Roger Penrose and Black Holes

Changhai Lu
Science Writer

ABSTRACT

Black holes have been a hot topic in recent years partly due to the successful detections of gravitational waves from pair merges mostly involving black holes. It is therefore not too great a surprise that the 2020 Nobel Prize in Physics went to black hole researchers: the Royal Swedish Academy of Sciences announced on October 6 2020 that English mathematician and mathematical physicist Roger Penrose had been awarded half of the prize "for the discovery that black hole formation is a robust prediction of the general theory of relativity"; the other half of the prize was shared by German astrophysicist Reinhard Genzel and American astronomer Andrea Ghez "for the discovery of a supermassive compact object at the centre of our galaxy". In this article, we will give a brief introduction to Penrose's research which, as we will see, has a certain unique peculiarity among the achievements that have won Nobel Prizes in Physics.

INTRODUCTION

Let's start, as background, with a quick review of the early history of black holes. The origin of the concept of black holes is often attributed to English natural philosopher John Michell. In 1783, Michell deduced from Newtonian gravity that a star with same density as the sun but hundreds of times bigger in diameter will have a gravity so strong that even light cannot escape. Such a star will therefore look "black" or "dark" to observers far away, and was quite appropriately called a "dark star" by Michell.

Michell's "dark star" is fairly simple in terms of both theory and concept. Nowadays, even a high school student should be able to deduce without much effort that for a star with mass M to become a "dark star", in terms of Newtonian gravity its radius must be no greater than $2GM/c^2$. An impressive fact about this result is that $2GM/c^2$ happens to be the so-called Schwarzschild radius of the simplest (i.e. non-rotating and not charged) black hole in the modern sense (namely, according to general relativity). But despite this impressive equality in radius, the modern black hole has little in common with Michell's "dark star". In fact, even the equality in radius is nothing but a misleading coincidence, since its meaning in modern black hole theory is not the measurable distance from its center to its surface as in Michell's "dark star". To quote Penrose himself [6], "the notion of a black hole really only arises from the particular features of general relativity, and it does not occur in Newtonian theory".

One might ask: how exactly a black hole "arises from the particular features of general relativity"? The answer goes back to a German physicist named Karl Schwarzschild. In January 1916, shortly before his premature death, and less than two months after Einstein published his field equation of general relativity, Schwarzschild found an exact solution, now called Schwarzschild's solution.

General relativity is, in a sense, a theory about spacetime. Schwarzschild's solution describes a particular spacetime configuration. This spacetime is spherically symmetric, and had two striking features both of which led to mathematical difficulties: one occurred at the center of the spherical symmetry, and the other occurred on a sphere now named the "event horizon", whose radius is the Schwarzschild radius that we have mentioned above. It took physicists many years, not without trouble and hesitation, to gradually gain an understanding of these two features. It turned out that the feature at the center of the spherical symmetry is a truly nasty one, now called a "singularity", and is associated with pathological properties such as the divergence of spacetime curvature. The feature on event horizon, however, reflects nothing but a defective choice of coordinate system, and does not pose any essential problem.

The singularity and the event horizon are the two major features of black holes. The discovery, therefore, of Schwarzschild's solution in which both features are present (and collectively called a Schwarzschild black hole) can be more or less treated as a prediction of black holes by general relativity — not a *robust* one though, since what is really "predicted" is only the fact that general relativity *is capable of* describing a black hole. But Schwarzschild's solution alone *cannot* tell us whether there exists any real physical process that can lead to actual black hole *formation* — without such a process, black holes would remain a bizarre theoretical concept without physical relevance.

FORMATION OF BLACK HOLE

So the question now becomes: is there any real physical process that can actually produce a black hole? American physicist J. Robert Oppenheimer and his student Hartland Snyder made some progress towards answering this question in 1939. Oppenheimer and Snyder studied the collapse that will inevitably happen when a star runs out of its nuclear fuel and no longer has sufficient radiative pressure to balance gravity. What they found was: assuming no force exists to stop collapse (which is simplistic but turned out to be correct for sufficiently large stars), when observed by a static observer outside the star, the collapse will slow down owing to relativistic effects and eventually freeze when approaching the event horizon. Whoever "abhors" black holes might be tempted into thinking that black holes cannot form because of the freeze. But such a freeze tells us no more about whether a black hole can form than a video tape broken in the middle tells us what can happen afterwards — it limits what a particular observer can see, but not what can actually happen. In fact, Oppenheimer and Snyder explicitly showed that when you switch to an observer that collapses with the star (the so-called comoving observer), the star will collapse into a singularity *in finite time* and that the event horizon causes no delay.

