Black hole and Singularity Theorems

Sudipta Sarkar

Indian Institute of Technology, Gandhinagar, Gujarat 382355, India.
Email: sudiptas@iitgn.ac.in

Sudipta Sarkar is a faculty at the physics discipline of Indian Institute of Technology, Gandhinagar. His research interests are Gravitation, Black hole physics and Quantum field theory in curved space-time.

Abstract

Prof. Roger Penrose was awarded one-half of the Nobel prize in physics 2020 "for the discovery that black hole formation is a robust prediction of the general theory of relativity." Penrose's landmark work on singularity theorems and black holes are the foundation of our understanding of these exotic objects. In this article, I describe the historical background of the singularity theorems and their impact on contemporary gravitational physics research.

"*The black holes of nature are the most perfect macroscopic objects there are in the universe: the only elements in their construction are our concepts of space and time.*"

–Subrahmanyan Chandrasekhar

This wonderful quotation by Subrahmanyan Chandrasekhar describes the most intriguing feature of the black holes; they are one of the most non-trivial manifestations of the causal structure of space-time, determined by the Einstein's theory of general relativity (GR). Our quest to understand these objects begins as early as 1916 when Karl Schwarzschild solved the vacuum Einstein's equations of general relativity and obtained the first spherically symmetric solution - describing the space-time outside a spherical object like our Sun. This is the relativistic analogue of the gravitational field around a point particle and the Schwarzschild solution is usually written in the familiar spherically symmetric coordinate system (t, r, θ, ϕ) in the form of the space-time metric [1],

$$ds^2 = -\left(1 - \frac{2GM}{c^2 r}\right) c^2 dt^2 + \frac{dr^2}{\left(1 - \frac{2GM}{c^2 r}\right)} + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right) \quad (1)$$

This metric is the description of the structure of the space-time around a spherical body of mass M. Notice two crucial points about this metric, set mass M to zero, and you are left with the usual flat Minkowski metric in spherical polar coordinates. Also, setting $r \longrightarrow \infty$, the metric again becomes flat. These features are indeed expected: the first shows that in the absence of any mass/energy, the universe is just an ordinary $3+1$ dimensional Minkowski space-time and the second reflects the fact that at a large distance from the mass, the metric again becomes flat. Apart from these two expected features, we also notice that something interesting must be happening at two locations, $r = 2GM/c^2$ and also at $r = 0$. At both these locations, the metric components diverge. Now, in general, the divergence of a metric component may merely reflect the fact that we are choosing a bad coordinate system, and it is only a coordinate singularity. Then, we can 'cure' the divergence by adopting a different coordinate system. A simpler example is the use of the usual spherical polar coordinate system on a sphere, where metric component diverges at the north pole. But a different coordinate system would eliminate this apparent discontinuity. After all, there is hardly anything discontinuous at the north pole. Also, general relativity allows us to describe physics from any coordinate system, assuring that the choice of the coordinates can not change the physics. It turns out that the singularity at $r = 2M$ (we have now set G = c = 1) is akin to this type of discontinuity. It can be removed by adopting a better coordinate system. The singularity at $r = 0$ is rather different. If we calculate a space-time scalar like $R_{abcd}R^{abcd}$, it is finite at $r = 2M$ but diverges at $r = 0$. This shows that $r = 0$ in actually a true singularity of the underlying spacetime manifold, where the curvature scalers diverge. Any cleaver choice of the coordinates can not remove a true singularity. it is an unavoidable feature of the Schwarzschild solution.

What about the surface at $r = 2M$? This surface also has an interesting feature. If we adopt a better coordinate system and study the behaviour of the particle and light trajectories, we can immediately see that inside $r = 2M$; all future-directed paths are always in the direction of the decreasing r [1]. The surface, while being locally regular, globally acts like a surface of no return - once a particle crosses it from outside, it can never return. This surface is known as the Schwarzschild radius or the event horizon. The event horizon is a causal boundary, which hides its interior from the rest of the universe.

