universe and so makes the age problem worse (the pressure increases the active gravitational mass and there are no pressure gradients to resist the collapse).

This is the basic singularity theorem, on which further elaborations are built. How can one avoid the singularity? It is clear that shear anisotropy makes the situation worse. On the face of it, there are five possible routes to avoid the conclusion: a positive cosmological constant; acceleration; vorticity; an energy condition violation or alternative gravitational equations. We consider them in turn.

(a) *Cosmological constant* $\Lambda > 0$. In principle, this could dominate the matter and turn the universe around. However in practice this cannot happen because we have seen galaxies and quasars up to a red-shift of about 5, implying that the universe has expanded by at least a ratio of 5 to 6 to the present time. This means that if it bounced, then at the time of the turnaround, the density would have been greater than the present density by a factor of at least $5^3 = 125$. So the cosmological constant would have to be equivalent to a larger energy density, in order to dominate the Raychaudhuri equation then. We would certainly have detected so large a cosmological constant today, and have not done so, as can be deduced from present estimates of its magnitude. Further, if we accept that the microwave background radiation indicates that the universe has expanded by at least a factor of 1000, the argument is overwhelming; the cosmological constant would have to be equivalent to more than 10^9 times the present matter density to dominate the Raychaudhuri equation then! There is no way we could have avoided detecting this.

(b) *Pressure inhomogeneity* (acceleration) and

(c) *Rotational anisotropy* (the effect of 'centrifugal force'). Both of these involve abandoning the FLRW geometry. On the face of it, they could succeed. However, the powerful Hawking–Penrose singularity theorems [6,7] strongly restrict the allowable cases where they might in fact succeed, because of the microwave background radiation observations which show, for universes which are approximately Robertson–Walker, that the conditions of those more general theorems hold [6].

(d) *Violation of the energy condition.* The above result depended on the energy condition

$$\mu + 3p/c^2 \geq 0$$

which is obeyed by all normal matter. However, the false vacuum equation $\mu + p/c^2 = 0$ violates this, and can in principle cause a turnaround of the universe, avoiding an initial singularity. Nevertheless, we do not expect this equation to become relevant until temperatures of at least 10^{12} K. Thus even if violating the energy condition could enable us to avoid the initial singularity, the turnaround would only take place under extraordinarily extreme conditions when quantum effects are dominant. Hence we can rephrase the conclusion: a viable non-singular universe model cannot obey the laws of classical physics at all times in the past.

(e) *Other gravitational field equations.* Finally, we have of course assumed Einstein's field equations here. Alternative theories of gravity will certainly allow singularity violation, effectively by introducing negative energy terms into the Raychaudhuri equation. In particular, at very early times quantum gravitational effects will become important, almost certainly causing effective energy condition violations.

Thus the prediction of a singularity is a classical prediction. Physically, we may take it as the prediction that, as we follow the evolution back into the past, the universe cannot avoid entering the quantum-gravity domain. We do not yet have any reliable idea of what this implies.

In Newtonian theory, the discussion is as above except for one important point: then rotation *can* enable the universe to avoid the initial singularity (unlike in general relativity). This is shown by the existence of spatially homogeneous rotating and expanding but shear-free Newtonian universe models, in which the rotation spins up to enable the universe to avoid the initial singularity, whereas such universes cannot exist in relativity theory [11].

5. Evaluation today

We obtain very useful information by evaluating the Raychaudhuri equation at the present time. To do this, we define some useful parameters as follows. The *deceleration parameter* is

$$q_0 = -\left(\frac{\ddot{\ell}}{\ell}\right)_0 \frac{1}{H_0^2} \tag{11}$$

which is a dimensionless version of the second derivative $\ddot{\ell}$, with the sign chosen so that a positive value corresponds to deceleration. The total energy density is represented by the dimensionless *density parameter*

$$\Omega_0 = \frac{\kappa\mu_0}{3H_0^2}\ . \tag{12}$$

Similarly pressure and the cosmological constant are represented by

$$\Omega_{\mathrm{p}} = \frac{\kappa p_0}{3H_0^2 c^2}, \quad \Omega_\Lambda = \frac{\Lambda_0}{3H_0^2}.$$

It then follows directly from the Raychaudhuri equation that

$$2q_0 = \frac{4}{3}\left(\frac{\sigma_0^2}{H_0^2} - \frac{\omega_0^2}{H_0^2}\right) - \frac{2\dot{u}^a{}_{;a}}{3H_0^2} + \Omega_0 + 3\Omega_{\mathrm{p}} - \Omega_\Lambda. \tag{13}$$

If the rotation, shear, acceleration and pressure terms are small today compared with the others, as is highly plausible, then

$$2q_0 \simeq \Omega_0 + \Omega_\Lambda, \tag{14}$$

where the error is of the order of the magnitude of the terms neglected in passing from the previous equation. This becomes an exact equation in an FLRW universe with vanishing pressure. If $\Lambda = 0$, then this reduces to

$$2q_0 \simeq \Omega_0, \tag{15}$$

which again is exact in an FLRW universe with vanishing pressure, but this is no longer believed to be true. The Newtonian discussion is the same, except that there is no pressure contribution to (13).

Equation (14) is a direct relation between the deceleration and density parameters which is pivotal in observational cosmology. Dynamical estimates and lensing observations suggest $\Omega_{\mathrm{darkmatter}} \simeq 0.3$ with $\Omega_{\mathrm{baryons}} \simeq 0.04$. Supernovae observations of q_0, with this equation, then indicate that $\Omega_\Lambda \simeq 0.7$. This is the famed dark energy; the fact that its physical nature is unknown is a core problem for present day cosmology.

6. Conclusion

The Raychaudhuri equation is central to cosmology, as is made clear by all the applications discussed above. I have not discussed here its higher dimensional versions, which are of course just as important in higher dimensional gravity theories. Raychaudhuri's paper [1] contains other interesting results not discussed here (see §4), and was entitled "Relativistic cosmology, I". It has always been a matter of regret to me that Paper II was never written.

References

[1] A Raychaudhuri, Relativistic cosmology, I, *Phys. Rev.* **98**, 1123 (1955). Reprinted as a 'Golden Oldie' in *GRG* **32**, 749 (2000)

[2] O Heckman, *Theorien der Kosmologie* (Springer, 1942)

[3] J Ehlers, *Akad. Wiss. Lit. Mainz, Abhandl. Math.-Nat. Kl.* **11**, 793 (1961), translation: J Ehlers, *Gen. Relativ. Gravit.* **25**, 1225 (1993)

[4] E T Whittaker, On Gauss' theorem and the concept of mass in general relativity, *Proc. R. Soc. London* **A149**, 384 (1935)

[5] A S Eddington, On the Instability of Einstein's spherical world, *Mon. Not. R. Astron. Soc.* **90**, 668 (1930)

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

[7] F J Tipler, C J S Clarke and G F R Ellis, Singularities and horizons: a review article, in: *General relativity and gravitation: One hundred years after the birth of Albert Einstein* edited by A Held (Plenum Press, New York, 1980) vol. 2, p. 97

[8] G F R Ellis, in: *General Relativity and Cosmology, Proceedings of the XLVII Enrico Fermi Summer School* edited by R K Sachs (Academic Press, New York, 1971) p. 104

[9] A Einstein, Kosmologisched Betrachtungen der allgemeinen Relativitätsthorie, *Preuss. Akad. Wiss. Sitz* (1917) p. 142

[10] Since one cannot perturb the actual universe, one must understand 'instability' in terms of the difficulty in setting exactly the correct initial conditions for a static universe rather than its instability to perturbation from a previous state.

[11] G F R Ellis, The dynamics of pressure-free matter in general relativity, *J. Math. Phys.* **8**, 1171 (1967)

PRAMANA
— journal of
physics

© Indian Academy of Sciences

Vol. 69, No. 1
July 2007
pp. 23–29

Singularity: Raychaudhuri equation once again

NARESH DADHICH
Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Pune 411 007, India
E-mail: nkd@iucaa.ernet.in

Abstract. I first recount Raychaudhuri's deep involvement with the singularity problem in general relativity. I then argue that precisely the same situation has arisen today in loop quantum cosmology as obtained when Raychaudhuri discovered his celebrated equation. We thus need a new analogue of the Raychaudhuri equation in quantum gravity.

Keywords. Cosmology; Raychaudhuri equation; Universe; quantum gravity; loop quantum gravity; loop quantum cosmology.

PACS Nos 04.20.Jb; 04.2.Cv; 98.80.Dr

1. Singularity and AKR

It would not be far from the truth to say that A K Raychaudhuri (AKR) had a fascinatingly engaging love affair with the notion of a spacetime singularity at the two ends of his research career. One of his early concerns was to construct a model of a collapsing homogeneous dust ball (he was unaware of Oppenheimer–Snyder collapse) and show that nothing prevented the ball from collapsing down to the centre $r = 0$ and thereby demystify the so-called Schwarzschild singularity at $r = 2M$ [1]. Then he addressed the most pertinent question of his time: Is the cosmic singularity predicted by the FRW model an artifact of the homogeneity and isotropy of space or not? As he explained in his reminiscences [2], inspired by the famous Gödel solution, he was looking for a rotating non-singular solution without closed time-like lines. In the process, he discovered his celebrated equation [3] which made the singularity analysis free of these restrictions. That ultimately led to the powerful Hawking–Penrose singularity theorems [4] which established in a very general setting the inevitability of the occurrence of singularities in Einstein gravity under reasonable energy and causality conditions.

The other end phase began in the mid-1990s. The singularity theorems reigned supreme, particularly since the observation of CMBR [5] had pointed to a singular birth of the Universe in a Big Bang. Nothing could be happier and more persuasive than the observation verifying the prediction of theory. This gave rise to a general belief that singularities were inevitable in general relativity (GR) so long as the dynamics were governed by Einstein's equations and more over positive energy and causality conditions were respected. However, this belief was shaken by Senovilla's discovery in 1990 of a singularity-free cosmological solution [6] which did not violate the energy and causality conditions. How could such a thing happen? It brought forth the main suspect in the proofs of the singularity theorems. Apart from the self-evident assumptions, the theorems also required the existence of a closed trapped surface. This last requirement is certainly not so obvious and self-evident as the other assumptions. That gravity should become so strong in some bound region of space that even light could not escape from it is a very limiting assumption. Indeed, where gravity should become how strong ought to be determined by the field equations rather than by prescription. The said assumption may, however, be reasonable and justifiable for the case of the gravitational collapse of an isolated body. We know from the study of stellar structure that a sufficiently massive body could, after the exhaustion of its nuclear fuel, ultimately undergo indefinite collapse and thereby reach the trapped surface limit. In the case of Big Bang cosmology what is required is not a trapped surface but instead sufficient amount of matter distribution for the focusing of non-space-like trajectories at a finite proper time in the past. It is a different matter that the amount required to thermalise the cosmic background radiation is indeed sufficient for the convergence of trajectories in the past [4]. Though this limitation in terms of either a trapped surface or a sufficient amount of energy density was known to experts in the field, it was not talked about much, perhaps in the belief that a singularity-free solution would never be found.

In the early 1990s, L K Patel, Ramesh Tikekar and myself obtained some singularity-free cosmological solutions [7], in particular one with a stiff fluid equation of state $\rho = p$. During this period, I had several discussions with AKR. We both shared the view that the assumption of the existence of a closed trapped surface almost amounted to putting in a singularity. Nothing could come out of a closed trapped surface nor could the collapse be halted or reversed inside it without violating energy and causality conditions. Thus a singularity would become inevitable. He was also not happy with the genericity condition. In his view it was too complicated and physically not very illuminating. In the ICGC meeting at IUCAA in December 1995, Jose Senovilla and I had discussions with him and he then came out with an insightful comment. He opined that the vanishing of the space average of physical and kinematic parameters was required for singularity-free solutions. We were both struck by this comment which showed a new direction. (See Senovilla's account in this volume [8].)

AKR had started thinking about singularity-free cosmological solutions, but was not yet quite taken up by them. In November 1996, there was an IUCAA sponsored workshop on Inhomogeneous Cosmological Models in North Bengal University, Siliguri. I spoke there on singularity-free cosmological solutions. During the talk, AKR asked several probing questions and we had a very lively and engaging discussion. It was indicative of his thought process in trying to understand and resolve intricate and involved conceptual and physical issues. It took him a couple of years before he could start working on the question of the avoidance of the cosmic singularity. This showed his work ethic – deep and long period of contemplation and thought before taking up a problem. This venture took him once again to the question of singularity theorems. He argued that the existence of singularity-free cosmological solutions should be recognized and proposed the vanishing of the spacetime averages of all the scalars appearing in the Raychaudhuri equation as a necessary condition for their existence [9–11]. He later proved a new singularity theorem in which he replaced the occurrence of a closed trapped surface by the non-vanishing of the space averages of all scalars occurring in the Raychaudhuri equation [11]. The vanishing of such space averages was shown to be the key to singularity-free cosmological models.

The last paper that AKR wrote was in 2004. In this he attempted to deduce the Ruiz–Senovilla family [12] of non-singular solutions for a non-rotating perfect fluid from very general considerations. His procedure was novel, though not mathematically rigorous [13]. A very large family of singularity-free cosmological models including some counter-examples to his paper [13] had also been found [14]. It is however known that for an imperfect fluid it is easy to construct non-singular and even oscillating models [15,16]. The real challenge is, in fact, in obtaining rotating perfect fluid solutions. Apart from increasing the mathematical complexity, rotation brings in the question of the occurrence of closed time-like lines and consequently causality violation. We have the well-known rotating Gödel Universe which has closed time-like lines. Recall that it was precisely the Gödel solution which had set him on the singularity trail. Right from the beginning in the early 1950s, his main aim was to find a rotating fluid solution, hopefully free of any singularity as well as of any closed time-like line. Instead, he discovered his equation. The question remained open and unsolved, however. In fact, AKR returned once again to it at the end. It is undoubtedly one of the most challenging open problems in classical gravity today. Ironically, he began and also breathed his last with it on 18 June 2005.

AKR had the profound insight to have identified the key feature of non-singular solutions, namely the vanishing of space averages of physical parameters. It was the interaction between him and Jose Senovilla which led to the formulation of this conjecture, though each of them had a different perspective on it. AKR's attempt to prove it [11] was not entirely satisfactory and Senovilla has now proved it ultimately [8]. This result should rightly be called the Raychaudhuri–Senovilla theorem. It was perhaps the limitation of mathematical and analytical tools that AKR had at his command which came in the way of his proving the theorem rigorously. Yet, had it not been for his insight, the theorem might not have been formulated. Therefore, it is to AKR's credit that he showd the right path in understanding singularity-free cosmological solutions. If he had the benefit of the right kind of mathematical backup in the mid-1950s, he could conceivably have arrived at the famous singularity theorems. Once again, I believe that the limiting factor was mathematical technology. With utmost reverence and affection, I would like to acknowledge this fact in the true spirit of AKR which embodied academic and intellectual honesty and objectivity.

After recording the story of my understanding and perception of the AKR-singularity saga, let me change gears in the next section to argue that his equation in a new avatar is once again badly needed.

2. Equation once again

A singularity marks the limiting point of a physical theory. It is enigmatic and calls for a new theory. In Einstein's GR, gravity is nothing but the curvature of spacetime. A gravitational singularity thus means the breakdown of spacetime structure itself and hence the end of everything.

The forces of Newtonian gravity as well as Maxwell's electric field diverge and are singular at the central location of the mass/charge point. This singularity does not disturb the spacetime background. Rather, it indicates the limit of validity of the theory. For the electric field, we go over to quantum electrodynamics to overcome the classical singularity. For gravity, apart from addressing the singularity, we also need a new theory for the more basic requirement of making it fully universal. So we have Einstein's theory of gravitation, namely GR. But in this new theory, the Newtonian singularity not only persists but also attains a more profound all-encompassing proportion. One therefore needs a quantum theory of gravity which has to address the question of singularity in the spacetime structure itself.

GR made two profound predictions, one of the Black Hole and the other of the Big Bang. Both harboured singularities. In the former case it is hidden behind an event horizon and hence is inaccessible to an external observer. Though the Schwarzschild solution was obtained in 1916 immediately after GR was propounded, its full import as the representation of a static Black Hole was not realized as late as the late 1960s. A star, collapsing under its own gravity, will go on collapsing indefinitely upon the exhaustion of all its nuclear fuel and eventually hit the central singularity $r = 0$. The latter would be encompassed by a Black Hole event horizon from which nothing could come out. Penrose in 1969 pronounced that any singularity occurring in a gravitational collapse will always be covered by a Black Hole and this is known as the Cosmic Censorship Conjecture. Oppenheimer and Snyder considered the collapse of a homogeneous dust cloud. Their conclusion is that it collapses down to a singularity covered by a Black Hole.

In 1924, Friedmann obtained a non-static solution to Einstein's equations representing an expanding model of the Universe. In 1929, Hubble's observation of receding galaxies lent observational strength to this model. It was indeed a wonderful marriage of theory and observation. The matter distribution in the Universe was assumed to be homogeneous and isotropic. It predicted that the Universe, which was now expanding, would have had a singular beginning in a hot Big Bang when all matter was concentrated within a very small point-like region.

An important question then arose. Was this singularity an artifact of the symmetries of matter distribution, to wit homogeneity and isotropy, or a generic feature of Einstein's gravity? That is when AKR came on stage and formulated the singularity issue in all its generality and obtained his celebrated equation in 1953 [3]. The equation brought to the fore the new feature that shear as well as pressure contribute positively to gravity, while rotation goes the other way, as expected. Inspired by the Raychaudhuri equation, Penrose, Hawking and Geroch then proved in the mid-1960s their powerful singularity theorems [4] under very general conditions to establish that singularities are inevitable in GR so long as some reasonable energy and causality conditions are satisfied.

Within the classical framework, just as in Maxwell electrodynamics, there is no way to avoid a singularity in Einstein's gravity. That is why one of the main goals of any quantum gravity theory is to address the singularity question. We do not yet have a full-fledged theory of quantum gravity. There are two main attempts. One is string theory which is based on particle physics. The other is the canonical quantization scheme of loop quantum gravity which is based on GR. We shall follow developments in the latter since it directly addresses the singularity issue.

The loop quantum gravity (LQG) idea rests on an important breakthrough achieved by Abhay Ashtekar in 1986. He discovered new variables in which Einstein's equations take a polynomial form [17]. Since then he has spearheaded this approach. This effort is, however, pursued by a comparatively small but highly committed and talented team of researchers [18]. Even in the absence of a full theory, it is instructive and insightful to apply this developing theory to idealized special cases and probe for possible signatures of quantum gravity effects in astrophysical and cosmological observations. Such applications, howsoever tentative, serve as good testbeds for the evolving theory with regard to its right orientation and direction. With that in view, Martin Bojowald and others have, for the past few years, been examining cosmological applications of loop quantum gravity. Such efforts have led to the subject of loop quantum cosmology (LQC) where one considers the symmetry-reduced mini-superspace and then carries out loop quantum calculations for specific problems of Big-Bang cosmology, cosmic microwave background radiation and gravitational collapse [19–24]. The most pertinent question is whether one could have some observational imprint or signature of quantum gravity effects.

Recently, there have been a couple of LQC-based calculations to look for quantum effects in astrophysical and cosmological scenarios. The first observable effect of LQC was studied on CMBR in 2003 by Shinji Tsujikawa *et al* [24]. It turns out that quantum gravity effects could avoid the Big Bang singularity and there could be a causal passage through it. Its imprint could therefore be seen on the CMBR spectra. In a recent paper, Abhay Ashtekar *et al* further illuminate on the quantum nature of the Big Bang [20]. In this context, it is worth recalling one of the first attempts made by T Padmanabhan and J V Narlikar in the 1980s towards avoiding the Big Bang singularity by quantizing conformal degree of freedom [25]. The pertinent question is how to justify in the extreme high energy regime highly restricted degrees of freedom or the reduced mini-superspace of LQC? Both considerations suffer from this lack of justification. It should however be noted that, though there is no rigorous derivation of LQC from LQG, the former does have a good theoretical backup with proper caveats. There has also been a consideration of the collapse of a homogeneous scalar field by R Goswami *et al* where LQC effects make the central singularity evaporate away as radiation [23]. An observational signature of such a quantum evaporation of a naked singularity may be a pulse of intense radiation such as a gamma ray burst (GRB).

We now come back to the old question: Is such an avoidance of the Big Bang or of a collapse singularity an artifact of the symmetry-reduced mini-superspace or is it generic to LQC? Near the singularity, curvatures are divergingly high. So it would not be possible to truthfully sustain the assumption of a reduced mini-superspace. Once again, we need another AKR today to find a new avatar of his equation. In other words, a new Raychaudhuri equation is urgently required in loop quantum gravity. It may, like the old one, show the way to new general quantum singularity (avoidance) theorems.

It is true that LQC deals with highly restrictive and idealized cases. One should, however, note that such idealized models have the uncanny knack of picking up the physical essence and innate characteristic of real life phenomena. There are several such examples in gravitational theories. The most famous one is, of course, that of the FRW model predicting a Big Bang singularity and similarly the Oppenheimer–Snyder collapse of a homogeneous dust ball. Despite being a highly idealized case, the model does carry quite truthfully the signature of a general collapse phenomenon. Similar is the case of the Schwarzschild interior solution with a uniform density assumed for the interior of a star. Such an assumption is physically unacceptable since a uniform density would give rise to an infinite sound speed. Nevertheless, the model picks up all the essential features of the stellar interior correctly. It is remarkable that, even when such considerations are highly restricted, idealized and not even entirely physically acceptable, they often correctly indicate general as well as generic features. Such may as well be the case for the highly idealized LQC toy models indicating the quantum avoidance of the Big Bang and collapse singularities. As already emphasized, what is needed is a Raychaudhuri equation for quantum gravity. It may bring forth some new features like the development of a negative pressure as the singularity is approached. Until then we have to make do with tentative results which may only be indicative of what the full quantum gravity will ultimately establish.

Finally, all this should be most satisfying and pleasing to one man, Abhay Ashtekar. He made the path-breaking discovery two decades ago of his famous new variables which set things on track, leading to loop quantum gravity. This idea has now matured sufficiently to make contact with observations. It has indeed been a long and ardous journey. But, at the end of the day, nothing could please one more than to see the clicking of what one had set out to do. Admittedly, we are far from a complete theory of quantum gravity and LQG has a long way to go yet. What is important is that it seems to be on the right track [26]. Recent works of Smolin [27], Bianchi *et al* [28] and Livine and Speziale [29] on graviton propagators in LQG and Bilson-Thompson *et al* [30] on the origin of the standard model of particle physics from the quantum nature of geometry are indeed very exciting.

Acknowledgement

I thank Jose Senovilla for recounting and reflecting on our discussion with AKR and also thank Parampreet Singh and Pankaj Joshi for clarifying comments. I also thank my fellow editor of this volume, Probir Roy for reading the manuscript and making changes which have made it more impactful.

References

[1] A K Raychaudhuri, *Phys. Rev.* **84**, 166 (1951)
[2] A K Raychaudhuri, *A little reminiscence in singularities, black holes and cosmic censorship* edited by P S Joshi (IUCAA, 1996)
[3] A K Raychaudhuri, *Phys. Rev.* **90**, 1123 (1955)
[4] S W Hawking and G F R Ellis, *The large scale structure of space-time* (Cambridge University Press, Cambridge, 1973)

[5] A A Penzias and R W Wilson, *Astrophys. J.* **142**, 419 (1965)
[6] J M M Senovilla, *Phys. Rev. Lett.* **64**, 2219 (1990)
[7] N Dadhich, L K Patel and R Tikekar, *Pramana – J. Phys.* **44**, 303 (1995)
[8] J M M Senovilla, *Pramana – J. Phys.* **69(1)**, 31 (2007)
[9] A K Raychaudhuri, *Phys. Rev. Lett.* **80**, 654 (1998)
[10] A K Raychaudhuri, A fresh look at the singularity problem, in: *The Universe* edited by N Dadhich and A K Kembhavi (Kluwer, Amsterdam, 2000)
[11] A K Raychaudhuri, *Mod. Phys. Lett.* **A15**, 319 (2000)
[12] E Ruiz and J M M Senovilla, *Phys. Rev.* **D45**, 1995 (1992)
[13] A K Raychaudhuri, *Gen. Relativ. Gravit.* **36**, 343 (2004)
[14] L Fernandez-Jambrina and L M Gonzalez-Romero, *Phys. Rev.* **D66**, 024027 (2002)
[15] N Dadhich, *J. Astrophys. Astron.* **18**, 343 (1997)
[16] N Dadhich and A K Raychaudhuri, *Mod. Phys. Lett.* **A14**, 2135 (1999)
[17] A Ashtekar, *Phys. Rev. Lett.* **57**, 2244 (1986)
[18] A Ashtekar, *Gravity, geometry and the quantum*, gr-qc/0605011
[19] M Bojowald, *Loop, Quantum Cosmology, Living Rev. Rel.* **8**, 11 (2005), gr-qc/0601085
[20] A Ashtekar, T Pawlowski and P Singh, *Phys. Rev. Lett.* **96**, 141301 (2006); *Phys. Rev.* **D73**, 124038 (2006); **D74**, 084003 (2006)
[21] A Ashtekar and M Bojowald, *Class. Quant. Grav.* **22**, 3349 (2005); **23**, 391 (2006)
[22] G Date and G M Hossain, *Phys. Rev. Lett.* **94**, 011302 (2005)
[23] R Goswami, P S Joshi and P Singh, *Phys. Rev. Lett.* **96**, 031302 (2006)
[24] S Tsujikawa, P Singh and R Maartens, *Class. Quant. Grav.* **21**, 5767 (2004)
[25] T Padmanabhan, in *Highlights in gravitation and cosmology* edited by B R Iyer, A Kembhavi, J V Narlikar and C V Vishveshwara (Cambridge University Press, Cambridge, 1988), p. 156
[26] A Ashtekar, *Nature Phys.* **2**, 725 (2006)
[27] C Rovelli, *Phys. Rev. Lett.* **97**, 151301 (2006)
[28] E Bianchi, L Modesto, C Rovelli and S Speziale, *Class. Quant. Grav.* **23**, 6989 (2006)
[29] E Livine and S Speziale, *J. High Energy Phys.* **0611**, 092 (2006)
[30] S O Bilson-Thompson, F Markopoulou and L Smolin, *Quantum gravity and the standard model*, hep-th/0603022

PRAMANA
— journal of
physics

Vol. 69, No. 1
July 2007
pp. 31–47

A singularity theorem based on spatial averages*

J M M SENOVILLA†
Física Teórica, Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain

Abstract. Inspired by Raychaudhuri's work, and using the equation named after him as a basic ingredient, a new singularity theorem is proved. Open non-rotating Universes, expanding everywhere with a non-vanishing spatial average of the matter variables, show severe geodesic incompleteness in the past. Another way of stating the result is that, under the same conditions, any singularity-free model must have a vanishing spatial average of the energy density (and other physical variables). This is very satisfactory and provides a clear decisive difference between singular and non-singular cosmologies.

Keywords. Raychaudhuri equation; singularity-free cosmologies; singularity theorems.

PACS Nos **04.20.Cv; 04.20.Dw; 98.80.Jk**

1. Introduction

In this paper I would like to present a result which confirms – at least partially – a 10-year-old conjecture: reasonable non-rotating cosmological models can be non-singular only if they are open and have vanishing spatial averages of the matter and other physical variables. This conjecture arose as a result of the interactions and discussions between Prof. A K Raychaudhuri (AKR from now on) and myself concerning the question of feasibility of singularity-free cosmological solutions

*The first instance of a non-singular cosmological solution, compatible with the energy and causality conditions and the Raychaudhuri equation, was given by Senovilla. This generated an intense interest in AKR on the nature and behaviour of such solutions. His initial proposal of the vanishing of the space-time average of the energy density (and other physical and kinematical quantities) as a condition for their existence got refined through many discussions with Dadhich and Senovilla. It evolved finally to the realization that the vanishing of the spatial average of the energy density is what distinguishes all these non-singular solutions, as stated in this theorem proved by Senovilla. In our opinion, this theorem, to which both Raychaudhuri and Senovilla have contributed, provides a succinct and complete characterization of the nature of all non-singular cosmological models – Editors.

†In memory of Amal Kumar Raychaudhuri (1924–2005).

(see [1,2]). This is a subject to which, as is well-known, AKR made fundamental pioneering contributions [3,4], and I influenced in a much more modest and lateral manner [5–7].

The next section contains a more or less detailed historical review of the antecedents, birth, and hazardous life of the conjecture. However, any reader, not interested in this historical appraisal, who only wishes to see and learn about the results and the new theorem – which are interesting on their own – may go directly to §§3 and 4. The former contains a brief derivation of the celebrated Raychaudhuri equation and some discussion on the definition of spatial averages; the latter presents the new theorem and its implications. A brief discussion and an Appendix have been placed at the end of the paper.

2. History of a conjecture

This is divided into three parts: the appearance of the first non-singular solution and subsequent developments are summarized in §2.1; the importance of the fresh new ideas brought in by AKR, and the formulation of the conjecture, are analyzed in §2.2; finally, the reaction by AKR about the conjecture, the sometimes chaotic discussions in publications-comments-replies, and some recent developments on the subject are reviewed in §2.3.

2.1 *Anadi Vishva*

In 1990, I published [5] which is considered to be the first 'interesting' singularity-free cosmological model. It is a spatially inhomogeneous solution of Einstein's field equations for a perfect fluid with a realistic equation of state $p = \varrho/3$. This solution came as a big surprise: it was widely believed that a model such as the one in [5] should be singular due to the powerful singularity theorems developed by Penrose, Hawking and others [8–10]. Therefore, the solution caused some impact (see e.g. [11]), and some discussions within the relativity community [12,13].

Shortly afterwards, the geodesic completeness of this solution was explicitly proven in [14], along with a list of its main properties. The solution turns out to be *cylindrically symmetric*, globally hyperbolic and everywhere expanding for half the history of the model. It satisfies the stronger energy requirements (the dominant and strict strong energy conditions) and is singularity-free (see the Appendix for a short summary). A detailed analysis of how the model fits in with the general conclusions of the singularity theorems was also performed in [14], and the solution was proven to be in full accordance with the theorems: in all versions of the theorems at least one of their hypotheses was not satisfied, usually the so-called boundary or initial condition (see [7]).

Almost simultaneously, another paper [15] extended the particular model in [5] to a general family of perfect-fluid solutions with an Abelian G_2-symmetry acting on spatial surfaces. This family was obtained under the assumption of separation of variables. It contains a diversity of models, with different properties, in particular a 2-parameter subfamily of geodesically complete, singularity-free, solutions – which include the original solution in [5] – as well as other relevant solutions such as the one in [16] (see also [17]). There were also solutions with time-like singularities, and solutions with no singularities in the matter variables but with time-like singularities in the Weyl conformal part of the curvature [15]. All such behaviour was somehow unexpected. As a matter of fact, one could prove that, within the subfamily of solutions in [15] with a barotropic equation of state and no initial singularity, the singularity-free solutions lie precisely at the boundary separating the solutions with only time-like Weyl singularities from the solutions with time-like singularities in both Weyl and matter variables. This was not very encouraging, since one could suspect the instability and the zero measure of the non-singular models.

Several generalizations of these solutions were later found by allowing the fluid to have heat flow [18] and scalar fields [19]. It was claimed in [19,20] that a particular 'inhomogeneization' procedure (by requiring the separation of variables) of the Robertson–Walker open cosmological models would lead to the non-singular models of [5,15]. This was later pursued in a review published by Dadhich in [21], where he intended to prove the uniqueness of those non-singular models.

Up to this point, all known non-singular solutions were (i) *diagonal*, that is to say, with the existence of a global coordinate chart adapted to the perfect fluid (so-called 'co-moving') such that the metric takes a diagonal form, (ii) *cylindrically symmetric*, and (iii) *separable* in comoving coordinates. The last feature meant that the metric components could be written as the product of a function of the separable time coordinate times a function of the separable radial coordinate. The first example of a non-diagonal non-singular perfect fluid model was presented in [22], though the solution had been previously published in quite a different context in [23]. This was a solution for a cylindrically symmetric stiff fluid (the equation of state is $p = \varrho$) or equivalently, for a massless scalar field. It was also separable but could be almost immediately generalized to a family of non-separable, non-diagonal, non-singular stiff fluid solutions in [24]. Other non-singular solutions followed (see e.g. [7,25,26]).

It was then shown in [6] that the general family in [15] as well as the whole class of Robertson–Walker cosmologies belong to a single unified wider class of cylindrically symmetric [26a], separable, diagonal (non-necessarily perfect) fluid solutions. Those depended on one arbitrary function of time – essentially the scale factor – and four free parameters selecting the openness or closeness of the models, the anisotropy of the fluid pressures, or the anisotropy and spatial inhomogeneity of the models. The physical properties of this general class were analyzed in detail [6,7], in particular the deceleration parameter, leading to natural inflationary models (without violating the strong energy condition), and the generalized Hubble law. The possibility of constructing realistic cosmological models by 'adiabatically' changing the parameters in order to start with a singularity-free model which at later times becomes a Robertson–Walker model was also considered in [6] and in §7.7 in [7].

Some interesting lines of research appeared in print in 1997–8. First, by keeping the cylindrical symmetry, a new diagonal but non-separable family of stiff fluid singularity-free solutions was presented in [27]. The family contained the same static limit as the solution in [22], thereby suggesting that they both form part of a more general class, perhaps of non-zero measure, of non-singular cylindrically symmetric stiff fluid models. Second, the role of shear in expanding perfect fluid models was analyzed in [28] proving that non-singular models with an Abelian spatial G_2-symmetry should be spatially inhomogeneous. This result was much improved and proven in a more general context in [29], showing in particular that the symmetry assumption was superfluous. And third, by giving up cylindrical symmetry, a family of non-singular (non-perfect) fluid solutions with spherical symmetry was presented in [30]. These models depend on one arbitrary function of time. Once again, they can avoid the singularity theorems due to the failure of the boundary/initial condition: there are no closed trapped surfaces. It was also shown in §7.8 of [7] that these models cannot represent a finite star, since this would require a place where the radial pressure vanished, which is impossible for appropriate selections of the arbitrary function of time. This is a property shared by all models mentioned so far in this subsection.

2.2 *Raychaudhuri comes into play: The conjecture*

In December 1995, I attended the International Conference on Gravitation and Cosmology (ICGC-95), held in Pune (India), where I had the chance to meet Prof. Raychaudhuri for the first time. I was impressed by his personality and accessibility, specially for a man of his age and reputation. But more importantly, I was deeply influenced by his remarks in a brief conversations that we – AKR, Naresh Dadhich and myself – had at that time. If I remember well, AKR mentioned averages in these informal conversations, but just by the way. This came at a critical time: in a short talk at the workshop on 'Classical General Relativity' (see [31]). I presented the above-mentioned combined Robertson–Walker plus non-singular general family [6] and its properties. I thought that the paper, which had already been accepted, would open the door for realistic models.

Even though AKR meant spacetime averages, I immediately realized the relevance of his idea, especially to discriminate between singular and non-singular cosmological models, but using purely spatial averages. As remarked at the end of the previous subsection, all known non-singular models were 'cosmological' in the sense that they could not describe a finite star surrounded by a surface of vanishing pressure. However, it can certainly happen that (say) the energy density falls off too quickly at large distances (this certainly occurred in all known singularity-free solutions). Thus one may raise the issue whether or not this will better describe the actual Universe or rather a weakly-localized object such as a very large galaxy. Of course, a good way to distinguish between these two possibilities is to use the spatial average of the energy density. Thus, I was inspired by AKR's remarks and believed that this was the right answer to the existence of non-singular models such as the one in [5]. I incorporated this view to the review [7] (see p. 821).

I met AKR for the second time in Pune again, on the occasion of the 15th International Conference on General Relativity and Gravitation (GR15), held in December 1997. Either at GR15 or in an informal seminar (I cannot exactly recall), I attended a talk where he discussed some of the non-singular models and made some comments about the importance of the averages of the physical quantities such as the energy density, the pressure, or the expansion of the fluid. In 1998, AKR proved [1] that, under some reasonable assumptions, open non-rotating non-singular models must have vanishing spacetime averages of the matter and kinematical variables. Later, it was shown [2,32] explicitly that, in the open Robertson–Walker models (with an initial singularity), the same spacetime averages vanish too. Actually, this holds true for most open spatially homogeneous models as well. Since this property is shared by all models, it cannot be used to decide between singularity-free and singular spacetimes.

In my comment, I stressed the following fact. I came to understand, after listening to AKR, that pure spatial averages (at a given instant of time) vanish in the known non-singular solutions, while they are non-vanishing in open Robertson–Walker models. This, together with a well-known singularity theorem for expanding globally hyperbolic models (Theorem 1 below), enabled the formulation of the following conjecture [2]:

> In every singularity-free, non-rotating, expanding, globally hyperbolic model satisfying the strong energy condition, spatial averages of the matter variables vanish.

This will be made precise and proven in §4 below.

2.3 *History of the conjecture*

My comment [2] was only submitted after electronic correspondence with AKR [32a]. It seems that he was initially skeptical about the use of purely spatial – and not spacetime – averages, for he also replied in [33] to our comments. Nevertheless, in an e-mail dated 9 September 1998, he mentioned that a letter proving the vanishing of spatial averages 'following his earlier method' had already been submitted for publication. This private announcement was followed, shortly thereafter, by (i) another paper by Dadhich and AKR [34] where they proved the existence of oscillatory non-singular models within the non-perfect fluid spherically symmetric family of [30] mentioned above, (ii) general theorems providing sufficient conditions for the geodesic completeness of general cylindrically symmetric spacetimes [35,36] and (iii) some work [37] showing the relevance that the singularity-free solutions might have in the fashionable String Cosmology (see also [38] and references therein).

The letter that AKR mentioned was published in [39]. However, in it global hyperbolicity was assumed without being mentioned (due to the assumption of the existence of global coordinates associated to a hypersurface-orthogonal time-like eigenvector field of the Ricci tensor). Moreover, the openness of some local coordinates was taken for granted. Further, the statement that the spatial average of the divergence of the acceleration associated to the time-like eigenvector field vanishes was not clearly proved. More importantly, the blow-up of the kinematical quantities of this eigenvector field was incorrectly related to the blow up of some Ricci scalar invariants [39a]. All in all, the result in [39], involving spatial averages, was not completely proved.

This led to some interesting works by others [40,41], where the existence of a wide class of singularity-free (geodesically complete) cylindrically symmetric stiff fluid cosmologies was explicitly demonstrated, and many solutions were actually exhibited. In these papers, the family of regular cylindrically symmetric stiff fluids was proven to be very abundant, allowing for arbitrary functions. Furthermore, strong support for the conjecture was also provided in the second of these papers [41]: the vanishing of the energy density (and pressure) of the fluid at spatial infinity on every Cauchy hypersurface was demonstrated to be a necessary requirement if the spacetime was to be geodesically complete. This was quite encouraging, and constituted the first serious advance towards the proof of the conjecture.

AKR may not have been aware of these important developments and results. He put out a preprint [42] to which Fernández-Jambrina [43] found a counter-example. A revised version was published in [44] which was also not fully free of the shortcomings, as neatly pointed out in [45]. AKR acknowledged these deficiencies in [46], yet he believed that some of his results were still valid.

Reference [46] was AKR's last published paper. In my opinion, after having identified the clue to non-singular models (i.e. averages), which led us all to the right track, he tried to prove more ambitious and challenging results which were perhaps beyond the techniques he was using. This, of course, does not in any way diminishes his fundamental contribution to the field of singularities in Cosmology, a subject in which, probably, the most important ideas came from his insight and deep intuition. It is in this sense that the spirit of the theorem to be proven in §4 should be credited to him. I can only hope that he would have welcomed the new results.

3. Spatial averages and the Raychaudhuri equation

3.1 *The Raychaudhuri equation*

As is known, the first result predicting singularities under reasonable physical conditions was published in 1955 – exactly the year of Einstein's demise – by AKR [3]. In this remarkable paper, he presented what is considered to be the first singularity theorem, and included a version (the full equation appeared soon after in [47], see also [4]) of the equation named after him which is the basis of later developments and of all the singularity theorems [7–10]. The Raychaudhuri equation can be easily derived from the general Ricci identity:

$$(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu) u^\alpha = R^\alpha_{\rho\mu\nu} u^\rho \,.$$

Contracting α with μ here, then with u^ν, one gets

$$u^\nu \nabla_\mu \nabla_\nu u^\mu - u^\nu \nabla_\nu \nabla_\mu u^\mu = R_{\rho\nu} u^\rho u^\nu,$$

where $R_{\mu\nu}$ is the Ricci tensor. Reorganizing by parts the first summand on the left-hand side, one derives

$$u^\nu \nabla_\nu \nabla_\mu u^\mu + \nabla_\mu u_\nu \nabla^\nu u^\mu - \nabla_\mu (u^\nu \nabla_\nu u^\mu) + R_{\rho\nu} u^\rho u^\nu = 0 \tag{1}$$

which is the Raychaudhuri equation. AKR's important contribution was to understand and explicitly show the fundamental physical implications of this simple geometrical relation.

Let us analyze some of these implications. Observe that in the case that u^μ defines a (affinely parametrized) geodesic vector field, then $u^\nu \nabla_\nu u^\mu = 0$ and the third term vanishes. The second term can then be rewritten by splitting

$$\nabla_\mu u_\nu = S_{\mu\nu} + A_{\mu\nu}$$

into its symmetric $S_{\mu\nu}$ and antisymmetric $A_{\mu\nu}$ parts, so that

$$\nabla_\mu u_\nu \nabla^\nu u^\mu = S_{\mu\nu} S^{\mu\nu} - A_{\mu\nu} A^{\mu\nu} .$$

Now the point is to realize two things. (i) If u^μ is time-like (and normalized) or null, then both $S_{\mu\nu} S^{\mu\nu}$ and $A_{\mu\nu} A^{\mu\nu}$ are non-negative, (ii) u_μ is also proportional to a gradient (therefore defining orthogonal hypersurfaces) if and only if $A_{\mu\nu} = 0$. In summary, for hypersurface–orthogonal geodesic time-like or null vector fields u^μ, one has

$$u^\nu \nabla_\nu \nabla_\mu u^\mu = -S_{\mu\nu} S^{\mu\nu} - R_{\rho\nu} u^\rho u^\nu ,$$

so that the sign of the derivative of the divergence or expansion $\theta \equiv \nabla_\mu u^\mu$ along the geodesic congruence is governed by the sign of $R_{\rho\nu} u^\rho u^\nu$. If the latter is non-negative, then the former is non-positive. In particular, if the expansion is negative at some point and $R_{\rho\nu} u^\rho u^\nu \geq 0$ then one can prove, by introducing a scale factor L such that $u^\mu \nabla_\mu (\log L) \propto \theta$ and noting that $S^\mu_\mu = \theta$, that necessarily the divergence will reach an infinite negative value in finite affine parameter (unless all the quantities are zero everywhere).

If there are physical particles moving along these geodesics, then clearly a physical singularity is obtained, since the mean volume decreases and the density of particles will be unbounded (see Theorem 5.1 in [7], p. 787). This was the situation treated by AKR for the case of irrotational dust. In general, no singularity is predicted, though, and one only gets a typical caustic along the flow lines of the congruence defined by u^μ. This generic property is usually called the focusing effect on causal geodesics. For this to take place, of course, one needs the condition

$$R_{\rho\nu} u^\rho u^\nu \geq 0 \tag{2}$$

which is a geometric condition and independent of the particular theory. However, in General Relativity, one can relate the Ricci tensor to the energy–momentum tensor $T_{\mu\nu}$ via Einstein's field equations ($8\pi G = c = 1$)

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = T_{\mu\nu} . \tag{3}$$

Here R is the scalar curvature and Λ the cosmological constant. Therefore, condition (2) can be rewritten in terms of physical quantities. This is why sometimes (2), when valid for all time-like u^μ, is called the time-like convergence condition, also the strong energy condition in the case where $\Lambda = 0$ [10]. One should bear in mind, however, that this is a condition on the Ricci tensor (a geometrical object) and therefore will not always hold (see the discussion in §6.2 in [7]).

The focusing effect on causal geodesics predicted by the Raychaudhuri equation was the fundamental ingredient needed to derive the powerful singularity theorems. Nevertheless, as remarked above, this focusing does not lead to singularities on its own in general. As a trivial example, observe that flat spacetime satisfies condition (2) trivially, but there are no singularities: the focusing effect simply leads to focal points or caustics of the geodesic congruences. This is why it took some time to understand the necessity of combining the focusing effect with the theory of existence of geodesics maximizing the interval (which necessarily cannot have focal points or caustics) in order to prove results on geodesic incompleteness, which is a sufficient condition in the accepted definition of a singularity. With the imaginative and fruitful ideas put forward by Penrose in the 1960s, later followed by Hawking, this led to the celebrated singularity theorems (see [7–10]).

As a simple but powerful example, which we shall later need to prove the new theorem, let us present the following standard singularity theorem (Theorem 9.5.1 in [48], Theorem 5.2 in [7]).

Theorem 1. *If there is a Cauchy hypersurface Σ such that the time-like geodesic congruence emanating orthogonal to Σ has an initial expansion $\theta|_\Sigma \geq b > 0$ and condition* (2) *holds along the congruence, then all time-like geodesics are past incomplete.*

The idea of the proof is simple [7,10,48]: since Σ is a Cauchy hypersurface, the spacetime is globally hyperbolic so that one knows that there is a maximal time-like curve from Σ to any point. From standard results any such maximal curve must be a time-like geodesic orthogonal to Σ without any point focal to Σ between Σ and the point. But the Raychaudhuri equation (1) implies that these focal points should exist in the past at a proper time less than or equal to a fixed value (given by $3/\theta|_\Sigma \leq 3/b$). As every causal curve crosses the Cauchy hypersurface Σ, no time-like geodesic can have length greater than $3/b$ back to the past.

3.2 *Spatial averages*

Let Σ be any space-like hypersurface in the spacetime and let $\boldsymbol{\eta}_\Sigma$ be the canonical volume element 3-form on Σ. The average $\langle f \rangle_S$ of any scalar f on a finite portion S of Σ is defined as

$$\langle f \rangle_S \equiv \frac{\int_S f\boldsymbol{\eta}_\Sigma}{\int_S \boldsymbol{\eta}_\Sigma} = [\mathrm{Vol}(S)]^{-1} \int_S f\boldsymbol{\eta}_\Sigma,$$

where $\mathrm{Vol}(S)$ is the volume of $S \subseteq \Sigma$. The spatial average on the whole Σ is defined as (the limit of) the previous expression when S approaches the entire Σ

$$\langle f \rangle_\Sigma \equiv \lim_{S\to\Sigma} \frac{\int_S f\boldsymbol{\eta}_\Sigma}{\int_S \boldsymbol{\eta}_\Sigma} \equiv [\mathrm{Vol}(\Sigma)]^{-1} \int_\Sigma f\boldsymbol{\eta}_\Sigma. \tag{5}$$

Obvious properties of these averages are

1. (linearity) For any $S \subseteq \Sigma$, any functions f, g and any constants a, b:

$$\langle af + bg \rangle_S = a \langle f \rangle_S + b \langle g \rangle_S .$$

2. For any $S \subseteq \Sigma$, $\langle f \rangle_S \leq \langle |f| \rangle_S$.
3. If $S \subseteq \Sigma$ such that its closure is compact (so that its volume is finite, $\mathrm{Vol}(S) < \infty$), then for any $f \geq b \geq 0$ on S, $\langle f \rangle_S \geq b$ and the equality holds only if f is constant, $f = b$, almost everywhere on S. In particular, if $f \geq 0$ on such an S, then $\langle f \rangle_S \geq 0$ and the equality holds only if f vanishes almost everywhere on S.
4. If $S \subseteq \Sigma$ does not have a finite volume (so that it cannot be of compact closure), then for any $f \geq 0$ on S, $\langle f \rangle_S \geq 0$ and the equality requires necessarily that $f \to 0$ when 'approaching the boundary' (i.e., when going to infinity). Conversely, if $f > 0$, f is bounded on S and f is bounded from below by a positive constant at most along a set of directions of zero measure, then $\langle f \rangle_S = 0$.
5. If $|f|$ is bounded on S ($|f| \leq M$), then $\langle f^2 \rangle_S \leq M \langle |f| \rangle_S$. In particular, $\langle |f| \rangle_S = 0$ implies that $\langle f^2 \rangle_S = 0$.
6. Similar results hold, of course, for negative and non-positive functions (just use $-f$).

4. The theorem

Let us consider any space-like hypersurface Σ in the spacetime, and let u^μ be its unit normal vector field (ergo time-like). The projector $h_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu$ defines the canonical first fundamental form of Σ. A classical result relates the intrinsic Riemannian structure of $(\Sigma, h_{\mu\nu})$ with the Lorentzian one of the spacetime, and in particular their respective curvatures. The main result is the Gauss equation, which reads (e.g. [10,48])

$$R_{\alpha\beta\gamma\delta} h^\alpha_\lambda h^\beta_\mu h^\gamma_\nu h^\delta_\tau = \bar{R}_{\lambda\mu\nu\tau} + K_{\lambda\nu} K_{\mu\tau} - K_{\lambda\tau} K_{\mu\nu},$$

where $K_{\mu\nu} = K_{\nu\mu} \equiv h^\beta_\mu h^\gamma_\nu \nabla_\beta u_\gamma$ is the second fundamental form of Σ, and $\bar{R}_{\lambda\mu\nu\tau}$ is the intrinsic curvature tensor of $(\Sigma, h_{\mu\nu})$. Observe that $u^\mu K_{\mu\nu} = 0$ and $u^\mu \bar{R}_{\lambda\mu\nu\tau} = 0$. Contracting all indices here one derives the standard result (e.g. §10.2 in [48])

$$K^2 = K_{\mu\nu} K^{\mu\nu} + 2 R_{\mu\nu} u^\mu u^\nu + R - \bar{R},$$

where $\bar{R}$ is the scalar curvature of Σ and $K \equiv K^\mu_\mu$. Using the Einstein's field equations (3) this can be rewritten in terms of the energy–momentum tensor as

$$K^2 = K_{\mu\nu} K^{\mu\nu} + 2 T_{\mu\nu} u^\mu u^\nu + 2\Lambda - \bar{R}. \tag{5}$$

Recall that $T_{\mu\nu} u^\mu u^\nu$ is the energy density of the matter content relative to the observer u^μ, and thus it is always non-negative. Note, also, that for any extension of u^μ outside Σ as a hypersurface-orthogonal unit time-like vector field (still called

u^μ), their previously defined kinematical quantities $\theta = \nabla_\mu u^\mu$ and $S_{\mu\nu} = \nabla_{(\mu} u_{\nu)}$ are simply

$$\theta|_\Sigma = K, \quad \left(S_{\mu\nu} + a_{(\mu} u_{\nu)}\right)\big|_\Sigma = K_{\mu\nu},$$

where $a^\mu = u^\nu \nabla_\nu u^\mu$ is the acceleration vector field of the extended u^μ.

Combining this with Theorem 1 one immediately deduces a very strong result concerning the average of the energy density:

Proposition 1. *Assume that* (1) *there is a non-compact Cauchy hypersurface* Σ *such that the time-like geodesic congruence emanating orthogonal to* Σ *is expanding and condition* (2) *holds along the congruence,* (2) *the spatial scalar curvature is non-positive on average on* Σ*:* $\langle \bar{R} \rangle_\Sigma \leq 0$, (3) *the cosmological constant is non-negative* $\Lambda \geq 0$ *and* (4) *the spacetime is past time-like geodesically complete. Then,*

$$\langle T_{\mu\nu} u^\mu u^\nu \rangle_\Sigma = \langle K_{\mu\nu} K^{\mu\nu} \rangle_\Sigma = \langle \bar{R} \rangle_\Sigma = \Lambda = 0 \,. \tag{6}$$

Remarks

- The first assumption requires spacetime to be globally hyperbolic (so that it is causally well-behaved). Also, the time-like convergence condition has to hold along the geodesic congruence orthogonal to one of the Cauchy hypersurfaces. Furthermore, the Universe is assumed to be non-closed (non-compactness assumption) and expanding everywhere at a given instant of time as described by the hypersurface Σ. All this is standard (Theorem 1).
- The second assumption demands that the space of the Universe (at the expanding instant) be non-positively curved on average. Observe that this still allows for an everywhere positively curved Σ (see the example in the Appendix). This is in accordance with our present knowledge of the Universe and with all indirect observations and measures, e.g. the recent data [49] from WMAP.
- The third assumption is also in accordance with all theoretical and observational data [49]. Observe that the traditional case with $\Lambda = 0$ is included. Notice, however, that the time-like convergence condition (2) is also assumed. This may impose very strict restrictions on the matter variables if $\Lambda > 0$.
- The second and third assumptions could be replaced by milder ones such as $\langle \bar{R} \rangle_\Sigma \leq 2\Lambda$, allowing for all signs in both Λ and $\langle \bar{R} \rangle_\Sigma$. The conclusion concerning the vanishing of the averaged energy density would be unaltered, as well as the next one in (6), but the last two equalities in (6) should be replaced by $\langle \bar{R} \rangle_\Sigma = 2\Lambda$.
- Probably, the fourth condition may be relaxed substantially, as one just needs that the set of past-incomplete geodesics be not too big. There are some technical difficulties, however, to find a precise formulation of the mildest acceptable condition.
- The conclusion forces the second and third assumptions to hold in the extreme cases and, much more importantly, it implies that the energy density of the matter on Σ has a vanishing spatial average. This conclusion was the main goal in this paper.

Proof. From Theorem 1 and the fourth hypothesis it follows that $0 < \theta|_\Sigma = K$ cannot be bounded from below by a positive constant. Furthermore, the existence of a complete maximal time-like curve (which must be a geodesic without focal points) from the Cauchy hypersurface Σ to any point to the past implies that, actually, K can be bounded from below away from zero only along a set of directions of zero measure. Point 4 of the list of properties for averages implies that $\langle\theta\rangle_\Sigma = \langle K\rangle_\Sigma = 0$, and point 5 in the same list provides then $\langle K^2\rangle_\Sigma = 0$. Taking averages on formula (5) and using point 1 in that list one arrives at

$$\langle K_{\mu\nu}K^{\mu\nu}\rangle_\Sigma + 2\langle T_{\mu\nu}u^\mu u^\nu\rangle_\Sigma + 2\Lambda - \langle\bar{R}\rangle_\Sigma = 0.$$

Then, given that all the summands here are non-negative, the result follows.

This result can be made much stronger by using, once again, the Raychaudhuri equation. To that end, we need a lemma first.

Lemma 1. *If the energy–momentum tensor satisfies the dominant energy condition and $\langle T_{\mu\nu}u^\mu u^\nu\rangle_\Sigma = 0$ for some unit time-like vector field u^μ, then all the components of $T_{\mu\nu}$ in any orthonormal basis $\{e^\mu_\alpha\}$ have vanishing average on Σ*

$$\langle T_{\mu\nu}e^\mu_\alpha e^\nu_\beta\rangle_\Sigma = 0, \quad \forall\alpha,\beta = 0,1,2,3. \tag{7}$$

Proof. The dominant energy condition implies that [10,50], in any orthonormal basis $\{e^\mu_\alpha\}$ (where e^μ_0 is the time-like leg)

$$T_{\mu\nu}e^\mu_0 e^\nu_0 \geq \left|T_{\mu\nu}e^\mu_\alpha e^\nu_\beta\right|, \quad \forall\alpha,\beta = 0,1,2,3,$$

so that, by taking any orthonormal basis with $u^\mu = e^\mu_0$, points 2 and 6 in the list of properties of spatial averages lead to eq. (7) in those bases. As any other orthonormal basis is obtained from the selected one by means of a Lorentz transformation – so that the components of $T_{\mu\nu}$ in the new basis are linear combinations, with bounded coefficients, of the original ones – the result follows.

The combination of this lemma with Proposition 1 leads to the following result.

Proposition 2. *Under the same assumptions as in Proposition 1, if $T_{\mu\nu}$ satisfies the dominant energy condition then not only the averages shown in eqs* (6) *and* (7) *vanish, but furthermore*

$$\langle u^\mu\nabla_\mu\theta - \nabla_\mu a^\mu\rangle_\Sigma = \langle v^\mu\nabla_\mu\vartheta\rangle_\Sigma = \langle R_{\mu\nu}e^\mu_\alpha e^\nu_\beta\rangle_\Sigma = 0, \tag{8}$$

where v^μ is the unit time-like geodesic vector field orthogonal to Σ, $\vartheta = \nabla_\mu v^\mu$ its expansion, u^μ is any hypersurface-orthogonal unit time-like vector field orthogonal to Σ, and $a^\mu = u^\nu\nabla_\nu u^\mu$ its acceleration.

Remarks

- The hypersurface-orthogonal vector field u^μ may represent the mean motion of the matter content of the Universe. Observe that, thereby, an acceleration of the cosmological fluid is permitted. This is important, since such an acceleration is related to the existence of pressure gradients, and these forces oppose gravitational attraction.

- The expression $u^\mu\nabla_\mu\theta$ can be seen as a 'time derivative' of the expansion θ: a derivative on the transversal direction to Σ. In particular, $v^\mu\nabla_\mu\vartheta$ is the time derivative, with respect to proper time, of the expansion for the geodesic congruence orthogonal to Σ. This proper time derivative of ϑ does have a vanishing average on Σ. Notice, however, that the generic time derivative of the expansion θ does not have a vanishing average in general: this is governed by the average of the divergence of the acceleration.
- Evidently, $a^\mu u_\mu = 0$ so that a^μ is space-like and tangent to Σ on Σ. Furthermore, one can write

$$\nabla_\mu a^\mu = (h^{\mu\nu} - u^\mu u^\nu)\nabla_\mu a_\nu = h^{\mu\nu}\nabla_\mu a_\nu + a_\mu a^\mu.$$

Letting $\vec{a}$ represent the spatial vector field $a^\mu|_\Sigma$ on Σ, this implies

$$\nabla_\mu a^\mu|_\Sigma = \mathrm{div}_\Sigma\vec{a} + \vec{a}\cdot\vec{a},$$

where div_Σ stands for the three-dimensional divergence within $(\Sigma, h_{\mu\nu})$ and $\cdot$ is its internal positive-definite scalar product; hence $\vec{a}\cdot\vec{a} \geq 0$ and this vanishes only if $\vec{a} = \vec{0}$. Note that the average $\langle \mathrm{div}_\Sigma\vec{a}\rangle_\Sigma$ will vanish for any reasonable behaviour of $a^\mu|_\Sigma$, because the integral in the numerator leads via Gauss theorem to a boundary (surface) integral 'at infinity', which will always be either finite or with a lower-order divergence than vol(Σ). Therefore, by taking averages of the previous expression one deduces

$$\langle\nabla_\mu a^\mu\rangle_\Sigma = \langle\vec{a}\cdot\vec{a}\rangle_\Sigma = \langle a_\nu a^\nu\rangle_\Sigma \geq 0.$$

In other words, the first conclusion in (8) can be rewritten in a more interesting way as $\langle u^\mu\nabla_\mu\theta - a_\nu a^\nu\rangle_\Sigma = 0$, or equivalently

$$\langle u^\mu\nabla_\mu\theta\rangle_\Sigma = \langle a_\nu a^\nu\rangle_\Sigma \geq 0.$$

Observe that, if these averages do not vanish, this implies in particular that there must be regions on Σ where $u^\mu\nabla_\mu\theta$ is positive.

Proof. From Proposition 1 it follows that necessarily $\Lambda = 0$. Hence in any orthonormal basis $\{e^\mu_\alpha\}$ eq. (3) implies

$$R_{\mu\nu}e^\mu_\alpha e^\nu_\beta = T_{\mu\nu}e^\mu_\alpha e^\nu_\beta - \frac{1}{2}T^\rho_\rho\eta_{\alpha\beta},$$

where $(\eta_{\alpha\beta})$ =diag$(-1,1,1,1)$. By taking averages on Σ here and using Lemma 1, one deduces

$$\langle R_{\mu\nu}e^\mu_\alpha e^\nu_\beta\rangle_\Sigma = 0, \quad \forall\alpha,\beta = 0,1,2,3$$

which are the last expressions in (8).

The Raychaudhuri equations (1) for u^μ and v^μ are respectively

$$u^\nu\nabla_\nu\theta + \nabla_\mu u_\nu\nabla^\nu u^\mu - \nabla_\mu a^\mu + R_{\mu\nu}u^\mu u^\nu = 0, \tag{9}$$

$$v^\nu\nabla_\nu\vartheta + \nabla_\mu v_\nu\nabla^\nu v^\mu + R_{\mu\nu}v^\mu v^\nu = 0. \tag{10}$$

Obviously $u^\mu|_\Sigma = v^\mu|_\Sigma$ and it is elementary to check that $\theta|_\Sigma = \vartheta|_\Sigma$ and

$$\nabla_\mu u_\nu \nabla^\nu u^\mu|_\Sigma = \nabla_\mu v_\nu \nabla^\nu v^\mu|_\Sigma = K_{\mu\nu}K^{\mu\nu}.$$

From (9) and (10) one thus gets

$$(u^\nu \nabla_\nu \theta - \nabla_\mu a^\mu)|_\Sigma = v^\nu \nabla_\nu \vartheta|_\Sigma = -K_{\mu\nu}K^{\mu\nu} - R_{\mu\nu}u^\mu u^\nu|_\Sigma \leq 0$$

so that $u^\nu \nabla_\nu \theta - \nabla_\mu a^\mu$ is everywhere non-positive on Σ. Taking averages on Σ, using the second equation in (6) and the previous result on the averages of the Ricci tensor components one finally gets

$$\langle u^\nu \nabla_\nu \theta - \nabla_\mu a^\mu \rangle|_\Sigma = \langle v^\nu \nabla_\nu \vartheta \rangle|_\Sigma = 0.$$

That ends the proof.

The main theorem in this paper is now an immediate corollary of the previous propositions.

Theorem 2. *Assume that* (1) *there is a non-compact Cauchy hypersurface* Σ *such that the time-like geodesic congruence emanating orthogonal to* Σ *is expanding and the time-like convergence condition* (2) *holds along the congruence,* (2) *the spatial scalar curvature is non-positive on average on* Σ*:* $\langle \bar{R} \rangle_\Sigma \leq 0$, (3) *the cosmological constant is non-negative* $\Lambda \geq 0$, (4) *the energy–momentum tensor satisfies the dominant energy condition.*

If any single one of the following spatial averages

$$\Lambda,\ \langle\theta\rangle_\Sigma,\ \langle\vartheta\rangle_\Sigma,\ \langle\theta^2\rangle_\Sigma,\ \langle\vartheta^2\rangle_\Sigma,\ \langle K_{\mu\nu}K^{\mu\nu}\rangle_\Sigma,\ \langle\bar{R}\rangle_\Sigma,\ \langle v^\mu \nabla_\mu \vartheta\rangle_\Sigma,$$
$$\langle u^\mu \nabla_\mu \theta - \nabla_\mu a^\mu\rangle_\Sigma,\ \langle T_{\mu\nu} e^\mu_\alpha e^\nu_\beta\rangle_\Sigma,\ \langle R_{\mu\nu} e^\mu_\alpha e^\nu_\beta\rangle_\Sigma \quad \forall \alpha, \beta = 0, 1, 2, 3$$

does not vanish, then the spacetime is past time-like geodesically incomplete.

As before, notice that the first average in the second row can be replaced by $\langle u^\mu \nabla_\mu \theta - a_\mu a^\mu \rangle_\Sigma$. Therefore, another sufficient condition is that $u^\mu \nabla_\mu \theta$ is non-positive everywhere but its average be non-zero.

5. Conclusions

Observe the following implication of the theorem. Under the stated hypotheses, a non-vanishing average of any component of the energy–momentum tensor (or of the Ricci tensor) – such as the energy density, the different pressures, the heat flux, etc. – leads to the existence of past singularities. It is quite remarkable that one does not need to assume any specific type of matter content (such as a perfect fluid, scalar field, ...); only the physically compelling and well-established dominant energy condition is required. The theorem is valid for 'open' models, as the Cauchy hypersurface is required to be non-compact. This is, however, no real restriction, because for closed models there are stronger results [7–10,48]. As a matter of fact, closed expanding non-singular models require the violation of the strong energy condition [7]. There is no hope of physically acceptable closed non-singular models satisfying the time-like convergence condition (2) (without this condition, there are some examples [7,26]).

Let me stress that the conclusion in Theorem 2 is quite strong: it tells us that the incompleteness is to the past. Besides, I believe that one can in fact prove a stronger theorem such that the geodesic time-like incompleteness is universal to the past.

The main implication of the above results is this: a clear, decisive, difference between singular and regular (globally hyperbolic) expanding cosmological models is that the latter must have a vanishing spatial average of the matter variables. Somehow, one could then say that the regular models are not cosmological, if we believe that the Universe is described by a more or less not too inhomogeneous distribution of matter. This is, on the whole, a very satisfactory result.

Appendix: The singularity-free model of [5]

In cylindrical coordinates $\{t, \rho, \varphi, z\}$ the line-element reads

$$\begin{aligned} ds^2 = &\cosh^4(at)\cosh^2(3a\rho)(-dt^2 + d\rho^2) \\ &+\frac{1}{9a^2}\cosh^4(at)\cosh^{-2/3}(3a\rho)\sinh^2(3a\rho)d\varphi^2 \\ &+\cosh^{-2}(at)\cosh^{-2/3}(3a\rho)dz^2, \end{aligned} \tag{11}$$

where $a > 0$ is a constant. This is a cylindrically symmetric (the axis is defined by $\rho \to 0$) solution of the Einstein's field equations (3) (with $\Lambda = 0$ for simplicity) for an energy–momentum tensor describing a perfect fluid: $T_{\mu\nu} = \varrho u_\mu u_\nu + p\,(g_{\mu\nu} + u_\mu u_\nu)$. Here ϱ is the energy density of the fluid given by

$$\varrho = 15a^2\cosh^{-4}(at)\cosh^{-4}(3a\rho),$$

and

$$u_\mu = \left(-\cosh^2(at)\cosh(3a\rho), 0, 0, 0\right)$$

defines the unit velocity vector field of the fluid. Observe that u^μ is not geodesic (except at the axis), the acceleration field being

$$a_\mu = (0, 3a\tanh(3a\rho), 0, 0).$$

The fluid has a realistic barotropic equation of state relating its isotropic pressure p to ϱ by

$$p = \frac{1}{3}\varrho.$$

This is the canonical equation of state for radiation-dominated matter and is usually assumed to hold at early stages of the Universe. Note that the energy density and the pressure are regular everywhere, and one can in fact prove that the spacetime (11) is completely free of singularities and geodesically complete [14]. For complete discussions on this spacetime, see [14] and §7.6 in [7]. One can nevertheless see that the focusing effect on geodesics takes place fully in this spacetime. This does not lead to any problem with the existence of maximal geodesics between any pair of chronologically related points (see the discussion in [7], pp. 829–830).

Spacetime (11) satisfies the strongest causality condition: it is globally hyperbolic, any t =const. slice is a Cauchy hypersurface. All typical energy conditions, such as the dominant or the (strictly) strong ones (implying in particular condition (2) with the strict inequality) also hold everywhere. The fluid expansion is given by

$$\theta = \nabla_\mu u^\mu = 3a \frac{\sinh(at)}{\cosh^3(at)\cosh(3a\rho)}. \tag{12}$$

Thus this Universe is contracting for half of its history ($t < 0$) and expanding for the second half ($t > 0$), having a rebound at $t = 0$ which is driven by the spatial gradient of pressure, or equivalently, by the acceleration a_μ. Observe that the entire Universe is expanding (that is, $\theta > 0$) everywhere if $t > 0$. Note that this is one of the assumptions in Propositions 1, 2 and Theorem 2. It is however obvious that, for any Cauchy hypersurface Σ_T given by $t = T$ =constant, the average $\langle\theta\rangle_{\Sigma_T} = 0$. As one can check for the explicit expression (12), θ is strictly positive everywhere but not bounded from below by a positive constant because $\lim_{\rho\to\infty}\theta = 0$. Observe, however, that for finite ρ one has $\lim_{z\to\infty}\theta > 0$ and finite.

Similarly, one can check that the scalar curvature of each Σ_T is given by

$$\bar{R} = 30a^2\cosh^{-4}(aT)\cosh^{-4}(3a\rho) > 0$$

which is positive everywhere. However, $\langle\bar{R}\rangle_{\Sigma_T} = 0$, and analogously $\langle\varrho\rangle_{\Sigma_T} = \langle p\rangle_{\Sigma_T} = 0$. Observe also that

$$a_\mu a^\mu = 9a^2 \frac{\sinh^2(3a\rho)}{\cosh^4(at)\cosh^4(3a\rho)}$$

and thus $\langle a_\mu a^\mu\rangle_{\Sigma_T} = 0$. This implies that in this case $\langle u^\mu\nabla_\mu\theta\rangle_{\Sigma_T} = 0$. The sign of $u^\mu\nabla_\mu\theta - \nabla_\mu a^\mu$ is negative everywhere, as can be easily checked:

$$u^\mu\nabla_\mu\theta = 3a^2 \frac{1 - 3\tanh^2(at)}{\cosh^4(at)\cosh^2(3a\rho)},$$

$$\nabla_\mu a^\mu = 3a^2 \frac{\cosh^2(3a\rho) + 5}{\cosh^4(at)\cosh^4(3a\rho)}.$$

All these are in agreement with, and illustrate, Theorem 2, Propositions 1 and 2, and their corresponding remarks.

This simple model shows that there exist well-founded, well-behaved classical models which expand everywhere, satisfying all energy and causality conditions, and are singularity-free. However, as we have just seen, the model is somehow not 'cosmological' to the extent that the above-mentioned spatial averages vanish.

Acknowledgments

I am thankful to Naresh Dadhich and Leonardo Fernández-Jambrina for many interesting conversations over the years. I am indebted to Matthias Blau and Harald Jerjen for pointing out a flaw in the proof of a previous version of the theorem. Financial support under grants FIS2004-01626 and no. 9/UPV 00172.310-14456/2002 is acknowledged.

References

[1] A K Raychaudhuri, *Phys. Rev. Lett.* **80**, 654 (1998)
[2] J M M Senovilla, *Phys. Rev. Lett.* **81**, 5032 (1998)
[3] A K Raychaudhuri, *Phys. Rev.* **98**, 1123 (1955)
[4] A K Raychaudhuri, *Phys. Rev.* **106**, 172 (1957)
[5] J M M Senovilla, *Phys. Rev. Lett.* **64**, 2219 (1990)
[6] J M M Senovilla, *Phys. Rev.* **D53**, 1799 (1996)
[7] J M M Senovilla, *Gen. Relativ. Gravit.* **30**, 701 (1998)
[8] R Penrose, *Phys. Rev. Lett.* **14**, 57 (1965)
[9] S W Hawking and R Penrose, *Proc. R. Soc. London* **A314**, 529 (1970)
[10] S W Hawking and G F R Ellis, *The large scale structure of spacetime* (Cambridge University Press, Cambridge, 1973)
[11] J Maddox, *Nature (London)* **345**, 201 (1990)
[12] P Joshi, *Phys. Rev. Lett.* **67**, 2109 (1991)
[13] J M M Senovilla, *Phys. Rev. Lett.* **67**, 2110 (1991)
[14] F J Chinea, L Fernández-Jambrina and J M M Senovilla, *Phys. Rev.* **D45**, 481 (1992)
[15] E Ruiz and J M M Senovilla, *Phys. Rev.* **D45**, 1995 (1992)
[16] A Feinstein and J M M Senovilla, *Class. Quant. Grav.* **6**, L89 (1989)
[17] N van den Bergh and J E F Skea, *Class. Quant. Grav.* **9**, 527 (1992)
[18] L K Patel and N K Dadhich, *Class. Quant. Grav.* **10**, L85 (1993)
[19] N K Dadhich, L K Patel and R Tikekar, *Pramana – J. Phys.* **44**, 303 (1995)
[20] N K Dadhich, R Tikekar and L K Patel, *Curr. Sci.* **65**, 694 (1993)
[21] N K Dadhich, in *Inhomogeneous cosmological models* edited by A Molina and J M M Senovilla (World Scientific, Singapore, 1995) pp. 63–74
[22] M Mars, *Phys. Rev.* **D51**, R3989 (1995)
[23] P S Letelier, *J. Math. Phys.* **20**, 2078 (1979)
[24] J B Griffiths and J Bicák, *Class. Quant. Grav.* **12**, L81 (1995)
[25] M Mars, *Class. Quant. Gravit.* **12**, 2831 (1995)
[26] M Mars and J M M Senovilla, *Class. Quant. Grav.* **14**, 205 (1997)
[26a] Recall that all Robertson–Walker geometries are cylindrically symmetric
[27] L Fernández-Jambrina, *Class. Quant. Grav.* **14**, 3407 (1997)
[28] N K Dadhich and L K Patel, *Gen. Relativ. Gravit.* **29**, 179 (1997)
[29] A Barnes, *Gen. Relativ. Gravit.* **30**, 515 (1998)
[30] N K Dadhich, *J. Astrophys. Astron.* **18**, 343 (1997)
[31] N K Dadhich and M A H MacCallum, Report on the Workshop on Classical General Relativity, in *Gravitation and Cosmology (Proceedings of ICGC-95)* edited by S Dhurandhar and T Padmanabhan (Kluwer Academic, Dordrecht, 1997) pp. 279–296
[32] A Saa, *Phys. Rev. Lett.* **81**, 5031 (1998)
[32a] I thank Dr Narayan Banerjee for helping with AKR's e-mails

[33] A K Raychaudhuri, *Phys. Rev. Lett.* **81**, 5033 (1998)
[34] N K Dadhich and A K Raychaudhuri, *Mod. Phys. Lett.* **A14**, 2135 (1999)
[35] L Fernández-Jambrina and L M González-Romero, *Class. Quant. Grav.* **16**, 953 (1999)
[36] L Fernández-Jambrina, *J. Math. Phys.* **40**, 4028 (1999)
[37] M Giovannini, *Phys. Rev.* **D59**, 083511 (1999)
[38] M Gasperini and G Veneziano, *Phys. Rep.* **373**, 1 (2003)
[39] A K Raychaudhuri, *Mod. Phys. Lett.* **A15**, 391 (2000)
[39a] The blowing-up of the expansion of the Ricci time-like eigendirection can take place due to many different reasons: a caustic, the failure of hypersurface orthogonality or of local coordinates, the failure of the vector field to be a Ricci eigendirection, or a change in the algebraic type of the Ricci tensor, among others. There are well-known examples of hypersurface-orthogonal time-like congruences showing this behaviour even in flat spacetime
[40] L Fernández-Jambrina and L M González-Romero, *Phys. Rev.* **D66**, 024027 (2002)
[41] L Fernández-Jambrina and L M González-Romero, *J. Math. Phys.* **45**, 2113 (2004)
[42] A K Raychaudhuri, http://xxx.lanl.gov/abs/gr-qc/0202056
[43] L Fernández-Jambrina, http://xxx.lanl.gov/abs/gr-qc/0202088
[44] A K Raychaudhuri, *Gen. Relativ. Gravit.* **36**, 343 (2004)
[45] L Fernández-Jambrina, *Gen. Relativ. Gravit.* **37**, 421 (2005)
[46] A K Raychaudhuri, *Gen. Relativ. Gravit.* **37**, 425 (2005)
[47] A Komar, *Phys. Rev.* **104**, 544 (1956)
[48] R M Wald, *General relativity* (The University of Chicago Press, Chicago, 1984)
[49] see http://lambda.gsfc.nasa.gov/product/map/dr2/parameters.cfm
[50] J M M Senovilla, *Class. Quant. Grav.* **17**, 2799 (2000)
[51] R Penrose, *Techniques of differential topology in relativity*, Regional Conference Series in Applied Math. **7** (SIAM, Philadelphia, 1972)
[52] J M M Senovilla, Singularity-free Spacetimes, in *Rotating objects and relativistic physics* edited by F J Chinea and L M González-Romero, *Lect. Notes in Phys.* **423** (Springer, Berlin, 1993), pp. 185–193
[53] F J Tipler, C J S Clarke and G F R Ellis, in *General relativity and gravitation: One hundred years after the birth of Albert Einstein* edited by A Held (Plenum Press, New York, 1980)

PRAMANA — journal of physics © Indian Academy of Sciences Vol. 69, No. 1 July 2007 pp. 49–76

The Raychaudhuri equations: A brief review

SAYAN KAR[1] and SOUMITRA SENGUPTA[2]
[1]Department of Physics and Centre for Theoretical Studies, Indian Institute of Technology, Kharagpur 721 302, India
[2]Department of Theoretical Physics, Indian Association for the Cultivation of Science, Jadavpur, Kolkata 700 032, India
E-mail: sayan@phy.iitkgp.ernet.in; sayan@cts.iitkgp.ernet.in; tpssg@mahendra.iacs.res.in

Abstract. We present a brief review on the Raychaudhuri equations. Beginning with a summary of the essential features of the original article by Raychaudhuri and subsequent work of numerous authors, we move on to a discussion of the equations in the context of alternate non-Riemannian spacetimes as well as other theories of gravity, with a special mention on the equations in spacetimes with torsion (Einstein–Cartan–Sciama–Kibble theory). Finally, we give an overview of some recent applications of these equations in general relativity, quantum field theory, string theory and the theory of relativisitic membranes. We conclude with a summary and provide our own perspectives on directions of future research.

Keywords. Raychaudhuri equations; general relativity; alternate theories of gravity.

PACS Nos 04.20.-q; 04.50.+h; 11.25.-w

1. The beginnings

About half a century ago, general relativity (GR) was young (just forty years old!), and even the understanding of the simplest solution, the Schwarzschild, was incomplete. Cosmology was virtually in its infancy, despite the fact that the Friedmann–Lemaitre–Robertson–Walker (FLRW) solutions had been around for quite a while. The question about the then-known exact solutions of GR, which worried the serious relativist quite a bit, concerned their singular nature. Both the Schwarzschild and the cosmological solutions were singular. It is well-known that the creator of GR, Einstein himself, was quite worried about the appearance of singularities in his theory. Was there a way out? Was it correct to believe in a theory which had singular solutions? Were singularities inevitable in GR?

It was during these days in the early 1950s, Raychaudhuri began examining some of these questions in GR. One of his early works during this era involved the construction of a non-static solution of the Einstein equations for a cluster of radially moving particles in an otherwise empty space [1]. A year before, he had also written an article related to condensations in an expanding Universe [2] where he dealt with cosmological perturbations (in a sense, this article deals with what is today known as structure formation). Subsequent to these papers, in 1955, appeared *Relativistic Cosmology I* [3], which contains the derivation of the now-famous Raychaudhuri equation.

Fifty years hence, the Raychaudhuri equations have been discussed and analysed in a variety of contexts. Their rise to prominence was largely due to their use (through the notion of geodesic focusing) in the proofs of the seminal Hawking–Penrose singularity theorems of GR. Today, the importance of this set of equations, as well as their applicability in diverse scenarios, is a well-known fact.

This article is a brief review on these equations. We shall deal with some selected aspects in greater detail. We, of course, would like to emphasize that there are many topics which we leave untouched or, barely touched. We hope to do justice to these in a later, and more extensive article.

The overall plan of this article is as follows. In the remainder of this section, we shall recall the basic ideas and results in the 1955 paper and the singularity theorems. The next section introduces (with illustrative examples) the kinematical quantities (expansion, rotation and shear) which govern the characteristics of geodesic flows and also outlines the derivation and consequences of the equations. Section 3 considers the equations in alternative Riemannian and non-Riemannian theories of gravity ($R+\beta R^2$ theory and the Einstein–Cartan–Sciama–Kibble (ECSK) theory, in particular). In §4, we give a glimpse of the diverse uses of these equations in contexts within, as well as outside the realm of GR. Finally, we present a summary and provide our perspectives on possible future work.

1.1 *The original 1955 paper*

The derivation of the Raychaudhuri equation, presented in the 1955 article, is somewhat different from the way it is presented in standard textbooks today. It must however be mentioned, that in a subsequent paper in 1957 [4], Raychaudhuri presented further results which bear a similarity with the modern approach to the derivation. Let us now briefly summarise the main points of the original derivation of Raychaudhuri.

(a) Raychaudhuri's motivation behind this article is almost entirely restricted to cosmology. He assumes that the Universe is represented by a time-dependent geometry but does *not* assume homogeneity or isotropy at the outset. In fact, one of his aims is to see whether non-zero rotation (spin), anisotropy (shear) and/or a cosmological constant can succeed in avoiding the initial singularity.

(b) The entire analysis is carried out in the comoving frame (in the context of cosmological line elements) – the frame in which the observer is at rest in the fluid.

(c) The quantity R_4^4 (spacetime coordinates in the 1955 paper are labeled as x^1, x^2, x^3, x^4 with the fourth one being time), is evaluated in two ways – once using the Einstein equations (with a cosmological constant Λ) and, again, using the geometric definition of R_4^4 in terms of the metric and its derivatives. In the second way of writing this quantity, Raychaudhuri introduces the definitions of shear and rotation.

(d) Finally, equating the two ways of writing R^4_4 the equation for the evolution of the expansion rate is obtained. Note that the definition of expansion given in this paper refers to the special case of a cosmological metric.
(e) Apart from obtaining the equation, the article also arrives at the focusing theorem (though it is not mentioned with this name) and some additional results (the last section).

In the same year (1955), Heckmann and Schucking [5], while dealing with Newtonian cosmology arrived at a set of equations, one of which is the Raychaudhuri equation (in the Newtonian case). Prompted by this work, Raychaudhuri re-derived his equations in a somewhat different way in an article where he also showed that Heckmann and Schucking's work for the Newtonian case could be generalised without any problems to the fully relativistic scenario. It must also be noted that Komar [6], a year after Raychaudhuri's article appeared, obtained conclusions similar to what is presented in Raychaudhuri's article. Raychaudhuri pointed this out in a letter published in 1957 [7].

Subsequently, in 1961, Jordan *et al* wrote an extensive article on the relativistic mechanics of continuous media where the derivation of the evolution equations of shear and rotation seem to appear for the first time [8]. Furthermore, for null geodesic flows, the kinematical quantities: expansion, rotation and shear (related to the so-called optical scalars) and the corresponding Raychaudhuri equations, were first introduced by Sachs [9].

The Raychaudhuri equation is sometimes referred to as the Landau–Raychaudhuri equation. It may be worthwhile to point out precisely, the work of Landau, in relation to this equation. Landau's contribution appears in his treatise *The Classical Theory of Fields* [10] and is also discussed in detail in [6,11]. Working in the synchronous (comoving) reference frame, Landau defines a quantity $\chi_{\alpha\beta} = \partial\gamma_{\alpha\beta}/\partial t$, where $\gamma_{\alpha\beta}$ is the 3-metric. Subsequently, using the fact that $\chi^\alpha_\alpha = (\partial/\partial t)\gamma$, where γ is the determinant of the 3-metric, he writes down an expression for R^0_0 and then, an inequality $(\partial/\partial t)\chi^\alpha_\alpha + \frac{1}{6}\left(\chi^\alpha_\alpha\right)^2 \geq 0$. While deriving the inequality, Landau implicitly assumes the strong energy condition (though it is not mentioned with this name). Then, using it, he is able to show that γ must necessarily go to zero within a finite time. However, he mentions quite clearly that this does not imply the existence of a physical singularity in the sense of curvature. Though Landau's work captures the essence of focusing, he does not explicitly mention geodesic focusing. Moreover, he does not introduce shear and rotation or write down the complete equation for the expansion.

Even though it was mentioned in [11] and [8], Raychaudhuri's contribution found its true recognition only after the seminal work of Hawking and Penrose which appeared a decade later. It was at that time, along with the proofs of the singularity theorems, the term Raychaudhuri equations came into existence in the physics literature.

1.2 *The singularity theorems*

It must be mentioned that Raychaudhuri did point out the connection of his equations to the existence of singularities in his 1955 article. However, more general results (based on global techniques in Lorentzian spacetimes) appeared in the form of singularity theorems following Penrose's work [12] and, then, Hawking's contributions [13,14]. The crucial element of the singularity theorems is that the existence of singularities is proved by using a minimal set of assumptions (loosely speaking, these are: Lorentz signature metrics and causality, the generic condition on the Riemann tensor components, the existence of trapped surfaces and energy conditions on matter). In fact, a precise definition of 'what is a singularity?' first appeared in the works of Hawking and Penrose. The notion of geodesic incompleteness and its relation to singularities (not necessarily curvature singularities) was also born in their work. One should also realise that the focusing of geodesics arrived at by Raychaudhuri and discussed in much detail in later articles by other authors could be completely benign (irrespective of any actual singularity being present in the manifold). Thus, a singularity would always imply focusing of geodesics but focusing alone cannot imply a singularity (also pointed out by Landau [10]). We refrain from discussing the singularity theorems any further here – excellent discussion on global aspects in gravitation as well as the Hawking–Penrose theorems are available in [15–17].

2. The geometry and physics of the equations

Let us now review the basic ingredients and the derivation of the equations. First, of course, we need to know – what do these equations deal with? In a sentence, one may say that they are concerned about the kinematics of flows. Flows are generated by a vector field – they are the integral curves of the given vector field. These curves may be geodesic or non-geodesic, though the former is more useful in the context of gravity. Thus, a flow is a congruence of such curves – each curve may be time-like or null or, in the Euclidean case, have tangent vectors with a positive definite norm. One does not, in the context of these equations, ask, how the flow is generated. In other words, we are more interested, in deriving the kinematic characteristics of such flows. The evolution equations (along the flow) of the quantitites that characterise the flow in a given background spacetime, are the Raychaudhuri equations. Historically speaking, it is the equation for one of the quantitites (the expansion), which is termed as the Raychaudhuri equation. However, in this article, we will refer to the full set of equations as Raychaudhuri equations.

2.1 *Expansion, rotation, shear*

What quantities characterise a flow? If λ denotes the parameter labeling points on the curves in the flow, then, in order to characterise the flow, we must have different functions of λ. In other words, the gradient of the velocity field being a second rank tensor is split into three parts: the symmetric traceless part, the antisymmetric part and the trace. These define for us the shear, rotation and the expansion of the flow. Specifically,

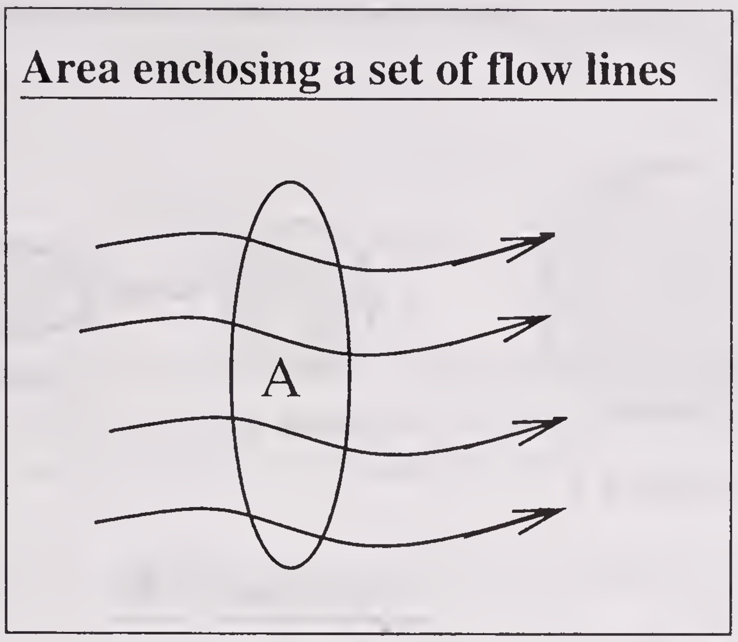


Figure 1. The cross-sectional area enclosing a congruence of geodesics.

$$\nabla_b v_a = \sigma_{ab} + \omega_{ab} + \frac{1}{n-1} h_{ab}\Theta, \tag{1}$$

where the symmetric, traceless part, the shear, is defined as $\sigma_{ab} = \frac{1}{2}(\nabla_b v_a + \nabla_a v_b) - (1/(n-1))h_{ab}\Theta$, the trace, expansion, is $\Theta = \nabla_a v^a$ and the antisymmetric rotation is given as, $\omega_{ab} = \frac{1}{2}(\nabla_b v_a - \nabla_a v_b)$. n is the dimension of spacetime, and $h_{ab} = g_{ab} \pm v_a v_b$ is the projection tensor (the plus sign is for time-like curves whereas the minus one is for space-like ones). Also, correspondingly, $v_a v^a = \mp 1$. Later, we shall discuss the case of null geodesic congruences briefly.

The geometric meaning of these quantities is shown through figures 1 and 2. The expansion, rotation and shear are related to the geometry of the cross-sectional area (enclosing a fixed number of geodesics) orthogonal to the flow lines (figure 1). As one moves from one point to another, along the flow, the shape of this area changes. It still includes the same set of geodesics in the bundle but may be isotropically smaller (or larger), sheared or twisted. The analogy with elastic deformations or fluid flow is, usually, a good visual aid for understanding the change in the geometry of this area. A recent, nice discussion is available in [18]. In refs [19,20] these quantities are explained in quite some detail.

2.2 *Examples*

It is useful to illustrate these quantities with a set of examples. We first choose to work with Schwarzschild spacetime. Our examples here will involve (i) rotation-free time-like geodesic flows and (ii) time-like flows with all three kinematical quantities non-zero. We focus on examples with non-zero shear and rotation because these are not usually available in standard texts on GR. Our choice of examples are primarily based on the problems suggested in a recent monograph by Poisson [18]. In a third example (iii) we discuss briefly a time-like geodesic flow in the FLRW Universe. Finally, in (iv) we briefly deal with geodesic flows in wormhole spacetimes.

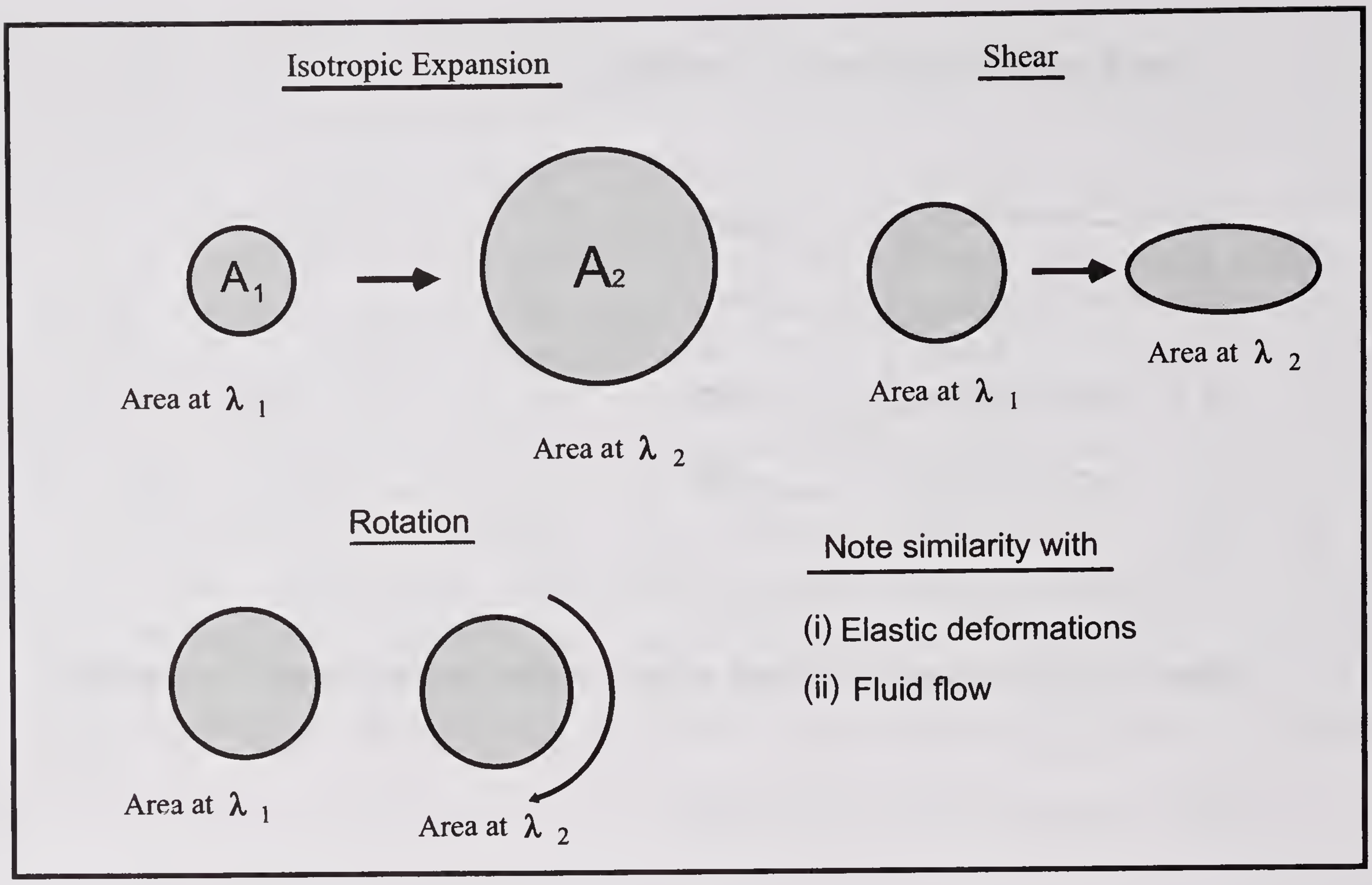


Figure 2. Illustrating expansion, rotation and shear.

(i) From the Frobenius theorem [18], we know that hypersurface orthogonal vector fields must necessarily have zero rotation though the shear can have non-zero components. In Schwarzschild spacetime, we construct a congruence which has the above properties (i.e. it is irrotational but has non-zero shear and expansion). Consider the vector field

$$u^a \partial_a = \frac{1}{1-(2M/r)}\partial_t \pm \sqrt{\frac{2M}{r}}\partial_r. \tag{2}$$

The geodesics corresponding to the above vector field are marginally bound (i.e. $u_t = -1$) and the upper and lower signs refer to outgoing and incoming geodesics. It is easy to show that the above vector field can be written as $u_a = \partial_a \phi$, where $\phi(x^a) =$ constant would represent the hypersurface with respect to which u^a is orthogonal. In fact, ϕ is the same as the new time T (the proper time as measured by a freely falling observer starting from rest at infinity and moving radially inward) used in the Painleve-Gullstrand representation [18] of the Schwarzschild line element.

It is easy to calculate the expansion, which turns out to be

$$\Theta = \pm\frac{3}{2}\sqrt{\frac{2M}{r^3}}. \tag{3}$$

Notice that the expansion is positive for outgoing and negative for incoming geodesics. We can also find the non-zero shear tensor components which are given as

$$\sigma_{tt} = \mp \frac{2M}{r^2}\sqrt{\frac{2M}{r}}; \quad \sigma_{rr} = \mp \frac{\sqrt{2M/r^3}}{(1-(2M/r))^2}, \tag{4}$$

$$\sigma_{tr} = \frac{2M/r^2}{1-(2M/r)}; \quad \sigma_{\theta\theta} = \pm\sqrt{\frac{Mr}{2}} = \frac{1}{\sin^2\theta}\sigma_{\phi\phi}. \tag{5}$$

One can then check that the Raychaudhuri equations (given in the next sub-section) hold with the above expressions for the shear and expansion.
(ii) We now move on to an example where all three kinematical quantities are non-zero.

Consider the following vector field in Schwarzschild spacetime:

$$u^a\partial_a \equiv \frac{1}{\sqrt{1-(3M/r)}}\left(\partial_t + \sqrt{\frac{M}{r^3}}\partial_\theta\right), \tag{6}$$

where M is the usual mass. It can be verified that the geodesics corresponding to this vector field are time-like and they are circular (r is a constant). We intend to calculate the expansion, rotation and shear of the above vector field. The expansion is given as

$$\Theta = \cot\theta\sqrt{\frac{M/r^3}{1-(3M/r)}}. \tag{7}$$

Notice that the expansion is positive in the northern hemisphere and negative in the southern hemisphere.

Following the definition, one can show that the rotation tensor for this vector field is given as

$$\omega_{tr} = \frac{M}{4r^2}\frac{1-(6M/r)}{(1-(3M/r))^{3/2}} = \sqrt{\frac{M}{r^3}}\omega_{r\theta} \tag{8}$$

and the shear tensor is

$$\begin{aligned}
\sigma_{tt} &= \frac{M}{r^3}\sigma_{\theta\theta} = -\frac{M}{2r^3\sin^2\theta}\frac{(1-(2M/r))}{(1-(3M/r))}\sigma_{\phi\phi} \\
&= -\sqrt{\frac{M}{r^3}}\sigma_{t\theta} = \frac{M}{r}\frac{(1-(2M/r))^2}{(1-(3M/r))}\sigma_{rr} \\
&= -\frac{1}{3}\cot\theta\sqrt{\frac{M^3}{r^5}}\frac{(1-(2M/r))}{(1-(3M/r))^{3/2}} \\
\sigma_{tr} &= \sqrt{\frac{M}{r^3}}\sigma_{r\theta} = -\frac{3M}{4r^2}\frac{(1-(2M/r))}{(1-(3M/r))^{3/2}}.
\end{aligned} \tag{9}$$

One can further verify that here too, the Raychaudhuri equations (given below) are satisfied for the above quantities. One can also evaluate $\sigma^2 - \omega^2$ and show that it is positive for $r > 3M$. The expansion is defined for this domain of r ($>3M$) and can diverge to negative infinity (focusing) at $\theta = \pi$ (south pole).

(iii) As a third example, we now quickly discuss the expansion, rotation and shear with respect to the vector field $u^a \partial_a = \partial_t$ in the standard cosmological line element (FLRW). The shear and rotation are identically zero. The expansion is given as

$$\Theta = 3\frac{\dot{a}}{a} = \frac{1}{\sqrt{a^6}}\frac{\mathrm{d}}{\mathrm{d}t}\left(\sqrt{a^6}\right). \tag{10}$$

Note that a^6 is the volume of the expanding 3-space. Hence the term 'volume' expansion is also used in the literature. Raychaudhuri, in his original article, defined the expansion in this way. However, his treatment did include non-zero shear and rotation because he did not assume, to start with, maximally symmetric metrics on spatial slices representing R^3, S^3 or H^3. It may be mentioned here that the equation for the expansion reduces to the equation for $(\ddot{a}/a) = (4\pi G/3)(\rho + 3p)$.

(iv) Our final example concerns the case of geodesic flows in a traversable wormhole [21]. It is known that traversable wormholes require energy-condition-violating matter. These are non-singular spacetimes where the spatial slices resemble two asymptotically flat regions connected by a throat. Thus a geodesic congruence passing through the throat from one asymptotic region to the other would necessarily tend to focus first, not quite reach a focal point, and then defocus. For a typical wormhole, the line element is given as

$$\mathrm{d}s^2 = -\chi^2(l)\mathrm{d}t^2 + \mathrm{d}l^2 + r^2(l)\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right), \tag{11}$$

where, for a wormhole, $\chi(l)$ is non-zero and finite for all l and $r(l = 0) = b_0$ with $r(l \to \pm\infty) \sim l$. It is easy to check that the expansion is proportional to r'/r (the prime denoting a derivative with respect to l). Thus, the expansion never becomes negative infinity and thus there is no focusing. One can work out the expansion for a typical example using $r(l) = \sqrt{b_0^2 + l^2}$ (the so-called Ellis wormhole) [21].

2.3 *The equations and the focusing theorem*

We now turn towards writing down the evolution equations for the expansion, shear and rotation along the flow representing a time-like geodesic congruence. A fact worth mentioning here is that, these evolution equations (and their generalisations) are essentially geometric statements and are independent of any reference to the Einstein field equations.