Now, none of these may matter in the actual physical situation. Consider the Sun, an approximate spherical object of mass $M \approx 10^{30}$ kg which gives a Schwarzschild radius $r_s \approx 3$ km. So the location of the Schwarzschild radius is well inside the Sun where vacuum Schwarzschild solution is no longer applicable. Nevertheless, there are situations in our universe where there exists a massive object compressed beyond its Schwarzschild radius. The first suggestion for the existence by such objects came from the pioneering study of the stellar structure by Subrahmanyan Chandrasekhar and subsequent analysis of the gravitational collapse by Oppenheimer, Snyder, and Datt [2, 3, 4]. All these works indicated that at the end of the lifetime of a massive star, where the nuclear fuel for the fusion runs out, gravity takes over leading to a catastrophic collapse. Then the question arises, whether the end stage of the gravitational collapse is always a black hole, endowed with two essential features of the Schwarzschild geometry, the event horizon and the central singularity? This is a crucial question - the existence of the space-time singularity indicates something radical. Consider an object, possibly a rocket ship falling into the black hole, though it will never come back, we can still in principle describe the path using the laws of general relativity. But, near the singularity, the gravitational tidal forces starts increasing and eventually the rocket ship encounters infinite tidal force. The laws of general relativity can not describe the final fate of the vessel. The gravitational singularity implies the end of the space-time and a fundamental incompleteness of the classical general relativity. So, it is vital to understand if this is a generic feature of gravitational collapse, or merely the characteristics of the assumption of spherical symmetry. After all, a physical gravitational collapse is never spherically symmetric.

So far, what we have described is the situation of theoretical gravitational physics till the 1960s. By then, the properties of the Schwarzschild solution was well understood, particularly after the important discovery of a well defined coordinate system, known as the Kruskal-Szekeres coordinates [5, 6]. Also, the study of the gravitational collapse indicated that the formation of the singularity might be unavoidable. But, there was no general proof. It was not clear what mathematical tools are required for such a proof. Einstein's equations are a set of hyperbolic, non-linear, partial differential equations. It is challenging to find any generic property of the solutions. The study of the few exact available solutions was hardly helpful. These solutions are derived assuming high degree of symmetry.

At the same time, the astronomers start observing 'quasi-stellar object' or the quasars in the sky. These quasars radiate enormous power; the most powerful quasars have luminosities thousands of times greater than a galaxy such as the Milky Way. The measurement of a high red-shift supports an interpretation of a very distant object with extraordinarily high luminosity and power output, far beyond any object seen so far. The existence of these quasars implied energies that were far in excess of any known energy conversion processes, including nuclear fusion [7]. The only possible explanation was that the energy source is the radiation of matter from an accretion disc falling onto a supermassive black hole, converting the gravitational energy into heat and light. So, the question of the singularity was not only a topic of academic and mathematical interest but also relevant for astrophysical objects in our universe.

This is the historical context of Roger Penrose's first paper on Singularity Theorems, published in Physical Review Letters in 1965 [8]. The title of the paper is 'Gravitational collapse and space-time singularities'. The motivation of the article was clear in the introduction which says,

"*The question has been raised as to whether this singularity is, in fact, simply a property of the high symmetry assumed. The matter collapses radially inwards to the single point at the center, so that a resulting space-time catastrophe there is perhaps not surprising. Could not the presence of perturbations which destroy the spherical symmetry alter the situation drastically?*."

The paper provided a definitive answer to this question as,

"*It will be shown that, after a certain critical condition has been fulfilled, deviations from spherical symmetry cannot prevent space-time singularities from arising.*"

This work signals a remarkable development in the conceptual understanding of general relativity, probably as important as the theory itself. To fully comprehend the scope of this result, we first start with the definition of the gravitational singularity used in the theorem. Intuitively, one expects that divergences of any geometrical quantity would be a characteristic of space-time singularities. However, there are several problems with this definition. Such curvature singularities do not belong to the spacetime manifold which is by definition constituted of only regular points. So, a different criterion is used, which is mathematically better suited and also physically more relevant. The trick is to use the notion of 'incomplete' curves. A geodesic curve, in particular a time-like geodesic curve, represents the world line of a physical particle in the space-time. A null curve is the path of light rays. So, suppose there are curves, time-like or null, which cannot be continued regularly within the space-time even at finite values of their canonical parameter. In that case, we call this a sufficient condition for the space-time to be singular. This 'geodesic incompleteness' is the fundamental criterion for establishing the existence of a singularity [9].

Next, what are those 'critical conditions' which led to the existence of a space-time singularity? The most important one is the idea of a *closed trapped surface*. In the absence of gravity, the causal structure of the space-time is expressed using the light cone. At an instant of time, consider a two dimensional closed surface in flat space-time. Such a surface has two light-like directions normal to the surface at each point. Thus we can distinguish two future-directed families of null geodesics emerging from the surface. Think of the surface as a sphere, then one family of the light ray is moving outwards and the other directed inwards. Therefore, after an infinitesimal time, we can think of two new spheres, the outgoing one with a slightly larger area and the ingoing one with a smaller area than the original surface. In this case, the original surface is called an untrapped surface.