The modern (textbook) way to derive these equations (see [15]) is as follows. Consider the quantity $v^c\nabla_c B_{ab}$ (where $B_{ab} = \nabla_b v_a$). Evaluate this as an identity and then split it into its trace, antisymmetric and symmetric traceless parts. The equations that emerge are the ones given below (for $n \equiv$ the dimension of spacetime $= 4$).

$$\frac{\mathrm{d}\Theta}{\mathrm{d}\lambda} + \frac{1}{3}\Theta^2 + \sigma^2 - \omega^2 = -R_{ab}v^a v^b, \tag{12}$$

$$\frac{\mathrm{d}\sigma_{ab}}{\mathrm{d}\lambda} = -\frac{2}{3}\Theta\sigma_{ab} - \sigma_{ac}\sigma^c_b - \omega_{ac}\omega^c{}_b + \frac{1}{3}h_{ab}\left(\sigma^2 - \omega^2\right)$$
$$+C_{cbad}v^c v^d + \frac{1}{2}\tilde{R}_{ab}, \tag{13}$$

$$\frac{\mathrm{d}\omega_{ab}}{\mathrm{d}\lambda} = -\frac{2}{3}\Theta\omega_{ab} - 2\sigma^c{}_{[b}\omega_{a]c}, \tag{14}$$

where $\sigma^2 = \sigma_{ab}\sigma^{ab}$, $\omega^2 = \omega_{ab}\omega^{ab}$, C_{cbad} is the Weyl tensor and the quantity $\tilde{R}_{ab} = h_{ac}h_{bd}R^{cd} - \frac{1}{3}h_{ab}h_{cd}R^{cd}$.

There are a few points to note here. Firstly, one must realise that these are *not* equations but, essentially, identitites. Hence, in some references [22–24] we find the usage Raychaudhuri identity or Codazzi–Raychaudhuri identity (in the context of surface congruences to be discussed later) which is, indeed, rigorously correct. The identities, however become equations once we use the Einstein equations or any other geometric property (e.g. Einstein space, or vacuum, etc.) as an extra input. However, we shall continue to use the term equations in this article.

Furthermore, the equations are coupled, nonlinear and first order. The equation for the expansion is of central interest (in the context of the singularity theorems) and it is rather straightforward to analyse. In mathematical parlance, it (the equation for the expansion) is known as a Riccati equation. Such equations can be transformed into a second-order linear form (more precisely a Hill-type equation or a harmonic oscillator equation with a time-varying frequency) [25,26]. Redefining $\Theta = 3(F'/F)$ one gets

$$\frac{\mathrm{d}^2 F}{\mathrm{d}\,\lambda^2} + \frac{1}{3}(R_{ab}v^a v^b + \sigma^2 - \omega^2)F = 0. \tag{15}$$

The analysis of the expansion equation can be done using the above form. One notes that the expansion Θ is nothing but the rate of change of the cross-sectional area orthogonal to the bundle of geodesics. Therefore, the expansion approaching negative infinity implies a convergence of the bundle, whereas a value of positive infinity would imply a complete divergence. What are the conditions for convergence? Firstly, for convergence we must have an initially negative expansion. Finally, with F' negative we must end up at a zero of F (at a finite λ), in order to have a negatively infinite expansion [25,26]. Thus, the criterion for the existence of zeros in F at finite values of the affine parameter is what is required for convergence. Using the well-known Sturm comparison theorems in the theory of differential equations one can show that convergence occurs if

$$R_{ab}v^a v^b + \sigma^2 - \omega^2 \geq 0. \tag{16}$$

Thus, rotation defies convergence, while shear assists it. The equation for the evolution of the rotation ω_{ab}, has a trivial solution $\omega_{ab} = 0$. The criterion for convergence then becomes particularly simple for such hypersurface orthogonal congruences (zero rotation): $R_{ab}v^a v^b \geq 0$. This leads to geodesic focusing.

If we make use of the Einstein field equations and rewrite the Ricci tensor in terms of the energy–momentum tensor $R_{ab} = T_{ab} - \frac{1}{2}g_{ab}T$ then the so-called time-like convergence condition becomes a condition on matter stress energy. This, given as, $\left(T_{ab} - \frac{1}{2}g_{ab}T\right) v^a v^b \geq 0$ is known as the strong energy condition (SEC). For a diagonal T_{ab} (with $T_{00} = \rho, T_{aa} = p_a$) we must have $\rho + p_a \geq 0, \rho + \sum_a p_a \geq 0$ if the SEC is to be obeyed. In other words, geodesic focusing encodes the simple statement that if matter is attractive, geodesics must be eventually drawn towards each other. This seemingly trivial statement is proved via the focusing theorem.

In the late seventies, Tipler [25,26] realised that the assumption of the SEC imposed to prove focusing and hence the existence of singularities could be further weakened. Among other results, he was able to show how, in the proof of the Hawking–Penrose theorem one might replace SEC by the weak energy condition (WEC: $T_{ab}v^a v^b \geq 0 \ \forall$ non-space-like v^a). Tipler also introduced in his article, for the first time, the notion of an averaged energy condition (the averaged strong and weak energy conditions (ASEC and AWEC)) which are global in nature. For instance, the AWEC is obtained by integrating the WEC along a non-space-like geodesic and gives a number ($\int_{\lambda_1}^{\lambda_2} T_{ab}v^a v^b \mathrm{d}\lambda \geq 0$). The question of whether a violation of the energy conditions could lead to a non-singular solution was also addressed by him. One should mention here that a few years before Tipler's work, Bekenstein [27] and Murphy [28] had proposed certain non-singular spacetimes. These models turned out to be singular, following the above-mentioned generalisations of the singularity theorems. Tipler also quotes, in his paper, the well-known Epstein–Glaser–Yaffe theorem in quantum field theory where it is said that there can exist a quantum state with respect to which the expectation value of the stress–energy tensor can be negative [29]. However, he does mention that the existence of one state with respect to which the $\langle T_{00}\rangle$ is less than zero cannot really lead to the prevention of singularities. Of late, however, energy-condition violations have been discussed in great detail in the context of wormholes [21], dark energy [30], braneworld models [31] and semi-classical gravity (quantum field theory in curved spacetimes) [32]. The experimentally observed Casimir effect (the feeble attraction between parallel, conducting capacitor plates) [33] has been cited as an example of the existence of negative energy density (though, truly speaking this effect is concerned with negative pressures as opposed to actual negative energy densities).

The quest for further weakening the criteria for geodesic focusing and thereby making the singularity theorems stronger was continued later through the work of Borde [34] and Roman [35]. Further work on issues related to geodesic focusing, singularities, energy conditions and causality violations have been carried out in [36–40].

2.4 *Null geodesic congruences*

The Raychaudhuri equations for null geodesic congruences were first derived by Sachs [9] in 1961. Let us briefly recall the salient features of these equations. It must be mentioned, however, that these equations are not very different (in structure as well as consequences thereof) from the equations for time-like congruences.

The central issue in the case for null geodesic congruences is the construction of the transverse parts of the deviation vector and the spacetime metric. Assuming an affine parametrisation in the sense $\mathrm{d}x^a = k^a \mathrm{d}\lambda$ with $k^a k_a = 0$ and $k^a \xi_a = 0$ (with ξ^a being the deviation vector) we realise that we are in trouble because of the above two orthogonality relations. Naively writing $h_{ab} = g_{ab} + k_a k_b$ will not work here ($k^a h_{ab} \neq 0$). The transverse metric is thus constructed by introducing an auxiliary null vector N^a with $k_a N^a = -1$. (the choice of -1 is by convention, the essence is that the quantity must be non-zero). If we choose $k_a = -\partial_a u$ ($u = t - x$) then we can have $N_a = -\frac{1}{2}\partial_a v$ and hence $h_{ab} = g_{ab} + k_a N_b + k_b N_a$. This satisfies $k^a h_{ab} = 0$ and $N^a h_{ab} = 0$. Note that h_{ab} now is entirely two-dimensional. Keeping this transverse metric in mind we can proceed in the same way as for the time-like case by constructing $\hat{B}_{ab} = \nabla_b k_a$. We quote below the equation for the expansion:

$$\frac{\mathrm{d}\hat{\Theta}}{\mathrm{d}\lambda} + \frac{1}{2}\hat{\Theta}^2 + \hat{\sigma}^2 - \hat{\omega}^2 = -R_{ab}k^a k^b, \tag{17}$$

where the hatted quantitites are the expansion, rotation and shear for the null geodesic congruence. The focusing theorem for null geodesic congruences follows in the same way as for time-like congruences, with the null convergence condition $R_{ab}k^a k^b \geq 0$ being the requirement. Using Einstein equations one can obtain the so-called null energy condition $T_{ab}k^a k^b \geq 0$. Similar to the case for time-like congruences, we have corresponding equations for the evolution of shear and rotation for null geodesic congruences too. These are available in [15,18].

2.5 *The acceleration term and non-affine parametrisations*

The discussions above were exclusively for geodesic congruences. In the non-geodesic (i.e. time-like or null congruences) case, it is obvious that there will be differences. We state below, how the equation for the expansion changes via the addition of the so-called acceleration term. The equation is now given as

$$\frac{\mathrm{d}\Theta}{\mathrm{d}\lambda} + \frac{1}{3}\Theta^2 + \sigma^2 - \omega^2 - \nabla_a(v^b \nabla_b v^a) = -R_{ab}v^a v^b, \tag{18}$$

where the fifth term on the LHS is the acceleration term. Notice that this term is zero for geodesic congruences (zero acceleration). The average distance between the world lines in the flow is changed due to the divergence of the acceleration. The non-geodesic character of the flow also affects the equation for the shear.

For a non-affine parametrisation of null geodesic congruences, with $k^a \nabla_a k^b = \kappa k^b$ (κ, a constant, defined through the above equation) the definition of the expansion changes: $\Theta = \nabla_a k^a - \kappa$ and the Raychaudhuri equation for the expansion takes the form

$$\frac{\mathrm{d}\hat{\Theta}}{\mathrm{d}\lambda} - \kappa\Theta + \frac{1}{2}\hat{\Theta}^2 + \hat{\sigma}^2 - \hat{\omega}^2 = -R_{ab}k^a k^b. \tag{19}$$

The new feature here is the presence of the linear (in Θ) term. The conclusions on geodesic focusing however do not change, except for differences in the values of the expansion.

2.6 *A theorem for non-rotating singularity-free Universes: Raychaudhuri's later papers*

Over the entire period of more than forty years, Raychaudhuri did not work much on the equations which bear his name today. Interestingly, he came back to have a look at them once again in the late nineties with a series of papers [41,42]. In these articles, he constructed a theorem on non-rotating singularity-free Universes. The main inspiration behind these papers was the singularity-free cosmological solutions due to Senovilla [43], which created a lot of interest and curiosity among relativists in the 1990s. In fact, the work of Raychaudhuri sets out to show that these solutions may not quite be physically relevant, though surely, mathematically correct. Let us now briefly recall the theorem of Raychaudhuri.

The basic premise of [41,42] is the use of spacetime averages of quantitites, defined as

$$\langle \chi \rangle = \left[\frac{\int_{-x_0}^{x_0} \int_{-x_1}^{x_1} \int_{-x_2}^{x_2} \int_{-x_3}^{x_3} \chi \sqrt{|g|} \mathrm{d}^4 x}{\int_{-x_0}^{x_0} \int_{-x_1}^{x_1} \int_{-x_2}^{x_2} \int_{-x_3}^{x_3} \sqrt{|g|} \mathrm{d}^4 x} \right]_{\lim x_{0,1,2,3} \to \infty} . \tag{20}$$

Raychaudhuri shows that for a singularity-free non-rotating Universe, open in all directions, the spacetime average of all stress–energy invariants, including the energy density vanishes. The proof is worked out using the spacetime averages of the scalars that appear in the Raychaudhuri equation for the expansion. Following the statement of the theorem, Raychaudhuri claimed that an observationally consistent Universe cannot have a zero average density and hence one must necessarily give up the hope of having singularity-free solutions. Subsequent to Raychaudhuri's work, Saa and Senovilla [41] wrote a couple of comments which were mainly concerned with the converse of Raychaudhuri's theorem and the question whether spacetime averages could really be a property which can distinguish between singular and non-singular models. In fact, one must note that the theorem does not say that if the spacetime averages are zero the spacetime must be non-singular. Therefore, the theorem cannot really be used to distinguish between singular and non-singular cosmological models. Furthermore, Senovilla, in his comment, made a conjecture that spatial and not spacetime averages could be a distinguishing property between singular and non-singular models. More precisely, his claim was that for non-singular, non-rotating, globally hyperbolic and everywhere expanding models where SEC holds, the spatial averages of all stress–energy invariants must vanish. Therefore, if such spatial averages do not vanish then the model must be singular. More details and a recent proof of the conjecture is available in the article by Senovilla in this volume [44]. Despite the conflict between whether spatial or spacetime averages was the crucial distinguishing factor between singular and non-singular models, it goes without saying that Raychaudhuri's idea about averages being a deciding factor will surely be remembered as a lasting contribution apart from his coveted equations.

3. The equations in different geometries and different theories of gravity

The Raychaudhuri equations discussed in the previous sections do not change as long as we respect the Riemannian (pseudo-Riemannian) metric structure of space or spacetime. However, in an alternative theory of gravity, the Einstein field equations are of course different and hence the relation between the energy–momentum tensor and other geometric quantities do change. This results in a modification of the consequences that arise while analysing these equations. On the other hand, if the usual Riemannian structure of spacetime changes, such as in the case of spaces with torsion (Einstein–Cartan–Sciama–Kibble theory), then the Raychaudhuri equations are surely different. We shall give an example of both these scenarios in the following two subsections. The equations in spacetimes with torsion is discussed in relatively greater detail because, this is one scenario where we actually notice a generalisation of the usual Raychaudhuri equations.

3.1 *Metric theories with symmetric connections*

As mentioned above, for theories with a symmetric connection, the Raychaudhuri equations do not change, though the RHS of the expansion equation when written using matter stress–energy can be very different from what it is in GR. This change can surely affect geodesic focusing. It must be noted here that any change in the conclusions about geodesic focusing in this case is inherently due to the new solutions (spacetime geometries) of the modified Einstein equations for a given stress–energy. The above-mentioned aspects have been analysed in the context of the Brans–Dicke and Hoyle–Narlikar theories [45–47] as well as the $R+\beta R^2$ theory [48], on which we focus our attention below.

The study of the Raychaudhuri equations in the presence of higher curvature terms in the action has been a subject of interest for a long time [49]. The equation for the expansion, in a spacetime background solution of $R+\beta R^2$ gravity in the presence of matter ($\rho \propto a(t)^{-n}$) has been analysed in [50]. The strong energy condition (SEC) is examined and it is shown that the condition for a Big-Bang singularity changes in such a scenario. Recall that in GR, the singularity theorem indicates the inevitable presence of singularities when the SEC (as well as some other conditions) are satisfied. However, at the microscopic scale, in the evolution of the expansion of a geodesic congruence, quantum effects are expected to show up – which, in turn, may completely change classical predictions. One such quantum effect leads to the appearance of higher curvature terms, namely, quadratic gravity [48]. Even though it might seem that the analysis in [50] for quadratic gravity is classical, it is, in a broader sense, a semi-classical analysis. Classical solutions of quadratic gravity have been studied to explore the nature of singularities and it has been shown that the Big-Bang singularity may be avoided [51]. In [50], quadratic gravity, i.e. $R+\beta R^2$ gravity, was considered as a backreaction effect on pure Einstein's gravity. It was found that the SEC for $R+\beta R^2$ gravity is different from that of Einstein gravity. We briefly summarise the results of [50] below.

In GR, the SEC follows from the use of Einstein's equations through the relation

$$R_{ab}v^a v^b = 8\pi G\left[T_{ab} - \frac{1}{2}T g_{ab}\right] v^a v^b \geq 0 \tag{21}$$

for all time-like v^a. This, as mentioned before, leads to focusing.

The author in [50] is primarily interested in the cosmological singularity. To analyse the effects of the quadratic terms one only needs to find the new expression for the Ricci tensor R_{ab} in terms of the energy–momentum tensor T_{ab}. Therefore, we require the modified field equations for quadratic gravity. It is known that this equation is

$$-G_{ab} + 16\pi G\beta\left(\frac{1}{2}R^2 g_{ab} - 2RR_{ab} - 2R_{;n}^{;n} g_{ab} + 2R_{;a;b}\right) = -8\pi G T_{ab}. \tag{22}$$

In a perturbative analysis, we are primarily interested in the first-order contribution from βR^2 to the Raychaudhuri equation. After some algebra one arrives at

$$\begin{aligned} R_{ab}v^a v^b = &\left[\tilde{G}\left(T_{ab} - \frac{1}{2}T g_{ab}\right) + 2\beta\tilde{G}^2\right. \\ &\times\left.\left(\frac{1}{2}T^2 g_{ab}\tilde{G} - 2TT_{ab}\tilde{G} - T_{;n}^{;n} g_{ab} - 2T_{;a;b}\right)\right] v^a v^b + O(\beta^2), \end{aligned} \tag{23}$$

where $\tilde{G} = 8\pi G$.

Comparing the above equation with that in Einstein gravity, we find that an effective energy–momentum tensor appears in the RHS of the above equation. Assuming a matter-dominated Universe with a characteristic dependence on the scale factor (i.e., $\rho = \rho_0/a^n$) and a local conservation of T_{ab} leads to $p = [(n-3)/3](\rho_0/a^n)$. Thus, finally, we have

$$\begin{aligned} R_{ab}v^a v^b = &-\frac{n-2}{2}\tilde{G}\rho_n \\ &+\beta\tilde{G}^2\left[3n(n-1)(n-4)\tilde{G}\rho_n^2 + 6n^2(n-4)ka^{-2}\rho_n\right] \\ &+(A^2-1)\left[-\frac{n}{3}\tilde{G}\rho_n + \beta\tilde{G}^2\left[2n^2(n-4)\tilde{G}\rho_n^2\right.\right. \\ &\left.\left.+4n(n+2)(n-4)ka^{-2}\rho_n\right]\right], \end{aligned} \tag{24}$$

where A is a constant taking values in $(1,\infty)$, $v^a = At^a + \sqrt{A^2-1}x^a$ and $\{t^a, x^a, y^a, z^a\}$ are eigenvectors of T_b^a. The above expression is the modified term which appears in the RHS of the Raychaudhuri equation for the expansion. It is thereafter analysed in the context of different values of n and the possibilities of avoiding the Big–Bang singularity or its inevitable occurrence are pointed out for different cases.

We repeat once again that for any modified theory of gravity (e.g. Brans–Dicke, Einstein–Gauss–Bonnet in higher dimensions, induced gravity, low energy effective stringy gravity etc.) as long as the Riemannian structure of spacetime is respected, changes in conclusions related to geodesic focusing can arise only through the modified field equations and its use in the convergence condition.

3.2 *The Raychaudhuri equation in spacetimes with torsion*

The symmetric nature of the affine connection is one of the underlying assumptions of Riemannian geometry. This fact is also assumed while constructing GR. An antisymmetric connection may originate from the presence of spin matter fields in spacetime leading to a transition from the $\mathbf{V}_4$ to $\mathbf{U}_4$ manifolds [52]. This asymmetric part is known as torsion. Generalisation of the theory of gravity in such a spacetime with torsion was proposed by Einstein–Cartan–Sciama–Kibble (ECSK) [53]. Absence of any experimental signature of torsion however is the primary criticism of such models although there has been a renewed theoretical interest in the context of superstring theories [54] where spacetime torsion appears in the form of a massless string mode. The third rank field strength of the massless second rank antisymmetric tensor field of string theory (known as the Kalb–Ramond field) is identified with spacetime torsion in the low energy limit of the theory. Torsion also appears naturally in a theory of gravity where twistors are used [55] as well as in the supergravity scenario where torsion, curvature and matter fields are treated in an analogous way [56]. It has been shown in several articles [57] that in string inspired models such a background with torsion results in a departure from the experimentally predicted values of the well-known phenomena like gravitational lensing, perihelion precession of planetary orbits, gravitational redshift, rotation of the plane of polarization of the distant galactic radiowaves etc. All the above arguments, and several more, compel us to include torsion in any comprehensive theory of gravity. In the context of geodesic congruences it has further been shown that the kinematical quantities – shear, rotation, acceleration, expansion and their evolution equations are modified by the presence of torsion. This naturally leads to a generalisation of the Raychaudhuri equations in the presence of torsion leading to a more general understanding of the phenomenon of geodesic focusing. Here we give a brief review of the work presented in [58] where the role of torsion in modifying the Raychaudhuri equations and some of its implications have been discussed.

The torsion tensor ${T_{ab}}^c$ which is the antisymmetric part of the affine connection Γ^c_{ab}, is given by

$$ {T_{ab}}^c = \frac{1}{2}\left(\Gamma^c_{ab} - \Gamma^c_{ba}\right) \equiv \Gamma^c_{[ab]}\,, \tag{25} $$

where $a, b, c = 0, \ldots, 3$.

In GR, ${T_{ab}}^c$ is postulated to be zero.

From the torsion tensor one constructs the contortion tensor as

$$ {K_{ab}}^c = -{T_{ab}}^c - {T^c}_{ab} + {T_b}^c{}_a = -{K_a}^c{}_b. \tag{26} $$

This leads to a general expression for the affine connection given as

$$\Gamma^c_{ab} = \{^{\,c}_{ab}\} - K_{ab}{}^{c}, \tag{27}$$

where $\{^{\,c}_{ab}\}$ is the symmetric part of the connection (Christoffel symbols). With this modified connection the commutator of the covariant derivatives of a scalar field ϕ is

$$\tilde{\nabla}_{[a}\tilde{\nabla}_{b]}\phi = -T_{ab}{}^{c}\tilde{\nabla}_c\phi; \tag{28}$$

which is zero in the absence of torsion.

Similarly for a vector v^a the commutator of the derivative gives

$$\tilde{\nabla}_{[a}\tilde{\nabla}_{b]}v^c = R_{abd}{}^{c}v^d - 2T_{ab}{}^{d}\tilde{\nabla}_d v^c, \tag{29}$$

where the Riemann tensor is defined as

$$R_{abc}{}^{d} = \partial_a\Gamma^d_{bc} - \partial_b\Gamma^d_{ac} + \Gamma^d_{ae}\Gamma^e_{bc} - \Gamma^d_{be}\Gamma^e_{ac}. \tag{30}$$

The contribution of torsion, to the Riemann tensor, is explicitly given through the following expression:

$$R_{abc}{}^{d} = R_{abc}{}^{d}(\{\}) - \nabla_a K_{bc}{}^{d} + \nabla_b K_{ac}{}^{d} + K_{ae}{}^{d}K_{bc}{}^{e} - K_{be}{}^{d}K_{ac}{}^{e}, \tag{31}$$

where $R_{abc}{}^{d}(\{\})$ is the tensor generated by the Christoffel symbols. The symbols $\tilde{\nabla}$ and ∇ have been used to indicate the covariant derivative with and without torsion respectively.

From eq. (31), the expressions for the Ricci tensor and the curvature scalar are

$$R_{ab} = R_{ab}(\{\}) - 2\nabla_a T_c + \nabla_b K_{ac}{}^{b} + K_{ae}{}^{b}K_{bc}{}^{e} - 2T_e K_{ac}{}^{e} \tag{32}$$

and

$$R = R(\{\}) - 4\nabla_a T^a + K_{ceb}K^{bce} - 4T_a T^a \tag{33}$$

where

$$T_a = T_{ab}{}^{b}. \tag{34}$$

T_a can be either time-like, space-like or light-like.

3.2.1 *Contributions of torsion to shear, expansion, vorticity and acceleration.* Beginning with the behaviour of fluids and moving on to the initial singularity problem in cosmological models, the Raychaudhuri equations have been shown to be the key equation to explore the role of torsion in such diverse phenomena [59–62]. In [63] an inflationary model of the Universe in the context of ECSK theory was considered. References [64] and [65] addressed the crucial role of Raychaudhuri equation in the context of a gauge invariant formalism for cosmological perturbations in theories with torsion. We now explain the kinematical quantities mentioned before, in a scenario with torsion.

Torsion in a spacetime modifies the definition of kinematical quantities. The covariant derivative of the four-velocity U_a [16] can be decomposed as

$$\tilde{\nabla}_a U_b = \tilde{\sigma}_{ab} + \frac{1}{3} h_{ab} \tilde{\Theta} + \tilde{\omega}_{ab} - U_a \tilde{a}_b, \tag{35}$$

where $h_{ab} = g_{ab} + U_a U_b$ and

$$\tilde{\Theta} = \tilde{\nabla}_a U^a = \Theta - 2T^c U_c, \tag{36}$$

$$\tilde{\sigma}_{ab} = h_a^c h_b^d \tilde{\nabla}_{(c} U_{d)} = \sigma_{ab} + 2h_a^c h_b^d K_{(cd)}{}^e U_e, \tag{37}$$

$$\tilde{\omega}_{ab} = h_a^c h_b^d \tilde{\nabla}_{[c} U_{d]} = \omega_{ab} + 2h_a^c h_b^d K_{[cd]}{}^e U_e, \tag{38}$$

and the acceleration

$$\tilde{a}_c = U^a \tilde{\nabla}_a U_c = a_c + U^a K_{ac}{}^d U_d. \tag{39}$$

The quantities without the tilde are the values of the corresponding expressions in a spacetime without torsion.

3.2.2 *The Raychaudhuri equation.* Using the identity for the four-velocity U_a $(U_a U^a = -1)$,

$$U^b \tilde{\nabla}_c \tilde{\nabla}_b U_a = \tilde{\nabla}_c (U^b \tilde{\nabla}_b U_a) - \tilde{\nabla}_c U^b \tilde{\nabla}_b U_a \tag{40}$$

and from eq. (30)

$$U^b \tilde{\nabla}_c \tilde{\nabla}_b U_a = U^b \tilde{\nabla}_b \tilde{\nabla}_c U_a + R_{cba}{}^d U_d U^b - 2U^b T_{ab}{}^c \tilde{\nabla}_d U_c \tag{41}$$

we find the equation

$$\begin{aligned} &\frac{1}{3} h_{ca} \tilde{\Theta} + \tilde{\sigma}_{ca} + \tilde{\omega}_{ca} - U_c \tilde{a}_a \\ &\quad = \tilde{\nabla}_c \tilde{a}_a - \left(\frac{1}{9} h_{ca} \tilde{\Theta} + \frac{2}{3} \tilde{\Theta} \tilde{\sigma}_{ca} + \frac{2}{3} \tilde{\Theta} \tilde{\omega}_{ca} + 2\tilde{\sigma}_c^b \tilde{\omega}_{ba} \right. \\ &\qquad\quad \left. + \tilde{\sigma}_c^b \tilde{\sigma}_{ba} + \tilde{\omega}_c^b \tilde{\omega}_{ba} - \frac{1}{3} U_c \tilde{\Theta} \tilde{a}_a - U_c \tilde{a}^b \tilde{\sigma}_{ba} - U_c \tilde{a}^b \tilde{\omega}_{ba} \right) \\ &\qquad\quad - R_{cba}{}^d U_d U^d - 2U^b T_{ab}{}^c \left(\frac{1}{3} h_{dc} \tilde{\Theta} + \tilde{\sigma}_{dc} + \tilde{\omega}_{dc} - U_d \tilde{a}_c \right). \end{aligned} \tag{42}$$

Contracting the indices in eq. (42), one obtains a general expression for the Raychaudhuri equation for the expansion, in the presence of torsion as

$$\begin{aligned} \dot{\tilde{\Theta}} = {} & \tilde{\nabla}_c \tilde{a}^c - \frac{1}{3} \Theta^2 - \tilde{\sigma}^{ab} \tilde{\sigma}_{ab} + \tilde{\omega}^{ab} \tilde{\omega}_{ab} - R_{ab} U^a U^b \\ & - 2U^b T_{ab}{}^d \left(\frac{1}{3} h_d^a \tilde{\Theta} + \tilde{\sigma}_d^a + \tilde{\omega}_d^a - U_d \tilde{a}^a \right). \end{aligned} \tag{43}$$

This is the most general form of Raychaudhuri equation for the expansion in the presence of torsion. Simpler versions of this equation have been discussed in

[61,62,66,67]. It is obvious that there would be corresponding equations for shear and rotation too, which we do not mention here. Interestingly, it can be shown that if we have torsion as a phantom field (a scalar field with a negative kinetic energy term) through the dual of the third rank tensor field strength of a string-inspired Kalb–Ramond field, one can get a positive contribution in the RHS of the above equation, which, in turn, may be useful in eliminating singularities.

4. The equations in diverse contexts

4.1 *General relativity and relativistic astrophysics*

We have already discussed one of the main uses of the Raychaudhuri equations – the geodesic focusing theorem. Though it is an entirely geometric result, its use is mostly confined to the domain of GR. It is also quite obvious that the null version of these equations are useful in the study of gravitational lensing. The focal point in this case is nothing but the intersection of trajectories representing light rays and is known as the caustic of the bundle of trajectories. The optical scalars (Sachs scalars) are the quantitites of interest and their evaluation enables us to understand the nature of the null geodesic flow (light ray bundles). Is there anything else one can say apart from focusing, which will be relevant for lensing. Some of these issues have been addressed in [68]. For example, it is possible to rewrite the Raychaudhuri equation for null geodesics in a form involving the angular diameter distance d_A:

$$\frac{d''_A}{d_A} = -\frac{1}{2}R_{00} - \frac{1}{2}\sigma^2, \tag{44}$$

where $(\mathrm{d}/\mathrm{d}\lambda)\ln d_A = \frac{1}{2}\Theta$. Thus, for the case of zero shear one can determine d_A entirely in terms of geometry. In [68] the authors have asked the question 'how do voids affect light propagation?' and made use of the above relation to provide an answer.

A second example within relativistic astrophysics is a study of crack formation [69,70] using this equation. This is an interesting application, quite unique – but not pursued much later. The basic question here is – when will a spherical object develop cracks? Appearance of total radial forces of different signs in different regions in a perturbed configuration leads to the occurrence of such cracking, which, in turn leads to local anisotropy of fluid/emission of incoherent radiation. Using the Einstein and Raychaudhuri equations it is possible to express the net radial force R (after perturbation) in terms of $\mathrm{d}\Theta/\mathrm{d}\lambda$, p_i, ρ, g_{ij}. In this way a criterion for cracking can be obtained.

In 1995, Jacobson [71] in a rather unique paper demonstrated how one might view the Einstein field equation as a thermodynamic equation of state. In this calculation, Jacobson starts out by arguing that energy flux across a causal horizon is some kind of heat flow and the entropy of the system beyond is proportional to the area of the horizon. The heat flux δQ is related to the stress–energy tensor whereas the area variation is related to the expansion of a bundle of null geodesics. He then makes use of the Raychaudhuri equation in its linearised form (ignoring the Θ^2 term) to write down the expansion as an integral over the $R_{ij}k^ik^j$. Thereafter, with the input of the entropy–area relation he arrives at the Einstein equation as an equation of state. It is important to note here that Jacobson uses the geometric content of the Raychaudhuri equation and views it as more fundamental than the Einstein equation, which, in his approach is a derived relation. Extensions of Jacobson's work in the setting of non-equilibrium thermodynamics have appeared recently [72].

Scattered around the literature, are innumerable instances where the equations have been used in the context of GR and astrophysics. Prominent among them is its use in black hole physics – in studying the properties of black holes and in deriving the laws of black hole mechanics (see [18] and references therein). Apart from black holes, we mention, *en passant*, a few other situations which we thought were interesting: (i) use in a fluid-flow description of density irregularities in cosmology [73], (ii) quantum gravitational optics and an effective Raychaudhuri equation [74], (iii) magnetic tension and gravitational collapse [75], (iv) the effective Einstein and Raychaudhuri equations derived from higher-dimensional warped braneworld models [76].

4.2 *The Capovilla–Guven equations for relativistic membranes*

An obvious question, which was never asked till the work of Capovilla and Guven [77] appeared in the scene is: what happens if we consider a congruence of extremal surfaces instead of geodesics (extremal curves)? Are there similar Raychaudhuri equations?

To address this issue, we must first find out how to generalise the notions of expansion, rotation and shear for the case of a family of surfaces. Unlike a curve, a surface is parametrised by more than one parameter. Thus, it is natural to imagine an expansion, a rotation and a shear along each of these independent directions. Introducing a separate label for the surface coordinates, we find that we now have Θ^a, Σ^a_{ij} and Ω^a_{ij} (here $i, j, ...$, denote the indices representing the normals to the surface). The relation between the i, j indices and the spacetime indices (say μ, ν) is given through the embedding of the surface. Let us now make things more concrete and explicit.

Define a D-dimensional surface in a N-dimensional background through an embedding $x^\mu = x^\mu(\xi^a)$. E^μ_a constitute the tangent vector basis chosen such that $g_{\mu\nu}E^{\mu a}E^{\nu b} = \eta_{ab}$ (a, b run from 1 to D). $n^{\mu i}$ are the normals, with $g_{\mu\nu}n^{\mu i}n^{\nu j} = \delta^{ij}$ (i, j run from 1 to $N - D$). Also $g_{\mu\nu}n^{\mu i}E^{\nu a} = 0$. K^{abi} are the $N - D$ extrinsic curvatures (one along each normal direction). The embedded surface is minimal provided $\mathrm{Tr}(K_i) = 0$.

With the above definitions, one can follow the derivation for the case of geodesic curves and obtain the equation for the generalised expansion (assuming zero values for Σ_{ij} and Ω_{ij}):

$$\nabla^a\Theta_a + \frac{1}{N-D}\Theta_a\Theta^a + \left(M^2\right)^i_i = 0, \tag{45}$$

where

$$\left(M^2\right)^i_i = K^{abi}K_{abi} + R_{\mu\nu\rho\alpha}E^{\mu a}E^\rho_a n^{\nu i}n^\alpha_i. \tag{46}$$

Note that the above equation is a partial differential equation – the reason being that we require more than one parameter to describe a surface. Structurally, the equation is similar to the original Raychaudhuri equation for the expansion – one can easily note this by comparing the two. However, there are crucial differences – one of which involves the appearance of extrinsic curvature terms. It is possible to rewrite this equation in a second order form by choosing $\Theta_a = \nabla_a F$. We obtain

$$\nabla_a \nabla^a F + \left(M^2\right)_i^i F = 0 \tag{47}$$

which resembles a variable-mass wave equation on a D-dimensional surface with a non-trivial induced line element. For special cases, one can work out focusing criteria, although the notion of focusing will be largely different for surface congruences. For the case of string world-sheets, it is possible to make use of the conformal character of any two-dimensional line element and rewrite the equation as

$$\frac{\partial^2 F}{\partial \sigma^2} - \frac{\partial^2 F}{\partial \tau^2} + \Omega^2(\tau, \sigma) \left(M^2\right)_i^i F = 0, \tag{48}$$

where τ, σ are the coordinates on the string world sheet and the induced metric is $\mathrm{d}s^2 = \Omega^2(-\mathrm{d}\tau^2 + \mathrm{d}\sigma^2)$.

It can be shown [78] that the criterion for focusing is given as (where ${}^2R$ is the Ricci scalar of the worldsheet):

$$-{}^2R + R_{\mu\nu} E^{\mu a} E^{\nu}_a > 0 \tag{49}$$

which is a requirement for the function F to have zeros. However, there are many questions related to focusing of surfaces which remain un-answered. Some of these have been recently addressed in [79]. Earlier references on examples of solutions of the Raychaudhuri equations for surface congruences and related issues are available in [23,24,80].

It must be mentioned here that the above-generalised equation for the expansion vector field is a subset of a full set of equations which include those for the generalised shear and rotation too. Moreover, the analysis is entirely for extremal membranes of Nambu–Goto type. Here too, if the action changes, we find that the equations change – an example is available in [81].

4.3 *Quantum field theory*

In recent times, the kinematic quantitites (expansion, shear and rotation), as well as the Raychaudhuri equations, have appeared, quite unexpectedly, in the context of quantum field theory. We outline below, some of these scenarios briefly.

4.3.1 *The Langevin–Raychaudhuri equation.* A couple of years ago, Borgman and Ford investigated gravitational effects of quantum stress tensor fluctuations [82,83]. They showed that these fluctuations produce fluctuations in the focusing of a bundle of geodesics. An explicit calculation using the Raychaudhuri equation, treated as a Langevin equation (ignoring the Θ^2 term by a smallness assumption converts the Raychaudhuri equation to the form $\mathrm{d}\Theta/\mathrm{d}\lambda = f(\lambda)$, a Langevin-type equation) was performed to estimate angular blurring and luminosity fluctuations of the images of distant sources. The stress-tensor flucttutations were obtained assuming the case of a massless, minimally coupled scalar field in a flat background in a thermal state. Scalar field fluctuations drove the Ricci tensor fluctuations (via the semi-classical Einstein equations), which, in turn led to fluctuations in the expansion Θ. These authors also made some numerical estimates for the quantity $\Delta L/L$ (the fractional luminosity fluctuation) and pointed out possible astrophysical situations (gamma-ray bursts, for example), where an observable value of this quantity might exist. However, it is true that their work has many assumptions (ignoring shear, rotation as well as the Θ^2 term) and much further analysis is required to have a clearer picture of the possibility of actually seeing such effects in the real Universe. Nevertheless, the idea, on the whole, is novel and interesting in its own right and provides us with another application of the Raychaudhuri equation in a very different scenario.

4.3.2 *RG flows in theory (coupling) space.* Till now we have been looking at geodesic flows in spacetime. However, one might also contemplate geodesic flows in fictitious spaces where a metric is defined. Such spaces may correspond to certain physical scenarios. We highlight one example here in order to show how ubiquitous the notions of expansion, rotation, shear and the Raychaudhuri equation are.

For a given Lagrangian field theory, the set of couplings can form a space which is known as theory/coupling space. A curve in such a space will therefore represent a flow of couplings. In such a space, we can define a distance function using two-point functions integrated over physical spacetime. Such a definition of metric goes back to the well-known Cramer–Rao metric in probability theory and has also been highlighted in the context of field theory by Zamolodchikov as well as O'Connor–Stephens (see [84] and references therein).

Once we have a metric, we can use it to define derivatives of a vector field. The vector field of interest here is the so-called β-function vector field, which generates the RG flow. The covariant derivative of this vector field when split into the usual trace, anti-symmetric and symmetric traceless parts define the usual expansion, rotation and shear. However, we also know that the β-function vector field must satisfy the conformal Killing condition (this is the geometric form of the RG equation). These facts together imply the result that the focal point of the congruence generated by the β-function vector field must necessarily be a fixed point (i.e. all β_a ('a' labels the coupling space coordinates) vanish there) [84].

This line of thought is useful in obtaining generic results in field theory – the understanding of shear and rotation in the context of RG flows might provide a better geometric understanding of RG flows.

4.3.3 *Holography, c-function and the Raychaudhuri equation for the expansion.* The holographic principle has recently played a crucial role in our understanding of quantum aspects of gravity. The principle states that the information of gravity degrees of freedom in a D-dimensional volume is encoded in a quantum field theory defined on the $(D-1)$-dimensional boundary of this volume. As an immediate realisation of this, it is shown that the renormalisation group flow equation for the β-function of a four-dimensional quantum gauge field theory defined on the boundary of a five-dimensional volume can be described by geodesic congruences in a scalar-coupled five-dimensional gravitational theory. Such a gauge-gravity duality was proposed in a more general framework through the Maldacena conjecture and can be elegantly described through the renormalisation group equation with the bulk coordinate (or the holographic coordinate) as the renormalisation group parameter. It is shown that if the central charge or the c-function in a quantum field theory evolves monotonically under RG then the holographic principle indicates that in the corresponding dual gravity in five dimensions, the picture is realised through the Raychaudhuri equation governing the monotonic flow of the expansion parameter Θ for the geodesic congruences in the gravity sector [85–87]. The central charge of the boundary theory, or the c-function, is a measure of the degrees of freedom of the theory. As a consequence it is also a measure of the entropy of the black hole in the dual gravity sector which is related to the area of the horizon through the Bekenstein–Hawking formula. The effective central charge is a function of the couplings of the theory, which monotonically decreases as one flows to lower energies through the RG equation. The fixed point described by the boundary conformal field theory corresponds to the extrema of this function. It turns out that the null geodesic congruences can be used as a probe to decode the holographically encoded informations.

Consider a D-dimensional spacetime with a negative cosmological constant where the metric g_{ab} is foliated by appropriate choices of constant time surfaces. If we now focus on a $(D-2)$-dimensional spacetime surface M at a fixed time t, then one may construct a null vector field m^a on the light sheets (consisting of spacetime points scanned by null geodesic congruences) which is orthogonal to M such that $m^a n_a = -1$, where n_a are tangents to the geodesic. The metric on M is given by

$$h_{ab} = g_{ab} + n_{(a}m_{b)}. \tag{50}$$

Defining,

$$B_{ab} = D_b n_a \tag{51}$$

such that $B_{ab} = B_{ba}$ (Frobenius theorem ensures the symmetry property) and

$$\Theta = \operatorname{Tr} B \tag{52}$$

it can be shown from the Raychaudhuri equation for the expansion that the geodesics converge since,

$$\Theta \leq 0. \tag{53}$$

The question that arises now is – how is the information on the light sheet encoded in M? Assuming that the spacetime admits a time-like Killing vector field, we may visualise the time flow of M to constitute the $(D-2)+1 = (D-1)$-dimensional boundary of the D-dimensional bulk. The c-function of the corresponding dual gauge theory sits on the boundary. The RG flow to lower energy scales is then associated to the motion along the converging congruences of null geodesics described by Raychaudhuri equation for the expansion. In this sense, the null geodesics act as a probe for the dual theory at a lower energy scale. Pictorially, it may be described as follows. The lower the energy scale, the RG takes us into the gauge theory, deeper inside the bulk we move along null geodesic congruences following Raychaudhuri equation for the expansion. The caustic point where the geodesics meet correspond to the fixed point of the RG flow. This establishes a remarkable significance of the Raychaudhuri equation for the expansion in describing the holographic principle.

As described previously the c-function tells us about the degree of freedom of the system, which, in turn, is related to black hole entropy through the area law. Therefore, the c-function is naturally related to the area function $A(r)$ as

$$c(r) = \frac{A(r)}{4}. \tag{54}$$

In the context of the recently developed attractor mechanism [88] in describing black holes in a scalar coupled gravitational theory, one uses this flow of the c-function to decode some interesting properties of both supersymmetric and non-supersymmetric attractors. The Raychaudhuri equation for the expansion once again plays a crucial role. It is shown that in any spherically symmetric scalar coupled static asymptotically flat solution, $c(r)$ decreases monotonically as one moves radially inward from infinity. It is further shown that the minimum value of $c(r)$ corresponds to the entropy of the horizon [89]. In order to establish the c-theorem in this context once again the Raychaudhuri equation for the expansion is used. Taking a congruence of null geodesics, the expansion is given by

$$\Theta = \frac{\mathrm{d} \ln A}{\mathrm{d}\lambda}, \tag{55}$$

(where λ is the affine parameter). Using the null energy condition,

$$T_{ab}k^a k^b \geq 0 \tag{56}$$

and the Raychaudhuri equation, we get

$$\frac{\mathrm{d}\Theta}{\mathrm{d}\lambda} \leq 0. \tag{57}$$

The c-theorem can thus be proved on very general grounds.

In summary, these rather novel applications of the equations in the context of quantum field theory seems to assert once again the generality and the ubiquitous nature of the equations.

5. Summary and scope

We have reviewed, in this article, the Raychaudhuri equations and its applications in diverse contexts. It must be admitted that since its inception, the equations have used extensively to enhance our understanding of various situations within, as well as outside the scope of GR. The primary reason behind its wide-ranging applicability is (as emphasised several times in this article) the fact that the equations encode geometric statements about flows. Since flows appear in many different contexts in physics, it goes without saying, that the equations will be useful in furthering our understanding in different ways.

Among the applications within classical GR, the utility of the equations in understanding the singularity problem, is surely the most prominent one. In astrophysics, we have quoted the use of the equations in the context of lensing, cracking of self-gravitating objects. At a more fundamental level, Jacobson's work exploits the equation and, in some sense, makes use of its geometric nature to arrive at a new way of looking at the Einstein equations.

On the other hand, in a non-Riemannian spacetime, we have seen how the equations change – in particular, in spacetimes with torsion. Torsion appears in disguise in string theory and is therefore not totally irrelevant today, even though observations might have ruled out the Einstein–Cartan–Sciama–Kibble theory long ago.

Furthermore, what has come as a pleasant surprise in recent times, is the utility of these equations (and the quantities that appear in the equations: expansion, shear and rotation) in the context of quantum field theory. In addition, the generalisation of these equations to surface congruences is also an interesting development, despite the fact that not much has been done on such generalised Raychaudhuri (Capovilla–Guven) equations.

Despite the large body of work on or based on the Raychaudhuri equations, there still remain many unanswered questions. In conclusion, we list a few of them. The choice is surely biased by our own perspective.

(a) Normally, in fluid mechanics, the independent variable is the velocity field. The Raychaudhuri equations are for the gradient of the velocity field – they are one derivative higher. We mentioned before that they are essentially identities and become equations when we choose a geometric property (e.g. vacuum, Einstein space etc.) or use Einstein's field equations of GR. Is it possible to set up and solve the initial value problem for this coupled system of equations such that we know the characteristics of a geodesic flow in a given spacetime geometry once we specify the initial conditions on expansion, shear and rotation? Similar analysis is done while deriving the focusing theorem but can we do it by analysing the full system of equations analytically/numerically. Following this approach one might be able to reconstruct the gradient of the velocity field and also obtain the velocity field itself.

(b) In the context of RG flows, is it possible to understand the role of shear and rotation and use it to improve our geometric analysis of such flows?

(c) For surface congruences, can we obtain equations for families of null surfaces? Or, for surfaces defined by the extremality of actions other than Nambu–Goto (rigidity corrections, Willmore functionals etc.)?

(d) In quantum mechanics, we talk about the flow of probability. Imagining this as a fluid, can we obtain expansion, rotation and shear of such flows and then write down conclusions based on the corresponding Raychaudhuri equations?

(e) Electrodynamics, is, after all, the physics of the electric and magnetic vector fields. Suppose we construct the gradient of the electric/magnetic fields, split it up into expansion, rotation and shear and write down Raychaudhuri equations for the so-called Faraday lines of force! What does such an analysis tell us?

To sum up, we would like to conjecture that wherever there are vector fields describing a physical/geometrical quantity, there must be corresponding Raychaudhuri equations. We believe that there are many avatars of the same equations, of which, till today, we are aware of only a few.

Acknowledgements

The authors thank the editors (N Dadhich, P S Joshi and Probir Roy) of this special volume, for giving them the opportunity to write this review. They also specially thank Probir Roy and Anirvan Dasgupta for their comments and a careful reading of the manuscript. This work was partially supported by funds from a DST, Government of India project (Grant No. SR/S2/HEP-10/2005).

References

[1] A Raychaudhuri, *Phys. Rev.* **89**, 417 (1953)
[2] A Raychaudhuri, *Phys. Rev.* **86**, 90 (1952)
[3] A Raychaudhuri, *Phys. Rev.* **98**, 1123 (1955)
[4] A Raychaudhuri, *Z. Astrophysik* **43**, 161 (1957)
[5] O Heckmann and E Schucking, *Z. Astrophysik* **38**, 95 (1955)
[6] A Komar, *Phys. Rev.* **104**, 544 (1956)
[7] A Raychaudhuri, *Phys. Rev.* **106**, 172 (1957)
[8] J Ehlers, *Gen. Relativ. Gravit.* **25**, 1225 (1993); English translation of original German article by P Jordan, J Ehlers, W Kundt and R K Sachs, *Proceedings of the Mathematical–Natural Sciences Section of the Mainz Academy of Sciences and Literature*, Nr. 11, 792 (1961)
[9] R K Sachs, *Proc. R. Soc. (London)* **A264**, 309 (1961); *ibid* **A270**, 103 (1962)
[10] L Landau and E M Lifshitz, *Classical theory of fields* (Pergamon Press, Oxford, UK, 1975)
[11] E M Lifshitz and I M Khalatnikov, *Adv. Phys.* **12**, 185 (1963)
[12] R Penrose, *Phys. Rev. Lett.* **14**, 57 (1965)
[13] S W Hawking, *Phys. Rev. Lett.* **15**, 689 (1965)
[14] S W Hawking, *Phys. Rev. Lett.* **17**, 444 (1966)
[15] R M Wald, *General relativity* (University of Chicago Press, Chicago, USA, 1984)
[16] S W Hawking and G F R Ellis, *The large scale structure of spacetime* (Cambridge University Press, Cambridge, UK, 1973)
[17] P S Joshi, *Global aspects in gravitation and cosmology* (Oxford University Press, Oxford, UK, 1997)
[18] E Poisson, *A relativist's toolkit: the mathematics of black hole mechanics* (Cambridge University Press, Cambridge, UK, 2004)
[19] G F R Ellis, in *General relativity and cosmology,* International School of Physics, Enrico Fermi–Course XLVII (Academic Press, New York, 1971)
[20] I Ciufolini and J A Wheeler, *Gravitation and inertia* (Princeton University Press, Princeton, USA, 1995)
[21] M Visser, *Lorentzian wormholes from Einstein to Hawking* (AIP Press, USA, 1995)
[22] T Frankel, *Gravitational curvature* (W H Freeman, USA, 1979)
[23] E Zafiris, *J. Geom. Phys.* **28**, 271 (1998)
[24] B Carter, *Contemp. Math.* **203**, 207 (1997)
[25] F J Tipler, *Phys. Rev.* **D17**, 2521 (1978)
[26] F J Tipler, *J. Diff. Equ.* **30**, 165 (1978)
[27] J D Bekenstein, *Phys. Rev.* **D11**, 2072 (1975)
[28] G L Murphy, *Phys. Rev.* **D8**, 4231 (1973)
[29] H Epstein, V Glaser and A Jaffe, *Nuovo Cimento* **36**, 1016 (1965)

[30] R R Caldwell, M Kamionkowski and N N Weinberg, *Phys. Rev. Lett.* **91**, 071301 (2003)
S M Carroll, M Hoffman and M Trodden, *Phys. Rev.* **D68**, 023509 (2003)
P Frampton, *Phys. Lett.* **B555**, 139 (2003)
V Sahni and Y Shtanov, *J. Cosmol. Astropart. Phys.* **0311**, 014 (2003)
[31] L Randall and R Sundrum, *Phys. Rev. Lett.* **83**, 4690 (1999)
C Csaki, *TASI Lectures on Extra Dimensions and Branes*, hep-ph/0404096
[32] N D Birrell and P C W Davies, *Quantum fields in curved space* (Cambridge University Press, Cambridge, UK, 1982)
[33] H B G Casimir, On the attraction between two perfectly conducting plates, *Proc. Kon. Nederl. Akad. Wetenschap* **51**, 793 (1948)
[34] A Borde, *Class. Quant. Grav.* **4**, 343 (1987)
[35] T Roman, *Phys. Rev.* **D37**, 546 (1988)
[36] F J Tipler, *Ann. Phys.* **108**, 1 (1977); *ibid. Phys. Rev. Lett.* **37**, 879 (1976)
[37] C Chicone and P Ehrlich, *Manuscripta Math.* **31**, 297 (1980)
[38] T Roman and P G Bergmann, *Phys. Rev.* **D28**, 1265 (1983)
[39] T Roman, *Phys. Rev.* **D33**, 3526 (1986)
[40] M Visser, *Phys. Rev.* **D47**, 2395 (1993)
[41] A K Raychaudhuri, *Phys. Rev. Lett.* **80**, 654 (1998)
A Saa, *Phys. Rev. Lett.* **81**, 5031 (1998)
J M M Senovilla, *Phys. Rev. Lett.* **81**, 5032 (1998)
A K Raychaudhuri, *Phys. Rev. Lett.* **81**, 5033 (1998)
[42] A K Raychaudhuri, *Mod. Phys. Lett.* **A15**, 391 (2000)
[43] J M M Senovilla, *Phys. Rev. Lett.* **64**, 2219 (1990)
E Ruiz and J M M Senovilla, *Phys. Rev.* **D45**, 1995 (1992)
[44] J M M Senovilla, *Pramana – J. Phys.* **69(1)**, 31 (2007)
[45] A K Raychaudhuri and S Banerji, *Zeit. fur. Astrophysik* **58**, 187 (1964)
[46] S Banerji, *Phys. Rev.* **D9**, 877 (1974)
[47] A K Raychaudhuri, *Prog. Theor. Phys.* **53**, 1360 (1975)
[48] J H Kung, *Phys. Rev.* **D52**, 6922 (1995)
[49] V Muller and H J Schmidt, *Gen. Relativ. Gravit.* **17**, 769 (1985)
D Page, *Phys. Rev.* **D36**, 1607 (1987)
J D Barrow and A Ottewill, *J. Phys.* **A16**, 2757 (1983)
[50] J H Kung, *Phys. Rev.* **D53**, 3017 (1996)
[51] B Whitt, *Phys. Lett.* **B145**, 176 (1984)
J C Alonso, F Barbero, J Julve and A Tiemblo, *Class. Quant. Grav.* **11**, 865 (1994)
M Mijic *et al*, *Phys. Rev.* **D34**, 2934 (1986)
[52] J A Schouten, *Ricci calculus* (Springer-Verlag, Berlin, 1954)
[53] F W Hehl, P von der Heyde, G D Kerlick and J M Nester, *Rev. Mod. Phys.* **48**, 393 (1976)
[54] M B Green, J H Schwarz and E Witten, *Superstring theory* (Cambridge University Press, Cambridge, UK, 1987)
R Hammond, *Il Nuovo Cimento* **B109**, 319 (1994); *Gen. Relativ. Gravit.* **28**, 419 (1986)
V De Sabbata, *Ann. der Physik* **7**, 419 (1991)
V de Sabbata, *Torsion, string tension and quantum gravity*, in 'Erice 1992', *Proceedings, String Quantum Gravity and Physics at the Planck Energy Scale*, p. 528
Y Murase, *Prog. Theor. Phys.* **89**, 1331 (1993)
[55] P S Howe, A Opfermann and G Papadopoulos, *Twistor spaces for QKT manifold*, hep-th/9710072

P S Howe and G Papadopoulos, *Phys. Lett.* **B379**, 80 (1996)
V de Sabbata and C Sivaram, *Il Nuovo Cimento* **A109**, 377 (1996)
[56] G Papadopoulos and P K Townsend, *Nucl. Phys.* **B444**, 245 (1995)
C M Hull, G Papadopoulos and P K Townsend, *Phys. Lett.* **B316**, 291 (1993)
B D B Figueiredo, I Damiao Soares and J Tiomno, *Class. Quant. Grav.* **9**, 1593 (1992)
L C Garcia de Andrade, *Mod. Phys. Lett.* **A12**, 2005 (1997)
O Chandia, *Phys. Rev.* **D55**, 7580 (1997)
P S Letelier, *Class. Quant. Grav.* **12**, 471 (1995); **12**, 2221 (1995)
R W Kühne, *Mod. Phys. Lett.* **A12**, 2473 (1997)
D K Ross, *Int. J. Theor. Phys.* **28**, 1333 (1989)
[57] S Kar, S SenGupta and S Sur, *Phys. Rev.* **D67**, 044005 (2003)
S Kar, P Majumdar, S SenGupta and S Sur, *Class. Quant. Grav.* **19**, 677 (2002)
S SenGupta and S Sur, *Phys. Lett.* **B521**, 350 (2001)
S Kar, P Majumdar, S SenGupta and A Sinha, *Eur. Phys. J.* **C23**, 357 (2002)
P Majumdar and S SenGupta, *Class. Quant. Grav.* **16**, L89 (1999)
S SenGupta and S Sur, *Europhys. Lett.* **65**, 601 (2004)
[58] S Capozziello, G Lambiase and C Stornaiolo, *Ann. der Phys.* **10**, 713 (2001), gr-qc/0101038
[59] R de Ritis, M Lavorgna and C Stornaiolo, *Phys. Lett.* **A95**, 425 (1983)
[60] A Trautman, *Nature (Phys. Sci.)* **242**, 7 (1973)
[61] J Stewart and P Hajicek, *Nature (Phys. Sci.)* **244**, 96 (1973)
[62] G Esposito, *Fortschritte der Physik* **40**, 1 (1992)
[63] M Demianski, R de Ritis, G Platania, P Scudellaro and C Stornaiolo, *Phys. Rev.* **D35**, 1181 (1987)
[64] G F R Ellis and M Bruni, *Phys. Rev.* **D40**, 1804 (1989)
G F R Ellis, J Hwang and M Bruni, *Phys. Rev.* **D40**, 1819 (1989)
G F R Ellis, M Bruni and J Hwang, *Phys. Rev.* **D42**, 1035 (1990)
[65] D Palle, *Nuovo Cimento* **B114**, 853 (1999)
[66] J Tafel, *Phys. Lett.* **A45**, 341 (1973)
[67] A K Raychaudhuri, *Theoretical cosmology* (Clarendon Press, Oxford, UK, 1979)
[68] N Sugiura, K Nakao, D Ida, N Sakai and H Ishihara, *Prog. Theor. Phys.* **103**, 73 (2000)
[69] L Herrera, *Phys. Lett.* **A165**, 206 (1992)
[70] A di Prisco, E Fuenmayor, L Herrera and V Varela, *Phys. Lett.* **A195**, 23 (1994)
[71] T Jacobson, *Phys. Rev. Lett.* **75**, 1260 (1995)
[72] C Eling, R Guedens and T Jacobson, *Phys. Rev. Lett.* **96**, 121301 (2006)
[73] D Lyth and M Mukherjee, *Phys. Rev.* **D38**, 485 (1988)
[74] N Ahmadi and M Nouri-Zonoz, *Quantum gravitational optics: Effective Raychaudhuri equation*, gr-qc/0605009
[75] C Tsagas, *Magnetic tension and gravitational collapse*, gr-qc/0311085
[76] R Maartens, *Phys. Rev.* **D62**, 084023 (2000)
[77] R Capovilla and J Guven, *Phys. Rev.* **D52**, 1072 (1995)
S Kar, *Phys. Rev.* **D52**, 2036 (1995)
S Kar, *Phys. Rev.* **D53**, 2071 (1996)
[78] S Kar, *Phys. Rev.* **D55**, 7921 (1997)
[79] S Kar, Focusing of branes in warped backgrounds, *Ind. J. Phys.* (to appear in the special issue commemorating A K Raychaudhuri) (2006)
[80] S Kar, *Phys. Rev.* **D54**, 6408 (1996)

[81] S Kar, Generalised Raychaudhuri equations for strings and membranes, *Proceedings of IAGRG–XVIII* (Matscience, Madras, India, Feb. 15–17, 1996) Matscience Report no. 117
[82] J Borgman and L H Ford, *Phys. Rev.* **D70**, 064032 (2004)
[83] J Borgman, *Fluctuations of the expansion: The Langevin–Raychaudhuri equation*, Ph.D. Thesis, Tufts University, USA (2004)
[84] S Kar, *Phys. Rev.* **D64**, 105017 (2001)
[85] V Balasubramanian, E Gimon and D Minic, *J. High Energy Phys.* **05**, 014 (2000)
[86] V Sahakian, *Phys. Rev.* **D62**, 126011 (2000)
[87] E Alvarez and C Gomez, *Nucl. Phys.* **B541**, 441 (1999); *ibid. Holography and the c-theorem*, hep-th/0009203
[88] M Cvetic and A A Tseytlin, *Phys. Rev.* **D53**, 5619 (1996)
A Stromonger, *Phys. Lett.* **B383** (1996)
S Ferrara and R Kallosh, *Phys. Rev.* **D54**, 1525 (1996)
S Ferrara, G W Gibbons and R Kallosh, *Nucl. Phys.* **B500**, 75 (1997)
[89] K Goldstein, R P Jena, G Mandal and S P Trivedi, *J. High Energy Phys.* **02**, 053 (2006)

PRAMANA
— journal of
physics

© Indian Academy of Sciences

Vol. 69, No. 1
July 2007
pp. 77–92

Black hole dynamics in general relativity

ABHAY ASHTEKAR[1,2]
[1]Institute for Gravitational Physics and Geometry, Physics Department,
Penn State University, University Park, PA 16802, USA
[2]Institute for Theoretical Physics, University of Utrecht, Princetonplein 5,
3584 CC Utrecht, The Netherlands
E-mail: ashtekar@gravity.psu.edu

Abstract. Basic features of dynamical black holes in full, non-linear general relativity are summarized in a pedagogical fashion. Qualitative properties of the evolution of various horizons follow directly from the celebrated Raychaudhuri equation.

Keywords. Black holes; event; isolated and dynamical horizons; Raychaudhuri equation.

PACS Nos 04.25.Dm; 04.70.Bw

1. Introduction

Ever since the mid-seventies when I first came across Professor Raychaudhuri's work, I have admired his remarkable creativity and his deep insights into the unique interplay between mathematics and physics that underlies general relativity. I met him only on a few occasions at conferences. Although these encounters were brief, like many others I was immediately struck by his simplicity, earnestness and above all his passion for science. Last year I also had the good fortune of seeing the documentary on his life made by IUCAA and Vignan Prasar. Through it I developed a better understanding of how he lived, where he taught and what inspired him. My admiration for this remarkable scientist has grown even further. There is a large body of physicists he molded both through his physics lectures and by the way he lived. It is easy to understand the unusually deep reverence they have always expressed. Personally, it is an honour and a pleasure to contribute to this volume.

Discussions of Professor Raychaudhuri's contributions to general relativity – especially of his celebrated equation – generally emphasize singularity theorems and cosmology. This is true, for example, of the excellent documentary I just mentioned. However, the Raychaudhuri equation has also had a deep impact on black hole physics. Perhaps the reason why this aspect is not as widely known is that much of the literature in this area has focused on properties of stationary black holes. But in Nature, black holes are rarely in equilibrium. They grow by swallowing stars and galactic debris as well as electromagnetic and gravitational radiation. These processes are generally modelled using perturbations of Kerr solutions. However, to understand how black holes form and grow in violent astrophysical processes including mergers, one must consider the dynamical regime of the full, non-linear general relativity. And it is here that one can see the Raychaudhuri equation in action, as it elegantly determines the qualitative features of the formation and growth of horizons in exact general relativity.

In the rest of this section, I will set the stage by briefly recalling the Raychaudhuri equation for null congruences. In §2, I will discuss event horizons and in §3, quasi-local horizons.

Fix a 4-dimensional space-time $(\mathcal{M}, g_{ab})$ and a closed, space-like 2-surface S in it. At each point of S there are exactly two future pointing null normals to it, generally denoted by ℓ^a and n^a. Consider the 3 surfaces $\mathcal{N}^{\pm}$ spanned by null geodesics tangential to ℓ^a and n^a. In a finite neighborhood of S, these null geodesics will not develop caustics. Let us restrict ourselves to this neighborhood so that $\mathcal{N}^{\pm}$ can be taken to be smooth 3-surfaces. By construction they are null. Physically, one can think of them as follows. Let us suppose S is illuminated just for an instant. Then $\mathcal{N}^{\pm}$ are spanned by the resulting light fronts. If (M, g_{ab}) were Minkowski space, for example, cross-sections of $\mathcal{N}^{+}$ would represent the outgoing, expanding light fronts and those of $\mathcal{N}^{-}$ the ingoing, contracting ones.

In what follows we return to general space-times. For definiteness, let us focus on ℓ^a and (with a slight abuse of terminology) denote the geodesic congruence obtained from it also by ℓ^a. If v denotes the affine parameter along the null geodesics tangential to ℓ^a, and if we set $v = 0$ on S, then 2-surfaces S_v defined by $v = \text{const}$ are diffeomorphic to S and represent the evolution of S along $\mathcal{N}^{+}$. We will refer to these S_v as *cross-sections* of $\mathcal{N}^{+}$. Next, let us denote by $\tilde{q}_a{}^b$ the projection operator into the tangent space of S_v. Thus $\tilde{q}_a{}^b \ell^a = 0$, $\tilde{q}_a{}^b n^a = 0$ and $\tilde{q}_a^b t^a = t^b$ for all vectors tangential to S_v. $\tilde{q}_{ab} = \tilde{q}_a{}^m g_{bm}$ is the intrinsic metric on S_v of signature $+,+$. While the discussion of this section can be readily extended to more general null congruences which may not emanate from a closed 2-surface S, this slightly restricted case makes the discussion more intuitive.

Since ℓ^a is a geodesic vector field, its acceleration $\ell^a \nabla_a \ell^b$ vanishes. Also, since ℓ^a is the null normal to a 3-surface, if the surface is specified by $f = \text{const}$, then $\ell_a \propto \nabla_a f$, whence $\nabla_{[a}\ell_{b]} = V_{[a}\ell_{b]}$ for some 1-form field V_a. Therefore

$$\omega_{ab} := \tilde{q}_a{}^m \tilde{q}_b{}^n \, \nabla_{[m}\ell_{n]} = 0. \tag{1}$$

The tensor field ω_{ab} is called the twist of ℓ^a. Geometrically, its vanishing is the necessary and sufficient condition that the null geodesic congruence ℓ^a is surface-forming. Physically, this can be interpreted as saying that the light rays represented by ℓ^a have no 'rotation'. Since $\omega_{ab} = 0$, the projection of $\nabla_a \ell_b$ into S_v is symmetric.

Our null congruence ℓ^a has no acceleration or twist. The remaining two fields constructed from its derivatives are called *expansion* and *shear*. The expansion Θ is defined as

$$\Theta = \tilde{q}^{ab} \, \nabla_a \ell_b \, . \tag{2}$$

As the name suggests it captures the rate of increase of the area of cross-sections S_v of $\mathcal{N}^{+}$. More precisely, denote by $\tilde{\epsilon}_{ab}$ the area 2-form on cross-sections S_v (defined by 2-metrics $\tilde{q}_{ab}$). Then $\mathcal{L}_{\ell}\tilde{\epsilon}_{ab} = \Theta\tilde{\epsilon}_{ab}$. Hence if at all we consider a small area element A of S_v any point p of $\mathcal{N}$ and follow it along the geodesic flow generated by ℓ^a, we have $\Theta = (1/A)(\mathrm{d}A/\mathrm{d}v)$.

While the expansion is obtained by taking the trace of the projection of $\nabla_a \ell_b$ into S, the shear is obtained by taking the trace-free part.

$$\sigma_{ab} = \tilde{q}_a{}^m \tilde{q}_b{}^n \nabla_m \ell_n - \frac{1}{2}\Theta\, \tilde{q}_{ab}. \tag{3}$$

Again as its name indicates σ_{ab} represents how the cross-sections S_v are sheared with respect to one another. The amount of shear on a cross-section S_v is a heuristic measure of the flux of gravitational radiation across it [1,2].

Since we know ℓ^a only on the 3-surface $\mathcal{N}^+$, its directional derivatives $t^a \nabla_a \ell_b$ are well-defined only when t^a is tangential to $\mathcal{N}^+$. All these are captured in the acceleration, the twist, expansion and shear. Finally, note that to calculate the twist ω_{ab}, the expansion Θ and the shear σ_{ab} of a cross-section S_v, one only needs to specify the vector field ℓ^a on that cross-section[1]. This fact will be important especially in §3. These three quantities are sometimes referred to as the *optical scalars* associated with the null congruence ℓ^a.

With these preliminaries out of the way, we can now introduce the Raychaudhuri equation for the null congruence introduced above. Having defined the optical scalars, a natural question is how they 'evolve' along the given null congruence and, in particular, which properties of space-time curvature feature in these evolution equations. For our purposes, it will suffice to restrict ourselves to the evolution of Θ. We have (see, e.g., [3,4]):

$$\frac{\mathrm{d}\Theta}{\mathrm{d}v} \equiv \ell^a \nabla_a \Theta = -\frac{1}{2}\Theta^2 - \sigma_{ab}\sigma^{ab} - R_{ab}\ell^a \ell^b\,, \tag{4}$$

where R_{ab} is the Ricci tensor of the space-time metric g_{ab}. This is the Raychaudhuri equation for our (twist-free) null congruence. Matter fields one normally uses in general relativity satisfy the null energy condition: $T_{ab}k^a k^b \geq 0$ for all null vector fields k^a. In this case, we automatically have $R_{ab}\ell^a \ell^b \geq 0$. Thus, remarkably, in physically interesting situations, the right side is negative definite. This property of null congruences has an extremely rich set of physical consequences, making Raychaudhuri equation a key element in the investigation of a variety of strong field phenomena in general relativity. In the next two sections we will see examples of this striking interplay between geometry and physics.

To conclude this section, I would like to note a simple – but surprisingly powerful – consequence of the Raychaudhuri equation. The idea is to exploit the fact that in physically interesting cases, we have

---

[1]For pedagogical simplicity, I introduced $\mathcal{N}^+$ using geodesics emanating from a cross-section S. If we are given only a null surface $\mathcal{N}^+$ and a specific choice of its null normal ℓ^a, the optical scalars can be defined at a point p of $\mathcal{N}^+$ by choosing just a 2-dimensional, space-like subspace of the tangent space of $\mathcal{N}^+$ at p; global cross-sections S_v are not needed.

$$\frac{d\Theta}{dv} + \frac{1}{2}\Theta^2 \le 0 \qquad \Leftrightarrow \qquad \frac{d\Theta^{-1}}{dv} \ge \frac{1}{2}. \tag{5}$$

Therefore, if the initial value Θ_o is negative, i.e., the congruence ℓ^a is initially converging, we immediately have $\Theta(v) \le -2/[2|\Theta_o^{-1}| - v]$. Thus, $\Theta(v)$ must diverge in an affine parameter $v \le 2/|\Theta_o|$. This is a rather trivial mathematical consequence of (4), obtained just by solving an ordinary differential equation. However, it has a multitude of very interesting consequences.

2. Event horizons

To capture the intuitive notion that a black hole is a region from which signals cannot escape to the asymptotic part of space-time, one needs a precise definition of future infinity. The standard strategy is to use Penrose's conformal boundary $\mathcal{I}^+$ [5]. It is a future boundary: No point of the physical space-time lies to the future of any point of $\mathcal{I}^+$. It has topology $\mathbb{S}^2 \times \mathbb{R}$ and it is null (assuming that the cosmological constant is zero). In Minkowski space-time, one can think of $\mathcal{I}^+$ as the 'final resting place' of all future directed null geodesics. More precisely, the chronological past $I^-(\mathcal{I}^+)$ of $\mathcal{I}^+$ is the entire Minkowski space[2].

Given a general asymptotically flat space-time $(\mathcal{M}, g_{ab})$, one first finds the chronological past $I^-(\mathcal{I}^+)$ of $\mathcal{I}^+$. If it is not the entire space-time, then there is a region in $(\mathcal{M}, g_{ab})$ from which one cannot send causal signals to infinity. When this happens, one says that the space-time admits a black hole. More precisely, *black-hole region* $\mathcal{B}$ of $(\mathcal{M}, g_{ab})$ is defined as

$$\mathcal{B} = \mathcal{M} \setminus I^-(\mathcal{I}^+)\,, \tag{6}$$

where the right side is the set of points of $\mathcal{M}$ which are not in $I^-(\mathcal{I}^+)$. The boundary $\partial\mathcal{B}$ of the black hole region is called the *event horizon* and is denoted by $\mathcal{E}$ [6]. $I^-(\mathcal{I}^+)$ is often referred to as the asymptotic region and $\mathcal{E}$ is the boundary of this region within physical space-time[3]. I would like to emphasize that these notions use only asymptotic flatness and causal structure of the physical space-time. Since we have not introduced a Killing vector anywhere in the discussion, there is no restriction to stationarity. In particular, the event horizon $\mathcal{E}$ is not the set of points at which some Killing vector becomes null. Logically, it is distinct from the Killing horizon even when a Killing vector exists[4].

---

[2] $I^-(\mathcal{I}^+)$ is the set of all points in the physical space-time from which there is a future directed time-like curve to a point on $\mathcal{I}^+$ in the conformally completed space-time. The term 'chronological' refers to the use of time-like curves. A curve which is everywhere time-like or null is called 'causal'.

[3] By a time reversal, i.e. by replacing $\mathcal{I}^+$ with $\mathcal{I}^-$ and I^- with I^+, one can define a white hole region $\mathcal{W}$. However, here I will consider only black holes.

[4] The only general relation between Killing vectors and event horizons is that if g_{ab} does admit a Killing vector, it must necessarily be tangential to $\mathcal{E}$. This follows from the fact that, since the event horizon is determined solely by the space-time metric, it is left invariant by every isometry of g_{ab}.

If a black hole region $\mathcal{B}$ exists, there are several *a priori* possibilities. A space-time may contain a single black hole while another may contain multiple black holes. At first glance one would also expect that one or more black holes may coalesce to form a single black hole in the distant future or a single black hole may bifurcate, giving rise to multiple black holes in the future. To obtain a precise formulation of these ideas and analyse if these possibilities are actually realised, one has to consider space-times in which there is a well-defined representation of 'instants of time'.

An appropriate notion is that of a *Cauchy surface* defined as follows: A 3-dimensional sub-manifold $\bar{M}$ of $\mathcal{M}$ is said to be achronal if no two points on it can be joined by a causal curve. A Cauchy surface $\bar{M}$ is an achronal sub-manifold of $(\mathcal{M}, g_{ab})$ such that every inextendible causal curve in $(\mathcal{M}, g_{ab})$ intersects $\bar{M}$ once and only once. A Cauchy surface represents 'an instant of time'. A space-time which admits a Cauchy surface is said to be globally hyperbolic. In such space-time, it is possible to predict the entire future and past using information specified only on the Cauchy surface $\bar{M}$. Finally, a globally hyperbolic space-time can be foliated by a family of Cauchy surfaces, each representing an instant of time.

Fix a globally hyperbolic space-time $(\mathcal{M}, g_{ab})$. Suppose it admits a black hole region $\mathcal{B}$. The intersection of $\mathcal{B}$ with a Cauchy surface $\bar{M}$ may have several disjoint components. Each then represents a separate black hole at that instant of time. With this general framework at hand, we can now ask for dynamics of black holes. How do they evolve? Can black holes merge? Can they bifurcate? If $\bar{M}'$ is a Cauchy surface to the future of $\bar{M}$, one can show that the number of disjoint components of $\bar{M}' \cap \mathcal{B}$ in the causal future of $\bar{M} \cap \mathcal{B}$ must be less than or equal to those of $\bar{M} \cap \mathcal{B}$ [3]. In this sense, then, black holes can merge but cannot bifurcate.

The next question one can ask is that of time evolution of the event horizon $\mathcal{E}$. Here, there is a striking result due to Hawking [3,6,7] which brought out an unforeseen relation between general relativity and thermodynamics. He showed that, under time evolution, the area of the event horizon cannot decrease. The horizon area is thus analogous to entropy. The Raychaudhuri equation discussed in §1 lies at the heart of this result and is thus an integral part of the foundation on which the rich literature on black hole thermodynamics rests.

Let $(\mathcal{M}, g_{ab})$ admit a black hole region B. As noted above, the event horizon $\mathcal{E}$ is a null 3-surface. Let us suppose that we can introduce a local coordinate v such that a small piece of $\mathcal{E}$ is given by $v = \text{const}$. Set $\ell_a = \nabla_a v$. Then ℓ_a is normal to $\mathcal{E}$ and, since $\mathcal{E}$ is null it is also tangential to $\mathcal{E}$. By definition, $\nabla_{[a}\ell_{b]} = 0$ and, on $\mathcal{E}$, $\ell^a \ell_a = 0$. Hence for any vector field t^a tangential to $\mathcal{E}$, we have $t^b \ell^a \nabla_a \ell_b = t^b \ell^a \nabla_b \ell_a = (1/2)\, t^b \nabla_b \ell \cdot \ell = 0$. Thus ℓ^a is a geodesic vector field. While this explicit and pedagogical argument assumes the existence of a function v with certain properties, this assumption is not necessary (and is not satisfied by the portion of $\mathcal{E}$ depicting a merger). Using general facts about causal structure one can show that $\mathcal{E}$ is a null 3-surface, ruled by future inextendible null geodesics whose expansion cannot become infinite at any point of $\mathcal{E}$ [3,4,6].

As in §1, let us denote the expansion of ℓ^a by Θ. As we saw, an immediate consequence of the Raychaudhuri equation (4) is that if Θ were to become negative at a point p of $\mathcal{N}$, it would become infinite within a finite affine parameter on the null geodesic through p. From the property of $\mathcal{E}$ just mentioned, this cannot happen if geodesics generated by ℓ^a are assumed to be complete [7]. Under this assumption then Θ must be non-negative everywhere on $\mathcal{N}$. Now suppose a cross-section S_2 of $\mathcal{E}$ is to the future of a cross-section S_1. Then, because of the relation between the expansion Θ of ℓ^a and area elements of cross-sections discussed in §1, we must have $a_{S_2} \geq a_{S_1}$. Thus, in any dynamical process the change Δa in the horizon area is always non-negative. This result is known as the *second law of black hole mechanics.* It is analogous to the second law of thermodynamics, the horizon area playing the role of entropy. Note however that the black hole is not treated as a closed system and, in the context of classical relativity considered here, area is defined without reference to any coarse graining.

This argument, originally given in [7], assumes that the geodesics ruling the event horizon are complete. This assumption can be replaced by another with direct physical interpretation [3,6]. Heuristically the new assumption is that there are no naked singularities in the asymptotic region $I^-(\mathcal{I}^+)$. More precisely, one assumes that there $\mathcal{M}$ admits a globally hyperbolic region containing $I^-(\mathcal{I}^+) \cup \mathcal{E}$. Now suppose that Θ becomes negative on some cross-section S of $\mathcal{E}$. Then, there exists a slight outward deformation S' of S which lies in the asymptotic region $I^-(\mathcal{I}^+)$ on which Θ is negative. Now one can again use the Raychaudhuri equation and properties of $I^-(\mathcal{I}^+)$ to arrive at a contradiction. Thus, Θ must be non-negative on $\mathcal{E}$. The argument of the last paragraph now implies the second law of black hole mechanics.

To summarize, the Raychaudhuri equation is a key ingredient in the derivation of the result that the area of the event horizon cannot decrease as it 'evolves to the future'.

3. Quasi-local horizons

3.1 *Limitations of the notion of an event horizon*

Event horizons and their properties have provided considerable insight into the dynamics of black holes. However, the notion has two severe limitations.

First, while the notion neatly encodes the idea that asymptotic observers cannot 'look into' a black hole, it is too global for many applications. For example, since it refers to null infinity, it cannot be used in spatially compact space-times. Clearly, one should be able to analyze black hole dynamics also in such space-times. Asymptotic flatness and the notion of $\mathcal{I}^+$ is used also in other contexts, in particular to discuss gravitational radiation in full, non-linear general relativity [5,8]. However, there $\mathcal{I}^+$ is used to facilitate the imposition of boundary condition and make notions such as '$1/r^n$-fall-off' precise. Concepts such as the 'Bondi-mass' and explicit expressions, e.g., of fluxes of gravitational energy across $\mathcal{I}^+$ are completely insensitive to geometry in the space-time interior. Situation with event horizons is quite different because they refer to the *full chronological past of* $\mathcal{I}^+$. As a consequence, by changing the geometry in a small – say Planck scale region – around the singularity, one can change the event horizon dramatically and even make it disappear [9]! And there is no reason to trust the space-time geometry provided by classical general relativity very close to the singularity. These considerations cast a long shadow on the physical meaning and relevance of event horizons.

Second, the notion is teleological; let us speak of a black hole *only after we have constructed the entire space-time.* Thus, for example, an event horizon may well be developing in the room you are now sitting *in anticipation* of a gravitational collapse that may occur in this region of our galaxy a million years from now. Clearly, when astrophysicists say that they have discovered a black hole in the center of our galaxy, they are referring to something much more concrete and quasi-local than an event horizon.

Because of these features, the notion of an event horizon is often not so useful in practice. A striking example is provided by numerical simulations of gravitational collapse leading to the formation of a black hole, or, of binary black holes which merge. Here the main task is to *construct* the space-time describing black hole dynamics. As one evolves initial data, one needs to know, at each time step, if a black hole has formed and, if so, where it is. Since space-time is the *end-product* of this procedure the event horizon cannot be located even in principle until after the simulation is complete. This teleological nature of the event horizon makes it totally unsuitable to detect black holes *during* these simulations. One needs another, quasi-local notion that can sense the presence of the black hole at each time step and home-in on its approximate location.

Finally, let us return to Hawking's area law discussed in the last section. Because of its similarity to the second law of thermodynamics, it has been extremely influential in fundamental physics. However, it is only a 'qualitative' result; it does not provide an explicit formula for the amount by which the area increases in any given physical process. One might hope that the change in area is related, in a direct manner, to the flux of matter fields and gravitational radiation falling into the black hole. Although this expectation seems natural on physical grounds, it is *impossible* to establish such a result for event horizons. For, the event horizons $\mathcal{E}$ can form and grow in a flat region of space-time in anticipation of a future collapse. *In this region, its growth occurs even though there is absolutely nothing falling across* $\mathcal{E}$. This is not just a qualitative idea but can be readily realized in general relativity. An explicit example is provided by the Vaidya solution where a spherically symmetric null fluid falls in from past null infinity $\mathcal{I}^-$ and collapses to form a black hole (see figure 1). In the distant past, before the in-fall has begun, space-time is flat and, if the in-fall lasts only for a finite time, in the distant future it is isometric to the Schwarzschild space-time. In between is the dynamical region during which the black hole forms and grows in mass. As figure 1 shows, the event horizon first forms and grows in the flat part of the space-time.

These considerations motivate the introduction and analysis of alternate, quasi-local notions to describe the non-linear dynamics of black holes. The rest of this section will be devoted to this task.

3.2 *Marginally trapped surfaces and tubes*

Since analytical solutions to Einstein's equations encoding dynamical black holes are only handful and represent idealised situations, bulk of the work in this area

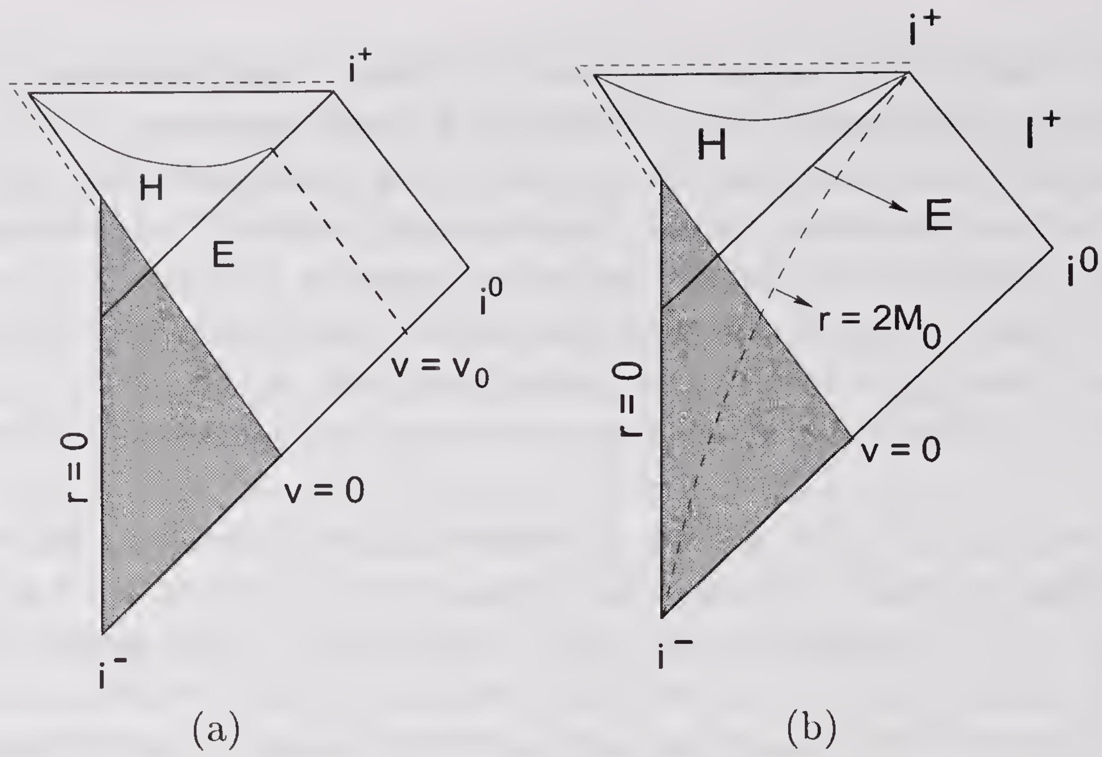


Figure 1. Black hole formation through the gravitational collapse of a null fluid. The figures show Penrose diagrams of Schwarzschild–Vaidya metrics for which there is no incoming radiation from $\mathcal{I}^-$ for retarded time $v \leq 0$. Space-time is flat in the past of $v = 0$ (i.e., in the shaded region). In figure (**a**), the incoming flow stops at a retarded time $v = v_o$. Space-time is isometric to the Schwarzschild space-time to the future of $v = v_o$. The space-like dynamical horizon H and the event horizon E meet tangentially at $v = v_o$. In figure (**b**) the incoming radiation continues till $v = \infty$ (but tapers off so that the total incoming energy is finite). Thus, the energy flux into the black hole mimics a more realistic collapse. The space-like dynamical horizon H, the null event horizon E and the time-like surface $r = 2M_o$ (represented by the dashed line) all meet tangentially at i^+. In both figures, the event horizon originates in the shaded flat region while the dynamical horizon exists only in the curved region.

uses numerical techniques. Let me begin by sketching the standard procedure used in these simulations to get out of the quandary described in §3.1.

The first notion we need is that of marginally trapped 2-surfaces. In §1 we saw that if we instantaneously illuminate a compact, space-like 2-surface S in Minkowski space, there are two light fronts: the outgoing ones sweep out an expanding null surface $\mathcal{N}^+$ and the ingoing ones sweep out a contracting null surface $\mathcal{N}^-$. As before, let us use the convention that ℓ^a are the outgoing ones, future directed null normal to S and n^a the ingoing ones. Further, let us now normalize these vectors by demanding $\ell^a n_a = -2$. Since ℓ^a is normal to light fronts which are expanding, its expansion is positive $(\Theta_{(\ell)} > 0)$ and since n^a is normal to contracting light fronts, its expansion is negative $(\Theta_{(n)} < 0)$. Since light can escape from S to infinity, S is said to be *untrapped.* In Minkowski space, every 2-surface S is untrapped.

By contrast, black hole is a region from which light cannot escape. In this case, both light fronts would be ingoing, i.e. expansions of both null normals would be negative. This leads us to the first definition. A compact 2-surface S in a space-time $(\mathcal{M}, g_{ab})$ is said to be (outer) *trapped* if $\Theta_{(n)} < 0$ *and* $\Theta_{(\ell)} < 0$. The limiting case – which separates the trapped and untrapped 2-surfaces – occurs when the outgoing normal ℓ^a has zero expansion while the ingoing normal n^a has negative expansion, as usual. This brings us to the notion we set out to define. A compact

2-surface S is said to be (future, outer) *marginally trapped* if $\Theta_{(\ell)} = 0$ and $\Theta_{(n)} < 0$ on it[5].

Note that notions of trapped and marginally trapped surfaces are quasi-local. They do require us to distinguish between outgoing and incoming normals. But apart from this qualitative property, the notions refer only to geometry in the immediate neighborhood of S. They do not depend on asymptotic flatness, causal structure or, indeed, any property of the rest of the space-time. Thus, unlike the event horizon, these notions are neither global nor teleological. In particular you can be rest assured that there is no trapped or marginally trapped surface in the room you are now working in.

The idea in numerical simulations is to track the dynamics of black holes by following the evolution of marginally trapped surfaces. One generally uses the initial value formulation of Einstein's equations which can be summarized as follows. Let us consider just source-free general relativity since inclusion of sources does not add new conceptual elements. Fix a 3-manifold $\bar{M}$, topologically R^3. The initial data for general relativity consists of a pair $(\bar{q}_{ab}, \bar{K}_{ab})$ on $\bar{M}$ where $\bar{q}_{ab}$ is a metric of signature +,+,+ and $\bar{K}_{ab}$ is a symmetric tensor field. These fields have to satisfy constraints – the time-time and space-time components of Einstein's equations. In numerical simulations this is typically ensured using elliptic solvers (since the constraints are elliptic equations). The space–space part of Einstein's equations are then used to evolve these data numerically. The result is a 1-parameter family of pairs $(\bar{q}_{ab}(t), \bar{K}_{ab}(t))$. Next, consider the 4-manifold $\mathcal{M} := \bar{M} \times \mathbb{R}$ which, by construction is naturally foliated by copies of $\bar{M}$, denoted $\bar{M}_t$, each representing an 'instant of time'. The existence and uniqueness theorems for Einstein's equations imply that there is a 4-metric g_{ab} on $\mathcal{M}$ of signature −,+,+,+ such that each $\bar{M}_t$ is a Cauchy slice of $(\mathcal{M}, g_{ab})$ with $\bar{q}_{ab}(t)$ as its positive definite metric and $\bar{K}_{ab}(t)$ its extrinsic curvature.

To implement this procedure in the black hole context, one must first identify in the initial data one or several black holes. As I emphasized before, one cannot use the notion of the event horizon because its definition is teleological. *The idea is to use the quasi-local notion of marginally trapped surfaces.* But even this strategy seems difficult at first. For, as we just saw, the notion refers to the expansions $\Theta_{(\ell)}$ and $\Theta_{(n)}$ of null normals and these quantities seem to require the knowledge of the space-time metric which one does not yet have. However, it turns out possible to recast the definitions using just the Cauchy data. Let us first consider a space-time $(\mathcal{M}, g_{ab})$ with a Cauchy slice $\bar{M}$ admitting a marginally trapped surface S. Then, if $\bar{t}^a$ denotes the unit time-like normal to $\bar{M}$ and $\bar{r}^a$ the unit outward normal to S within $\bar{M}$, we can express the two null normals as $\ell^a = \bar{t}^a + \bar{r}^a$ and $n^a = \bar{t}^a - \bar{r}^a$. Then, using the definition of $\Theta_{(\ell)}$ and $\Theta_{(n)}$ and the expressions

$$\bar{K}_{ab} = \bar{q}_a{}^m \bar{q}_b{}^n \nabla_m \bar{t}_n \qquad \text{and} \qquad \tilde{K}_{ab} = \tilde{q}_a{}^m \tilde{q}_b{}^n \nabla_m \bar{r}_n \tag{7}$$

[5] I have included the parenthetical adjectives 'future' and 'outer' to facilitate comparisons with the more detailed discussions of [10–15]. They can be ignored for the rest of this article.

of the extrinsic curvature $\bar{K}_{ab}$ of $\bar{M}$ in $\mathcal{M}$ and of the extrinsic curvature $\tilde{K}_{ab}$ of S in $\bar{M}$ it follows immediately that

$$\Theta_{(\ell)} = \tilde{q}^{ab}(\bar{K}_{ab} + \tilde{K}_{ab}) = \bar{K} - \bar{r}^a \bar{r}^b \bar{K}_{ab} + \tilde{K}, \tag{8}$$

$$\Theta_{(n)} = \tilde{q}^{ab}(\bar{K}_{ab} - \tilde{K}_{ab}) = \bar{K} - \bar{r}^a \bar{r}^b \bar{K}_{ab} - \tilde{K}. \tag{9}$$

Therefore, given the Cauchy data $(\bar{M}, \bar{q}_{ab}, \bar{K}_{ab})$, it is possible to search for 2-surfaces S on which $\Theta_{(\ell)} = 0$ and $\Theta_{(n)} < 0$. In practice one restricts oneself to the physically interesting case of 2-surfaces which are topologically S^2. There may be one or more such surfaces on $\bar{M}$. Each is taken to represent a black hole at the 'initial instant of time' represented by $\bar{M}$. As explained above, evolution produces a 1-parameter family of Cauchy slices $\bar{M}_t$ and Cauchy data thereon. One then searches for marginally trapped surfaces on each $\bar{M}_t$. Efficient numerical programs – called apparent horizon finders – are available to find these surfaces (see [16] and references therein). The 3-dimensional world-tubes obtained by 'stacking' these marginally trapped 2-surfaces depict the dynamical evolutions of interest.

This brings us to the next definition. *A marginally trapped tube M is a 3-dimensional sub-manifold (possibly with boundary) of $(\mathcal{M}, g_{ab})$ which admits a foliation by marginally trapped surfaces S.* Note that this definition does not refer to a space-time foliation, i.e., does not require that M be generated by a Cauchy evolution. The marginally trapped tube is a 3-dimensional sub-manifold in its own right. This abstraction removes the excess baggage and makes the notion conceptually simpler to use [2,11,12,15] (see §3.3). However, the marginally trapped tubes that arise in numerical relativity do result from space-time foliations as explained above.

Are there any constraints on the marginally trapped tubes that a space-time can admit? These would provide insights into how black holes evolve in full general relativity. The first natural question is: What is the signature of the intrinsic metric on these tubes? Are they time-like, null or space-like? One's first impulse would be to say that they must be time-like because they are obtained by time evolution of space-like 2-surfaces S. However, this expectation is too naive. Take Minkowski space and consider the null cone C of a point p or the 'mass-shell' H based at p (i.e. spanned by all position vectors X relative to p satisfying $X \cdot X = -1$). If we slice Minkowski space by $t =$ constant planes, we obtain a foliation of C and H by 2-spheres, which can be thought of as related by time evolution. Yet, C is null and H is space-like. Similarly, marginally trapped tubes can have any signature (see figure 2).

In fact the signature is determined by a mixture of physics and geometry through the *Raychaudhuri equation*. Let M be a marginally trapped tube in $(\mathcal{M}, g_{ab})$ and denote by V^a a vector field which is tangential to M and everywhere orthogonal to its foliation by marginally trapped surfaces S and which furthermore preserves this foliation. Since V^a is orthogonal to the leaves S, there exists a function f such that $V^a = \ell^a - fn^a$. Since $\ell \cdot n = -2$ we have $V \cdot V = 4f$. Therefore M is respectively, space-like, null or time-like, depending on whether f is positive, zero or negative. Now, the definition of V^a immediately implies $\mathcal{L}_V \Theta_{(\ell)} = 0$, whence, $\mathcal{L}_\ell \Theta_{(\ell)} = f \mathcal{L}_n \Theta_{(\ell)}$. Therefore, the Raychaudhuri equation (4) for ℓ^a implies

$$f\,\mathcal{L}_n\,\Theta_{(\ell)} = -\tilde{q}^{ab}\tilde{q}^{mn}\sigma_{am}\sigma_{bn} - R_{ab}\ell^a\ell^b, \tag{10}$$

where, as before, σ_{ab} is the shear of ℓ^a (and ℓ^a is extended off S using the geodesic equation to evaluate $\mathcal{L}_n\Theta_\ell$). Because the right side is negative definite, this equation can be used to determine the sign of f and hence the signature of M [10].

The shear term is a measure of the energy carried by gravitational waves across M [1,2]. Similarly, through Einstein's equation, the second term on the right-hand side of (10) can be interpreted as a measure of the flux of matter energy across M [1,2]. (In both cases, 'energy' is defined with respect to ℓ^a.) Thus, f vanishes – i.e., M is null – only if there is no flux of matter or gravitational energy across M. Such a null marginally trapped tube is called *a non-expanding horizon* [12,17]. As the name suggests, in this case the area of any cross-section S of M is the same; the horizon does not expand or contract. The null normal ℓ^a to 2-spheres S is now also tangential to M. It is easy to show that the intrinsic metric on M (of signature 0,+,+) is Lie dragged by ℓ^a. Thus the intrinsic geometry of the marginally trapped tube is 'time independent', whence the tube represents a black hole *in equilibrium.* In every stationary black hole space-time the event horizon $\mathcal{E}$ is a non-expanding horizon. (Because of stationarity we are trivially spared of teleology.) However, there also exist more general examples. A class of Robinson–Trautman metrics provides an instructive illustration [18]. This space-time admits a non-expanding horizon but the space-time is dynamical and carries gravitational radiation in every of its neighborhoods, although none falls across the horizon! This example illustrates how surprising features can arise even in exact solutions. Non-expanding horizons also emerge as the future asymptotic limits in situations which are of direct physical interest. When black holes form through gravitational collapse, or pre-existing black holes merge, the marginally trapped tube asymptotes to a non-expanding horizon in the distant future [2,12]. In numerical simulations, this limit is reached rather rapidly [14] because the back-scattered radiation quickly becomes so small that it is indistinguishable from zero within numerical precision. Finally, the vanishing of right side of (10) is only a necessary condition for M to be a non-expanding horizon because so far we have allowed for the possibility that $\mathcal{L}_n\Theta_\ell$ itself may vanish. However, in numerical simulations one does not consider all mathematically viable possibilities but restricts oneself to physically interesting situations. In all simulations I am aware of the condition that has also proved to be sufficient: The marginally trapped tubes become null as soon as the right side of (10) vanishes.

If the right side of (10) is non-zero, on the other hand, M can be either time-like or space-like. Since the right side is negative definite, the sign of f is determined by whether $\mathcal{L}_n\Theta_{(\ell)} > 0$ or $\mathcal{L}_n\Theta_{(\ell)} < 0$. In the first case, an infinitesimal inward deformation along n^a of the marginally trapped surfaces S on M would produce an untrapped surface while in the second case it would produce a trapped surface. If M is time-like, it is called a *time-like membrane* (TLM). In this case M does not have the connotation of a horizon because light can travel from either of its sides to the other. If M is space-like, it is called a *dynamical horizon.* Both these possibilities are realized in numerical simulations [13,14].

The general situation is captured in a recent mathematical existence result [19]. Let $(\mathcal{M}, g_{ab})$ be a globally hyperbolic, smooth space-time on which the null energy condition holds. Let a Cauchy surface M admits a marginally trapped surface S

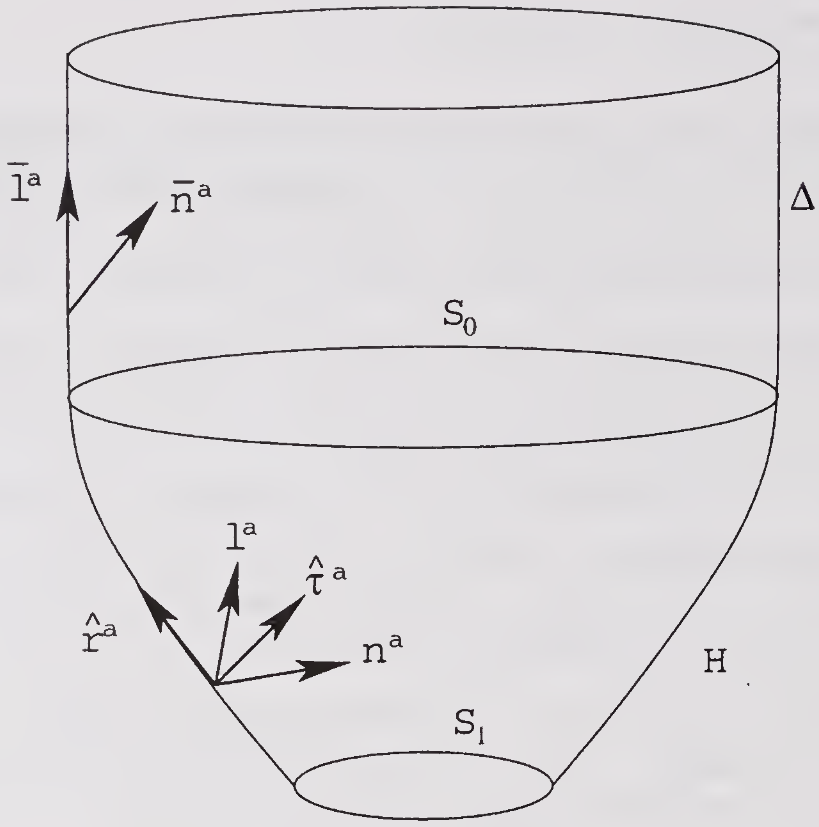


Figure 2. H is a dynamical horizon, foliated by marginally trapped surfaces S. $\hat{\tau}^a$ is the unit time-like normal to H and $\hat{r}^a$ the unit space-like normal within H to the foliations. Although H is space-like, motions along $\hat{r}^a$ can be regarded as time evolution with respect to observers at infinity. In this respect, one can think of H as a hyperboloid in Minkowski space and S as the intersection of the hyperboloid with space-like planes. In the asymptotic future H joins on to a non-expanding horizon Δ, mimicking the situation in numerical relativity.

at which the right side of (10) is non-zero. Suppose furthermore that S is stable in the sense that there exists an infinitesimal outward deformation which makes S untrapped. Then, S 'evolves' to a marginally trapped tube M which is space-like (at least) in a neighborhood of S. Note that this conclusion is arrived at assuming only 'stability' which, in practice, is substantially weaker than the condition $\mathcal{L}_n\Theta_{(\ell)} < 0$ referred to above. In particular, even when the right side of (10) is non-zero on a small portion of the 2-sphere S, M is guaranteed to be space-like 'all the way around' in a neighborhood of S (including the region where the right side may vanish).