The situation is presented in the Figure 1, where S is the closed surface, k^a is the outward null normal, and n^a is the inward normal. The area is increasing in the outward direction and decreasing in the inward one. Now, what Penrose

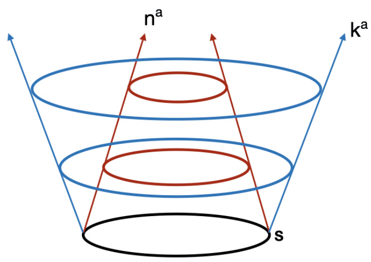

Figure 1: Untrapped surface.

proposed that in the presence of a strong gravitational field, the situation can be radically different. The gravity affects the causal structure of space-time, bends the light rays and leads to a future trapped surface, for which area decreases in both outward and inward null directions! The situation is depicted in the Figure 2. It is possible to show that such sur-

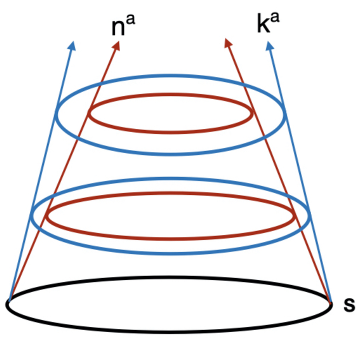

Figure 2: Trapped surface.

faces do exist inside the event horizon of the Schwarzschild solution [1]. But the general notion of a trapped surface is independent of the symmetry of the space-time manifold. This is a mathematically rigorous concept to represent a finite region of space-time which is causally disconnected from the rest of the universe.

After the introduction of the trapped surfaces, we can now state the Penrose singularity theorem as described in his 1965 paper [8]:

Penrose's singularity theorem: *If the space-time contains a non-compact Cauchy hyper-surface and a closed future-trapped surface, and if null energy condition holds, then there are future incomplete null geodesics.*

To understand this statement, let us first focus on the null energy condition. This is a relativistic generalization of the concept of the positivity of energy. The condition says that for every future-pointing null vector field k^a, matter stress-energy tensor T_{ab} satisfies the condition $T_{ab}k^a k^b \geq 0$. For a perfect fluid of energy ρ and pressure P, the null energy condition implies $\rho + P \geq 0$. We expect every reasonable matter to obey this condition. Naively speaking, the use of this energy condition is to ensure that gravity always remains an attractive force.

The existence of a non-compact Cauchy hyper-surface is a technical condition to enforce causality. A Cauchy hyper-surface can be understood as an "instant of time" that provides good initial value conditions for the entire space-time [9]. The existence of such a surface ensures that there is no causality violating *closed time-like curves*. There is no guarantee that such a condition will always be valid. Nevertheless, in a reasonable physical situation, we expect all these conditions to hold. Then the theorem guarantees the existence of a future incomplete null geodesics. This shows that the future of evolution must be singular.

It is essential to realize the power and the scope of this theorem. There is no assumption of any symmetry. The only requirements are some technical conditions related to positivity of energy and causality. Once these conditions are satisfied, and trapped surfaces form under gravitational collapse, the future singularity is inevitable. The spac-etime can not be extended beyond this singularity, and this signals a complete failure of the classical general relativity. We should also mention that the proof of this theorem uses a fundamental equation derived by Indian relativist Amal Kumar Raychaudhuri [10]. The Raychaudhuri equation describes the evolution of the rate of change of the area of a closed two surface. This equation is essential to establish the geodesic incompleteness and the existence of the singularity [1].

The first singularity theorem by Penrose was a landmark result, leading to further generalizations. Immediately after the publication of the Penrose's work, Hawking realized that a past version of the closed trapped surfaces must be

[1] Penrose did not refer to Raychaudhuri in his 1965 work, but later dualy recognized the importance of Raychaudhuri's work in the subsequent papers.

present in any expanding universe close to being spatially homogeneous and isotropic [11]. This started a series of works by Hawking and Ellis which show the inevitability of an initial singularity in our past if Einstein's equations hold and some reasonable conditions are met [12, 13]. The cosmological version of the singularity theorem established that the universe must have evolved from a singular past - commonly knows as the Big Bang! The theory of general relativity breaks down at the beginning of the universe and at the centre of a black hole. Later, Hawking & Penrose together provided further general versions of these theorems where weaker assumptions are used [14, 15]. The Hawking-Penrose result conclusively proved that the end state of a generic gravitational collapse must be singular, beyond the scope of classical general relativity.

There is still an important question related to the formation of the trapped surfaces. It is not a priori evident that closed trapped surfaces can always form in a gravitational collapse starting from initial conditions in which no such surfaces are present. This is ultimately settled by Christodoulou when he solved the long-standing problem of evolutionary formation of trapped surfaces in vacuum [16, 17].