In numerical simulations the following behavior has been observed [13,14]. Marginally trapped tubes appear to be smooth. In a gravitational collapse of a star, there are no marginally trapped surfaces in the distant past. Suppose the first marginally trapped surface appears on the Cauchy surface M_0 corresponding to $t = t_0$. Then in its immediate future, the marginally trapped tube has a cusp – or, an '$\bigcup$ shape' – each point of $\bigcup$ denoting a marginally trapped 2-sphere. As time evolves, the outer branch expands and the inner branch contracts. Both branches are space-like to begin with. The outer branch remains space-like – i.e., is a dynamical horizon – and asymptotes to a non-expanding horizon in the distant future. For black hole dynamics, this is the physically interesting component. The inner branch becomes time-like – i.e., a time-like membrane – after a while. While its time-evolution is interesting in its own right, as mentioned above, it is not of direct significance for black hole dynamics. In black hole mergers, in the asymptotic past when the black holes are very far apart there are two distinct marginally trapped tubes which are null. They become space-like as the black holes approach each other. On some Cauchy surface M_o a new marginally trapped surface forms enclosing the two distinct ones. The evolution of this common marginally trapped tube follows the same steps as the marginally trapped tube formed in a gravitational collapse, described above. In particular, the physically interesting outer branch of the tube is a dynamical horizon and it asymptotes a non-expanding horizon in the distant future.

3.3 *Dynamical horizons*

Discussion of §3.2 suggests that a quantitative handle on black hole dynamics can be obtained by examining the evolution of dynamical horizons in detail. This task is now being carried out in some detail in numerical simulations. In particular, programs are now available to monitor the change in the multipole moments [20] of the horizon geometry and analyze their approach to the asymptotic values corresponding to the Kerr horizon [14]. In this section I will conclude with an analytical result. We saw in §3.1 that while Hawking's theorem tells us that the area of the event horizon $\mathcal{E}$ cannot decrease, the teleological nature of $\mathcal{E}$ makes it impossible to relate the growth to the matter and gravitational waves falling into the black hole. Dynamical horizons on the other hand are well suited to obtain such a relation. Since we seek a *quantitative* relation, the Raychaudhuri equation will no longer suffice. Now we have to use Einstein's equations in detail.

Let us begin by fixing notation. Let $(\mathcal{M}, g_{ab})$ be a space-time satisfying the dominant energy condition. Thus, given any future directed causal vector V^a, the stress-energy T_{ab} is such that $-T^a{}_bV^b$ is also a future directed causal vector. Let $\mathcal{H}$ be a dynamical horizon, foliated by marginally trapped surfaces S. One can show that the foliation is unique [21]. Let $q_a{}^b$ and $\tilde{q}_a{}^b$ denote the projection operators on the tangent spaces to $\mathcal{H}$ and S respectively, $\hat{\tau}^a$ the unit future pointing normal to $\mathcal{H}$ and $\hat{r}^a$ the unit outward normal within $\mathcal{H}$ to the marginally trapped surfaces S. We first note an immediate consequence of the definition. Since $2\hat{r}^a = \ell^a - n^a$, $\Theta_{(\ell)} = 0$ and $\Theta_{(n)} < 0$ on each S, it follows that

$$\tilde{K} = \tilde{q}^{ab} D_a \hat{r}_b = \frac{1}{2}\tilde{q}^{ab}\nabla_a(\ell_b - n_b) > 0, \tag{11}$$

where D is the intrinsic derivative operator on $\mathcal{H}$, compatible with q_{ab}. Since the trace $\tilde{K}$ of the extrinsic curvature of S measures the rate of change of its area along $\hat{r}^a$, (11) implies that the area of marginally trapped surfaces increases monotonically. This is the dynamical horizon analog of Hawking's area law for event horizons. A more non-trivial task is to obtain an explicit expression for the change of area.

For this one has to use constraints – i.e., the time-time and space-time parts of Einstein's equations – on $\mathcal{H}$. Consider the portion $\Delta\mathcal{H}$ of $\mathcal{H}$ bounded by two marginally trapped surfaces S_1 and S_2. A series of rather straightforward manipulations of constraint equations leads to a surprising relation [2]:

$$\left(\frac{R_2}{2G} - \frac{R_1}{2G}\right) = \int_{\Delta H} T_{ab}\hat{\tau}^a N\ell^b \, \mathrm{d}^3V + \frac{1}{16\pi G} \times \int_{\Delta H} N \left\{\tilde{q}^{am}\tilde{q}^{bn}\, \sigma_{ab}\sigma_{mn} + 2\tilde{q}_{ab}\zeta^a\zeta^b\right\} \mathrm{d}^3V =: \mathcal{F}_{\mathrm{matt}} + \mathcal{F}_{\mathrm{gw}}. \tag{12}$$

Here $\zeta^a = \tilde{q}^{am}\hat{r}^b\nabla_b\ell_a$ is a vector field tangential to S; R is the 'area-radius', the area a_S of any marginally trapped surface being given by $a_S = 4\pi R^2$; and $N = |DR|$. Equation (12) provides an explicit, quantitative formula, relating the area change to processing happening at $\mathcal{H}$. The striking feature of (12) is that the integrand of each term on the right side is positive definite.

Let us now interpret the various terms appearing in this equation. The left side gives us the change in the horizon 'radius' caused by the dynamical process under which the horizon evolves from S_1 to S_2. The first integral on the right side of this equation is the flux $\mathcal{F}_{\mathrm{matt}}$ of matter energy (associated with the vector field $N\ell^a$) across $\Delta\mathcal{H}$. Since $N\ell^a$ is null and $\hat{\tau}^a$ time-like, by dominant energy condition this quantity is guaranteed to be non-negative. Since the second term is purely geometrical and emerged as the 'companion' of the matter term, it is natural to interpret it as the flux $\mathcal{F}_{\mathrm{gw}}$ of energy in the gravitational waves across $\Delta\mathcal{H}$. The presence of the shear term $|\sigma|^2$ in the flux formula (12) is expected from weak field considerations. The term $|\zeta|^2$, on the other hand, is two orders higher than the shear term in the weak field expansion [22]. Thus, it captures some genuinely non-linear, strong field physics. The interpretation of the last two, geometric terms as flux of energy carried by gravitational waves is supported by several independent considerations [2,12].

To summarize, it is remarkable that Einstein's equations provide a direct local relation between the growth of area of a dynamical horizon $\mathcal{H}$ and the flux of in-falling matter and gravitational waves. Furthermore, as a by-product a gauge invariant expression emerged for the energy-flux of gravitational waves across $\mathcal{H}$. As at null infinity, it is local to the cross-section and positive definite. This is surprising because in general relativity no such expressions exists for general 3-surfaces in the strong field regime. Indeed, had $\mathcal{H}$ been replaced by a 3-manifold which is foliated by either untrapped *or* trapped 2-surfaces, the procedure would fail to provide a positive definite expression. The fact that a satisfactory expression exists precisely when we try to calculate the energy-flux falling into a black hole is another illustration of the extraordinary interplay between geometry and physics contained in general relativity.

4. Discussion

Although a systematic mathematical theory of black holes was introduced over thirty years ago, in the ensuing decades most of the attention was focused on stationary black holes, specifically on properties of the Kerr family and perturbations thereof. In the fully dynamical regime, we only had qualitative results involving event horizons, most of them dating back to the seventies. While they spurred much activity in the area of black hole thermodynamics and associated fundamental physics, they have not been directly useful in unravelling the quantitative physics of black hole formation and growth. Furthermore, because of the global and teleological nature of event horizons these results have had limited practical applications, e.g., to numerical simulations.

Over the last three years this status-quo has begun to change. In particular, new properties of marginally trapped surfaces and tubes have been discovered, motivating the introduction of quasi-local horizons. These in turn are now providing detailed insights into the non-linear processes dictating black hole dynamics. As with event horizons, the Raychaudhuri equation plays a key role in determining the qualitative features of marginally trapped tubes. However, one can finally go beyond and use the detailed form of Einstein's equations to track, in quantitative detail, how black holes form and grow. This endeavour continues Professor Raychaudhuri's tradition of uncovering and exploiting new facets of the extraordinary interplay between geometry and physics that is hidden in general relativity.

Over the last three years, rapid progress could occur largely because of a synergistic interchange between mathematical and numerical relativity communities. If this interplay continues, the subject of dynamical black holes is likely to enrich both areas. By uncovering new structures and their properties, these efforts could further refine and clarify the intuitive scenarios that have emerged over the past three decades. More importantly, they could continue to open new avenues, leading to the discovery of unforeseen features and forcing us to revise older scenarios in important ways.

Acknowledgments

This work was supported in part by the NSF grant PHY-0456913, the Alexander von Humboldt Foundation, the Krammers Chair program of the University of Utrecht and the Eberly research funds of Penn State.

References

[1] A Ashtekar and B Krishnan, Dynamical horizons: energy, angular momentum, fluxes and balance laws, *Phys. Rev. Lett.* **89**, 261101 (2002)
[2] A Ashtekar and B Krishnan, Dynamical horizons and their properties, *Phys. Rev.* **D68**, 104030 (2003)
[3] S W Hawking and G F R Ellis, *Large scale structure of space-time* (Cambridge University Press, Cambridge, 1973)
[4] R M Wald, *General relativity* (University of Chicago Press, Chicago, 1984)
[5] R Penrose, Zero rest mass fields including gravitation, *Proc. R. Soc. (London)* **284**, 159 (1965)
[6] S W Hawking, *The event horizon in black holes* edited by B S DeWitt and C M DeWitt (North-Holland, Amsterdam, 1972)
[7] S W Hawking, Gravitational radiation from colliding black holes, *Phys. Rev. Lett.* **26**, 1344 (1971)

[8] A Ashtekar, *Asymptotic quantization* (Bibliopolis, Naples, 1987), also available at http://igpg.gravity.psu.edu/research/asymquant-book.pdf
[9] P Hajicek, Origin of Hawking radiation, *Phys. Rev.* **D36**, 1065 (1987)
[10] S Hayward, General laws of black hole dynamics, *Phys. Rev.* **D49**, 6467 (1994), gr-qc/9306006
[11] B Krishnan, Isolated horizons in numerical relativity, Ph.D. Dissertation (The Pennsylvania State University, 2002)
[12] A Ashtekar and B Krishnan, Isolated and dynamical horizons and their applications, *Living Rev. Rel.* **10**, 1 (2004), gr-qc/0407042
[13] I Booth, B Brits, J A Gonzalez and C Van Den Broeck, Marginally trapped tubes and dynamical horizons, *Class. Quantum Grav.* **23**, 413 (2006), gr-qc/0506119
[14] E Schnetter, B Krishnan and F Beyer, Introduction to dynamical horizons in numerical relativity, *Phys. Rev.* **D74**, 024028 (2006), gr-qc/0604015
[15] I Booth, Black hole boundaries, *Can. J. Phys.* **83**, 1073 (2005), gr-qc/0508107
[16] J Thornberg, A fast apparent horizon finder for 3-dimensional Cartesian grids in numerical relativity, *Class. Quantum Grav.* **21**, 743 (2004)
[17] A Ashtekar, C Beetle and J Lewandowski, Geometry of generic isolated horizons, *Class. Quantum Grav.* **19**, 1195 (2002), gr-qc/0111067
[18] P T Chruściel, On the global structure of Robinson–Trautman space-time, *Proc. R. Soc. (London)* **A436**, 299 (1992)
[19] L Andersson, M Mars and W Simon, Local existence of dynamical and trapping horizons, *Phys. Rev. Lett.* **95**, 111102 (2005), gr-qc/0506013
[20] A Ashtekar, J Engle, T Pawlowski and C Van Den Broeck, Multipole moments of isolated horizons, *Class. Quantum Grav.* **21**, 2549 (2004)
[21] A Ashtekar and G Galloway, Some uniqueness results for dynamical horizons, *Adv. Theo. Math. Phys.* **9**, 1 (2005), gr-qc/0503109
[22] I Booth and B Fairhurst, The first law for slowly evolving horizons, *Phys. Rev. Lett.* **92**, 011102 (2005), gr-qc/0505049

PRAMANA
— journal of
physics

Vol. 69, No. 1
July 2007
pp. 93–108

String theory and cosmological singularities

SUMIT R DAS
Department of Physics and Astronomy, University of Kentucky, Lexington, KY 40506, USA
E-mail: das@pa.uky.edu

Abstract. In general relativity space-like or null singularities are common: they imply that 'time' can have a beginning or end. Well-known examples are singularities inside black holes and initial or final singularities in expanding or contracting universes. In recent times, string theory is providing new perspectives of such singularities which may lead to an understanding of these in the standard framework of time evolution in quantum mechanics. In this article, we describe some of these approaches.

Keywords. String theory; cosmological singularities.

PACS Nos **11.25.Pm; 11.25.Tq**

1. Introduction

More than fifty years ago, Raychaudhuri initiated what is arguably one of the most exciting developments in the classical theory of gravity. The culmination of this development was the formulation and proof of singularity theorems – and taught us that the theory of general relativity in fact precisely predicts the limits of its validity. These theorems showed that singularities appear under rather general conditions and led to the question: what are we supposed to do when the space-time has a singularity? While it has been clear for a long time that the very notions of space and time needs revision near these singularities where quantum effects of gravity become important, it is still not clear what structure could replace space-time at the fundamental level. In recent years, however, string theory is beginning to provide a concrete and convincing framework for a quantitative analysis near a class of such singularities.

Some kinds of singularities are easier to understand than others. Time-like singularities are regions of space which exist for all times – and very often they do not appear mysterious once one understands that the singular nature of the space-time is caused by some object sitting there. A proper microscopic description of this object would lead to a resolution of such singularities. *Space-like* or *null* singularities are much more difficult to understand. These singularities are not located at some point – one cannot look around and *see* that they are there. Rather they just *happen*. A well-known example is the singularity inside a neutral black hole

– all observers who have crossed the horizon will encounter this singularity in finite proper time. They cannot 'look ahead' and see that there is a singularity and skirt around them. Another example is the Big Bang singularity of an expanding universe. This, too, just *happened* at some time in the past.

In string theory, the basic degrees of freedom are one-dimensional extended (rather than point-like) objects in the regime where the string coupling constant is small. The very fact that strings have an extension leads to a resolution of some kinds of singularities – a classic example is an *orbifold* singularity. Because of a finite extension, strings can happily propagate on orbifold space-times, while the time evolution of particles is necessarily singular. This in fact can – and does – happen at the *classical* level. There are other time-like singularities which can be cured in string theory at the quantum level – this too is reasonable since after all string theory is a consistent quantum mechanical description of gravity.

In this article I will describe some recent attempts to understand space-like singularities. One reason why it is difficult to understand a space-like singularity is that this signifies a 'beginning' or 'end' of time. If we believe that the laws of quantum mechanics can be applied to gravity, this is a troublesome concept. Even though the notion of time in a quantum theory of gravity is rather subtle, no one has been able to make sense of a situation where time just stops. Note that this has nothing to do with the fact that such boundaries of time are often associated with the fact that space-time curvatures diverge at these points. Consider for example normal flat space-time and we simply cut out one half of it by putting in a boundary at some time $t = 0$. This is a trivial example of a geodesically incomplete space-time. The quantum dynamics of some field in this space-time is problematic – one has to impose some final state conditions, and there is no reason why the standard Hamiltonian cannot evolve the system beyond $t = 0$. The problem becomes more pronounced if in addition the space-time has curvature singularities at initial or final times.

The key to this understanding is the fact that in string theory dynamical space-time is not fundamental, but an *emergent* concept. Over the past decade we have been able to understand the underlying structure from which space-time emerges in *certain circumstances*. Surprisingly this fundamental structure is a gauge theory without dynamical gravity. And even more suprisingly it turns out that this gauge theory lives in a *lower* number of dimensions. Thus a theory without gravity (like a gauge theory) can encode all the details of a theory with dynamical gravity which however lives in higher number of dimensions – a phenomenon which is now called *holography*.

In this article I will discuss this approach to cosmological singularities which has been reviewed in [1,2]. Instead of trying to compile an exhaustive list of references, I have cited a few review articles which should be regarded as a guide to the original literature in the field. This is by no means the only approach to space-like and null singularities in string theory. Early attempts to address the question in perturbative string theory is reviewed in [3]. Attempts to study perturbative strings in exact time-dependent backgrounds and the problems encountered therein are reviewed in [4] and [2]. More recently, there has been some progress in understanding how a phase of 'tachyon condensation' can replace space-like singularities in worldsheet formulations of string theory. This is reviewed in [5].

2. Faces of holography

The key idea evolved from the classic work of 't Hooft in the 1970s. By mid 1970s it was clear that the correct theory for strong interactions is QCD – a non-abelian gauge theory with gauge group $SU(3)$ coupled to quarks in the fundamental representation. However the quarks are confined into mesons and baryons. It was soon realized that this happens because in QCD electric fields cannot spread out – rather they form flux tubes between quarks and antiquarks, forming mesons; or a flux tube closing onto itself forming a glueball; or three flux tubes emerging from three quarks of different colors joining at a vertex, forming a baryon. This was a satisfying picture of confinement. It qualitatively explained why experimentally hadrons appeared to behave like strings – which gave rise to string theory in the first place. This begs the question: what is the coupling constant of strings, g_s in terms of QCD quantities? This should be a dimensionless parameter, and QCD does not have any free dimensionless parameter! In a landmark paper in 1974, 't Hooft argued that this dimensionless parameter can be discovered if one generalizes the gauge group in QCD from $SU(3)$ to $SU(N)$ and expand the theory in a power series expansion in $1/N$, with the combination $g_{\mathrm{YM}}^2 N$ regarded to be $O(1)$. Each term in this expansion is a sum of an infinite number of Feynman diagrams which may be thought of tiling a two-dimensional surface. If this surface has h handles and B boundaries the overall power of N is simply N^χ, where χ is the Euler characteristic $\chi = 2 - 2h - B$. This is precisely how a string perturbation theory ought to look like. A typical amplitude in string theory may be written as a sum over two-dimensional surface which represents the worldsheet of strings with each surface weighted by a factor of $g_s^{-\chi}$. Thus the string coupling constant is precisely $1/N$. In the limit $N \to \infty$ with $g_{\mathrm{YM}}^2 N$ fixed these strings are weakly coupled. A useful review of large-N may be found in [6].

Following the work of 't Hooft it was in fact realized that not just gauge theories, but any theory whose fields are in the adjoint representation of a large group like $SU(N)$ should give rise to a string theory in the large-N limit. However the precise details of this string theory turned out to be quite elusive.

Around the same time of 't Hooft's work, Yoneya, Scherk and Schwarz proposed that certain supersymmetric versions of string theory contain gravity and reduce to general relativity at low energies. The celebrated work of Green and Schwarz in 1984 indicated that such string theories in fact form quantum mechanically consistent ultraviolet completions of the supersymmetric generalization of general relativity, supergravity [7]. This indicated that if one can find descriptions of these string theories in terms of large-N gauge theories, the latter may provide a non-perturbative description of gravity itself.

Developments which took place throughout the late 1980s and 1990s led to concrete realizations of this idea – and led to a surprise. The underlying gauge theory lives in lower number of dimensions. This is now referred to as *holography* and ties up beautifully with an idea which emerged from black hole thermodynamics.

The first concrete formulation of holography was achieved in the theory of closed strings in $1 + 1$ dimensions [10–12]. Here the holographic theory is the gauged quantum mechanics of a single hermitian matrix – which is a gauge theory in $0 + 1$ dimensions. This is the first time it was realized that *space-time is an emergent* *concept*. At the same time it became clear that space-time itself is an approximate concept which breaks down in suitable circumstances. The closed string theory contains gravity (though there is no dynamical graviton) and gravitational interactions are encoded in the gauge theory in a subtle and interesting way.

The discovery of duality symmetry (for a review see [13]) and D-branes (for an introduction see e.g. [14]) led to another realization of holography. The low energy dynamics of a stack of N p-dimensional D-branes in string theory is a $U(N)$ gauge theory living on the $p+1$ dimensional world-volume. However these objects produce gravitational backgrounds. This indicates that gravitational phenomena in the *entire* space-time should be describable in this gauge theory. In fact the appropriate limit of the gauge theory is precisely 't Hooft's large-N limit. Perhaps the most striking result of this development has been the statistical explanation of black hole thermodynamics and Hawking radiation. A certain class of black holes appear as stacks of a large number of D-branes wrapped in internal directions and appear as states in the gauge theory. In certain cases, their degeneracies can be counted reliably. This leads to a statistical entropy which is in precise agreement with the results of Bekenstein and Hawking in semiclassical gravity. Furthermore, Hawking radiation of such black holes may be understood as usual quantum mechanical decay of excited states in this gauge theory and the decay rate is in precise agreement with the semiclassical luminosity. For reviews of black holes in string theory, see [15,16].

This connection between gauge theory and gravity is in fact a manifestation of a basic property of string theory which has been known since its inception: open–closed duality. In string theory processes involving open strings can also be viewed as processes involving closed strings. Now the low energy limit of open string theory is a gauge theory, while the low energy limit of closed strings contain gravity – so there should be connection here. The dynamics of D-branes are described by open strings which live on the brane – the gauge theory in the previous paragraph is in fact the low energy limit of this theory of open strings. It has been recently realized that the gauge theory – gravity connection uncovered in $1+1$ dimensions is also a result of open–closed duality. The matrix of gauged matrix quantum mechanics in fact represents the degrees of freedom of a bunch of D-particles in this theory. In the region near horizons of certain black holes, the theory of open strings can be truncated to its low energy limit – which leads to a precise connection between this gauge theory and gravity. This connection is known as the AdS/CFT correspondence since these near-horizon geometries are asymptotically anti de-Sitter space-times and the holographic gauge theory which may be thought of as living on the boundary is some deformation of a conformally invariant field theory. A useful review of this dual correspondence is [17]. Pretty much like the two-dimensional example, an extra space dimension is generated in the gauge theory, and at the same time the theory secretly contains dynamical gravity. In the AdS/CFT correspondence, the additional dimension arises out of the *renormalization group scale* of the quantum field theory. The RG equations of the field theory are essentially Raychaudhuri equation in the bulk (for connections of Raychaudhuri equation with AdS holography, see e.g. [18,19].

The open–closed duality applied to D-branes implies a rather different relation between gauge theories and closed string theories which are called 'matrix theories'. This follows from the fact that even though critical string theories live in 10 space-time dimensions, these are actually 11-dimensional theories in disguise. This 11-dimensional theory is called M-theory. String theories appear as Kaluza–Klein reductions of M-theory on a circle, the momentum along the circle appearing as usual as a charge in the 10-dimensional theory – this charge is in fact the charge associated with D0 branes. Now suppose we consider M-theory in an infinite boosted reference frame along a compact direction. In this frame the only states which survive are those which carry a quantized positive momentum J along the direction of boost. From the point of view of string theory, these are states which carry a large D0 brane charge J. Now, boosting along a compact spatial circle leads to a compact *null* circle. It turns out that one can go to a regime where the effective theory is the $U(J)$-gauged matrix quantum mechanics which describe the low energy limit of the theory of D0 branes. Furthermore, if additional dimensions are compact this matrix quantum mechanics effectively becomes low dimensional supersymmetric Yang–Mills theories on compact spaces. This Yang–Mills theory then describes string theory and hence gravity in 10-dimensional space-time. Like the previous two examples, space dimensions are 'manufactured' in the gauge theory. The reason why this gauge-string connection is different from the previous two examples is that here we do not need a 't Hooft limit. In fact a strong version of the matrix theory conjecture works for any finite J. A review of matrix theory is [20].

In these examples, a gauge theory leads to gravitational theories in higher number of *space* dimensions. The *time* was built in. Nevertheless, as we will see below, the notion of time perceived by the open strings, i.e. the gauge theory can be quite different from the notion of time perceived by closed strings which emerge out of the fundamental theory. There are in fact versions of matrix theory in which even time is manufactured – the holographic theory is simply a random matrix theory. For a review, see [21].

Except in the two-dimensional string theory example, the gauge-gravity correspondence remain conjectures rather than proven relationships. The common theme in all these realizations of holography is that the quantum dynamics of the system is unambiguously defined in the holographic gauge theory. The higher dimensional space-time and gravitational dynamics in it is always an *effective* description in a certain regime of parameters. By the same token, usual *perturbative* closed string theory would appear only in a certain corner of the parameter space. A 'proof' of the conjecture would involve a demonstration that in this corner of parameter space, the gauge theory indeed reproduces closed string perturbation theory. This is the sense in which there is such a proof for the two-dimensional model. It is confusing to ask what a 'proof' would mean at the non-perturbative level – since we do not know of an independent way to define closed strings non-perturbatively. In fact, it seems reasonable to take the open string theory as a fundamental definition of the theory itself. This theory has no dynamical space-time, no notion of geodesic incompleteness and so on – but does have a notion of time evolution of states. In a certain approximation, the theory can be re-interpreted as a theory of perturbative closed strings. In general, there is no such interpretation and therefore no notion of dynamical space-time.

This is the key point in our discussion of singularities. These holographic descriptions lead to a proposal for the structure which replaces space-time near singularities. *If there are situations where the fundamental dynamics defined in terms of the gauge theory makes sense in regions where the space-time appears singular one has the possibility that these singularities are simply problems with interpretation.*

3. Cosmologies in two-dimensional string theory

The first concrete relations between large-N gauge theories and string theories appeared in the late 1980s in the study of toy models of strings in small number of dimensions. The best understood model of this type is the description of string theory in one space and one time dimension [10–12]. The corresponding large-N theory is quantum mechanics of a single $N \times N$ hermitian matrix in a certain double scaling limit where one takes $N \to \infty$ and the coupling of the matrix quantum mechanics $G \to G_{\mathrm{c}}$ with a certain combination of $N\mu_{\mathrm{F}}$ held constant, where μ_{F} is a function of $(G - G_{\mathrm{c}})$. Recently it has been realized that this should be in fact a gauge theory. In the double scaling limit the action is given by

$$S = \int \mathrm{d}t \, \frac{1}{2}[(D_t M)^2 + M^2] \tag{1}$$

which has a gauge symmetry with gauge group $U(N)$. The covariant derivative $D_t = \partial_t + iA_t$ involves the gauge field A_t. In $0+1$ dimensions a gauge field can be set to zero by a gauge transformation and this results in a constraint which implies that all physical states must be *singlets* under $U(N)$.

In the singlet sector, the entire dynamics is encoded in the eigenvalues of the matrix, $\lambda_i(t)$. However the change of variables from the matrix to the eigenvalues results into a well-known measure factor which makes the wave function into a slater determinant. The eigenvalues may be therefore thought to be locations of N non-interacting fermions living in an inverted harmonic oscillator potential, with a single particle Hamiltonian given by

$$H = \frac{1}{2}[p^2 - x^2]. \tag{2}$$

Equivalently, the theory may be re-written in terms of the density of eigenvalues of the matrix $\partial_x \varphi(x,t)$ which is defined by

$$\partial_x \varphi(x,t) = \frac{1}{N} \mathrm{tr} \delta(M(t) - x \cdot I). \tag{3}$$

This field may be regarded as a bosonization of the fermions.

At the classical level the action of the collective field is given by

$$S = N^2 \int \mathrm{d}x \, \mathrm{d}t \, \left[\frac{1}{2} \frac{(\partial_t \varphi)^2}{(\partial_x \varphi)} - \frac{\pi^2}{6} (\partial_x \varphi)^3 - \left(\mu - \frac{1}{2} x^2 \right) \partial_x \varphi \right]. \tag{4}$$

The dimensionful parameter μ is a Lagrange multiplier which ensures that the total number of eigenvalues is fixed.

3.1 *Physics near the ground state*

This is of course a field theory in $1+1$ dimension, the spatial dimension arising out of the space of eigenvalues. What is remarkable is the fact that this two-dimensional theory in fact represents a massless particle with a relativistic dispersion relation – even though there was no such relativistic invariance to begin with. It is clear from (4) that for large N we can expand the field around the ground state classical solution which is

$$\partial_x \varphi_0 = \frac{1}{\pi}\sqrt{x^2 - 2\mu}\partial_t \varphi_0 = 0, \quad |x| > \sqrt{2\mu} \tag{5}$$

and vanishes for $|x| < \sqrt{2\mu}$. The quadratic part of the action for fluctuations

$$\eta(x,t) = \varphi(x,t) - \varphi_0(x) \tag{6}$$

is given by

$$S_\eta^{(2)} = \frac{1}{2}\int \mathrm{d}t\ \mathrm{d}x\ \sqrt{g} g^{\mu\nu}\partial_\mu \eta \partial_\nu \eta, \tag{7}$$

where the two-dimensional metric is

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \frac{\mathrm{d}x^2}{x^2 - 2\mu}. \tag{8}$$

The fluctuation obeys boundary conditions

$$\eta(\pm\sqrt{2\mu}, t) = 0 \tag{9}$$

so that $x = \pm\sqrt{2\mu}$ acts as a *mirror*. The classical interaction Hamiltonian is purely cubic when expressed in terms of the fluctuation field η and its canonically conjugate momentum Π_η,

$$H_\eta^{(3)} = \int \mathrm{d}x \left[\frac{1}{2}\Pi_\eta^2 \partial_x \eta + \frac{\pi^2}{6}(\partial_x \eta)^3\right]. \tag{10}$$

The global nature of the semiclassical space-time which is generated is transparent if we make a change of variables to 'Minkowskian' coordinates (τ, σ),

$$t = \tau, \quad x = \pm\sqrt{2\mu}\cosh\sigma. \tag{11}$$

In these coordinates

$$\mathrm{d}s^2 = -\mathrm{d}\tau^2 + \mathrm{d}\sigma^2 \tag{12}$$

and we have *two* fields $\eta_\pm(\tau, \sigma)$.

The interaction Hamiltonian in these coordinates becomes

$$H_3 = \int \mathrm{d}\sigma \frac{1}{2\sinh^2\sigma}\left[\frac{1}{2}\tilde{\Pi}_\eta^2 \partial_\sigma \eta + \frac{\pi^2}{6}(\partial_\sigma \eta)^3\right] \tag{13}$$

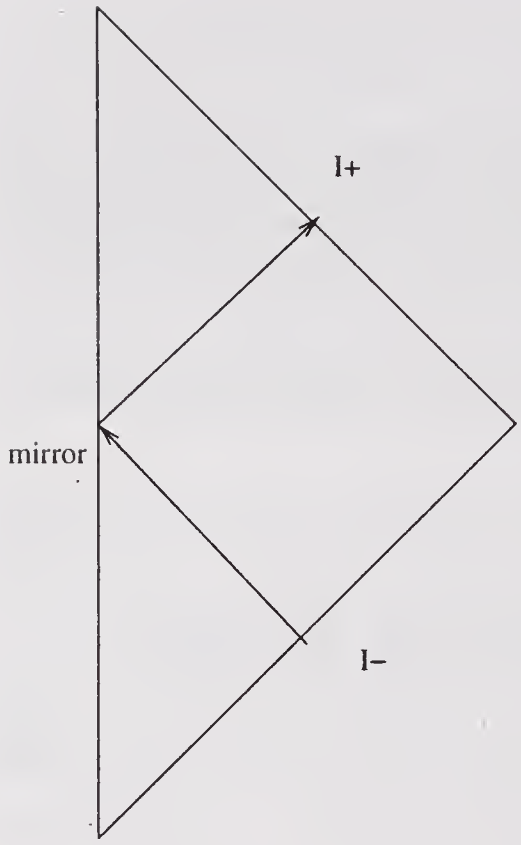


Figure 1. Penrose diagram of space-time produced by ground state solution showing an incoming ray getting reflected at the mirror.

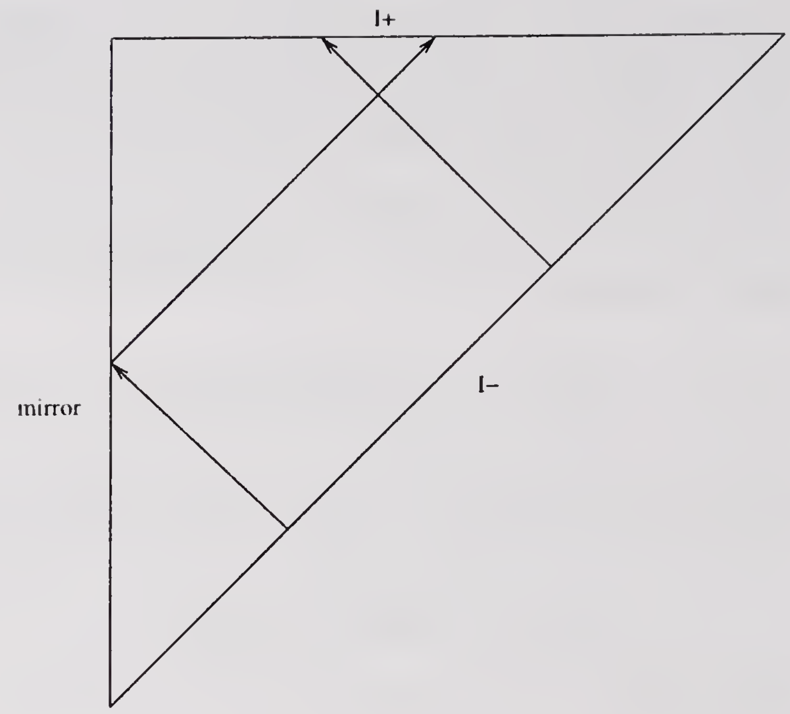


Figure 2. Penrose diagram for the closing hyperbola solution showing two classes of null rays.

so that the interactions are strong at the mirror $\sigma = 0$. In any case, the space-time generated is quite simple. The Penrose diagram is that of the two-dimensional Minkowski space with a mirror as shown in figure 1. The fluctuations are massless particles which come in from $\mathcal{I}^-_{\rm L,R}$, get reflected at the mirror and arrive at $\mathcal{I}^+_{\rm L,R}$. Interactions, which are strong at the mirror, lead to a non-trivial S-matrix.

Even though our theory looks like a usual quantum field theory in 1 + 1 dimensions, it is actually a 1 + 1 dimensional *closed* string theory in disguise. In 1 + 1-dimensions there are no transverse oscillations of a string, so that the only dynamical degrees of freedom is its center of mass – which corresponds to a usual field theory. Indeed the two massless scalar fields displayed above are related to only two dynamical fields of Type 0B two-dimensional string theory. The relation is however *nonlocal at the string scale* and this has important consequences. This slight non-locality in fact encodes the gravitational attraction between objects in this model and is instrumental in showing that the S-matrix obtained in the matrix model is in exact agreement with the S-matrix obtained in a perturbative string computation.

Finally, the interpretation of the model in terms of a scalar field in 1+1 dimensions is a good interpretation only when the coupling is weak. Fortunately the couplings are weak at $\mathcal{I}^\pm$ – this allows a definition of asymptotic states and therefore a S-matrix. Near the mirror, the coupling is strong. The fact that the dispersion relation is relativistic at the free level has no significance in this region and it is not useful to interpret this model as a relativistic model of bosons. The fermions, however, are free everywhere – the understanding of this model in terms of fermions is exact and valid everywhere.

3.2 *Time-dependent solutions with singularities* [1]

The model (2) has an infinite number of global symmetries which are symmetries of the action, but do not generically commute with the Hamiltonian. The symmetry group is called W_∞. If we act on the ground state with such a group element, e.g.

$$|\lambda\rangle = \exp[i\lambda Q]|\mu\rangle, \tag{14}$$

we obtain a new state of the theory with non-trivial time dependence which automatically satisfies the Schrödinger equation. However such a state would not be normalizable and would therefore not be included in the Hilbert space of the fluctuations around the ground state. Rather this would correspond to a new time-dependent background with a Hamiltonian

$$H' = \mathrm{e}^{-i\lambda Q} H \mathrm{e}^{i\lambda Q}. \tag{15}$$

At the semiclassical level some of these states are represented by classical solutions of the collective field theory (4), $\varphi_0(x,t)$. Our aim is to find out what kind of space-time is perceived by fluctuations around such a classical solution.

The strategy to do this is exactly the same as what we did for the ground state. We expand the field as in (6) and find the dynamics of $\eta(x,t)$. It turns out that at the free level, the fluctuations around any classical solution behave as massless scalars in $1+1$ dimensions with the space-time metric given by

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \frac{(\mathrm{d}x + \frac{\partial_t\varphi_0}{\partial_x\varphi_0}\mathrm{d}t)^2}{(\pi\partial_x\varphi_0)^2}. \tag{16}$$

In $1+1$ dimensions, all metrics are conformally flat – so we can always determine the global structure of the space-time by making a coordinate transformation to Minkowskian coordinates (τ,σ). The latter will be called the closed string space-time.

The interesting fact is that there are several classical solutions of this type which lead to closed string space-times which appear to have *space-like boundaries.* A simple example is given by

$$\partial_x\varphi_0 = \frac{1}{\pi(1+\mathrm{e}^{2t})}\sqrt{x^2-(1+\mathrm{e}^{2t})}, \quad \partial_t\varphi_0 = -\frac{x\mathrm{e}^{2t}}{1+\mathrm{e}^{2t}}\,\partial_x\varphi_0. \tag{17}$$

Here we have rescaled x and t to set $2\mu = 1$ and have also shifted the origin of t to absorb a parameter of the W_∞ transformation. This solution starts off like the ground state at $t \to -\infty$, but the regions in x space in which the fermions live is pushed out to infinity on both sides at late times. The fluctuations which correspond to small ripples on the Fermi sea, however, have a smooth time evolution. Any such fluctuation originating in the large $|x|$ region will get reflected and will be pushed by a moving mirror which will again end up in the large $|x|$ region.

The Minkowskian coordinates (τ,σ) in terms of which the metric is simply $\mathrm{d}s^2 = -\mathrm{d}\tau^2 + \mathrm{d}\sigma^2$ are given by the following transformations:

$$x = \cosh\sigma\sqrt{1+\mathrm{e}^{2t}}, \quad \mathrm{e}^{\tau} = \frac{\mathrm{e}^t}{\sqrt{1+\mathrm{e}^{2t}}}. \tag{18}$$

This immediately shows that as $-\infty < t < \infty$ the time τ has the range $-\infty < \tau < 0$. Since the dynamics of the matrix model ends at $t = \infty$ the resulting space-time appears to be geodesically incomplete with a space-like boundary at $\tau = 0$.

The Penrose diagram for this space-time is given in figure 2, which has a space-like $\mathcal{I}^+$ at $\tau = 0$, parametrized by σ. Note that the entire $\mathcal{I}^+$ is at $|x| = \infty$.

If such a space-time were encountered in the usual semiclassical gravity, one would have immediately completed this by attaching a region from $\tau = 0$ to $\tau = \infty$. However in this case this does not make sense. The fundamental dynamics of this model is given by matrix quantum mechanics which has a perfectly unitary time evolution in time t, and this time $t = \infty$ on $\mathcal{I}^+$. Extending the space-time beyond this would mean extending this fundamental time beyond $t = \infty$ and this does not make much sense. Recall that in this model space-time is a derived notion. What we are finding is that if we extrapolate this notion all the way to late times, we encounter a singularity which is space-like.

What is really happening in this model is that the interpretation of the model in terms of the usual relativistic two-dimensional space-time is breaking down at late times. This may be seen in several ways. Using the fact that the background is a W_∞ transform of the ground state, one can exactly calculate correlation functions of fermion operators and hence the correlators of fluctuations around the ground state. The result is that near $\mathcal{I}^+$ these differ significantly from the semiclassical values, indicating $\mathcal{I}^+$ is strongly coupled. The fermion time evolution is perfectly smooth over the entire range.

There are several backgrounds of this type. They all have the common feature that while the 'open string time', i.e. the time in terms of which the quantum dynamics is defined at the fundamental level, runs over the full range, the 'closed string time' appears to end at some point. There are of course time reversed solutions where the closed string time appears to begin abruptly in a caricature of a Big Bang. And in all these instances, the space-time interpretation of the holographic model is breaking down at these space-like boundaries, even though the underlying theory makes perfect sense.

4. Ten-dimensional Big Bangs from matrix theory

There are similar examples in ten-dimensional string theory, which can be described in matrix theory and its descendants.

4.1 *Matrix string theory*

Let us briefly review the holographic description of Type IIA strings in terms of a 1 + 1-dimensional supersymmetric Yang–Mills theory (for a useful review, see [22]). This connection arises from duality transformations on the standard ten-dimensional flat background with string frame metric

$$\mathrm{d}s^2 = 2\mathrm{d}x^+\mathrm{d}x^- + \mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x}, \tag{19}$$

with the null coordinate x^- compactified on a circle of radius R. The string coupling is g_s and the string length is l_s. Consider the sector of the theory with a momentum

$$p_- = J/R \tag{20}$$

along x^-.

Now, ten-dimensional string theory has a rich set of symmetries called duality symmetries which generally relate one kind of string theory with another kind. In this case the above background may be shown to be dual to a Type IIB string theory with a string coupling $\tilde{g}_s$ and string length $\tilde{l}_s$ given by

$$\tilde{g}_s = \frac{R}{g_s l_s}, \quad \tilde{l}_s^2 = \frac{g_s l_s^3}{R}; \tag{21}$$

living on a compact space-like circle of radius $\tilde{R}$ given by

$$\tilde{R} = \frac{l_s^2}{R}; \tag{22}$$

and carrying J units of D1-brane charge. The matrix theory conjecture in this case claims that a non-perturbative formulation of the theory is given by a $U(J)$ Yang–Mills theory with a dimensional coupling constant

$$g_{\mathrm{YM}} = \frac{R}{g_s l_s^2}. \tag{23}$$

The bosonic part of the action is given by

$$S = \int \mathrm{d}\tau \mathrm{d}\sigma \mathrm{tr} \left\{ \frac{1}{2} g_s^2 F_{\tau\sigma}^2 + \frac{1}{2}(D_\tau X^i)^2 - \frac{1}{2}(D_\sigma X^i)^2 + \frac{1}{4g_s^2}[X^i, X^j]^2 \right\}, \tag{24}$$

where $X^i, i = 1 \ldots 8$ are adjoint scalars and $F_{\tau\sigma}$ is the $U(J)$ gauge field strength. These fields live on a circle parametrized by σ and $0 < \sigma < \tilde{R} = l_s^2/R$. However the space-time of this Yang–Mills theory is not dynamical.

In the regime of weak string coupling $g_s \ll 1$ the potential term in (24) suppresses X^i which do not commute. Therefore in this limit the theory reduces to a theory of $8J$ scalar fields which may be chosen to be the diagonal components of the matrices X^i, which we denote by $X_a^i, a = 1 \ldots J$. Similarly, the gauge field strength is locally zero. However as we go around the σ circle, the gauge symmetry allows non-trivial boundary conditions which are characterized by conjugacy classes of the gauge group. For example, we can have

$$\begin{aligned} X_1^i(\sigma + 2\pi\tilde{R}) &= X_2^i(\sigma) \\ X_2^i(\sigma + 2\pi\tilde{R}) &= X_3^i(\sigma) \\ &\cdots \\ X_J^i(\sigma + 2\pi\tilde{R}) &= X_1^i(\sigma). \end{aligned} \tag{25}$$

With these boundary conditions, the action (24) becomes the action of eight massless scalars living on a circle of radius $J\tilde{R}$. This is precisely the worldsheet description of a single Type IIA string in the light cone gauge. In a similar way one could have boundary conditions

$$\begin{aligned}
&X_1^i(\sigma + 2\pi\tilde{R}) = X_2^i(\sigma)\\
&X_2^i(\sigma + 2\pi\tilde{R}) = X_3^i(\sigma)\\
&\ldots\\
&X_K^i(\sigma + 2\pi\tilde{R}) = X_1^i(\sigma)\\
&X_{K+1}^i(\sigma + 2\pi\tilde{R}) = X_{K+2}^i(\sigma)\\
&X_{K+2}^i(\sigma + 2\pi\tilde{R}) = X_{K+3}^i(\sigma)\\
&\ldots\\
&X_J^i(\sigma + 2\pi\tilde{R}) = X_{K+1}^i(\sigma)
\end{aligned} \tag{26}$$

and one would have the worldsheet action for *two* strings. It is clear that for a given J one would have various sectors of the theory which describes J strings. Furthermore, the commutator interaction is capable of describing in a precise fashion the joining and splitting of these strings.

Therefore in the $g_s \ll 1$ regime, this two-dimensional gauge theory describes a second quantized theory of closed strings in *ten space-time dimensions*, and therefore gravitational interactions in ten dimensions as well. It is important to realize that this happens only in this regime, which by virtue of (23) is the strongly coupled regime of the gauge theory. The fields X^i metamorphize into transverse space-time coordinates. For generic g_s all the non-abelian excitations are important and there is no such ten-dimensional interpretation and therefore no clear interpretation in terms of a usual dynamical space-time.

4.2 *A matrix Big Bang*

It turns out that there is a remarkably simple modification of flat space which provides a useful model of a cosmological singularity [23]. This is a background which still has a flat string frame metric, but has in addition a dilaton which is linear in the null coordinate x^+,

$$\Phi = -Qx^+. \tag{27}$$

For $Q > 0$, the point $x^+ = -\infty$ is in fact a *null Big Bang singularity.* This is because the Einstein frame metric (i.e. in terms of which the low energy effective theory is the Einstein–Hilbert action) is geodesically incomplete, with geodesics reaching $x^+ = -\infty$ in a finite proper time. The effective string coupling $g_{\rm eff} = e^{\Phi}$ is infinite at this point. In a similar way for $Q < 0$ we have a big crunch.

The logic which leads to a $1+1$-dimensional Yang–Mills theory in the $p_- = J/R$ sector now leads to a theory whose bosonic action is

$$S = \int \mathrm{d}\tau\, \mathrm{d}\sigma\, \mathrm{tr} \left\{ \frac{1}{2} g_{\mathrm{s}}^2 \mathrm{e}^{-2Q\tau} F_{\tau\sigma}^2 + \frac{1}{2}(D_\tau X^i)^2 - \frac{1}{2}(D_\sigma X^i)^2 \right.$$
$$\left. + \frac{1}{4g_{\mathrm{s}}^2} \mathrm{e}^{2Q\tau} [X^i, X^j]^2 \right\}. \tag{28}$$

Thus the effect of the non-trivial dilaton is to make the coupling constant of the Yang–Mills theory time-dependent. It is useful to rewrite this action (28) as follows:

$$S = \int \mathrm{d}\tau\, \mathrm{d}\sigma\, \sqrt{h}\, \mathrm{tr} \left\{ \frac{1}{4} g_{\mathrm{s}}^2 h^{ab} h^{cd} F_{ac} F_{bd} + \frac{1}{2} h^{ab} (D_a X^i)(D_b X^i) \right.$$
$$\left. + \frac{1}{4g_{\mathrm{s}}^2} [X^i, X^j]^2 \right\}, \tag{29}$$

where h_{ab} denotes a two-dimensional metric

$$\mathrm{d}s_2^2 = h_{ab} \mathrm{d}\xi^a \mathrm{d}\xi^b = \mathrm{e}^{2Q\tau}[-\mathrm{d}\tau^2 + \mathrm{d}\sigma^2]. \tag{30}$$

Thus the theory may be viewed as one with *constant* couplings, but in a non-trivial space-time (30) – which is the Milne universe, or the future light cone of the origin. We would expect that in an appropriate limit the action (29) represents a string theory living in this Milne space-time with eight other transverse directions. It is clear from (30) that the space-time defined by (τ, σ) with $-\infty < \tau < \infty, 0 < \sigma < 2\pi\tilde{R}$ has a conical singularity at $\tau = -\infty$ and this singularity may be reached in a finite proper time. One way to see the geodesic incompleteness of the space-time is to go to the Minkowskian coordinates for the metric

$$T = \mathrm{e}^{Q\tau} \cosh\sigma, \quad X = \mathrm{e}^{Q\tau} \sinh\sigma,$$
$$\mathrm{d}s_2^2 = -\mathrm{d}T^2 + \mathrm{d}X^2. \tag{31}$$

The situation is quite similar to the two-dimensional backgrounds of the previous section. There is an *open string time* τ which runs over the full range. However the closed string time T seems to begin at $T = 0$ at a null Big Bang.

From the action (28) we see that at $\tau \to \infty$, i.e. $T \to \infty$ the Yang–Mills coupling is strong. Therefore we expect that this regime would describe perturbative IIB strings in the manner discussed above. However, as $\tau \to -\infty$, i.e. at the 'Big Bang' the Yang–Mills coupling is *weak*. This means that there is nothing which suppresses X^i which are non-commuting and all the $8J^2$ degrees of freedom are equally important. This means that there is no interpretation in terms of second quantized strings and no interpretation of the fields as coordinates in eight transverse space dimensions. The Yang–Mills theory, however, continues to make perfect sense. In fact the coupling of the theory goes to *zero* as we approach this 'singularity' and nothing is obviously wrong with a bunch of free fields !

What has happened is similar to the two-dimensional example. There is a certain regime of the parameters of the holographic theory where a space-time interpretation is valid. If we forcibly extrapolate this interpretation to early or late times, it appears that from the point of view of the closed string theory there is a singularity. However it is precisely in this region that it is illegal to ascribe a space-time interpretation: the holographic gauge theory is what it is and makes perfect sense.

4.3 *Zooming onto the Big Bang*

At the place where the closed string theory appears to perceive a Big Bang singularity the non-abelian nature of the theory becomes important. To get some idea of the nature of non-abelian excitations it is useful to consider strings moving on a time-dependent gravitational wave rather than flat space [24]. Fortunately, matrix theory may be formulated in a class of such gravitational waves whose string frame metric and dilaton are given by

$$\begin{aligned} \mathrm{d}s^2 &= 2\mathrm{d}x^+\mathrm{d}x^- - \left[\left(\frac{\mu}{3}\right)^2 \vec{x}^{\,2} + \left(\frac{\mu}{6}\right)^2 \vec{y}^{\,2}\right](\mathrm{d}x^+)^2 + \mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x}\,\mathrm{d}\vec{y}\cdot\mathrm{d}\vec{y}, \\ \Phi &= -Qx^+, \end{aligned} \tag{32}$$

where $\vec{x} = (x^1 \ldots x^3), \vec{y} = (y^1 \ldots y^5)$. There is, in addition, a background 5-form field strength also proportional to μ. The resulting matrix theory is a deformation of the theory described in the previous subsection, additional terms involving μ.

The introduction of a gravitational wave as above introduces a new length scale $1/\mu$ into the Yang–Mills theory. It turns out that the dimensionless ratio μ/G_{YM} acts as a semiclassical parameter and when $\mu \gg G_{\mathrm{YM}}$ *classical* solutions of the Yang–Mills theory are relevant. Among such classical solutions are highly non-abelian configurations called *fuzzy spheres*

$$X^i = S(\tau)J^i, \quad i = 1, 2, 3, \tag{33}$$

where $S(\tau)$ is a real function of τ and J^i are generators of a N-dimensional representation of $SU(2)$. These are called spheres since the Casimir condition implies that

$$\sum_{i=1}^{3}(X^i)^2 = \frac{N^2-1}{4}S^2(\tau)I_{N\times N} \tag{34}$$

which would represent a sphere if the X^i were real numbers rather than matrices. They are called fuzzy because X^i are not real numbers. Clearly, in the presence of a large number of fuzzy spheres, the matrices are far from being diagonal and the usual interpretation of the theory as a string theory in conventional ten-dimensional space-time is invalid. These fuzzy spheres are really discretized versions of D-branes.

The time dependence of the radius of the fuzzy sphere, $S(\tau)$ is determined by the dynamics of the YM theory. An examination of the action shows that at early times (near the Big Bang) such fuzzy spheres rather than strings proliferate. As time evolves, the size of these fuzzy spheres diminish and at late times they become zero size objects leaving only strings.

A Type IIB version of this model [25] exhibits another interesting aspect of the region near the singularity. It is well-known that quantum field theory in time-dependent backgrounds exhibit particle production. This results from the fact that typically the vacuum state at late times is not the vacuum at early times and vice versa. In these models, this same phenomenon leads to an interesting insight into the question of initial conditions. We saw that in this type of model one generically expects perturbative strings and hence conventional space-time at late times and non-abelian configurations at early times. The latter include D-branes which can be easily excited. The question we can now ask is the following: If we require that the state at late times contain only perturbative strings and nothing else, can we start with *any* conceivable initial state?

One might expect that the answer should be positive. After all at late times these D-branes and other non-abelian excitations are suppressed – we might expect that they all go away at sufficiently late times regardless of the initial state. Surprisingly the answer is negative. Because of particle production (or depletion) effects, the state which does not contain such D-branes at late times turns out to be a *squeezed state* of these D-branes at early times – with a thermal distribution of the number of D-branes with temperature $1/Q$. This means that if we require that at late time usual space-time emerges with no such remnant D-branes filling the vacuum – the initial state near the Big Bang must be close to this special squeezed state. It will be interesting to see whether this has some implication for the general question of initial conditions in cosmology.

5. Outlook

Raychaudhuri showed that a congruence of geodesics would tend to shrink and form singularities under generic conditions. These singularities clearly signal the breakdown of general relativity. It has been suspected that a proper theory of quantum gravity would 'resolve' such singularities, though it has never been clear what such a resolution might mean for space-like or null singularities.

The results from string theory are beginning to provide a concrete meaning to this by demonstrating that space-time is itself an approximate notion which emerges from more fundamental structures which form *holographic* descriptions of gravitational physics. In this article we have described how this happens in some versions of holography, viz. matrix models of two-dimensional strings and matrix theory of ten-dimensional strings. There have been some recent progress in understanding such cosmological singularities using the AdS/CFT correspondence as well [26,27]. These investigations have been in toy models of cosmology. However one would expect that we should be able to draw some general lessons which can be applied one day to realistic cosmology as well.

Acknowledgements

I would like to express my deep gratitude to AKR for giving me the first glimpse into the beauty of theoretical physics. He has been instrumental in shaping my taste in the subject. I thank the editors of this special issue for inviting me to contribute an article to honor him. I also thank my collaborators Joanna Karczmarek, Samir Mathur, Jeremy Michelson, K Narayan and Sandip Trivedi for enjoyable collaborations and my colleagues at Tata Institute of Fundamental Research and at University of Kentucky for numerous discussions. This work was supported in part by a National Science Foundation (USA) Grant PHY-0244811 and a Department of Energy (USA) contract DE-FG01-00ER45832.

References

[1] S R Das, *Mod. Phys. Lett.* **A20**, 2101 (2005)
[2] B Craps, *Big Bang models in string theory*, arXiv:hep-th/0605199
[3] M Gasperini and G Veneziano, *Phys. Rep.* **373**, 1 (2003), arXiv:hep-th/0207130
[4] H Liu, G W Moore and N Seiberg, *The challenging cosmic singularity*, arXiv:gr-qc/0301001
[5] E Silverstein, *Singularities and closed string tachyons*, arXiv:hep-th/0602230
[6] Sidney Coleman, $1/N$, *Proceedings of the 1979 International School of Subnuclear Physics* (1979) 0011
[7] Standard textbooks on String theory include [8] and [9]
[8] M Green, J Schwarz and E Witten, *Superstring theory* (Cambridge University Press, 1987) vols 1 and 2
[9] J Polchinski, *String theory* (Cambridge University Press, 1998) vols 1 and 2
[10] I R Klebanov, *String theory in two-dimensions*, arXiv:hep-th/9108019
[11] S R Das, *The one-dimensional matrix model and string theory*, arXiv:hep-th/9211085
[12] E J Martinec, *Matrix models and 2D string theory*, arXiv:hep-th/0410136
[13] A Sen, *An introduction to non-perturbative string theory*, arXiv:hep-th/9802051
[14] C V Johnson, *D-brane primer*, arXiv:hep-th/0007170
[15] S R Das and S D Mathur, *Ann. Rev. Nucl. Part. Sci.* **50**, 153 (2000), arXiv:gr-qc/0105063
[16] J R David, G Mandal and S R Wadia, *Phys. Rep.* **369**, 549 (2002), arXiv:hep-th/0203048
[17] O Aharony, S S Gubser, J M Maldacena, H Ooguri and Y Oz, *Phys. Rep.* **323**, 183 (2000), arXiv:hep-th/9905111
[18] V Balasubramanian, E G Gimon and D Minic, *J. High Energy Phys.* **0005**, 014 (2000), arXiv:hep-th/0003147
[19] R Bousso, *Rev. Mod. Phys.* **74**, 825 (2002), arXiv:hep-th/0203101
[20] W Taylor, *Rev. Mod. Phys.* **73**, 419 (2001), arXiv:hep-th/0101126
[21] H Aoki, S Iso, H Kawai, Y Kitazawa, A Tsuchiya and T Tada, *Prog. Theor. Phys. Suppl.* **134**, 47 (1999), arXiv:hep-th/9908038
[22] F Hacquebord, *Symmetries and interactions in matrix string theory*, arXiv:hep-th/9909227
[23] B Craps, S Sethi and E P Verlinde, *J. High Energy Phys.* **0510**, 005 (2005), arXiv:hep-th/0506180
[24] S R Das and J Michelson, *Phys. Rev.* **D72**, 086005 (2005), arXiv:hep-th/0508068
[25] S R Das and J Michelson, *Matrix membrane Big Bangs and D-brane production*, arXiv:hep-th/0602099
[26] S R Das, J Michelson, K Narayan and S P Trivedi, arXiv:hep-th/0602107
[27] C S Chu and P M Ho, *J. High Energy Phys.* **0604**, 013 (2006), arXiv:hep-th/0602054

PRAMANA
— journal of
physics

Vol. 69, No. 1
July 2007
pp. 109–118

Horizons in 2+1-dimensional collapse of particles

DIETER BRILL[1], PUNEET KHETARPAL[2] and VIJAY KAUL[1]
[1]Department of Physics, University of Maryland, College Park, MD 20782, USA
[2]Rensselaer Polytechnic Institute, Troy, NY 12180, USA
E-mail: brill@physics.umd.edu

Abstract. A simple, geometrical construction is given for three-dimensional spacetimes with negative cosmological constant that contain two particles colliding head-on. Depending on parameters like particle masses and distance, the combined geometry will be that of a particle, or of a black hole. In the black hole case the horizon is calculated. It is found that the horizon typically starts at a point and spreads into a closed curve with corners, which propagate along spacelike caustics and disappear as the horizon passes the particles.

Keywords. Black holes; gravitational collapse; horizon.

PACS Nos 04.20.Gz; 04.60.Kz; 04.70.Bw

1. Introduction

Event horizons are generated by null geodesics. They are therefore natural candidates for applying a time development equation of the Raychaudhuri type. But in the present context we focus on the non-smooth regions of horizons and other features that we describe by geometrical construction, a kind of virtual use of the Raychaudhuri equation. We hope that our material is nevertheless of interest to readers of this volume.

The presence of an event horizon is the defining characteristic of a black hole. When a black hole is formed, typically by gravitational collapse, an event horizon starts at some stage of the time development. It then spreads out increasing its area until it encompasses all the dynamical features, and eventually becomes stationary. The final black hole is then one of the small family of 'hairless' types. The most active and interesting period in the life of a black hole is the time near the formation of the horizon. Because the horizon is a global property of the spacetime and cannot be characterized locally, this interesting period is typically studied only in numerically generated spacetime regions that extend over a long time.

In two space- and one time-dimensions (2+1 D) the situation is much simplified for several reasons. There is a simple kind of matter that can collapse gravitationally and form black holes, namely point particles. The geometry of spacetimes with such collapsing matter is known exactly, at least in principle. It is also known that there are no gravitational waves emitted in the collapse that may allow the final black hole state to be reached only asymptotically in time. Thus, in the collapse of 2+1 D particles, the formative stage of the horizon lasts only a finite time and is over as soon as the horizon has passed all the particles. These simplifying features in a 2+1 D spacetime of course make it somewhat unrealistic as a model for 3+1 D collapse, but one would still expect many of the lessons one learns in 2+1 dimensions to survive in some form in 3+1 dimensions.

In this paper we focus on the horizon formation when two particles in 2+1 D collapse head-on. In order to form a true black hole, the particles must have sufficient mass–energy and there must be a negative cosmological constant. We first recall the equations of motion for test particles, and then discuss a model process for the case of a vanishing cosmological constant. The changes in the horizon's behavior during the active period for negative cosmological constant are shown to be relatively minor. We also mention the collapse of a particle into a black hole and the case of more than two collapsing particles.

2. Motion of test particles in anti-de Sitter space

A 2+1-dimensional (spinless) point particle is a spacetime with an angle deficit δ about the worldline of the particle. It involves identification by a finite rotation (by the deficit angle), where the fixed point set (axis) is the particle's worldline. For particles with general locations, this construction is possible only in homogeneous and isotropic spacetimes in which case the particle's worldline is necessarily a geodesic. We can construct such a spacetime from a given background in the following way. Choose a surface (usually totally geodesic time-like) containing the particle geodesic and rotate it about that geodesic by δ, remove the wedge volume swept out, and identify its boundaries by the rotation [1]. In the case of several particles (and masses δ not so large as to make the spacetime close up), there is an outer region where the wedges extend outwards from the particles and an inner region between the particles where the background spacetime is unchanged by the cut-and-paste. Within this inner region, the particles therefore move with respect to each other like geodesic test particles. In flat space ($\Lambda = 0$) this motion is simply the usual constant velocity motion that occurs when there is no interaction between the particles. In anti-de Sitter (adS) space ($\Lambda < 0$) the motion of such geodesics is not at constant relative velocity, but can be easily derived as follows.

We can choose the geodesic of one particle as an origin. In 'Schwarzschild' coordinates [2] the AdS metric (for unit negative curvature, $\Lambda = -1$) is

$$\mathrm{d}s^2 = -(1+q^2)\mathrm{d}t^2 + \frac{\mathrm{d}q^2}{1+q^2} + q^2\mathrm{d}\Omega^2. \tag{1}$$

Here $\mathrm{d}\Omega^2$ is the metric on the unit sphere; in 2+1 dimensions we simply have $\mathrm{d}\Omega^2 = \mathrm{d}\phi^2$. For geodesics with velocity $u^\mu = \mathrm{d}x^\mu/\mathrm{d}\tau$ the time- and rotational symmetries of (1) imply the conserved quantities

$$E = u_t = (1+q^2)\frac{\mathrm{d}t}{\mathrm{d}\tau} \qquad \text{and} \qquad L = u_\phi = q^2\frac{\mathrm{d}\phi}{\mathrm{d}\tau},$$

where E and L are the energy and angular momentum per unit mass. In terms of these quantities the normalization condition that τ be proper time becomes

$$-\frac{E^2}{1+q^2}+\frac{1}{1+q^2}\left(\frac{\mathrm{d}q}{\mathrm{d}\tau}\right)^2+\frac{L^2}{q^2}=-1$$

or

$$\left(\frac{\mathrm{d}q}{\mathrm{d}\tau}\right)^2+q^2+\frac{L^2}{q^2}=E^2+L^2-1.$$

But this is the same as the 'radial' conservation of energy equation for a simple harmonic oscillator of unit mass, unit spring constant, angular momentum L and energy E^2+L^2-1, in terms of the proper time τ. Since the angular equation $L=q^2\mathrm{d}\phi/\mathrm{d}\tau$ is also analogous to that of the 2- (or higher-) dimensional harmonic oscillator, the motion in q, ϕ coordinates, or in rectangular coordinates $x=q\cos\phi$, $y=q\sin\phi$, is that of a harmonic oscillator: x and y depend harmonically on τ, the orbits are ellipses, and all geodesics are periodic with period 2π.

This means, in particular, that two particles that will collide head-on at some instant in the future, reach a maximum distance from each other, and their spacetime is time-symmetric about this instant. Without loss of generality, we may therefore assume that the initial state of such a colliding-particle-spacetime is time-symmetric, i.e. a two-dimensional space-like surface of constant negative curvature and vanishing extrinsic curvature (totally geodesic hyperbolic 2-space H^2).

3. Minkowski space model

Although a flat spacetime does not allow horizons, there are null surfaces that share many of the properties of AdS horizons. In locally Minkowski space, the time-symmetric initial state of two particles, with zero relative velocity of the particles, corresponds to a static spacetime. We cut out two wedges, each having its edge at the location of a particle, and choose them to be symmetrically oriented with respect to the particles (figure 1a). The line joining the initial particles on the initial space-like surface (or the plane joining their two geodesics in spacetime) then divides the space into two congruent halves, and the full space is obtained by 'doubling' one of the halves, that is gluing two copies along the edges (figure 1b). This is rather like a Melitta coffee filter as it comes out of the box and before it has been spread into a 'cone'. The subsequent figures are understood to be doubled in this way. In the region outside the particles the initial surface (and all other surfaces $t=$ const.) is exactly conical; we will call it outer-conical, the 2D analog of 3D asymptotically flat.

From the figure and the angle sum of the plane triangle at the bottom of figure 1b (dotted), it follows immediately that the angle deficit δ_{T} measured at large distances for the equivalent single particle is the sum of the angle deficits of the two constituent particles. So the particle masses simply add, as expected when there is no gravitational interaction energy. The location of the single particle (tip of the

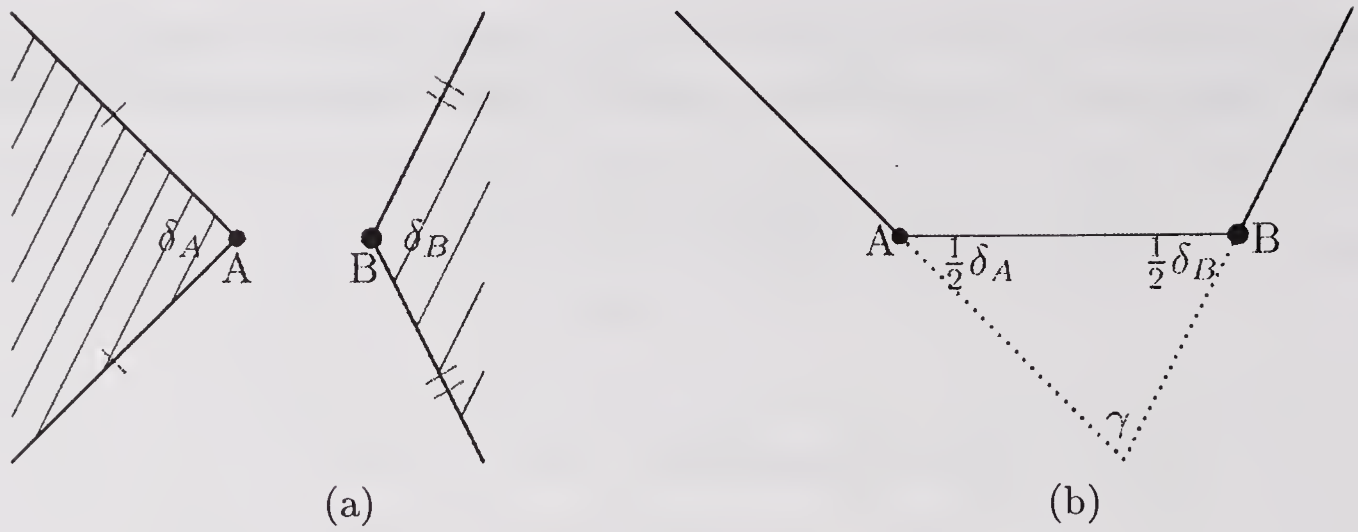


Figure 1. (**a**) A less massive and a more massive static particle, A and B, on the $t = 0$ space-like surface of three-dimensional Minkowski space. The cross-hatched wedges are removed from the space, and the boundaries with equal number of strokes are identified. The deficit angles $\delta_{\rm A}$, $\delta_{\rm B}$ measure the particle masses. (**b**) Half of figure 1a is sufficient to represent the space if it is understood that it represents two layers that are glued together at the boundary – the result of folding figure a along the line AB and identifying the boundaries as indicated. The total angle deficit $\delta_{\rm T} = 2\pi - 2\gamma$.

cone) is of course outside the two-particle spacetime. Only in the limit of small masses is it approximately at the center of mass between them.

Such a spacetime, of course, cannot represent a black hole, since null geodesics from all events eventually reach infinity. However, one can consider a congruence of null geodesics, and the wave-fronts normal to them at each Minkowski time, that reach infinity in the same way that the wave-front of a black hole's horizon reaches infinity: as a smooth curve of constant curvature without self-intersection. We will call such a congruence a pseudohorizon. Although in flat space such a wavefront can be arbitrarily translated, such is not the case in a conical space. The smooth wave-front having any (sufficiently small) constant curvature in an outer-conical space is *unique*. The center of these wave-fonts is the tip of the cone if the space is truly conical. For outer-conical spaces we can construct the true cone that analytically continues the outer parts. The wave-fronts that represent the pseudohorizon at different times are then simply those parts of circles centered at the tip of this true cone that lie in the actual space. In the half-figure of the static two-particle space this tip T is easily constructed by extending the slanted borders to a point, and the circles representing the horizon are drawn about this center (figure 2a). Figure 2b shows both halves joined at the line between the particles so that successive stages of the complete pseudohorizon can be seen. It consists of constant curvature wave-front sections W that propagate at the speed of light, interrupted by points of singularities S that travel at faster-than-light speed toward the particles. New generators enter the pseudohorizon along this space-like line of singularities. As the pseudohorizon crosses a particle it becomes a smooth wave-front (note that the last wave-fronts intersect the boundary of the wedges at right angles). It is clear the the pseudohorizon always has the topology of a circle and starts at a kind of 'center of mass' P between the particles, or at one of the particles if that particle has an angle deficit of more than π.

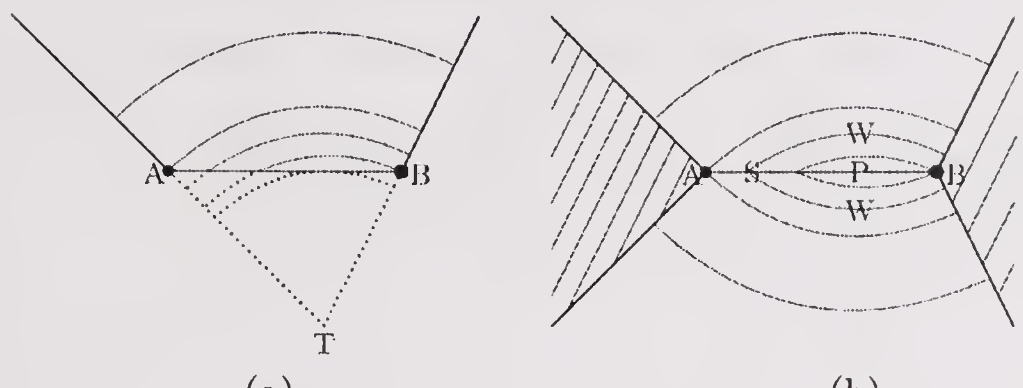


Figure 2. Construction of pseudohorizon for two static particles in flat space. **(a)** A wave-front that will be smooth and of constant curvature at late times starts at the tip T of the cone that is the continuation of the two-particle space's outer region. **(b)** Figure 2a doubled so that the full shape of the wave-fronts can be seen.

4. Two particles in anti-de Sitter space

Anti-de Sitter space, like Minkowski space, can be represented in static coordinates, eq. (1), as a time sequence of three-dimensional homogeneous spaces, which however have constant negative curvature. A point particle, centred at the origin, can be constructed as in Minkowski space by removing a time-like wedge of angle δ and identifying, for example $\phi = 0$ and $\phi = 2\pi - \delta$. If we define a rescaled angle $\varphi = (1 - \delta/2\pi)^{-1}\phi$, with the usual 2π periodicity, and a rescaled radial measure $r = (1 - \delta/2\pi)q$, the metric (1) becomes

$$\mathrm{d}s^2 = -(r^2 - m)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{r^2 - m} + r^2\,\mathrm{d}\varphi^2, \tag{2}$$

with the parameter $m = -(1 - \delta/2\pi)^2 < 0$.

These space-like surfaces are conveniently represented by Poincaré disks, yielding what has been called the 'sausage model' of AdS spacetime. In terms of the polar coordinates ρ, ϕ of the disk (related to q by $q = \frac{2\rho}{1-\rho^2}$, $\rho \le 1$) the spacetime metric has the 'sausage coordinate' [3] form

$$\mathrm{d}s^2 = -\left(\frac{1+\rho^2}{1-\rho^2}\right)^2 \mathrm{d}t^2 + \frac{4}{(1-\rho^2)^2}(\mathrm{d}\rho^2 + \rho^2\,\mathrm{d}\phi^2).$$

Our Minkowski space construction involved mainly straight lines, planes, and circles. The analogous objects in AdS space are geodesics, totally geodesic surfaces, and constant curvature curves. Geodesics and constant non-zero curvature curves are represented on the Poincaré disk by circles perpendicular and oblique to the boundary of the disk, respectively. Totally geodesic surfaces intersect the Poincré disks in geodesics, and in the time direction they execute the periodic motion described in §2.

A two-particle spacetime modeled on AdS space will in general be dynamic. We also saw, that this is so even in the test particle limit. We also saw, that for particles with angular momentum $L = 0$, there is always a time-symmetric moment. We will orient our sausage model so that this moment is one of the space-like coordinate surfaces, say $t = 0$. As in Minkowski space, we cut the spacetime into

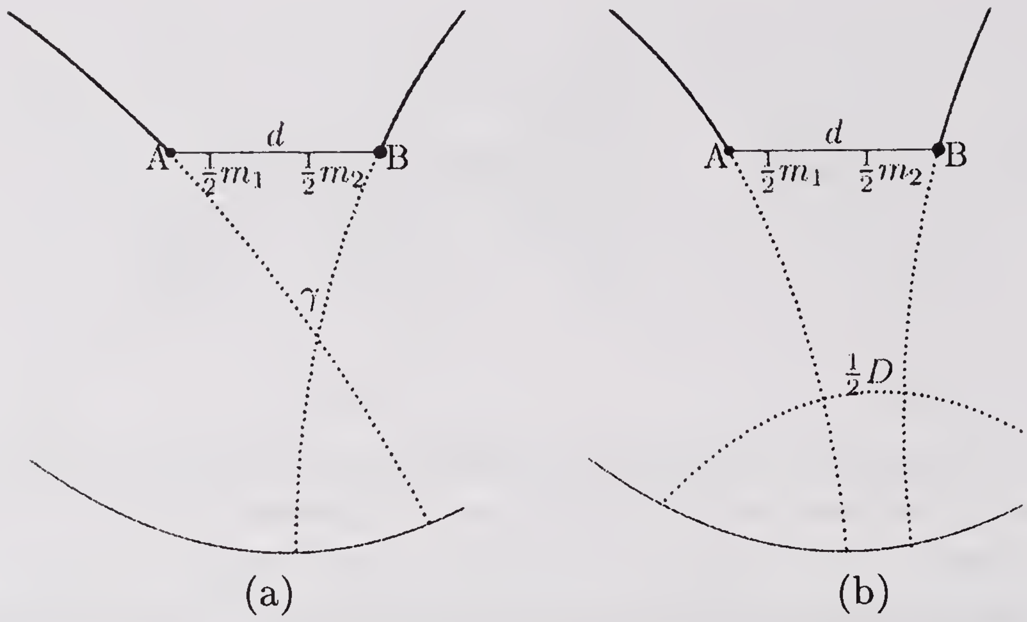


Figure 3. Two particles in anti-de Sitter space at the moment of their maximum separation d, shown on the Poincaré disk. The deficit angles are labeled by the corresponding masses. The disk is larger than the figure; only its bottom boundary is shown. (**a**) In the outer region the geometry is that of a single mass $M = 2\pi - 2\gamma$, which is more than the sum of the individual masses, corresponding to an increase in interaction energy as the masses are separated. (**b**) For larger particle masses or larger separation the outer geometry is no longer that of a single particle, but that of a black hole, characterized by its horizon circumference D.

two congruent halves along the geodesics joining the particles. The initial state of one half will look like figure 3a, the AdS analog of figure 1b, provided the particles' masses are not too large. (An advantage of the Poincaré disk representation is its conformal nature, so that angle deficits are shown faithfully. In the alternative 'Klein disk' representation, the figure at time symmetry would be indistinguishable from a Minkowski space figure, except for the disk boundary, because geodesics appear as straight lines on the Klein disk, but angles are distorted.) The equivalent single particle has no location in the two-particle spacetime, but its angle deficit can be defined from the behavior of the geometry at large distances. Hyperbolic trigonometry on the bottom triangle of figure 3a again shows that the mass M of the composite particle is given in terms of the masses m_1, m_2 of the constituents and their separation d by

$$\cos M = \cos m_1 \cos m_2 + \sin m_1 \sin m_2 \cosh d. \tag{3}$$

The presence of the factor $\cosh d$ may be viewed as the effect of the gravitational interaction energy between the particles.

The time development in terms of sausage time will show the approach of the particles to the centre of the figure and to each other, as expected from their geodesic motion in the space between them. They meet there at $t = \pi/2$, and the further time development cannot be ascertained without a law that determines the result of this collision (though in the literature some results are considered more natural than others [4]). Equation (3) in terms of angles is of course valid on each Poincaré slice of the 'sausage', but m_1 and m_2 are no longer the (constant) rest masses of the particles, containing a contribution from the kinetic energy. Thus the total mass M of the system remains unchanged as d changes. At large distances the system always looks like one static particle and so there is no horizon and no black hole formation.

It is, however, possible to choose the initial masses and the distance between them in such a way that eq. (3) cannot be fulfilled by a real M because the RHS is greater than 1. In this case the boundary geodesics that reach infinity in the figure (and which determine M by their angle of intersection, if they do intersect) are ultraparallel (figure 3b). The metric in the outer region, in terms of an angle coordinate with the usual 2π periodicity and a 'Schwarzschild' radial coordinate that gives the circumference of circles r = const. the usual value $2\pi r$, takes the form (3) with a positive m. This is the BTZ metric [5] for a non-rotating black hole, and m is its mass measured asymptotically. (By contrast, the particle masses m_1 and m_2 determined by deficit angles are measured locally.) Analogous to eq. (3) we have [6]

$$\cosh D = -\cos m_1 \cos m_2 + \sin m_1 \sin m_2 \cosh d, \tag{4}$$

where $D = 2\pi m$ is the circumference of the black hole throat at $r = m$. This is twice the minimum distance between the ultraparallel boundaries in figure 2b. Thus the problem, so difficult in 3+1 dimensions, to determine whether the given initial data will lead to black hole formation, is easily solved by considering the asymptotic dependence of the initial geometry and determining whether the total BTZ mass m is positive or negative.

If the metric in the outer region is that of a black hole, the initial position of the horizon is at the throat of the single black hole described by eq. (4). Since the outer region covers only a part of this equivalent black hole, its throat will generally not be part of the initial space, but as before we can show it by a kind of analytic continuation as in figure 3b. On later time slices, which are superimposed on the initial slice in figure 4, the horizon expands and the particles fall toward each other. The horizon enters the real spacetime at a point on the geodesic between the particles and spreads out with generally two singular points on that geodesic, which become smooth when the horizon crosses the particles. Thus this behaviour is qualitatively similar to that of the Minkowski space (figure 2a). A major difference is that the size of the horizon becomes constant once it has passed all the particles.

5. Generalizations

The starting 'point' of the horizon is not just the point on the line between the particles where it first appears in sausage time. Every point on that line contributes at some time a pair of generators (one for each half space) to the horizon. Because the horizon moves faster than the speed of light along this line, the line of the horizon's origin in spacetime is space-like. In a suitably different time slicing the horizon may therefore appear simultaneously everywhere along the line, and then spread out in all directions away from it [7], or it may start at one of the masses and have a singularity at only one point, which then runs to the other mass. (In fact, this is the case in sausage time if one particle's mass is sufficiently larger than the other's.)

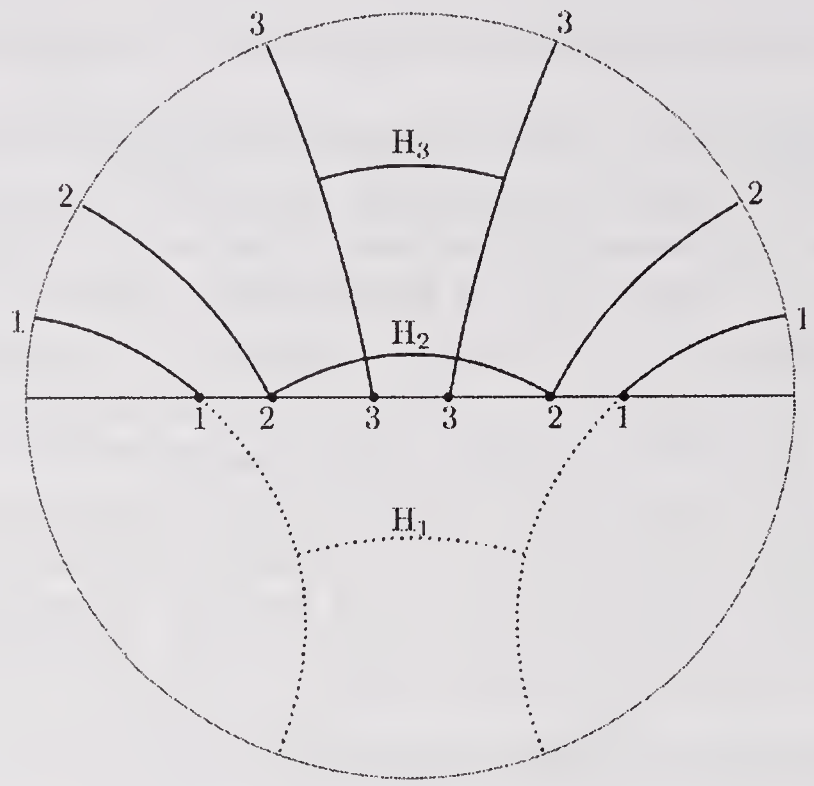


Figure 4. Three time slices in sausage coordinates of two-particle collapse and associated horizon, superimposed. The outer circle represents infinity. Only the half space is shown, the complete configuration at each time is obtained by reflection about the horizontal line and identifying the heavy curves that go to infinity. The two equal-mass particles are indicated by black dots at successive times 1, 2, 3. The initial, time-symmetric configuration is represented by the curves between by the four points labeled 1. There is no horizon initially, but the outer geometry, when continued inward without particle singularities, is shown by the dotted curves and would have a horizon at H_1. At the time labeled 2, the particles have approached each other, and the horizon H_2 has already propagated into the actual space and just reached the particles. In the third configuration, the particles are closer to collision, and the horizon H_3 surrounds the particles, propagating outward toward infinity. The horizon's circumference remains constant after the position H_2, though the Poincaré disk representation does not show its true size.

In our figures we can replace the geodesics that intersect at one or both of the particles by ultraparallel ones. Such initial states can be described as a particle falling into a black hole, or two black holes merging. For two black holes, the horizon of the final single black hole is always present on an initially time-symmetric surface [8]. For a black hole and a particle it is at least partially exposed. In the latter case the initial horizon is not that of the black hole alone, but it has a singularity and is thus prepared to become smooth by swallowing the particle. For two black holes (or the more extreme case of a black hole and a particle) the horizon's future time development is simply that of the resulting single black hole, with no trace of it having arisen from a collapse. However, in the past of the time-symmetric initial surface, the horizon consisted of two separate circular parts, each of which had singularities that moved toward each other and cancelled after they merged into a single circle.

If there are more than two particles collapsing to form a black hole, the main qualitative difference is the pattern formed by the horizon singularities. Let us follow the smooth, constant curvature horizon backwards in time from some late

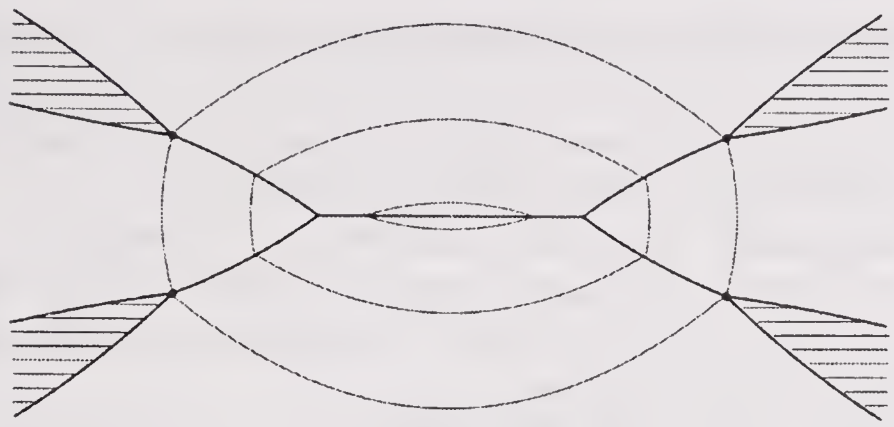


Figure 5. Development of the horizon in the collapse of four particles. The picture for each of the three times at which the horizon is shown was enlarged so that the particles appear at constant positions, but it is schematic only in certain respects. The heavy, V-shaped lines in the outer parts are to be identified as in previous figures to create the angle deficits. The inner heavy lines show the paths of the singular points. The lighter curves are stages of the horizon up to the time when it reaches the particles.

time. It contracts and remains smooth at increasing curvature until it crosses a particle. After crossing it acquires a discontinuity equal to the particle's angle deficit. Such discontinuities propagate and increase inward along geodesics from each particle. The lines of discontinuity meet in pairs and merge until a single line with two singularities propagating toward each other that eventually annihilate. Thus the branching tree pattern of the singularities tell the essential story of the horizon's development. An example for the case of four particles is shown in figure 5.

If the point particles are replaced by finite but concentrated matter distributions, the singular points of the horizon on space-like surfaces will be replaced by concentrations of high extrinsic curvature, running along a similar tree pattern.

In the case that the particles have non-zero angular momentum, they never meet and instead follow periodic orbits in the space between them. With appropriate initial conditions, they can nevertheless form a black hole, if one follows the usual custom in 2+1 D AdS geometries to regard as singular only the region of certain smooth time-like curves. This region has a gap through which the particles can pass at closest approach, and then separate again, albeit into another Universe, as is the case for the analytic extension of the Kerr geometry. A case of black holes on circular orbits has been discussed by DeDeo and Gott [9], and similar configurations have been recognized as a rotating BTZ wormhole by Holst and Matschull [10]. It will be interesting to explore the tree pattern of the horizon singularities in such cases.

References

[1] The analogous construction using a space-like geodesic and Lorentz boost is sometimes referred to as a tachyon, but it is equally appropriate to interpret it as the singularity inside a black hole, at least in spacetimes where such black holes exist

[2] Here the radial coordinate q is defined so that the circumference of a circle $q = \text{const.}$ is $2\pi q$. These coordinates are related to another common form of the AdS metric, $\mathrm{d}s^2 = -\cosh^2 r\,\mathrm{d}t^2 + \mathrm{d}r^2 + \sinh^2 r\,\mathrm{d}\Omega^2$ by the substitution $q = \sinh r$

[3] See S Holst, *Gen. Relativ. Gravit.* **28**, 387 (1996), where these coordinates are called 'Poincaré coordinates'
[4] Sören Holst, *Horizons and time machines*, Thesis in Theoretical Physics, Chapter 4, Department of Physics, Stockholm, 2000, ISBN 91-7265-054-0
[5] Bañados, Teitelboim and Zanelli, *Phys. Rev. Lett.* **69**, 1849 (1992)
Bañados, Henneaux, Teitelboim and Zanelli, *Phys. Rev.* **D48**, 1506 (1993)
[6] If one or both masses are replaced by black holes, a similar formula holds with trigonometric functions replaced by hyperbolic ones.
[7] H-J Matschull, *Class. Quant. Grav.* **16**, 1069 (1999)
[8] D Brill, *Phys. Rev.* **D53**, R4133 (1996)
A Steif, *Phys. Rev.* **D53**, 5527 (1996)
[9] S DeDeo and R Gott, *Phys. Rev.* **D66**, 084820 (2002)
[10] S Holst and H-J Matschull, *Class. Quant. Grav.* **16**, 3095 (1999)

PRAMANA
— journal of
physics

Vol. 69, No. 1
July 2007
pp. 119–135

On the genericity of spacetime singularities

PANKAJ S JOSHI
Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India
E-mail: psj@tifr.res.in

Abstract. We consider here the genericity aspects of spacetime singularities that occur in cosmology and in gravitational collapse. The singularity theorems (that predict the occurrence of singularities in general relativity) allow the singularities of gravitational collapse to be either visible to external observers or covered by an event horizon of gravity. It is shown that the visible singularities that develop as final states of spherical collapse are generic. Some consequences of this fact are discussed.

Keywords. Spacetime singularities; gravitational collapse; cosmology.

PACS Nos 04.20.Dw; 04.70.-s; 04.70.Bw

1. Introduction

After the advent of singularity theorems due to Hawking, Penrose and Geroch in the late nineteen sixties, the role of spacetime singularities as an inevitable feature of Einstein's theory of gravitation became clear. Such singularities occur in cosmology and in gravitational collapse. They also lead to situations where the gravitational field becomes ultra-strong and grows without any upper bound. The above-mentioned theorems further prove that the singularities manifest themselves in terms of the incompleteness of non-space-like geodesics in spacetime.

It is possible that such singularities represent the incompleteness of the theory of general relativity itself. Further, they may be resolved or avoided when quantum effects near them are included in a more complete theory of quantum gravity. Nevertheless, there is a key point here. Even if the final singularity is dissolved by quantum gravity, what is important is the inevitable occurrence of an ultra-strong gravity regime, close to the location of the classical singularity, in cosmology as well as during the dynamical processes involved in gravitational collapse. This affects the physics of the Universe. An example of such a situation is the big bang singularity of cosmology. Even though such singularities may be resolved through either quantum gravity effects [1], or features such as chaotic initial conditions, the effects of the super ultra-dense region of gravity that existed in the big bang epoch profoundly influence the physics and subsequent evolution of the Universe. Similarly, in the gravitational collapse of a massive star, the same theorems predict the occurrence of singularities which can be either visible to external observers or hidden behind the event horizon giving rise to a black hole. In the case of a collapse process leading to a visible or naked singularity, again the important issue is how the inevitably occurring super ultra-dense region would influence the physics outside.

The above discussion does not imply the absence of singularity-free solutions to Einstein's equations. There are many examples of such solutions (see e.g. [2] in this volume). In fact, the Gödel Universe is a cosmological model which is geodesically complete. Also, Minkowski spacetime is the vacuum solution which is geodesically complete. For an additional discussion of singularity-free models in general relativity, we refer to [3]. Again, in gravitational collapse, one could start with a matter cloud which is initially collapsing. However, there could be a bounce later and the cloud may then disperse away, at least in principle. In such a case, no spacetime singularity or ultra-strong gravity region needs to form.

Since its derivation in the early 1950s, the Raychaudhuri equation [4] has played a central role in the analysis of spacetime singularities in general relativity. Prior to the use of this equation to analyse collapsing and cosmological situations for the occurrence of singularities [5], most works on related issues had taken up only rather special cases with many symmetry conditions assumed on the underlying spacetime. Raychaudhuri, however, considered for the first time these aspects within the framework of a general spacetime without any symmetry conditions, in terms of the overall behaviour of the congruences of trajectories of material particles and photons propagating and evolving dynamically. This analysis of the congruences of non-space-like curves, either geodesic or otherwise, showed how gravitational focusing took place in the Universe, giving rise to caustics and conjugate points. Before general singularity theorems could be constructed, however, another important mathematical input was needed in addition to the Raychaudhuri equation. This was the analysis of the causality structure and general global properties of a spacetime manifold. This particular development took place mainly in the late 1960s (for a detailed discussion, see e.g. [6] or [7]). The work of Penrose, Hawking, and Geroch then combined these two important features, namely the gravitational focusing effects due to matter and causal structure constraints following from global spacetime properties, to obtain the singularity theorems.

We review these developments briefly in §2. We then point out that the main question here is that of genericity of such spacetime singularities, either in cosmology or in collapse situations. The singularity theorems, while proving the existence of geodesic incompleteness, by themselves provide no information either on the structure and properties of such singularities or on the growth of curvature in their vicinity. Some of these issues, together with the possible avoidance of singularities, are considered in §3. We then discuss the genericity aspects of visible singularities in §4. Section 5 contains some concluding remarks.

2. The occurrence of singularities

We discuss here the occurrence of spacetime singularities in some detail within a general spacetime framework, and point out the crucial role of the Raychaudhuri equation therein. The basic ideas involved in the proofs of the singularity theorems are reviewed, and what these theorems do not imply is pointed out.

We observe the Universe today to very far depths in space and time. While looking deep into space, say for example in diametrically opposite directions, regions with extremely distant galaxies are seen in each direction. Interestingly, these regions have quite similar properties in terms of their appearance and the homogeneity in their galactic spatial distribution. These regions are, however, so far away from each other that they have had no time to interact mutually. That is because the age of the Universe since the big bang has not been large enough for such interactions to have occurred in the past. Thus, within the big bang framework of cosmology, a relevant question arises: how come these regions have such similar properties? This is one of the major puzzles of modern cosmology today.

This observed homogeneity and isotropy of the Universe at large enough scales can be modelled by the Robertson–Walker geometry. The metric of the corresponding spacetime is given by

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + R^2(t)\left[\frac{\mathrm{d}r^2}{(1-kr^2)} + r^2\mathrm{d}\Omega^2\right]. \tag{1}$$

Here $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2$ is the metric on a two-sphere. The Universe is assumed to be spherically symmetric here. There is an additional assumption that the matter content of the spacetime is homogeneous and isotropic. That is, the matter density is the same everywhere in the Universe and also looks the same in all directions. Combining this geometry with the Einstein equations and solving the same, one is led to the Friedmann solution yielding a description of the dynamical evolution of the Universe. The latter requires that the Universe must have had a beginning at a finite time in the past. This is the epoch of the so-called big bang singularity. The matter density as well as the curvatures of spacetime diverge in the limit of approaching this cosmological singularity. This is where all non-spacelike geodesics are incomplete at a point in the past where spacetime comes to an end.

A similar occurrence of the formation of a spacetime singularity takes place when a massive star collapses freely under the force of its own gravity after the exhaustion of its nuclear fuel. If the mass of the star is small enough, the latter can stabilize as a white dwarf or a neutron star as its end-state. However, in case the mass is much larger, say of the order of tens of solar masses, a continual collapse is inevitable once there are no internal pressures left to sustain the star. Such a scenario was considered and modelled by Oppenheimer and Snyder [8], when they considered a collapsing spherical cloud of dust. Again, according to the equations of general relativity, a spacetime singularity of infinite density and curvature forms at the center of the collapsing cloud.

These singularities, after their discovery, were debated extensively by gravitation theorists. An important question that was persistently asked at this juncture was the following. Why should these models be taken so seriously when they assume so many symmetries? Perhaps such a spacetime singularity was arising as a result of these assumptions and could only occur in such a special situation. In other words, these singularities could be just some isolated examples. After all, the Einstein equations

$$R_{ab} - \frac{1}{2}Rg_{ab} = 8\pi T_{ab}, \tag{2}$$

governing the ever present force of gravity, are a complex system of second-order, non-linear, partial differential equations admitting an infinite space of solutions in which the models discussed above are isolated examples. In other words, the issue was the absence of any proof that singularities would always occur in a general enough gravitational collapse when a massive star dies, or in a generic enough cosmological model. In fact, there was widespread belief in the 1940s and 1950s that such singularities would be removed both from stellar collapse and from the cosmological beginning of the Universe (which are two very important physical situations, where gravity becomes a dominant force), once assumptions like the dust form of matter and the spherical symmetry of the model were relaxed and more general solutions to the Einstein equations were considered. This is where the work by Raychaudhuri on the gravitational focusing of matter or light in a spacetime, and that of Penrose, Hawking, and Geroch on the causal structure and global properties of a general spacetime, the culmination of which were the singularity theorems, became relevant. These theorems showed that spacetime singularities, such as those depicted in the above examples, manifested themselves in a large class of models under quite general physical conditions.

We now outline the basic idea and the chain of logic behind the proof of a typical singularity theorem, highlighting the role of the Raychaudhuri equation therein. There are several singularity theorems available which establish the non-space-like geodesic incompleteness for a spacetime under different sets of physical conditions. Each of these may be more relevant to one or the other specific physical situation, and may be applicable to different physical systems such as stellar collapse or the Universe as a whole. However, the most general of these is the Hawking–Penrose theorem [9], which is applicable to both the collapse situation and the cosmological scenario. Let us first briefly describe the main steps in the proof of this theorem. Using causal structure analysis, it is first shown that between certain pairs of events in spacetime there must exist time-like geodesics of maximal length. However, from both causal structure analysis and the global properties of a spacetime manifold M (which is assumed to satisfy the generic condition as well as a specific energy condition), it follows that a causal geodesic, which is complete with regard to both the future and the past, must contain pairs of conjugate points where the nearby null or time-like geodesics intersect. One is then led to a contradiction because the maximal geodesics mentioned above cannot contain any conjugate points, the existence of which would be against their maximality. Thus M itself must have non-space-like geodesic incompleteness. We restate the theorem more carefully and accurately below.

A spacetime (M, g) cannot be time-like and null geodesically complete if the following are satisfied:
(1) $R_{ij}K^iK^j \geq 0$ for all non-space-like vectors K^i;
(2) the generic condition is satisfied, that is, every non-space-like geodesic contains a point at which $K_{[i}R_{j]el[m}K_{n]}K^eK^l \neq 0$, where K is the tangent to the non-space-like geodesic;
(3) the chronology condition holds, and
(4) there exists in M either a compact achronal set without edge, or a closed trapped surface, or a point p such that for all past directed null geodesics from p, eventually the expansion parameter θ must be negative.

The first condition above is an energy condition. All classical fields have been observed to satisfy a suitable positivity of energy requirement. The second condition is a statement that all non-space-like trajectories do encounter some non-zero matter or stress–energy density somewhere during their entire path. The third is a global causality requirement to the effect that there are no closed time-like curves in the spacetime. Finally, in the last condition above, the first or the second part corresponds to a gravitational collapse situation, whereas the last part corresponds to gravitational focusing within a cosmological framework, where θ denotes the expansion for the congruence of non-space-like curves. If any of these conditions are satisfied then the theorem can go through.

The main idea of the proof is that one shows that the following three conditions cannot hold simultaneously: (a) every inextendible non-space-like geodesic contains pairs of conjugate points, (b) the chronology condition holds, (c) there exists an achronal set $\mathcal{S}$ in spacetime such that $E^+(\mathcal{S})$ or $E^-(\mathcal{S})$ is compact.

In the above, for any set $\mathcal{S}$, the future of the set is denoted by $I^+(\mathcal{S})$, which is the set of all those points which can be connected by future-directed time-like curves from any point in $\mathcal{S}$. The past of a set, denoted by I^- is defined similarly dually. An achronal set is a set of spacetime events, of which no two are chronologically related. The set $E(\mathcal{S})$ here denotes the edge of $\mathcal{S}$ which is a set of all points $x \in \mathcal{S}$ such that every neighbourhood of x.contains $y \in I^+(x)$ and $z \in I^-(x)$, such that there is a time-like curve that does not meet $\mathcal{S}$.