All these results provide us with the standard description of the gravitational collapse of a massive star as depicted in Figure 3. At the end of its life cycle, when the star had exhausted its nuclear fuel and if the mass of the star is more than a threshold, there is no physical force to counter the gravitational attraction. The star undergoes an implosion and collapses onto itself. The intense gravity distorts the light cone structure and a closed future trapped is formed. Then the singularity theorem ensures the continuation of the collapse till the star is crushed into a point of infinite density and curvature – a gravitational singularity.

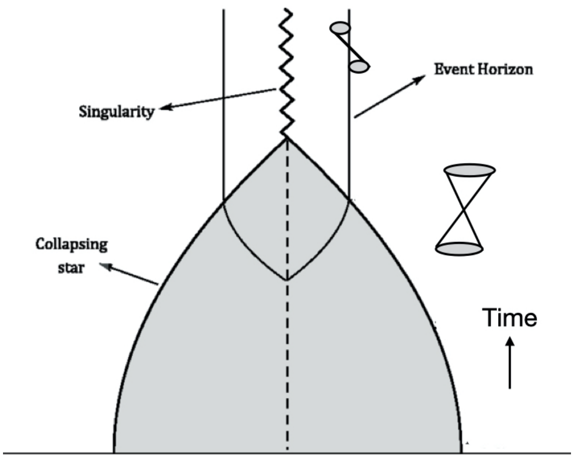

Figure 3: Gravitational collapse of a massive star.

Note that, the applicability of the singularity theorem never guarantee the formation of the event horizon. Whether the singularity is always hidden inside an event horizon is altogether a different question. It is again Penrose who proposed another intriguing hypothesis, known as the weak cosmic censorship [18]. The cosmic censorship hypothesis asserts that all singularities originated in gravitational collapse are always hidden within event horizons, so that no "naked singularities", visible to a distant observer, can ever occur in nature. The exact status of this conjecture is still far from clear [19, 20].

The singularity theorems are often quoted as one of the most significant theoretical accomplishments in general relativity, and mathematical physics. In particular, these theorems led to many radical new developments in theoretical relativity, and intriguing consequences related to the origin of the universe and the collapse of massive stars. These results indicate a fundamental incompleteness of classical general relativity – offering us a substantial challenge to find new physics. It is widely expected that quantum gravitational effects will provide us with a resolution, though the exact nature of the physics near the gravitational singularity remains elusive.

References

1. General Relativity by R. M. Wald (University of Chicago Press, 1984)
2. S. Chandrasekhar, Astrophys. J. **74**, 81-82 (1931)
3. J. R. Oppenheimer and H. Snyder, Phys. Rev. **56**, 455-459 (1939)
4. Über eine Klasse von Lösungen der Gravitationsgleichungen der Relativität. S. Datt, Z.Phys. bf108, 314 (1938).
5. M. D. Kruskal, Phys. Rev. **119**, 1743-1745 (1960)
6. G. Szekeres, Publ. Math. Debrecen **7**, 285-301 (1960)
7. G. A. Shields, Publ. Astron. Soc. Pac. **111**, 661 (1999)
8. R. Penrose, Phys. Rev. Lett. **14**, 57-59 (1965)
9. J. M. M. Senovilla and D. Garfinkle, Class. Quant. Grav. **32**, 124008(12) (2015) [arXiv:1410.5226 [gr-qc]].
10. A. Raychaudhuri, Phys. Rev. **98**, 1123-1126 (1955)
11. S. W. Hawking, Phys. Rev. Lett. **17**, 444-445 (1966)
12. S. Hawking and G. F. R. Ellis, Phys. Lett. **17**, 246-247 (1965)
13. G. F. R. Ellis and S. Hawking, Astrophys. J. **152**, 25 (1968)
14. S. W. Hawking and R. Penrose, Proc. Roy. Soc. Lond. A **314**, 529-548 (1970)
15. The Large Scale Structure of Space-Time by S. W. Hawking and G. F. R. Ellis (Cambridge University Press, USA, 1973)
16. D. Christodoulou, Commun. Pure Appl. Math. **44**, 339-373(3) (1991)
17. D. Christodoulou, The Twelfth Marcel Grossmann Meeting, pp. 24-34 (2012)
18. R. Penrose, Riv. Nuovo Cim. **1**, 252-276 (1969)
19. R. M. Wald, arXiv e-prints; arXiv:gr-qc/9710068
20. P. S. Joshi, Mod. Phys. Lett. A **17**, 1067-1079 (2002)