If the above is shown then the theorem is proved, because the condition (3) is same as (b), the condition (4) implies (c), and conditions (1) and (2) imply (a). First, we note that (a) and (b) imply the strong causality of the spacetime (see Proposition 6.4.6 of [5]). Strong causality is a causality condition stronger than the chronology which rules out the occurrence of closed time-like curves, and ensures that once a time-like path has left an event in the spacetime, it cannot return in the arbitrary vicinity of the same.

While for further details and definitions, we refer to [5] or [6], in the following we give a brief outline of the main argument for the proof. It can be shown that if $\mathcal{S}$ is a future trapped set, which essentially means that no non-space-like paths escape its own future, and if strong causality holds on $\overline{I^+(\mathcal{S})}$ then there exists a future endless trip γ such that $\gamma \subset \operatorname{Int} D^+(E^+(\gamma))$. Here the set D^+ denotes the future domain of dependence for a set. Now, one defines $T = \overline{J^-(\gamma)} \cap E^+(\mathcal{S})$, then T turns out to be past trapped and hence there exists λ, a past endless causal geodesic in $\operatorname{Int}(D^-(E^-(T))$. Then one chooses a sequence $\{a_i\}$ receding into the past on λ and a sequence $\{c_i\}$ on γ to the future. The sets $J^-(c_i) \cap J^+(a_i)$ are compact and globally hyperbolic. So there exists a maximal geodesic μ_i from a_i to c_i for each i. The intersections of μ_i with the compact set T have a limit point p and a limiting causal direction. The causal geodesic μ with this direction at p must have a pair of conjugate points. This is then shown to be contradictory to the maximality property of the geodesics stated above.

The Raychaudhuri equation plays a key role in the above argument. It implies the occurrence of conjugate points on the null or time-like geodesics under consideration. To state this in a somewhat different way, let us suppose that the spacetime manifold M is non-space-like geodesically complete. In that case, causal structure and the causality as well as regularity assumptions imply that there must exist maximal non-space-like geodesics between certain pairs of spacetime points or along the non-space-like geodesics which are orthogonal either to a space-like hypersurface or to a trapped spacetime surface. However, a suitable energy condition, such as the weak energy condition given by $T_{ab}V^aV^b \geq 0$, together with Einstein equations, implies that $R_{ab}V^aV^b \geq 0$, which in turn implies from the Raychaudhuri equation that conjugate points must develop along these maximal length geodesic trajectories. This contradicts the statement, dictated by causal structure requirements, that maximal geodesics cannot contain any conjugate points.

3. Singularity avoidance

It would be quite attractive if one could avoid any such singularity at the classical level itself. Efforts were made in that direction after singularities had been found in the special cases of FRW cosmologies and the Schwarzschild solution related to gravitational collapse. It was believed that, by going to situations which did not require those exact symmetries, one might get rid of spacetime singularities altogether. However, as discussed above, the singularity theorems show that this is not possible. Once we accept that geodesic incompleteness represents a genuinely singular behaviour, the only way to avoid spacetime singularities at a purely classical level within the framework of Einstein gravity would be to somehow break or violate one of the assumptions or conditions which have been used in proving these theorems. Let us consider this possibility in some detail. The assumptions used in proving various singularity theorems fall mainly into three classes, as pointed out above. The first is a reasonable causality condition on spacetime; it ensures an overall well-behaved global behaviour for the Universe. The second is some suitable energy condition that emphasises the positive nature of the energy density of classical matter fields. The final assumption is the occurrence of a suitable focusing effect in spacetime, as caused by the presence of matter fields. This occurrence is typically in the form of trapped surfaces present in situations of gravitational collapse or a suitable overall focusing in the cosmological situation.

First, consider a possible violation of the causality condition. Note that the Einstein equations as such do not demand causality of spacetime as a whole. In fact, they impose no constraints on the global topology of spacetime. Hence, without breaking any consistency requirements, one could consider causality violating spacetimes and see if somehow the occurrence of spacetime singularities could be avoided there. This issue has been analysed in some detail. It turns out that if causality is violated in any finite region of spacetime, that in its own right causes spacetime singularities in the form of geodesic incompleteness [10,11]. These results show that the occurrence of closed time-like lines in any finite region of spacetime necessarily causes singularities there. One could of course preserve causality in the Universe by disallowing closed non-space-like curves, but still admitting higher-order causality violations. The latter can occur from time-like paths coming back to arbitrarily close neighbourhoods of an event that they had left. Or, a spacetime (M, g) could be causal, but allowed to admit closed time-like curves with the slightest perturbation of the metric. This last situation is described as a violation of the stable causality condition. The above results show that such higher-order causality violations also give rise to spacetime singularities. One may then conclude that violating causality is not an effective alternative means to avoid spacetime singularities. Of course, as noted earlier, the Gödel Universe is geodesically complete and violates causality at every event of spacetime. So there might be a possibility that the violation of causality, introduced at every point of spacetime, may be able to avoid any singularity. However, such a Universe itself would be quite 'singular' in some sense. Indeed, because of this and other unphysical features, the Gödel Universe has never been regarded as a suitable model [5].

The second alternative, of violating the energy condition, has again been investigated in some detail (see e.g. [10] for a discussion on this). The essential conclusion here has been the following. So long as this condition, even if violated locally, holds in a spacetime averaged sense, the existence of singularities in the form of non-space-like geodesic incompleteness would be implied. Only if one violates the energy condition for almost all the events in the Universe making the energy density an overall negative quantity, could one avoid any possible singularity. This does not look to be physically achievable, at least for classical fields. So one could conclude that the violation of the energy condition does not offer an effective alternative in avoiding spacetime singularities. In other words, local violations of the energy condition do not affect the occurrence of singularities. Furthermore, no physical mechanisms have been observed or known which may allow such a violation of the energy condition in the Universe.

This brings us to the final and most important alternative, namely that of avoiding sufficient gravitational focusing in spacetime by breaking condition (4) above in the proof of the singularity theorem. This would amount to disallowing the formation of any trapped surfaces in gravitational collapse and similarly not having enough convergence in a cosmological scenario. Basically, if one wants to avoid all trappings, without violating the energy condition, the only option would be to have very little matter and stress–energy density, so that light rays just avoid getting sufficiently focused. This is the direction that some recent works appear to point to [2,12]. Similar results were obtained earlier, showing that if the space-like surfaces have sufficient matter in some non-vanishing average sense, then *all* non-space-like trajectories would be incomplete in the past [13,14]. For example, if the microwave background radiation is taken to be universal and also to have a global minimum in its energy density over a space-like surface, that will cause the necessary convergence; all non-space-like trajectories will then be incomplete in the past. Thus the avoidance of trapped surfaces or cosmological focusing would be one way not to have space-time singularities. However, if trapped surfaces did form in spherical or other more general gravitational collapse processes, the occurrence of singularities will be a generic phenomenon. We can elaborate this last point. Let such trapped surfaces develop as the collapsing system evolves dynamically. Could we generalize these conclusions for a non-spherically symmetric collapse? Are they valid at least for small perturbations from exact spherical symmetry. By using [5] the stability of Cauchy development in general relativity, one can show that the formation of trapped surfaces is a stable feature when departures from spherical symmetry are taken into account. The argument goes as follows. Consider a spherically symmetric collapse evolution from given initial data on a partial Cauchy surface S. Then trapped surfaces $\mathcal{T}$ are found to form in the shape of all spheres with $r < 2m$ in the exterior Schwarzschild geometry. The stability of Cauchy development then implies that, for all initial data sufficiently near the original data in the compact region $J^+(S) \cap J^-(\mathcal{T})$, trapped surfaces must still occur. The curvature singularity of a spherical collapse also turns out to be a stable feature as implied by the singularity theorems discussed above. The latter shows that closed trapped surfaces always imply the existence of a spacetime singularity under reasonably general conditions. In this sense, such singularities, when they occur, are really generic.

4. Genericity of naked singularities in spherical collapse

The above consideration points to a wide variety of circumstances under which singularities develop in general relativistic cosmologies and in many gravitational collapse processes. Singularity theorems imply the existence of vast classes of solutions to the Einstein equations that must contain spacetime singularities, as characterized by the conditions of these theorems, and of which the big bang singularity is one example. These theorems therefore imply that singularities must occur in Einstein's theory quite generically, i.e. under rather general physically reasonable conditions on the underlying spacetime. Historically, this implication considerably strengthened our confidence in the big bang model which is used extensively in cosmology today.

As mentioned earlier, the singularities are predicted to be either visible to external observers or hidden inside the event horizons of gravity. This is related to the causal structure in the vicinity of the concerned singularity and is particularly relevant to realistic gravitational collapse scenarios and to the basics of black hole physics. However, the singularity theorems provide no further information. Yet, we need more information on the structure of the singularities in terms of their visibility, curvature strengths and other such aspects. What is therefore called for is a detailed investigation of the dynamics of gravitational collapse within the framework of Einstein's theory. Extensive investigations have been made in the past couple of decades or so from such a perspective. These have dealt with dynamical gravitational collapse, mainly for spherical models but also in some non-spherical cases, investigating the structure and visibility of the concerned singularities. Depending on the initial conditions and the types of evolution of the collapse process allowed by the Einstein equations, the singularities of collapse can be either visible or covered (see, for example, [15–20] for some recent reviews and further references). For a detailed discussion of the gravitational collapse of radiation shells within a Vaidya metric and the related black hole and naked singularity geometries, we refer to [7].

Based on the work on spherical gravitational collapse done in [21–23], we consider here the genericity of naked singularities forming in such a process. Given the data on the initial density and pressure profiles of the collapsing cloud, classes of solutions to Einstein equations have been constructed which evolve to either a visible or a covered singularity, subject to the satisfaction of regularity and energy conditions. In other words, given the initial matter data on a space-like surface from which the collapse of the massive matter cloud evolves, the rest of the free functions, such as the velocities of the collapsing shells as well as the classes of evolution, can be chosen as allowed by the Einstein equations, subject to an energy condition and suitable regularity conditions. The latter take the collapse either to a black hole or a naked singularity in the final state, depending on the choice made. The basic formalism here can be summarized as follows. The form of matter considered is quite generic, which is any Type-I general matter field subject to an energy condition [5]. In the case of a black hole developing as the end-state of collapse, the spacetime singularity is necessarily hidden behind the event horizon of gravity. In contrast, a naked singularity develops when families of future directed non-space-like trajectories come out from the singularity. These trajectories can in principle communicate information to faraway observers in the spacetime. The existence of such families confirms the formation of a naked singularity, as opposed to a black hole end-state.

The eventual singularity, which is the singularity curve produced by the collapsing matter, is what we study. We show here that the tangent to this curve at the central singularity at $r = 0$ is related to the radially outgoing null geodesics from the singularity, if any. By determining the nature of the singularity curve and its relation to the initial data and the classes of collapse evolution, we are able to deduce whether the formation of any trapped surface during the collapse takes place before or after the singularity. It is this causal structure of the trapped region that determines the possible emergence or otherwise of non-space-like paths from the singularity. That then settles the final outcome of the collapse in terms of either a black hole or naked singularity. Several familiar equations of state for which extensive collapse studies have been made, such as dust or matter with only tangential or radial pressures and others, are special cases of our consideration.

The spacetime geometry within the spherically symmetric collapsing cloud is described by the general metric in the comoving coordinates (t, r, θ, ϕ) as

$$\mathrm{d}s^2 = -\mathrm{e}^{2\nu}\mathrm{d}t^2 + \mathrm{e}^{2\psi}\mathrm{d}r^2 + R^2(t,r)\mathrm{d}\Omega^2, \tag{3}$$

where $\mathrm{d}\Omega^2$ is the line element on a two-sphere. The matter fields, considered by us, belong to a broad class, called Type-I, where the energy–momentum tensor has one time-like and three space-like eigenvectors. This general class includes most physically reasonable matter, including dust, perfect fluids, massless scalar fields and such others. The stress–energy tensor for this class is given in a diagonal form: $T_t^t = -\rho, T_r^r = p_{\mathrm{r}}, T_\theta^\theta = T_\phi^\phi = p_\theta$. Here ρ, p_{r} and p_θ are the energy density, the radial pressure and the tangential pressure respectively. We also take the matter field to satisfy the weak energy condition. This means that the energy density measured by any local observer must be non-negative. So, for any time-like vector V^i, one must have $T_{ik}V^iV^k \geq 0$. The latter amounts to taking $\rho \geq 0$, $\rho + p_{\mathrm{r}} \geq 0$, $\rho + p_\theta \geq 0$. Now, for the metric (3), Einstein equations take the form (in the units $8\pi G = c = 1$):

$$\rho = \frac{F'}{R^2R'}; \quad p_{\mathrm{r}} = -\frac{\dot{F}}{R^2\dot{R}}, \tag{4}$$

$$\nu' = \frac{2(p_\theta - p_{\mathrm{r}})}{\rho + p_{\mathrm{r}}}\frac{R'}{R} - \frac{p_{\mathrm{r}}'}{\rho + p_{\mathrm{r}}}, \tag{5}$$

$$-2\dot{R}' + R'\frac{\dot{G}}{G} + \dot{R}\frac{H'}{H} = 0, \tag{6}$$

$$G - H = 1 - \frac{F}{R}, \tag{7}$$

where we have defined $G(t,r) = \mathrm{e}^{-2\psi}(R')^2$ and $H(t,r) = \mathrm{e}^{-2\nu}(\dot{R})^2$.

The arbitrary function $F = F(t,r)$ has the interpretation of the mass function for the cloud, giving the total mass in a shell of comoving radius r. The energy condition implies that $F \geq 0$. In order to preserve regularity at the initial epoch, we need to take $F(t_i, 0) = 0$, i.e. the mass function must vanish at the center of the cloud. As seen from eq. (4), there is a density singularity in the spacetime at $R = 0$, and one at $R' = 0$. However, the latter is due to shell-crossings and can be possibly removed [24] through a suitable extension of the spacetime. In any case, we are interested here only in the shell-focusing singularity at $R = 0$. This is a physical singularity where all the matter shells collapse to a zero physical radius. We can use the scaling freedom available for the radial coordinate r to write $R = r$ at the initial epoch $t = t_i$ from where the collapse commences. Introducing the function $v(t,r)$ by the relation

$$R(t,r) = rv(t,r), \tag{8}$$

we have $v(t_i, r) = 1$ and $v(t_{\mathrm{s}}(r), r) = 0$. The collapse situation is now characterized by the condition $\dot{v} < 0$. The time $t = t_{\mathrm{s}}(r)$ corresponds to the shell-focusing singularity $R = 0$ where all the matter shells collapse to a vanishing physical radius.

From the standpoint of the dynamic evolution of the initial data, prescribed at the initial epoch $t = t_i$, there are six arbitrary functions of the comoving shell radius r as given by, $\nu(t_i,r) = \nu_0(r), \psi(t_i,r) = \psi_0(r), R(t_i,r) = r\rho(t_i,r) = \rho_0(r), p_{\mathrm{r}}(t_i,r) = p_{\mathrm{r}_0}(r)$, $p_\theta(t_i,r) = p_{\theta_0}$. We note that not all of the initial data above are mutually independent, because from eq. (5) one gets a relation that gives the initial function $\nu_0(r)$ in terms of the rest of the other initial data functions. The initial pressures should have a physically reasonable behaviour at the center ($r = 0$) in that the pressure gradients should vanish there, i.e. $p'_{\mathrm{r}_0}(0) = p'_{\theta_0}(0) = 0$, and also the difference between the radial and the tangential pressures needs to vanish at the center, i.e. $p_{\mathrm{r}_0}(0) - p_{\theta_0}(0) = 0$. These conditions are necessary to ensure the regularity of the initial data at the center of the cloud. It is then evident from eq. (9) that $\nu_0(r)$ has the form

$$\nu_0(r) = r^2 g(r), \tag{9}$$

where $g(r)$ is at least a C^1 function of r for $r = 0$, and at least a C^2 function for $r > 0$. We see that there are five total field equations with seven unknowns, ρ, p_{r}, p_θ, ψ, ν, R, and F. Thus we have the freedom to choose two free functions. In general, their selection, subject to the weak energy condition and the given initial data for collapse at the starting surface, determines the matter distribution and the metric of the spacetime, and thus leads to a particular time evolution of the initial data.

Let us construct the classes of solutions to Einstein equations which give the collapse evolution from the given initial data. Consider first the general class of mass functions $F(t,r)$ for the collapsing cloud. These are given as

$$F(t,r) = r^3 \mathcal{M}(r,v), \tag{10}$$

$\mathcal{M}$ being any regular and suitably differentiable general function without further restrictions. It turns out from equations given below that the regularity and finiteness of the density profile at the initial epoch $t = t_i$ requires F to go as r^3 close to the center. It follows that the form of F above is not really any special choice, but is in the general class of mass functions consistent with the collapse regularity conditions. Equations (4) yield

$$\rho = \frac{3\mathcal{M} + r\left[\mathcal{M}_{,r} + \mathcal{M}_{,v}v'\right]}{v^2(v+rv')}; \quad p_{\mathrm{r}} = -\frac{\mathcal{M}_{,v}}{v^2}. \tag{11}$$

Thus the regular density distribution at the initial epoch is given by $\rho_0(r) = 3\mathcal{M}(r,1) + r\mathcal{M}(r,1)_{,r}$. It is evident in general that, as $v \to 0$, $\rho \to \infty$ and $p_{\mathrm{r}} \to \infty$, i.e. both the density and the radial pressure blow up at the singularity. One can, in fact, show the following. Given any regular initial density and pressure profiles for the matter cloud from which the collapse develops, there always exist velocity profiles for collapsing matter shells as well as classes of dynamical evolutions, as determined by the Einstein equations, which lead to a naked singularity or a black hole as the end-state of collapse, depending on the choice of the class.

We now provide classes of solutions to Einstein equations to this effect. Consider the class of velocity profiles as determined by the general function

$$\nu = A(t,R), \tag{12}$$

where $A(t,R)$ is any arbitrary, suitably differentiable function of t and R, with the initial constraint $A(t_i,R) = \nu_0(r)$. Using (12) in (6), we have

$$G(t,r) = b(r)\mathrm{e}^{2(A - \int A_{,t}\mathrm{d}t)}. \tag{13}$$

Here $b(r)$ is another arbitrary function of r which emerges on integrating the Einstein equations. A comparison with dust models enables one to interpret $b(r)$ as the velocity function for the shells. From eq. (9), the form of $A(t,R)$ is seen to be $A(t,R) = r^2 g_1(r,v)$, where $g_1(r,v)$ is a suitably differentiable function and $g_1(r,1) = g(r)$. Similarly, we have $A - \int A_{,t}\mathrm{d}t = r^2 g_2(r,v)$ and at the initial epoch $g_2(r,1) = g(r)$. Using (12) in (5), we obtain

$$2p_\theta = RA_{,R}(\rho + p_{\mathrm{r}}) + 2p_{\mathrm{r}} + \frac{Rp_{\mathrm{r}}'}{R'}. \tag{14}$$

In general, both the density and the radial pressure are known to blow up at the singularity. However, the above equation implies that the tangential pressure also does the same.

Writing

$$b(r) = 1 + r^2 b_0(r) \tag{15}$$

and using eqs (10), (12) and (13) in (7), we are led to

$$\sqrt{R}\dot{R} = -\mathrm{e}^{r^2 g_1(r,v)}\sqrt{(1+r^2 b_0)R\mathrm{e}^{r^2 g_2(r,v)} - R + r^3\mathcal{M}}. \tag{16}$$

Since we are considering a collapse situation, we have $\dot{R} < 0$. Defining a function $h(r,v)$ as

$$h(r,v) = \frac{\mathrm{e}^{2r^2 g_2(r,v)} - 1}{r^2} = 2g_2(r,v) + \mathcal{O}(r^2 v^2) \tag{17}$$

and using (17) in (16), we obtain after some simplifications:

$$\sqrt{v}\dot{v} = -\sqrt{\mathrm{e}^{2r^2(g_1+g_2)} v b_0 + \mathrm{e}^{2r^2 g_1}\left(vh(r,v) + \mathcal{M}(r,v)\right)}. \tag{18}$$

Integrating the above equation we have

$$t(v,r) = \int_v^1 \frac{\sqrt{v}\mathrm{d}v}{\sqrt{\mathrm{e}^{2r^2(g_1+g_2)} v b_0 + \mathrm{e}^{2r^2 g_1}\left(vh + \mathcal{M}\right)}}. \tag{19}$$

Note that r is treated as a constant in the above integration. Expanding $t(v,r)$ around the center, we have

$$t(v,r) = t(v,0) + r\mathcal{X}(v) + \mathcal{O}(r^2), \tag{20}$$

where the function $\mathcal{X}(v)$ is given by

$$\mathcal{X}(v) = -\frac{1}{2}\int_v^1 \mathrm{d}v \frac{\sqrt{v}(b_1 v + vh_1(v) + \mathcal{M}_1(v))}{(b_{00}v + vh_0(v) + \mathcal{M}_0(v))^{3/2}}, \tag{21}$$

with, $b_{00} = b_0(0)$, $\mathcal{M}_0(v) = \mathcal{M}(0,v)$, $h_0 = h(0,v)$, $b_1 = b_0'(0)$, $M_1(v) = \mathcal{M}_{,r}(0,v)$ and $h_1 = h_{,r}(0,v)$.

From the above, the time when the central singularity develops is given by

$$t_{\mathrm{s}_0} = \int_0^1 \frac{\sqrt{v}\mathrm{d}v}{\sqrt{b_{00}v + vh_0(v) + \mathcal{M}_0(v)}}. \tag{22}$$

The time for other shells to reach the singularity can be given by the expansion

$$t_{\mathrm{s}}(r) = t_{\mathrm{s}_0} + r\mathcal{X}(0) + \mathcal{O}(r^2). \tag{23}$$

It is now clear that the value of $\mathcal{X}(0)$ depends on the functions $b_0, \mathcal{M}$ and h, which in turn depend on the initial data at $t = t_i$ and on the dynamical variable v that evolves in time from a value $v = 1$ at the initial epoch to $v = 0$ at the singularity. Thus, a given set of initial matter distribution and the dynamical profiles including the velocity of shells completely determine the tangent at the center to the singularity curve. Equations (17)–(19), lead to

$$\sqrt{v}v' = \mathcal{X}(v)\sqrt{b_{00}v + vh_0(v) + \mathcal{M}_0(v)} + \mathcal{O}(r^2). \tag{24}$$

The end-state of collapse, in the form of either a black hole or a naked singularity, is determined by the causal behaviour of the apparent horizon, which is the boundary of trapped surfaces forming due to collapse, and is given by $R = F$. If the neighbourhood of the center gets trapped prior to the epoch of singularity, then it is covered and a black hole results. Otherwise, it would be a naked singularity with the non-space-like future directed trajectories escaping from it. It is then to be determined if there are any families of future directed non-space-like paths emerging from the singularity. To see this as well as to examine the nature of the central singularity at $R = 0$, $r = 0$, one has to consider the equation

$$\frac{\mathrm{d}t}{\mathrm{d}r} = \mathrm{e}^{\psi-\nu} \tag{25}$$

for outgoing radial null geodesics. For a further discussion on this, we refer to [21–23]. The main result that follows from this is that if the quantity $\mathcal{X}(0)$ is positive, one indeed gets radially outgoing null geodesics coming out from the central singularity, which then has to be naked. However, if $\mathcal{X}(0)$ is negative, we have a black hole solution, since there are no such trajectories coming out. If $\mathcal{X}(0)$ vanishes, then we have to take into account the next higher-order non-zero term in the singularity curve $t_{\mathrm{s}}(r)$, and a similar analysis can be carried out.

In the above discussion, the functions h and $\mathcal{M}$ have been expanded in r around $r = 0$ and only the first-order terms have been retained. The key point here is that the functions chosen for the classes of collapse evolution from the given initial data are such that the eventual singularity curve $t_{\mathrm{s}}(r)$ is expandable at the center $r = 0$, at least to first order. The well-known and extensively studied example of dust collapse (see e.g. [25] and references therein) does satisfy this feature. It thus forms a special sub-class of the above classes of solutions for spherical collapse. Various other collapse models studied earlier, such as collapse with a non-zero tangential pressure and others also satisfy such a behaviour of the singularity curve. For a detailed discussion of the differentiability conditions on the functions involved and of the classes of models that satisfy these, we refer to the papers cited above. The treatment given here applies to general classes of functions with the minimum differentiability conditions. However, in discussions of collapse these are sometimes assumed to be analytic and expandable with respect to r^2 with the argument that such smooth functions are physically more relevant. Such assumptions are in fact due to computational convenience, though one can mathematically exploit the freedom of definition and choice available. The formalism, outlined above, would of course work for such smooth functions, which constitute a special case of what has been considered. Consequently, one can have a smooth and differentiable singularity curve.

We thus see how the initial data determine the black hole and naked singularity phases as collapse end-states in terms of the available free functions. This happens because the quantity $\mathcal{X}(0)$ is determined by these initial and dynamical profiles, as given by (21). It is clear that, given any regular density and pressure profiles for the matter cloud from which the collapse develops, one can always choose velocity profiles such that the end-state of the collapse would be either a naked singularity or a black hole, and vice versa. It is interesting to note here that physical agencies, such as the spacetime shear within a dynamically collapsing cloud, could naturally give rise to such phases in gravitational collapse [26]. In other words, such physical factors can naturally delay the formation of any apparent horizon and trapped surfaces.

As stated above, we have worked here with the Type-I matter fields. This is a rather general form of matter which includes practically all known physically reasonable fields such as dust, perfect fluids, massless scalar fields and so on. However, note that we have assumed no explicit equation of state relating the density and pressure variables. Assuming a specific equation of state will fully close the system, all the concerned functions being then determined as a result of the development of the system from the initial data. We presently have little idea about the kind of equation of state that the matter should follow. This is particularly true when the matter is close to the collapse end-state, where we are dealing with ultra-high densities and pressures. One might as well be allowed to freely choose the property of the matter fields, as done above. It is nevertheless important to make the following point. The analysis given above does include several well-known classes of collapse models and equations of state. We had a choice of two free functions available and constructed general classes of collapse evolution using the mass function $F(r, v)$ and the metric function ν. Suppose now that, in addition to Einstein equations, an equation of state of the form $p_{\mathrm{r}} = f(\rho)$ (or alternatively, $p_\theta = g(\rho)$) is given. Then, as evident from eq. (4), there would be a constraint on the otherwise arbitrary function $\mathcal{M}$, specifying the required class, provided a solution to the constraint equation exists. Then the value of p_θ (or p_{r}) gets determined in terms of $\rho(t, r)$ from eq. (5), and the analysis for the black hole and naked singularity phases goes through as above.

For example, in the case of collapsing dust models, we have

$$p_{\mathrm{r}} = p_\theta = 0 \tag{26}$$

and the constraint equation yields

$$\mathcal{M}(r, v) = \mathcal{M}(r). \tag{27}$$

Next, for collapse with a constant (or zero) radial pressure, but with an allowed variable tangential pressure, the constraint equation leads to

$$\mathcal{M}(r, v) = f(r) - kv^3, \tag{28}$$

where k is the value of the constant radial pressure. The tangential pressure will then be given by

$$p_\theta = k + \frac{1}{2} A_{,R} R(\rho + k). \tag{29}$$

Along with these two well-known models, the analysis works for any other model in which any one of the pressures is specified by an equation of state and which permits a solution to the constraint equation on $\mathcal{M}$. This provides some insight into how these black hole and naked singularity phases come about as collapse end-states.

The above results provide us with information on the genericity aspects of naked singularities developing as the final states of spherical collapse. Consider a specific collapse evolution for which the quantity $\mathcal{X}(0)$ is positive, leading to the formation of a naked singularity as the end-state. As noted earlier, this value is decided by the initial density and pressure profiles and the class of dynamical evolutions as determined by the Einstein equations. The initial velocity profiles for the collapsing shells are controlled by the function $b_0(r)$ and the collapse evolutions $\mathcal{M}(r,v)$ and $h(r,v)$, which in turn depend on the initial data at $t = t_i$ and the dynamical variable v. Hence, by continuity, any arbitrarily small variation in any of these parameters will continue to yield a value for $\mathcal{X}(0)$ which is again positive. Thus if a collapse evolution creates a naked singularity, any sufficiently small variation in the initial matter and velocity profiles for the collapsing shell, or small variations in the dynamical evolutions functions will still preserve the final outcome of a naked singularity. In this sense the naked singularities of spherical collapse turn out to be completely generic.

We paraphrase the basic result discussed in this article. Given a starting set of regular data in terms of the density and pressure profiles at the initial epoch from which the collapse develops, there are sets of dynamical evolutions, and velocity profiles for the collapsing cloud, i.e. classes of solutions to the Einstein equations, which can evolve the given initial data to produce either a black hole or a naked singularity as the collapse end-state. The exact outcome depends on the choice of the rest of the free functions available. Indeed, the space of initial data can be divided into two distinct subspaces: one containing those that evolve into black holes and another containing those that lead to naked singularities. The latter are generic in the sense described above. In a way this brings out the stability of these collapse outcomes with respect to perturbations in the initial data set and in the choice of collapsing matter fields.

5. Concluding remarks

In this article we have discussed genericity aspects of spacetime singularities. It is seen that in Einstein gravity they occur generically, whether covered within event horizons or as visible to external observers. This is a consequence of the singularity theorems which make crucial use of the Raychaudhuri equation. It is possible that a future quantum theory of gravity may remove or resolve the final singularity of collapse or the initial one in cosmology. In such a scenario, what matters physically are the regions of ultra-strong gravity and spacetime curvature that develop as a result of the accompanying dynamical gravitational processes. This is true even if the very final singularity is removed through quantum effects. The following physical picture then emerges. Dynamical gravitational processes proceed and evolve to create ultra-strong gravity regions in the Universe. Once these form, strong curvature and quantum effects both come into their own in these regions. Quantum gravity then takes over and is likely to resolve the final singularity. Particularly interesting is the case when the singularities of collapse are visible. In such a situation, quantum gravity effects, taking place in those ultra-strong gravity regions, will in principle be accessible and observable to external observers. The consequences are likely to be intriguing [27].

Acknowledgment

I would like to express my deep gratitude to the late A K Raychaudhuri for his keen interest and extremely relevant comments on our work on gravitational collapse and spacetime singularities. It was a great pleasure knowing him and discussing with him, and to benefit from his penetrating insights. I like to warmly thank Probir Roy for several helpful comments, which clarified many important points.

References

[1] See for example, M Bojowald (2005), Living Rev. Rel. 8:11; and references therein, for details on loop quantum gravity effects near the big bang singularity
[2] J Senovilla, *Pramana – J. Phys.* **69(1)**, 31 (2007), gr-qc/0610127
[3] M Rayn and S Shepley, *Inhomogeneous cosmological models* (Princeton University Press, Princeton, 1975)
[4] A K Raychaudhuri, *Phys. Rev.* **98**, 1123 (1955)
See also *Pramana – J. Phys.* **69(1)**, 3 (2007). It gives an insight into how he arrived at his equation
[5] S W Hawking and G F R Ellis, *The large scale structure of space-time* (Cambridge University Press, Cambridge, 1973)
[6] R M Wald, *General relativity* (University of Chicago Press, Chicago, 1984)
[7] P S Joshi, *Global aspects in gravitation and cosmology* (Clarendon Press, OUP, Oxford, 1993)
[8] J R Oppenheimer and H Snyder, *Phys. Rev.* **56**, 455 (1939)
[9] S W Hawking and R Penrose, *Proc. R. Soc. London* **A314**, 529 (1970)
[10] F Tipler, C J S Clarke and G F R Ellis, in *General relativity and gravitation* edited by A Held (Plenum, New York, 1980) vol. 2
[11] P S Joshi, *Phys. Lett.* **A85**, 319 (1981)
[12] N Dadhich and A Raychaudhuri, *Mod. Phys. Lett.* **A14**, 2135 (1999)
[13] P S Joshi, *Pramana – J. Phys.* **15**, 225 (1980)
[14] P S Joshi, *Bull. Astron. Soc. India* **14**, 1 (1986)
[15] A Krolak, *Prog. Theor. Phys. Suppl.* **136**, 45 (1999)
[16] R Penrose in *Black holes and relativistic stars* edited by R M Wald (University of Chicago Press, 1998)
[17] P S Joshi, *Pramana – J. Phys.* **55**, 529 (2000)
[18] M Celerier and P Szekeres, gr-qc/0203094
[19] T Harada, H Iguchi and K Nakao, *Prog. Theor. Phys.* **107**, 449 (2002)
[20] R Giambo, F Giannoni, G Magli and P Piccione, *Commun. Math. Phys.* **235**, 545 (2003)
[21] P S Joshi and I H Dwivedi, *Class. Quant. Grav.* **16**, 41 (1999)
[22] P S Joshi and R Goswami, *Phys. Rev.* **D69**, 064027 (2004)
[23] R Goswami and P S Joshi, gr-qc/0608136 (2006). A discussion on generalization of gravitational collapse scenarios to higher dimensional spacetimes, and the differences that are caused due to increase in dimensions is contained here
[24] C J S Clarke, *The analysis of spacetime singularities* (Cambridge University Press, Cambridge, 1993)

[25] P S Joshi and I H Dwivedi, *Phys. Rev.* **D47**, 5357 (1993)
[26] P S Joshi, N Dadhich and R Maartens, *Phys. Rev.* **D65**, 101501 (2002). It is seen that while in the Vaidya collapse of radiation it is the time rate of collapse that governs the black hole and naked singularity final phases, in the dust case treated above it is the amount of inhomogeneity near the center that governs the final collapse outcome
[27] R Goswami, P S Joshi and P Singh, *Phys. Rev. Lett.* **96**, 031302 (2006)

PRAMANA
— journal of
physics

Vol. 69, No. 1
July 2007
pp. 137–145

On a Raychaudhuri equation for hot gravitating fluids

CHANDRASEKHER MUKKU, SWADESH M MAHAJAN* and BINDU A BAMBAH†
International Institute of Information Technology, Hyderabad 500 032, India
*Institute of Fusion Studies, University of Texas, Austin, Texas 78712, USA
†School of Physics, University of Hyderabad, Hyderabad 500 046, India
E-mail: mukku@iiit.ac.in

Abstract. We generalise the Raychaudhuri equation for the evolution of a self gravitating fluid to include an Abelian and non-Abelian hybrid magneto fluid at a finite temperature. The aim is to utilise this equation for investigating the dynamics of astrophysical high temperature Abelian and non-Abelian plasmas.

Keywords. Raychaudhuri equation; plasmas; magneto fluids.

PACS Nos 04.40.-b; 47.75.+f; 12.38.Mh

1. Introduction

It is perhaps one of the finest tributes to the simplicity and elegance of Einstein's theory of gravitation and to the human spirit that a laboratory scientific assistant, who later became a teacher sitting in Kolkata, was able to conceive the equation for the evolution of a gravitating fluid now known as Raychaudhuri's equation [1]. The equation served as a lemma for the Penrose–Hawking singularity theorems and for the study of exact solutions in general relativity [2,3]. It provides a simple validation of our expectation that gravitation should be a universal attractive force between any two particles in general relativity [4]. This equation has stood the test of time and has been generalised in many ways. It has found applications in modern theories of strings and membranes [5,6]. We attempt a more modest generalisation to tackle the statistical properties of a hot astrophysical plasma. This may be an electromagnetic or a chromomagnetic (quark gluon) plasma. We are guided here by a recent formalism that has been used to investigate the dynamics of a hot charged fluid in terms of a hybrid magneto-fluid [7]. The changes brought about to the Raychaudhuri equation by the introduction of statistical attributes associated with finite temperature are many and interesting. We illustrate some of these changes in the context of the evolution of gravitating non-Abelian plasmas [8] in the early Universe.

2. Raychaudhuri's equation for a 'unified' charged gravitating fluid

For a congruence of time-like vectors U^μ, the standard Raychaudhuri equation can be written down in the following form [9]:

$$\dot{\Theta} = \frac{q}{m}F^{\alpha}{}_{\gamma;\alpha}U^{\gamma} + \frac{q}{m}F^{\alpha}{}_{\gamma}\Omega^{g\gamma}{}_{\alpha} - \Sigma^{g}{}_{\alpha\beta}\Sigma^{g\beta\alpha} - \Omega^{g}{}_{\alpha\beta}\Omega^{g\beta\alpha} \\ -\frac{1}{3}\Theta^2 - 8\pi(t_{\mu\nu} - \frac{1}{2}tg_{\mu\nu})U^\mu U^\nu. \tag{1}$$

With our notation, $(U_\mu U^\mu = -1)$, the limit $F_{\mu\nu} \longrightarrow 0$ is the (geodesic) Raychaudhuri equation. For a geodesic congruence, $\Sigma^g_{\alpha\beta} = \frac{1}{2}(U_{\alpha;\beta} + U_{\beta;\alpha}) - \frac{1}{3}U^\gamma{}_{;\gamma}(g_{\alpha\beta} + U_\alpha U_\beta)$ is a symmetric, tracefree tensor and is called the shear tensor, $\Omega^g_{\alpha\beta} = \frac{1}{2}(U_{\alpha;\beta} - U_{\beta;\alpha})$ is called the vorticity tensor and $\Theta = U^\alpha{}_{;\alpha}$ is called the expansion scalar. Also, Einstein's equations have been used to replace the Ricci tensor by the energy-momentum tensor.

For the use of the Raychaudhuri equation to describe astrophysical plasmas in which gravitational effects are compatible with the fluid attributes, we have to account for the temperature and pressure of the fluid. For this, we incorporate into a small volume element of the fluid a statistical factor f which represents a temperature-dependent statistical attribute of the fluid, and is related to the enthalpy h, the scalar density in the rest frame n and the mass m of the fluid particles by the relation $h = mnf(T)$. When one does the kinetic theory of high temperature plasmas, $f(T)$ seems to emerge as the most useful variable to represent temperature effects. For relativistic plasmas, $h = mnK_3(m_s/T)/K_2(m_s/T)$ and $f(T)$ is purely a positive function of temperature. The velocity vector of the fluid is obtained as the average velocity of this small volume of the fluid and is written as $V^\mu = fU^\mu$. This drastically alters the character of the terms in the evolution equation. For example, now $V^\mu V_\mu = -f^2$, in contrast with $U^\mu U_\mu = -1$, and, unlike $U^\alpha U_{\alpha;\beta} = \frac{1}{2}(U^\alpha U_\alpha)_{;\beta} = 0$, now $V^\alpha V_{\alpha;\beta} = -f\partial_\beta f$. These terms significantly change the spatial terms of the Raychaudhuri equation and necessitate a generalisation to account for these statistical factors. Indeed, this statistical limit for the fluid velocity [7], unlike the particle limit, provides a natural factor for producing acceleration forces from within the fluid due to pressure and temperature gradients. In the unified magnetofluid picture one can write the equation of motion of a magneto-fluid with entropy σ as

$$T\partial^\nu\sigma = gM^{\mu\nu}U_\mu, \tag{2}$$

where

$$M^{\mu\nu} = F^{\mu\nu} + \frac{m}{g}S^{\mu\nu} \tag{3}$$

and

$$S_{\mu\nu} = \partial_\mu(fU_\nu) - \partial_\nu(fU_\mu) \tag{4}$$

are antisymmetric second rank 'flow' tensors defined in [7]. In this sense, $M_{\mu\nu}$ represents an anti-symmetric 'unified' field-flow tensor constructed from the kinematic (U^μ), statistical $(f(T))$ and electromagnetic $(F_{\mu\nu})$ attributes of the magneto-fluid.

While the statistical factors provide us with additional acceleration terms which alter the purely spatial character of the shear and vorticity tensors, there are some relations that remain unaltered. A trivial example is that V^μ remains orthogonal to the hypersurface with metric $h_{\alpha\beta}$ defined as the projector $h_{\alpha\beta} = g_{\alpha\beta} + U_\alpha U_\beta$. We notice that

$$\begin{aligned} V^\mu{}_{;\alpha\beta} &= (fU^\mu{}_{;\alpha} + U^\mu f_{;\alpha})_{;\beta} \\ &= f_{;\beta}U^\mu{}_{;\alpha} + U^\mu{}_{;\beta}f_{;\alpha} + U^\mu f_{;\alpha\beta} + fU^\mu{}_{;\alpha\beta}. \end{aligned} \tag{5}$$

It is easy to see that anti symmetrization in indices α, β allows us to write

$$V^\mu{}_{;\alpha\beta} - V^\mu{}_{;\beta\alpha} = f(U^\mu{}_{;\alpha\beta} + U^\mu{}_{;\beta\alpha}) = fR^\mu{}_{\sigma\alpha\beta}U^\sigma = R^\mu{}_{\sigma\alpha\beta}V^\sigma, \tag{6}$$

and its character is unchanged in the transition to a hot fluid.

The question that arises now is the following. Is it possible to define the generalizations of the standard definitions of the shear and vorticity tensors in a similar manner, and can they be constructed to be purely spatial tensors?

Let us decompose, following the standard procedure, $\tilde{B}_{\alpha\beta} = V_{\alpha;\beta}$ into its irreducible parts:

$$\tilde{B}_{\alpha\beta} = \tilde{\Sigma}_{\alpha\beta} + \tilde{\Omega}_{\alpha\beta} + \frac{1}{3}\tilde{\Theta}h_{\alpha\beta} + \frac{1}{f}V_\alpha\partial_\beta f. \tag{7}$$

The trace of $\tilde{B}_{\alpha\beta}$ defines the expansion scalar $\tilde{\Theta}$ via

$$\tilde{B}^\alpha{}_\alpha = \tilde{\Theta} + \frac{1}{f}V^\alpha\partial_\alpha f = V^\alpha{}_{;\alpha}, \tag{8}$$

where we have assumed that the generalized shear tensor $\tilde{\Sigma}_{\mu\nu}$ is symmetric and traceless. If we now define

$$\tilde{\Sigma}_{\alpha\beta} = \frac{1}{2}(V_{\alpha;\beta} + V_{\beta;\alpha}) - \frac{1}{3}\tilde{\Theta}h_{\alpha\beta} - \frac{1}{2f}(V_\alpha\partial_\beta f + V_\beta\partial_\alpha f), \tag{9}$$

then its trace

$$\tilde{\Sigma}^\mu{}_\mu = V^\mu{}_{;\mu} - \tilde{\Theta} - \frac{1}{f}V^\mu\partial_\mu f \tag{10}$$

goes to zero, because of eq. (8); the trace-free condition also reproduces the required definition of $\tilde{\Theta}$.

The generalised vorticity tensor $\tilde{\Omega}_{\alpha\beta}$ may also be written as

$$\tilde{\Omega}_{\alpha\beta} = \frac{1}{2}(V_{\alpha;\beta} - V_{\beta;\alpha}) - \frac{1}{2f}(V_\alpha\partial_\beta f - V_\beta\partial_\alpha f). \tag{11}$$

We see that the tensor $S_{\mu\nu}$ defined in eq. (4) allows us the following identification:

$$\tilde{\Omega}_{\alpha\beta} = \frac{1}{2}S_{\beta\alpha} - \frac{1}{2f}(V_\alpha\partial_\beta f - V_\beta\partial_\alpha f). \tag{12}$$

From an earlier work, we know that the gradient of f, for a perfect fluid, is related to the pressure gradient and to Lorentz force from the presence of an electromagnetic field [7]. So, unlike the standard particle picture of the shear and vorticity, the unified picture of shear and vorticity automatically includes accelerations coming from the internal forces of the fluids. This in turn ensures that the shear is now not a purely spatial tensor and the vorticity tensor behaves in a similar manner. It is the sum of the shear and vorticity tensors that leads us to the following simplification:

$$V^\alpha(\tilde{\Sigma}_{\alpha\beta} + \tilde{\Omega}_{\alpha\beta}) = 0. \tag{13}$$

This implies that

$$V^\alpha \tilde{B}_{\alpha\beta} = 0 \tag{14}$$

and

$$\tilde{B}_{\alpha\beta}V^\beta = (\tilde{\Sigma}_{\alpha\beta} + \tilde{\Omega}_{\alpha\beta} + \frac{1}{f}V_\alpha\partial_\beta f)V^\beta = V_{\alpha;\beta}V^\beta = \dot{V}_\alpha. \tag{15}$$

The decomposition of $V_{\alpha;\beta}$ into its irreducible components can therefore be written as

$$V_{\alpha;\beta} = \tilde{B}_{\alpha\beta} = \tilde{\Sigma}_{\alpha\beta} + \tilde{\Omega}_{\alpha\beta} + \frac{1}{3}\tilde{\Theta}h_{\alpha\beta} + \frac{1}{f}V_\alpha\partial_\beta f. \tag{16}$$

So, while the generalized expansion scalar, shear and vorticity tensors show evidence of internal fluid forces, the tensor, $\tilde{B}_{\alpha\beta} = V_{\alpha;\beta}$ itself shows similarity with the original tensor, $B_{\alpha\beta} = U_{\alpha;\beta}$.

From the definitions of $\tilde{\Theta}$, $\tilde{\Sigma}_{\mu\nu}$ and $\tilde{\Omega}_{\mu\nu}$, the following equations relate the scalar of expansion, shear and vorticity for a geodesic congruence in the particle view to those in the unified view of gravitating fluids with acceleration forces:

$$\begin{aligned} \tilde{\Theta} &= f\Theta, \\ \tilde{\Sigma}_{\mu\nu} &= f\Sigma^g{}_{\mu\nu}, \\ \tilde{\Omega}_{\mu\nu} &= f\Omega^g{}_{\mu\nu}, \end{aligned} \tag{17}$$

which then imply

$$\tilde{B}_{\mu\nu} = fB_{\mu\nu} + V_\mu\partial_\nu \ln f. \tag{18}$$

In analogy with the derivation of the standard Raychaudhuri equation [1], if we consider the parallely transported $\tilde{B}_{\alpha\beta;\gamma} = V_{\alpha;\beta\gamma}$, then from eq. (6), we get

$$\begin{aligned} \tilde{B}_{\alpha\beta;\gamma}V^\gamma &= (V_{\alpha;\gamma\beta}V^\gamma - R_{\alpha\sigma\gamma\beta}V^\sigma V^\gamma) \\ &= ((V_{\alpha;\gamma}V^\gamma)_{;\beta} - V_{\alpha;\gamma}V^\gamma{}_{;\beta} - R_{\alpha\sigma\gamma\beta}V^\sigma V^\gamma) \\ &= (\dot{V}_\alpha)_{;\beta} - \tilde{B}_{\alpha\gamma}\tilde{B}^\gamma{}_\beta - R_{\alpha\sigma\gamma\beta}V^\sigma V^\gamma). \end{aligned} \tag{19}$$

Taking the trace over the indices $\alpha\beta$ and using eq. (8) we have

$$(\tilde{\Theta} + \frac{1}{f}V^{\mu}\partial_{\mu}f)_{;\gamma}V^{\gamma} = (\dot{V^{\alpha}}_{;\alpha} - \tilde{B}_{\alpha\gamma}\tilde{B}^{\gamma\alpha} - R_{\sigma\gamma}V^{\sigma}V^{\gamma}). \tag{20}$$

We now use Einstein's equation to get

$$\begin{aligned}&\left(\tilde{\Theta} + \frac{1}{f}V^{\mu}\partial_{\mu}f\right)_{;\gamma} V^{\gamma} \\ &\quad = \left(\dot{V^{\alpha}}_{;\alpha} - \tilde{B}_{\alpha\beta}\tilde{B}^{\beta\alpha} - 8\pi\left(t_{\mu\nu} - \frac{1}{2}tg_{\mu\nu}\right)V^{\mu}V^{\nu}\right).\end{aligned} \tag{21}$$

The acceleration forces are two-fold, one coming from the EM forces within the fluid and the second from the fluid forces themselves (pressures, etc.). To examine what these are, we return to the equation of motion for a fluid with entropy σ in the unified picture eq. (2) and use the fact that

$$S_{\mu\nu} = \partial_{\mu}(f(T)U_{\nu}) - \partial_{\nu}(f(T)U_{\mu}) = V_{\nu;\mu} - V_{\mu;\nu}. \tag{22}$$

For an isentropic fluid,

$$U^{\mu}\partial_{\mu}\sigma = \sigma_{;\mu}V^{\mu} = 0. \tag{23}$$

Therefore, we can write

$$\begin{aligned}T\partial_{\nu}\sigma &= -TU^{\mu}[U_{\mu}\partial_{\nu}\sigma - U_{\nu}\partial_{\mu}\sigma] \\ &= -\frac{T}{f^2}V^{\mu}[V_{\mu}\partial_{\nu}\sigma - V_{\nu}\partial_{\mu}\sigma].\end{aligned} \tag{24}$$

Putting this together with (3), we find

$$-\frac{T}{f^2}V^{\mu}[V_{\mu}\partial_{\nu}\sigma - V_{\nu}\partial_{\mu}\sigma] = qU^{\mu}\left(F_{\nu\mu} + \frac{m}{q}S_{\nu\mu}\right) \tag{25}$$

or

$$mV^{\mu}S_{\mu\nu} = qV^{\mu}F_{\nu\mu} + \frac{T}{f}V^{\mu}[V_{\mu}\partial_{\nu}\sigma - V_{\nu}\partial_{\mu}\sigma]. \tag{26}$$

Therefore,

$$V^{\mu}V_{\nu;\mu} - V^{\mu}V_{\mu;\nu} = \frac{q}{m}\left(V^{\mu}F_{\nu\mu} + \frac{mT}{qf}V^{\mu}[V_{\mu}\partial_{\nu}\sigma - V_{\nu}\partial_{\mu}\sigma]\right) \tag{27}$$

and

$$\dot{V}_{\nu} = -f\partial_{\nu}f + \frac{q}{m}\left(V^{\mu}F_{\nu\mu} + \frac{mT}{qf}V^{\mu}[V_{\mu}\partial_{\nu}\sigma - V_{\nu}\partial_{\mu}\sigma]\right). \tag{28}$$

Defining a pure fluid factor $N_{\mu\nu}$ as

$$N_{\mu\nu} = \frac{T}{f}(\sigma_{;\mu}V_{\nu} - \sigma_{;\nu}V_{\mu}), \tag{29}$$

the acceleration vector simplifies to

$$\dot{V}_\nu = -f\partial_\nu f + \frac{q}{m}V^\mu\left(F_{\nu\mu} + \frac{m}{q}N_{\nu\mu}\right). \tag{30}$$

Defining

$$G_{\mu\nu} = F_{\mu\nu} + \frac{m}{q}N_{\mu\nu}, \tag{31}$$

substitution into (21) yields

$$\begin{aligned}(\tilde{\Theta} + \frac{1}{f}V^\mu\partial_\mu f)_{;\gamma}V^\gamma = \left(-f\partial^\alpha f + \frac{q}{m}V_\beta G^{\alpha\beta}\right)_{;\alpha} - \tilde{B}_{\alpha\beta}\tilde{B}^{\beta\alpha}\\ - 8\pi t_{\mu\nu}V^\mu V^\nu - 4\pi f^2 t,\end{aligned} \tag{32}$$

which simplifies to

$$\begin{aligned}\dot{\tilde{\Theta}} + \dot{\zeta} &= (-f\partial^\alpha f)_{;\alpha} + \frac{q}{m}V_{\beta;\alpha}G^{\alpha\beta} + \frac{q}{m}V_\beta G^{\alpha\beta}{}_{;\alpha}\\ &-\tilde{B}_{\alpha\beta}\tilde{B}^{\beta\alpha} - 8\pi t_{\mu\nu}V^\mu V^\nu - 4\pi f^2 t,\end{aligned} \tag{33}$$

where we have defined $V^\mu\partial_\mu \ln f = \zeta$. Using the relationship between $B_{\mu\nu}$ and $\tilde{B}_{\mu\nu}$, we find

$$\begin{aligned}\dot{\tilde{\Theta}} + \dot{\zeta} &= (-f\partial^\alpha f)_{;\alpha} + \frac{q}{m}V_{\beta;\alpha}G^{\alpha\beta} + \frac{q}{m}V_\beta G^{\alpha\beta}{}_{;\alpha}\\ &-fB_{\alpha\beta}fB^{\beta\alpha} - 2\tilde{B}^{\alpha\gamma}V_\gamma\partial_\alpha \ln f + \zeta^2 - 8\pi t_{\mu\nu}V^\mu V^\nu - 4\pi f^2 t\\ &= \frac{q}{m}V_{\beta;\alpha}G^{\alpha\beta} + \frac{q}{m}V_\beta G^{\alpha\beta}{}_{;\alpha} - fB_{\alpha\beta}fB^{\beta\alpha}\\ &+\zeta^2 - f^2(\partial^2 \ln f) - 8\pi t_{\mu\nu}V^\mu V^\nu - 4\pi f^2 t.\end{aligned} \tag{34}$$

Simplifying the left-hand side, we have

$$\begin{aligned}\dot{\tilde{\Theta}} &= \frac{q}{m}V_{\beta;\alpha}G^{\alpha\beta} + \frac{q}{m}V_\beta G^{\alpha\beta}{}_{;\alpha} - fB_{\alpha\beta}fB^{\beta\alpha} + \zeta^2\\ &-\dot{\zeta} - f^2(\partial^2 \ln f) - 8\pi t_{\mu\nu}V^\mu V^\nu - 4\pi f^2 t.\end{aligned} \tag{35}$$

On further simplification of the $B_{\mu\nu}B^{\nu\mu}$ terms, we have

$$\begin{aligned}\dot{\tilde{\Theta}} &= \frac{q}{2m}S_{\alpha\beta}G^{\alpha\beta} + \frac{q}{m}V_\beta G^{\alpha\beta}{}_{;\alpha} - \Big(\tilde{\Sigma}_{\alpha\beta}\tilde{\Sigma}^{\beta\alpha}\\ &+\tilde{\Omega}_{\alpha\beta}\tilde{\Omega}^{\beta\alpha} + \frac{1}{3}\tilde{\Theta}^2\Big) + \zeta^2 - \dot{\zeta} - f^2(\partial^2 \ln f)\\ &-8\pi t_{\mu\nu}V^\mu V^\nu - 4\pi f^2 t.\end{aligned} \tag{36}$$

This final form of the generalised Raychaudhuri equation, incorporating the statistical attributes of the gravitating hot plasma, reduces to the standard equation in the limit $f(T) \longrightarrow 1$.

3. A Raychaudhuri equation for the unified non-Abelian magneto fluid

In a recent paper [8], we had generalised the proposed unification of the Abelian magneto fluid [7] to a non-Abelian magneto fluid. The equations of motion for the magneto fluid clearly indicated the possibility of solitonic solutions which are normally absent in the Abelian case. The inherent non-linearities present in the Yang–Mills magneto fluid allow such possibilities and a natural question arises as to what will be the dynamics of a self-gravitating non-Abelian magneto fluid? This is particularly relevant in view of the compelling experimental evidence from the relativistic heavy-ion collider at Brookhaven National Laboratory (BNL) that the universe in its first few moments may have existed as a quark-gluon fluid. Since, large gravitational fields are also present in this epoch, it provides us with a motivation to give a generalisation of the Raychaudhuri equation for the non-Abelian magneto fluid.

The suggestion in [8] was that each worldline would now carry an internal index i labelling the non-Abelian species. Generalising the Lorentz force law to the non-Abelian case, we had derived the equation of motion for the magneto fluid in terms of a unified antisymmetric tensor, $M^i{}_{\mu\nu} = F^i{}_{\mu\nu} + \frac{m}{g} S^i{}_{\mu\nu}$, where $F^i{}_{\mu\nu}$ is the standard Yang–Mills field strength tensor while $S^i{}_{\mu\nu}$ is given by

$$\begin{aligned} S^i{}_{\mu\nu} = {}& \partial_\mu (f U^i{}_\nu) - \partial_\nu (f U^i{}_\mu) - igf[A_\mu, U_\nu]^i + igf[A_\nu, U_\mu]^i \\ & -imf^2[U_\mu, U_\nu]^i, \end{aligned} \tag{37}$$

with the gauge covariant derivative being defined by $\mathcal{D}_\mu = \partial_\mu - ig[A_\mu,]$. The non-Abelian fluid equations of motion corresponding to eq. (3) for the 'Yang–Mills magneto-fluid', with entropy σ are given by

$$T\partial^\nu \sigma = g M_a^{\mu\nu} U_{a\mu}. \tag{38}$$

For a non-Abelian magneto fluid velocity vector $U^i{}_\mu$, we can write

$$U^{i\mu}{}_{;\nu\sigma} - U^{i\mu}{}_{;\sigma\nu} = R^\mu{}_{\rho\nu\sigma} U^{i\rho}. \tag{39}$$

Examining $B^i{}_{\mu\nu} = U^i{}_{\mu;\nu}$,

$$B^i{}_{\mu\nu;\alpha} U_i^\alpha = (U^i{}_{\mu;\alpha} U_i^\alpha)_{;\nu} - R_{\mu\rho\nu\alpha} U^{i\mu} U_i^\rho - U^i{}_{\mu;\alpha} U_i{}^\alpha{}_{;\nu}. \tag{40}$$

From our earlier work [8], we can write

$$\mathrm{tr}\dot{U}_\mu = U^i{}_{\mu;\nu} U_i^\nu = \frac{g}{m} F^i{}_{\mu\nu} U_i^\nu \tag{41}$$

for the generalisation of the Lorentz force law in the particle picture of the non-Abelian magneto fluid; the trace is over the internal, gauge group indices ($i = 1 \ldots N = \dim(\text{gauge group})$). Clearly, since now, $U^i{}_\mu U_i^\mu = -N$, we can henceforth, assume that the $U^i{}_\mu$ are normalised so that $U^i{}_\mu U_i^\mu = -1$, i.e we assume that $U^i{}_\mu \longrightarrow U^i{}_\mu/\sqrt{N}$. With this proviso, we can define an orthogonal projection tensor as before

$$h_{\mu\nu} = g_{\mu\nu} + U^i{}_\mu U_{i\nu}. \tag{42}$$

It follows that

$$U^{j\mu} h_{\mu\nu} = U^{j\mu} g_{\mu\nu} + U^{j\mu} U^i{}_\mu U_{i\nu} = U^j{}_\nu - \delta^{ij} U_{i\nu} = 0 \tag{43}$$

and

$$h_{\mu\nu} U^{j\nu} = g_{\mu\nu} U^{j\nu} + U^i{}_\mu U_{i\nu} U^{j\nu} = U^j{}_\mu - \delta^{ij} U_{i\mu} = 0. \tag{44}$$

Equation (40) can be reduced to

$$B^i{}_{\mu\nu;\alpha} U_i{}^\alpha = (\frac{g}{m} F^i{}_{\mu\alpha} U_i{}^\alpha)_{;\nu} - R_{\mu\rho\nu\alpha} U^{i\mu} U_i{}^\rho - U^i{}_{\mu;\alpha} U_i{}^\alpha{}_{;\nu}. \tag{45}$$

Tracing over indices μ and ν, and defining $\Theta^i = B^i{}_\mu{}^\mu$, we find

$$\dot{\Theta}^i = (\frac{g}{m} F^{i\mu}{}_\alpha U_i{}^\alpha)_{;\mu} - R_{\rho\alpha} U^{i\alpha} U_i{}^\rho - B^i{}_{\mu\alpha} B_i{}^{\alpha\mu}. \tag{46}$$

Once again, we can decompose the tensor $B^i{}_{\mu\nu}$ into its irreducible parts and write

$$B^i{}_{\mu\nu} = U^i{}_{\mu;\nu} = \Sigma^i{}_{\mu\nu} + \Omega^i{}_{\mu\nu} + \frac{1}{3}\Theta^i h_{\mu\nu}. \tag{47}$$

It is to be noted here that we are considering a special sector of the magneto fluid dynamics, that of a non-interacting sector of the full non-Abelian fluid as is seen from our definition of the acceleration vector $a_\mu = U^i{}_{\mu;\nu} U_i{}^\nu$. From a fluid point of view, we are dealing with a multi-species model of the non-Abelian fluid. The full non-Abelian interactions of the fluid will be explored in a future study. Returning to the decomposition of $B^i{}_{\mu\nu}$, we write

$$\begin{aligned} \Sigma^i{}_{\mu\nu} &= \frac{1}{2}(U^i{}_{\mu;\nu} + U^i{}_{\nu;\mu}) - \frac{1}{3}\Theta^i h_{\mu\nu}, \\ \Omega^i{}_{\mu\nu} &= \frac{1}{2}(U^i{}_{\mu;\nu} - U^i{}_{\nu;\mu}). \end{aligned} \tag{48}$$

Clearly, these definitions give a decomposition of $B^i{}_{\mu\nu}$ into its irreducible components. As expected, the resulting Raychaudhuri equation, upon substitution of $B^i{}_{\mu\nu} B_i{}^{\nu\mu}$ into eq. (46), is a multi-species generalisation of the standard Raychaudhuri equation. This non-Abelian generalisation of Raychaudhuri's equation gives us a means of studying the shear and vorticity of quark-gluon astrophysical plasmas.

4. Conclusion

High temperature plasmas have an important role to play in the early universe. In what is known as the 'classical' or the 'radiation' epoch of the universe, both gravitational and high temperature effects are of equal importance. Thus, for the evolutionary dynamics of such a hot gravitating plasma, a generalisation of the

Raychaudhuri equation to include finite temperature fluid forces as well as electromagnetic effects was called for. Using a unified magneto-fluid approach to construct a generalised Raychaudhuri equation, we have attempted to respond to this call. Non-linear plasmas also find an application in cosmology through the (now compelling) evidence that a non-Abelian, non-linear quark gluon fluid existed in the early epochs of the universe. To deal with the evolution of this fluid when the gravitational effects are strong, we have laid the foundation of a generalised Raychaudhuri equation for the evolution of non-Abelian gravitating plasmas. Most interesting plasmas involve collective effects which are non-linear even in the special relativistic situation and the inclusion of gravity can only lead to more intriguing highly non linear phenomena. The investigation of further implications of the ideas, germinated in this paper, is a promising avenue for future research. The amalgamation and unification of ideas that cut across barriers always enriches physics. We feel that any work of this kind is a fitting tribute to a man whose physics has crossed international boundaries.

Acknowledgement

The research of one of the authors (SMM) was supported by the USDOE contract No. DE-FG02-04ER-54742.

References

[1] A Raychaudhuri, *Phys. Rev.* **98**, 1123 (1955)
[2] S Hawking, *Phys. Rev. Lett.* **15**, 689 (1965)
[3] R Penrose, *Rev. Mod. Phys.* **37**, 215 (1965)
[4] N Dadhich, *Derivation of the Raychaudhuri equation*, arXiv:gr-qc/0511123
[5] S Kar, *Phys. Rev.* **D53**, 2071 (1996)
[6] E Zafiris, *Phys. Rev.* **D58**, 043509 (1998)
[7] S M Mahajan, *Phys. Rev. Lett.* **90**, 035001 (2003)
[8] B A Bambah, S M Mahajan and C Mukku, *Phys. Rev. Lett.* **97**, 072301 (2006)
[9] A K Raychaudhuri, *Theoretical cosmology* (Clarendon Press, Oxford, 1979)

PRAMANA
— journal of
physics

Vol. 69, No. 1
July 2007
pp. 147–157

Raychaudhuri equation in quantum gravitational optics

N AHMADI[1] and M NOURI-ZONOZ[1,2]
[1]Department of Physics, University of Tehran, North Karegar Ave., Tehran 14395-547, Iran
[2]Institute for Studies in Theoretical Physics and Mathematics, P.O. Box 19395-5531 Tehran, Iran
E-mail: nahmadi@ut.ac.ir; nouri@theory.ipm.ac.ir

Abstract. The equation of Raychaudhuri is one of the key concepts in the formulation of the singularity theorems introduced by Penrose and Hawking. In the present article, taking into account QED vacuum polarization, we study the propagation of a bundle of rays in a background gravitational field through the perturbative deformation of Raychaudhuri's equation. In a sense, this could be seen as another semiclassical study in which geometry is treated classically but matter (which means the photon here) is allowed to exhibit quantum characteristics that are encoded in its coupling to the background curvature.

Keywords. Raychaudhuri equation; quantum gravitational optics; vacuum polarization.

PACS No. 04.62.+v

1. Introduction

Vacuum polarization is a quantum field-theoretic effect in which the photon exists as a virtual e^+e^- pair for a short time. This virtual transition assigns photons with an effective size of $\mathcal{O}(\lambda_c)$, where λ_c is the Compton wavelength of the electron [1]. In a curved background, the photon propagation will obviously be affected by the gravitational field if the typical scale of the space-time curvature L is comparable to λ_c. A simple analysis of null rays in classical general relativity implies that a curved space-time, compared to the flat case, could be treated as an optical medium with a refractive index [2]. The further addition of QED effects, in the context of semi-classical gravity, leads to interesting phenomena such as dispersive effects and the polarization-dependent propagation of photons. One of the main consequences is the modification of the light cone in such a way that QED photons do not propagate along null geodesics of the background geometry. They propagate instead along null geodesics of an effective geometry. Quantum characteristics of photon propagation in a curved background open up a whole new field. Coined quantum gravitational optics (QGO) [3], it started with the pioneering paper by Drummond and Hathrell and continued with detailed studies of the same effect for specific space-times [4–6]. The modification of the underlying space-time geometry, felt by photons due to QED interactions, has led to some unexpected results.

Here we shall study the geometric properties of *physical* null congruences which are families of physical null rays determined by the effective metric $\mathcal{G}_{\mu\nu}$. These are then distinct from the propertries of the *geometric* null congruences characterized by the space-time metric $g_{\mu\nu}$. Such congruences are special in the sense that the modification of the light cone in a local inertial frame (LIF) can be described as $\left(\eta_{(a)(b)} + \alpha^2\sigma_{(a)(b)}(R)\right) k^{(a)}k^{(b)} = 0$, where $\eta_{(a)(b)}$ is the Minkowski metric, α is the fine structure constant and $\sigma_{(a)(b)}(R)$ depends on the Riemann curvature at the origin of the LIF. In a one-loop approximation, these photons travel with unit velocity. For on-shell photons, some of the Newman–Penrose (NP) scalars play the role of optical scalars. Their propagation along the rays is governed by the known Raychaudhuri equation. We show that these scalars must be effectively modified to describe *physical* null congruences and find that a new set of NP scalars contribute to the definition of these *effective optical scalars*. Next, we discuss the modified Raychaudhuri equation which governs the propagation of these scalars.

2. Vacuum polarized photon propagation in a curved background

The effect of one-loop vacuum polarization on the photon propagation in a fixed curved background space-time is represented by the following effective action derived by Drummond and Hathrell [1]:

$$\Gamma = \int \mathrm{d}x\sqrt{-g}\Big[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{m^2}\Big(aRF_{\mu\nu}F^{\mu\nu} + bR_{\mu\nu}F^{\mu\lambda}F^{\nu}{}_{\lambda}$$

$$+cR_{\mu\lambda\nu\rho}F^{\mu\lambda}F^{\nu\rho}\Big)\Big]. \tag{1}$$

Here, $a = -\frac{1}{144}\frac{\alpha}{\pi}$, $b = \frac{13}{360}\frac{\alpha}{\pi}$ and $c = -\frac{1}{360}\frac{\alpha}{\pi}$ where α is the fine structure constant and m is the electron mass. The notable feature in the above action is the direct coupling of the electromagnetic field to the curvature tensor which in effect violates the strong equivalence principle. It should also be noted that since the photons are treated as test particles the backreaction has been assumed to be negligible. The effective equation of motion can be obtained by varying Γ with respect to A_ν as follows:

$$D_\mu F^{\mu\nu} + \frac{1}{m^2}\left(2bR^{\mu}{}_{\lambda}D_\mu F^{\lambda\nu} + 4cR^{\mu\nu}{}_{\lambda\rho}D_\mu F^{\lambda\rho}\right) = 0. \tag{2}$$

There are some approximations under which this equation of motion was obtained. The first one is the low frequency approximation in the sense that the derivation is only applicable to wavelengths $\lambda > \lambda_c$. By this approximation we ignore terms in the effective action involving higher-order field derivatives. The second is a weak-field approximation for gravity. This implicitly means that the wavelengths $\lambda_c \ll L$ are considered.

Employing the geometric optics approximation, the electromagnetic field can be written in the form $A_\mu = Aa_\mu e^{i\theta}$, where A is the amplitude and a_μ is the polarization vector. If we now identify the wave vector as $k_\mu = \partial_\mu\theta$, the equation governing the components of this vector field for a photon with space-like, normalized polarization vector i.e, $a_\mu a^\mu = -1$, can be read from (2) as

$$k^2 + \frac{2b}{m^2} R_{\mu\lambda} k^\mu k^\lambda - \frac{8c}{m^2} R_{\mu\nu\lambda\rho} k^\mu k^\lambda a^\nu a^\rho = 0. \tag{3}$$

Equation (3) is an effective light cone equation which re-expressed in terms of the Weyl tensor is given by

$$k^2 + \frac{(2b+4c)}{m^2} R_{\mu\lambda} k^\mu k^\lambda - \frac{8c}{m^2} C_{\mu\nu\lambda\rho} k^\mu k^\lambda a^\nu a^\rho = 0. \tag{4}$$

The corresponding momentum of the photon, i.e. p^μ, is the tangent vector to the light ray x^μ i.e, $p^\mu = \frac{\mathrm{d}}{\mathrm{d}s} x^\mu(s)$. Equation (4), being quadratic and homogenous, could be written in the following form:

$$\mathcal{G}^{\mu\nu} k_\mu k_\nu = 0. \tag{5}$$

This shows that, at this level of approximation, QGO can be characterized as a bimetric theory. The physical light cones are determined by the effective metric $\mathcal{G}_{\mu\nu}$, and are distinct from the geomerical null cones which are fixed by the space-time metric $g_{\mu\nu}$ (Note that indices are always raised and lowered using the space-time metric $g_{\mu\nu}$). Some implications of this bimetricity, including gravitational birefringence and superluminal speed of light, have been discussed in [7]. In the present study, we focus on modifications produced by vacuum polarization effects on the propagation of light rays in a curved background.

3. Abreast rays, optical scalars and Raychaudhuri equation

The physical congruence is specified by the integral curves γ of the vector field k^μ parametrized by the parameter u and scaled such that $\nabla_k u = 1$. The connecting vector field q^μ is introduced [8] in order to examine the relation between the neighbouring curves of the congruence. Defined along a particular curve γ, it characterizes the displacement from a point $P \in \gamma$ to a point P' on a neighbouring curve, where P and P' have the same parameter value u. Mathematically, this means that q^μ is being Lie transported along the curve by the vector field k^μ, i.e.

$$\mathcal{L}_k q^\mu = 0, \quad \text{or} \quad \nabla_k q^\mu = \nabla_q k^\mu. \tag{6}$$

Here the notation $\nabla_X = X^\alpha \nabla_\alpha$ for directional covariant derivative is used. Using the NP tetrad formalism and following closely the notations of [8] and [9], as a first step, we choose a null tetrad with l^μ to be the null vector along the photon momentum. Let a^μ and b^μ be space-like transverse vectors (later they will be identified with the polarization vectors). Define the complex null vectors m^μ and $\bar{m}^\mu$ by $m^\mu = \frac{1}{\sqrt{2}}(a^\mu + ib^\mu)$ and $\bar{m}^\mu = \frac{1}{\sqrt{2}}(a^\mu - ib^\mu)$ and add to this set another real null vector n^μ. All the null vectors satisfy the usual NP orthogonality conditions

$$l \cdot m = l \cdot \bar{m} = n \cdot m = n \cdot \bar{m} = 0, \tag{7}$$

and

$$l \cdot l = n \cdot n = m \cdot m = \bar{m} \cdot \bar{m} = 0. \tag{8}$$

The normalization conditions

$$l \cdot n = -m \cdot \bar{m} = 1 \tag{9}$$

are also imposed. We assign a vierbein $e^{\mu}_{(a)}$ to this tetrad [10] as

$$\begin{aligned} &e_{(1)} = e^{(2)} = l, \quad e_{(2)} = e^{(1)} = n, \\ &e_{(3)} = -e^{(4)} = m, \quad e_{(4)} = -e^{(3)} = \bar{m}, \end{aligned} \tag{10}$$

with the frame metric of the form

$$\eta_{(a)(b)} = e^{\mu}_{(a)} e_{\mu(b)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{11}$$

In the tetrad formalism, various quantities called 'Ricci rotation coefficients', $\gamma_{(c)(a)(b)}$, are employed to account for the tetrad covariant differentiation. These quantities can be defined as

$$e_{(a)\mu;\nu} = e^{(c)}_{\mu} \gamma_{(c)(a)(b)} e^{(b)}_{\nu}. \tag{12}$$

In the NP formalism, they are called 'spin coefficients' and are designated by the following special symbols [9]:

$$\begin{aligned} &\kappa = \gamma_{(3)(1)(1)}; \quad \rho = \gamma_{(3)(1)(4)}; \quad \epsilon = \frac{1}{2}\left(\gamma_{(2)(1)(1)} + \gamma_{(3)(4)(1)}\right); \\ &\sigma = \gamma_{(3)(1)(3)}; \quad \mu = \gamma_{(2)(4)(3)}; \quad \gamma = \frac{1}{2}\left(\gamma_{(2)(1)(2)} + \gamma_{(3)(4)(2)}\right); \\ &\lambda = \gamma_{(2)(4)(4)}; \quad \tau = \gamma_{(3)(1)(2)}; \quad \alpha = \frac{1}{2}\left(\gamma_{(2)(1)(4)} + \gamma_{(3)(4)(4)}\right); \\ &\nu = \gamma_{(2)(4)(2)}; \quad \pi = \gamma_{(2)(4)(1)}; \quad \beta = \frac{1}{2}\left(\gamma_{(2)(1)(3)} + \gamma_{(3)(4)(3)}\right). \end{aligned} \tag{13}$$

A general rule for obtaining the complex conjugate of any quantity is to replace the index (3), wherever it occurs, by the index (4) and vice versa.

The NP formalism is most advantageous when the null tetrads introduced can be completely tied to the geometry of the problem. Otherwise, some freedom remains in the choice of the basis as a result of subjecting it to the Lorentz transformation. This has the effect that many of the quantities involved in the calculation are of the same nature as gauge quantities whose values transform in certain ways as the basis frame is varied in accordance with the remaining freedom. In QGO, one null direction is singled out and that is the unperturbed photon momentum which is fixed in the direction of the null vector l^{μ}. Therefore the allowed transformations would be those which leave the l^{μ} direction unchanged and preserve the underlying orthogonality and normalization conditions. These are classified as follows:

$$\begin{aligned}
&\text{class I}: \quad l \longmapsto l; \quad m \longmapsto m + al; \quad \bar{m} \longmapsto \bar{m} + a^* l; \quad n \longmapsto n + a^* m + a\bar{m} + aa^* l; \\
&\text{class II}: \quad l \longmapsto A^{-1} l; \quad n \longmapsto An; \quad m \longmapsto \mathrm{e}^{i\theta} m; \quad \bar{m} \longmapsto \mathrm{e}^{-i\theta} \bar{m},
\end{aligned} \tag{14}$$

where a is a complex function and A and θ are two real functions on the manifold. We now proceed to write down eq. (6) in the above tetrad basis. Before doing so, let us employ all the gauge freedom that we have to fix the tetrad basis. As the tetrad frame is transformed, different tetrad components will be subject to changes. Starting with the l vector as the velocity vector along the geometric congruence, we have

$$\nabla_l l^\mu = (\epsilon + \epsilon^*)\, l^\mu - \kappa \bar{m}^\mu - \kappa^* m^\mu \quad \propto l^\mu \quad \Longrightarrow \kappa = 0. \tag{15}$$

Moreover, if they are affinely parametrized,

$$\nabla_l l^\mu = 0 \quad \Longrightarrow \kappa = \epsilon = 0. \tag{16}$$

If $\kappa = 0$, the latter requirement can be met by a class II rotation which will not affect the direction of l nor of the initial vanishing of κ. This is because we have

$$\begin{aligned}
&\gamma_{(c)(a)(b)} = e^k_{(c)} e_{(a)k;i} e^i_{(b)}, \\
&\kappa = \gamma_{(3)(1)(1)} = m^k l_{k;i} l^i \longmapsto \mathrm{e}^{i\theta} m^k (A^{-1} l)^i (A^{-1} l_k)_{;i} \\
&\qquad = A^{-2} \mathrm{e}^{i\theta} \kappa + \mathrm{e}^{i\theta} (m.l) A^{-1} (\nabla_l A^{-1}) = A^{-2} \mathrm{e}^{i\theta} \kappa,
\end{aligned}$$

$$\begin{aligned}
\epsilon &= \frac{1}{2}\left[\gamma_{(2)(1)(1)} + \gamma_{(3)(4)(1)}\right] = \frac{1}{2}\left[n^k l_{k;i} l^i + m^k \bar{m}_{k;i} l^i\right] \longmapsto \\
&\frac{1}{2}\left[A n^k \left(A^{-1} l_k\right)_{;i} A^{-1} l^i + \mathrm{e}^{i\theta} m^k \left(\mathrm{e}^{-i\theta} \bar{m}\right)_{k;i} A^{-1} l^i\right] \\
&= \frac{1}{2}\left[n^k l_{k;i} l^i A^{-1} + (n \cdot l)\, A^{-1}_{;i} l^i + m^k \bar{m}_{k;i} l^i A^{-1} - i\theta_{;i}\, (m \cdot \bar{m})\, A^{-1} l^i\right] \\
&= \frac{1}{2} A^{-1} \epsilon - \frac{1}{2} A^{-2} \nabla_l A + \frac{1}{2} i A^{-1} \nabla_l \theta.
\end{aligned} \tag{17}$$

By a suitable rotation of class I, it could be arranged so that $\pi = \gamma_{(2)(4)(1)}$ vanishes. Thus

$$\begin{aligned}
\pi = \gamma_{(2)(4)(1)} = n^k \bar{m}_{k;i} l^i &\longmapsto \left(n^k + a^* m^k + a\bar{m}^k + aa^* l^k\right) \left(\bar{m} + a^* l\right)_{k;i} l^i \\
&= \pi + 2a^* \epsilon + \left(a^*\right)^2 \kappa + \nabla_l a^*.
\end{aligned} \tag{18}$$

In a similar fashion, we find that the spin coefficients κ and ϵ transform as $\kappa \longmapsto \kappa$ and $\epsilon \longmapsto \epsilon + a^* \kappa$ in this class. So the gauge, that we are working in, is the one in which $\kappa = \epsilon = \pi = 0$. After such a rotation, the newly oriented vectors n, m and $\bar{m}$ will remain unchanged as they undergo parallel transport along l. This could be invoked through the following relations:

$$\begin{aligned}
&\nabla_l m^\nu = \pi^* l^\nu - \kappa n^\nu + (\epsilon - \epsilon^*)\, m^\nu, \\
&\nabla_l n^\nu = \pi^* \bar{m}^\nu + \pi m^\nu - (\epsilon + \epsilon^*)\, n^\nu,
\end{aligned} \tag{19}$$

with the right-hand sides vanishing in the above gauge. Now we have exploited all the gauge freedom that we had and are ready to write eq. (6) in the above gauge. We only consider states which propagate with a well-defined polarization. For two transverse polarizations expressed in terms of the vectors m^μ and $\bar{m}^\mu$, representing the left and the right handed circular polarizations, we have

$$\nabla_k q^\nu = \nabla_q \left[l^\nu - \frac{1}{m^2}(b+2c) R^{\lambda\nu} l_\lambda \right. \\ \left. \mp \frac{2c}{m^2} C^\nu{}_{\lambda\sigma\kappa} l^\sigma (m^\lambda \pm \bar{m}^\lambda)(m^\kappa \pm \bar{m}^\kappa) \right]. \tag{20}$$

On multiplying by $e^{(c)}_\nu$, we get

$$\nabla_k q^{(c)} = \left(\eta^{(c)(a)} + \frac{1}{m^2} A^{(c)(a)} \right) q^{(b)} \gamma_{(a)(1)(b)} + q_{(a)} \gamma^{(a)(c)(b)} k_{(b)}, \tag{21}$$

where

$$A^{(c)(a)} := -(b+2c) R^{(c)(a)} \mp (2c) M^{(c)(a)}, \\ M^{(c)(a)} := C^{(c)(3)(a)(3)} + C^{(c)(4)(a)(4)} \pm (C^{(c)(3)(a)(4)} + C^{(c)(4)(a)(3)}). \tag{22}$$

The specifically *gravitational* birefringence shows up here in $M^{(a)(c)}$ (we note that $A^{(c)(a)} = A^{(a)(c)}$). Equation (21) is a system of coupled differential equations describing the propagation of the tetrad components of the connecting vector along the velocity vector k^μ. The second term in (21) appears as a consequence of the propagation of the null tetrad along the *physical* ray (with the tangent vector k^μ) i.e. $q^\nu \nabla_k e^{(c)}_\nu$. We introduce different tetrad components of q^μ in the following way:

$$q^\mu = g l^\mu + \xi \bar{m}^\mu + \bar{\xi} m^\mu + h n^\mu. \tag{23}$$

Thus

$$q^{(1)} = q_{(2)} = g; \quad q^{(2)} = q_{(1)} = h; \quad q^{(3)} = -q_{(4)} = \bar{\xi}; \quad q^{(4)} = -q_{(3)} = \xi. \tag{24}$$

Explicitly for $c = 4$, eq. (21) gives

$$\nabla_k q^{(4)} = \left(\eta^{(4)(3)} + \frac{1}{m^2} A^{(4)(3)} \right) \\ [q^{(2)} \gamma_{(3)(1)(2)} + q^{(3)} \gamma_{(3)(1)(3)} + q^{(4)} \gamma_{(3)(1)(4)}] \\ + \frac{1}{m^2} A^{(4)(4)} [q^{(2)} \gamma_{(4)(1)(2)} + q^{(3)} \gamma_{(4)(1)(3)} + q^{(4)} \gamma_{(4)(1)(4)}] \\ + \frac{1}{m^2} A^{(4)(2)} q^{(b)} \gamma_{(2)(1)(b)} - \frac{1}{m^2} A^{(2)(b)} q^{(a)} \gamma_{(a)(3)(b)} \tag{25}$$

in which the last term is again the contribution due to null tetrad propagation along the physical ray discussed below eq. (22). In order to identify the geometrical and physical content of this equation, a 'covariant approach' in which any complicated gauge behaviour is completely avoided must be chosen. For example, in eq. (25), the term $A^{(4)(4)}q^{(2)}\gamma_{(4)(1)(2)}$, having a complicated gauge behaviour, transforms into

$$[A^{(4)(4)} - 2aA^{(4)(2)} + a^2 A^{(2)(2)}]q^{(2)}[\gamma_{(4)(1)(2)} + a^*\gamma_{(4)(1)(3)} + a\gamma_{(4)(1)(4)}] \tag{26}$$

under a class I transformation. To keep the formalism covariant, we remove the gauge-dependent terms by applying the extra condition that the neighbouring pair of rays obey the relation

$$q.k = 0. \tag{27}$$

We call such rays *abreast* [8]. This is a *primary* constraint and it means that

$$\begin{aligned} &O\left(\alpha^0\right): \quad q.l = 0 \Rightarrow h = 0; \\ &O\left(\alpha^1\right): \quad A_{(1)(a\neq 2)} = 0 \Rightarrow \Phi_{00} = \Phi_{01} = \Phi_{10} = \Psi_0 = \Psi_1 = 0. \end{aligned} \tag{28}$$

The parametrization independence of the abreast ray constraint can be seen by making the substitution $q^\mu \mapsto q^\mu + \Lambda k^\mu$. We also emphasize that, provided the abreast constraint (28) is satisfied, the tangent vector k^μ is given by

$$\begin{aligned} &k_{(2)} = 1 - \frac{1}{m^2}A_{(1)(2)}, \qquad k_{(a\neq 2)} = 0, \\ &k^2 = O\left(\alpha^2\right), \qquad \nabla_k k^\mu = O\left(\alpha^2\right). \end{aligned} \tag{29}$$

The abreast rays are *special* physical rays whose effective light cone condition, in a local inertial frame, is given by

$$\left(\eta_{(a)(b)} + \alpha^2\sigma_{(a)(b)}\left(R\right)\right)k^{(a)}k^{(b)} = 0. \tag{30}$$

This should be compared with the general light cone modification (3) induced by the effective action (1) which is of order α.

3.1 *Effective optical scalars*

To study the evolution of the cross sections of (physical) bundles of rays, through their so-called *optical scalars* [8], we start with the covariant deformation of ξ:

$$\nabla_k\xi = -\left(\bar{\xi}\sigma + \xi\rho\right) + \frac{1}{m^2}A^{(4)(3)}\left(\bar{\xi}\sigma + \xi\rho\right) + \frac{1}{m^2}A^{(4)(4)}\left(\bar{\xi}\rho^* + \xi\sigma^*\right). \tag{31}$$

The transverse distances from a specified ray γ are (frame) observer-independent. Therefore, for abreast rays, optical scalars bear explicit physical interpretations. To simplify the case we consider propagation along k of the area of a small triangle formed by the points o, ξ_1 and ξ_2 (see figure 1). Thus we have

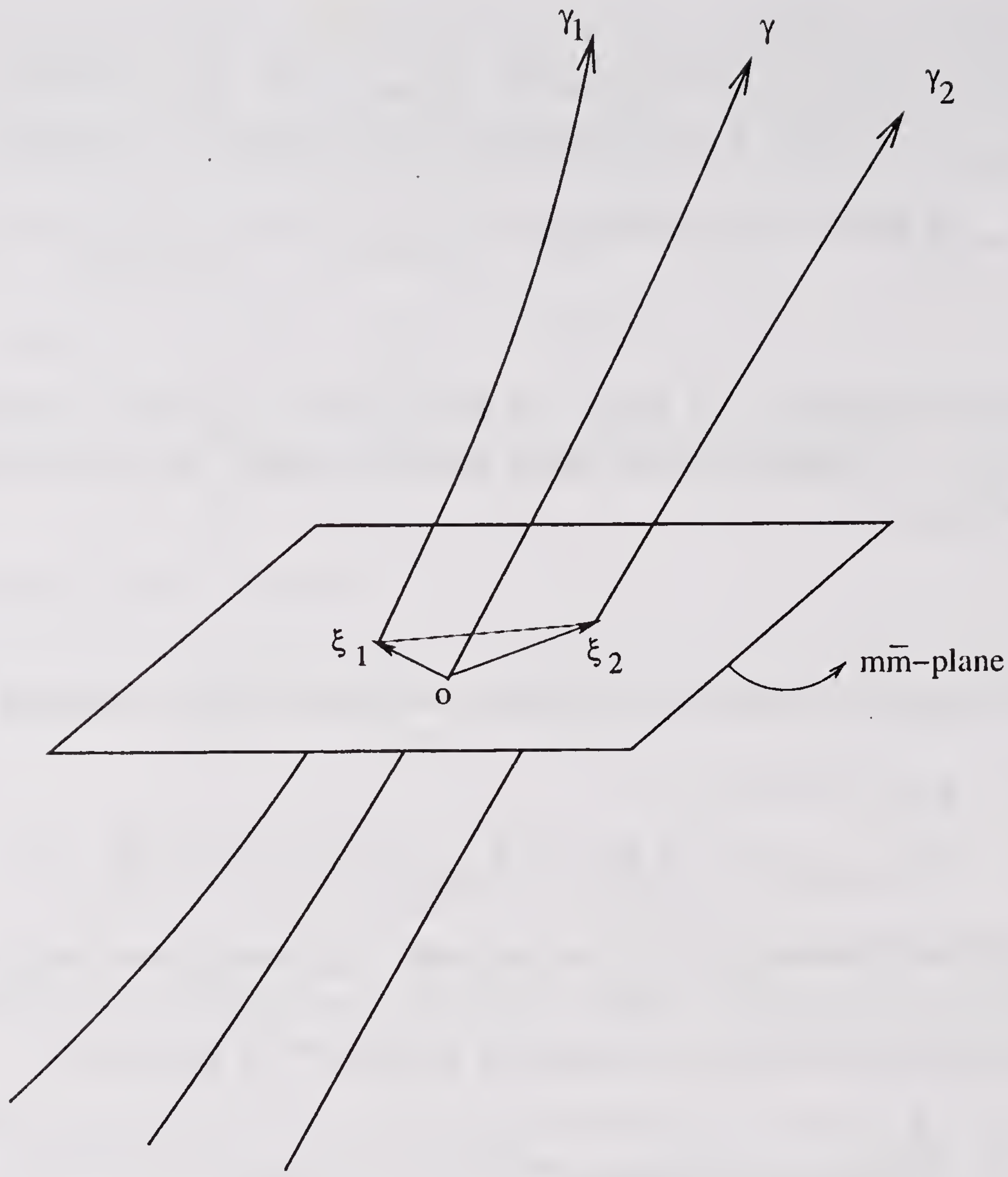


Figure 1. Three abreast rays of a congruence and the small triangle formed by two connecting vectors ξ_1 and ξ_2, connecting the ray γ to the rays γ_1 and γ_2 in the $m\bar{m}$-plane.

$$\begin{aligned}\nabla_k \delta A &= \nabla_k \left[\frac{i}{2}\left(\xi_1\bar{\xi}_2 - \xi_2\bar{\xi}_1\right)\right] \\ &= -\left[\rho - \frac{1}{m^2}\left(A^{(3)(4)}\rho + A^{(3)(3)}\sigma\right) + c.c\right]\delta A.\end{aligned} \tag{32}$$

The scalar quantity

$$-\mathrm{Re}\left[\rho - \frac{1}{m^2}\left(A^{(3)(4)}\rho + A^{(3)(3)}\sigma\right)\right] \tag{33}$$

is called the expansion parameter θ_{eff}. Its role as a measure of the pattern convergence (or divergence) is clear from equation (32). This equation shows that in an effective theory, unlike in the classical case, the local effect of σ can also change the area. Another useful parameter, the 'luminosity parameter' L, is defined along a bundle of rays as $L^2 \propto \delta A$ and is related to θ_{eff} as follows (using equation (32)):

$$\nabla_k L = \theta_{\text{eff}} L. \tag{34}$$

In other words, the expansion parameter is the logarithmic derivative of the luminosity parameter. An alternative formula for θ_{eff} is given by $\frac{1}{2}k^{\mu}{}_{;\mu}$.

Setting $\theta_{\rm eff}$ and the coefficient of $\bar{\xi}$ equal to zero in eq. (31), we obtain

$$\nabla_k \xi = -i\mathrm{Im}\left[\rho - \frac{1}{m^2}\left(A^{(3)(4)}\rho + A^{(3)(3)}\sigma\right)\right]\xi. \tag{35}$$

Thus it can also be claimed that $\mathrm{Im}\left[\rho - \frac{1}{m^2}\left(A^{(3)(4)}\rho + A^{(3)(3)}\sigma\right)\right]$ measures the twist in the bundle's cross-section and can be called the effective twist, $\omega_{\rm eff}$. The fact that this combination of scalars is a measure of the twist is consistent with the requirement that it measures the failure of k_μ to be hypersurface orthogonal. In the same spirit, if $\theta_{\rm eff} = \omega_{\rm eff} = 0$, the remaining part of eq. (31), i.e

$$\sigma + \frac{1}{m^2}\left(A^{(4)(3)}\sigma + A^{(4)(4)}\rho^*\right), \tag{36}$$

can be interpreted as the effective shear $\sigma_{\rm eff}$. This is a measure of the distortion in the shape of the bundle's cross-section so that a circular cross-section transforms into an elliptic one.

3.2 *Effective Raychaudhuri equation*

In order to study the variation of the above effective optical scalars along the physical ray, one can examine the second derivative of the connecting vector q^μ, propagating along the wave vector k_μ, through the operation of ∇_k on eq. (1):

$$\begin{aligned}\nabla_k\nabla_k q^\mu &= \nabla_k\left(\nabla_q k^\mu\right) = \nabla_q\nabla_k k^\mu + R_{\lambda\nu\sigma}{}^{\mu}k^\lambda q^\nu k^\sigma \\ &= R_{\lambda\nu\sigma}{}^{\mu}k^\lambda q^\nu k^\sigma + O\left(\alpha^2\right).\end{aligned} \tag{37}$$

On multiplying by $e^{(c)}_\mu$ and taking the components corresponding to the tetrads subjected to parallel transport, we get

$$\nabla_k\nabla_k q^{(c)} = R_{(1)(3)(1)}{}^{(c)}\bar{\xi} + R_{(1)(4)(1)}{}^{(c)}\xi + O\left(\alpha^2\right). \tag{38}$$

We have neglected here the terms, including the product of curvature components, which would be suppressed by $O(\frac{R}{m^2})$. From eq. (31), we obtain the useful (2×2) matrix form as

$$\nabla_k \mathbf{Z} = \mathbf{P}\mathbf{Z}, \tag{39}$$

where

$$\mathbf{Z} = \begin{pmatrix}\xi \\ \bar{\xi}\end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix}\theta_{\rm eff} - i\omega_{\rm eff} & \sigma_{\rm eff} \\ \sigma^*_{\rm eff} & \theta^*_{\rm eff} + i\omega^*_{\rm eff}\end{pmatrix}. \tag{40}$$

For abreast rays, eq. (39) reduces to

$$\nabla_k \mathbf{P} = -\mathbf{P}^2. \tag{41}$$

Equation (41), written out in full, leads to the directional propagation of the effective optical scalars as

$$\nabla_k \sigma_{\text{eff}} = -2\theta_{\text{eff}} \sigma_{\text{eff}}, \tag{42}$$
$$\nabla_k \omega_{\text{eff}} = -2\theta_{\text{eff}} \omega_{\text{eff}}, \tag{43}$$
$$\nabla_k \theta_{\text{eff}} = \omega_{\text{eff}}^2 - \theta_{\text{eff}}^2 - |\sigma_{\text{eff}}|^2, \tag{44}$$

the last of which is the effective Raychaudhuri equation whose physically important consequences are discussed below.

4. Discussion

In the derivation of the classical version of the Raychadhuri equation [11,12], the variation of the tangent vector $k_{\mu;\nu}$ is projected [13] on the plane spanned by m^μ and $\bar{m}^\mu$. Then the antisymmetric and symmetric parts of its trace are designated as expansion, vorticity and shear respectively. In the present modified version, we chose, through eq. (6), an equivalent method in which the variations of the connecting vector q^μ instead of the tangent vector have been considered and then the corresponding *effective* optical scalars have been defined. The invariants of the theory are easily seen in this way. For a luminosity parameter L, we see from eq. (43) that $\nabla_k \left(L^2 \omega_{\text{eff}}\right) = 0$. In other words, $q \cdot k$ and $L^2 \omega_{\text{eff}}$ are the constant geometric quantities along the congruence. Also, for two neighbouring rays of γ, whose connecting vectors q^μ and $\tilde{q}^\mu$ independently satisfy eq. (37), a symplectic invariant is assigned as follows:

$$q^\mu \nabla_k \tilde{q}_\mu - \tilde{q}^\mu \nabla_k q_\mu. \tag{45}$$

The constancy of expression (45) along γ is a consequence of the interchange symmetry of the curvature tensor and eq. (37):

$$\begin{aligned}\nabla_k \left(q^\mu \nabla_k \tilde{q}_\mu - \tilde{q}^\mu \nabla_k q_\mu\right) &= \left(\nabla_k q^\mu\right)\left(\nabla_k \tilde{q}_\mu\right) + q^\mu \nabla_k \nabla_k \tilde{q}_\mu \\ &\quad - \left(\nabla_k \tilde{q}^\mu\right)\left(\nabla_k q_\mu\right) - \tilde{q}^\mu \nabla_k \nabla_k q_\mu \\ &\approx q^\mu R_{\lambda\nu\sigma\mu} k^\lambda \tilde{q}^\nu k^\sigma - \tilde{q}^\mu R_{\lambda\nu\sigma\mu} k^\lambda q^\nu k^\sigma = 0. \end{aligned} \tag{46}$$

Considering only abreast rays, expression (45) can be written in the form

$$\begin{aligned}&Z^\top(\nabla_k \tilde{Z}) - (\nabla_k Z^\top)\tilde{Z} = Z^\top(P\tilde{Z}) - (PZ)^\top \tilde{Z} = Z^\top P \tilde{Z} - Z^\top P^\top \tilde{Z} \\ &(\bar{\xi}\xi)\begin{pmatrix} B - B^* & 0 \\ 0 & B^* - B \end{pmatrix}\begin{pmatrix} \tilde{\xi} \\ \bar{\tilde{\xi}} \end{pmatrix} = (\bar{\xi}\tilde{\xi} - \xi\bar{\tilde{\xi}})(B - B^*) \propto (\delta A)\omega_{\text{eff}} \propto L^2 \omega_{\text{eff}}. \end{aligned} \tag{47}$$

The above is, in effect, a restatement of $\nabla_k \left(L^2 \omega_{\text{eff}}\right) = 0$ in which $B = \theta_{\text{eff}} - i\omega_{\text{eff}}$ and $\top$ denotes Hermitian transpose.

These invariants reduce to the classical ones in the limit of zero perturbation and therefore there remains no anomaly in QGO. This is the case since we have applied the symmetries present at $O\left(\alpha^0\right)$ as a set of constraints and exploited them to find the covariant quantum corrections. Since the effective action, that we started with, has been derived in a gauge invariant manner [1], the results obtained here are also gauge invaraiant.

The classical limit of eq. (44) can be compared with that corresponding to the standard (purely general relativistic) Raychaudhuri equation, namely

$$\nabla_k \theta = \omega^2 - \theta^2 - |\sigma|^2 - \frac{1}{2} R_{(1)(1)}. \tag{48}$$

The last RHS term above is negative for all known forms of ponderable matter. For an initially contracting congruence, in the absence of rotation and shear (ρ real and $\sigma = 0$), it leads to the divergence of the expansion parameter ($\theta \to -\infty$). This is a necessary though not a sufficient condition for proving the singularity theorems [14]. In our case, a real ρ and vanishing σ are equivalent to $\omega_{\text{eff}} = \sigma_{\text{eff}} = 0$. Using (44), we see that during the propagation along the ray, it differs from zero only by $O\left(\alpha^2\right)$ [15],

$$\nabla_k \theta = -\theta^2 + \frac{1}{m^2} A_{(3)(4)} \theta^2 + O\left(\alpha^2\right). \tag{49}$$

On the other hand, the last RHS term in eq. (48) does not appear in our calculations due to the abreast ray conditions. Even if $R_{(1)(1)} = 0$, however, the correction terms calculated in this paper will contribute and their sign must be taken into account.

Acknowledgments

The authors would like to thank University of Tehran for supporting this project under the grants provided by the research council.

References

[1] I T Drummond and S J Hathrell, *Phys. Rev.* **D22**, 343 (1980)
[2] L D Landau and E M Lifshitz, *The classical theory of fields* (Pergamon Press, 1975)
[3] G M Shore, *Nucl. Phys.* **B460**, 379 (1996)
[4] R D Daniels and G M Shore, *Nucl. Phys.* **B425**, 634 (1994)
[5] R D Daniels and G M Shore, *Phys. Lett.* **B367**, 75 (1996)
[6] G M Shore, *Nucl. Phys.* **B605**, 455 (2001)
[7] G M Shore, *Contemp. Phys.* **44**, 503 (2003)
[8] R Penrose and W Rindler, *Spinors and space-time, Volume 2: Spinor and twistor methods in space-time geometry* (Cambridge University Press, 1993)
[9] S Chandrasekhar, *The mathematical theory of black holes* (Oxford University Press, 1993)
[10] Tetrad indices are enclosed in parantheses to be distinguished from tensor indices
[11] A K Raychaudhuri, *Phys. Rev.* **90**, 1123 (1955)
[12] R M Wald, *General relativity* (University of Chicago Press, 1984)
[13] The projection operator is $P_{\mu\nu} = m_\nu \bar{m}_\mu + \bar{m}_\nu m_\mu = g_{\mu\nu} + l_\mu n_\nu + n_\mu l_\nu$
[14] S W Hawking and G F R Ellis, *The large scale structure of space-time* (Cambridge University Press, 1973)
[15] For a more detailed calculation refer to [16]
[16] N Ahmadi and M Nouri-Zonoz, *Phys. Rev.* **D74**, 044034 (2006)