Does it mean that black hole formation can now be considered a *robust* prediction of general relativity? Not yet, for both Schwarzschild's solution and Oppenheimer and Snyder's study relied on a symmetry that cannot be strictly realized in the physical world: spherical symmetry. In fact, since general relativity is mathematically a highly complicated theory, almost all early efforts to find solutions relied on certain types of symmetry. For instance, the Kerr solution. which was found by New Zealand mathematician Roy Kerr in 1963 and is much more "realistic" than Schwarzschild's solution, relied on axial symmetry — a symmetry not as restricted as spherical symmetry but, nevertheless, still ideal enough to evade physical reality.

Usually, physicists are quite at ease with the fact that symmetries cannot be strictly realized in the physical world, not only because they rely on symmetry too much to dismiss, but also because it is commonly believed and widely validated that minor deviation from strict symmetry will only lead to minor discrepancy. But black hole formation became an exception, at least to some physicists, since it involves a singularity which will cause a breakdown of physical laws on which the very careers of physicists depend. No stake is higher than that, which makes no concept — not even symmetry — not sacrificable in order to save the stake. Some physicists, who blamed symmetry, therefore decided to tackle the problem by abandoning symmetry, in the hope of eliminating the singularity. Prominent Soviet physicists Evgeny Lifshitz and Isaak Markovich Khalatnikov were the main proponents of such efforts, and at a certain point in the 1960s they believed a proof had been obtained showing that the singularity would not arise once symmetry had been abandoned.

There are also numerous other doubts regarding singularity and black holes (for which singularity is a main feature), one of which came from the very person who established general relativity: Albert Einstein. But those other doubts are much less eloquent. For instance, the doubt Einstein himself cast is this: any circular motion around a Schwarzschild black hole at radius less than 1.5 times the Schwarzschild radius will require a speed greater than the speed of light. Since nothing can travel faster than the speed of light, black holes — Einstein so concluded — must not exist. This argument is surprisingly defective since the obvious and correct conclusion of this analysis should be: circular motion is not possible at such radius around a Schwarzschild black hole rather than black holes must not exist, just like, for example, if one cannot swim circularly (without been sucked in) near the center of a whirlpool, it doesn't mean whirlpools cannot exist.

Eloquent or not, it was in such an atmosphere of numerous doubts, that in the autumn of 1964, thirty-three year old Roger Penrose became deeply involved in black hole research.

PENROSE AND BLACK HOLES

Penrose completed a mathematics major and then obtained a Ph.D. in the field of geometry from St John's

College, Cambridge University in 1957. But even when he was still a mathematics student, Penrose developed interests in physics and astronomy under the influence of English astronomer Fred Hoyle and physicist Dennis Sciama. In fact, Hoyle and Sciama's influence is much more than a mere generic influence on interests. Hoyle attracted Penrose into his earliest research in astronomy; and it is in Sciama's circle that Penrose eventually met an excellent collaborator. Stephen Hawking, whose Ph.D. research was under Sciama's guidance and whose fame would skyrocket.

Both of these influences contributed to the achievement that eventually resulted in Penrose's Nobel Prize. Hoyle was a major advocate of the so-called steady-state model, which is a cosmological model that was abandoned by most astronomers in the 1960s when it was strongly disfavored by observations. In the 1950s when Penrose was influenced by Hoyle, however, the steady-state model still enjoyed certain popularity. But even then, while observational evidence was still lacking, an issue that Penrose called "an apparent contradiction between the steady-state model and standard general relativity" had already surfaced. One possible way out of the problem, as some cosmologists proposed, was to use a strategy quite similar to the one that singularity "deniers" would use, namely, to assume that the issue was caused by symmetry, and therefore abandoning symmetry would save the day. Penrose was attracted to the steady-state model and was serious enough to pursue this proposal. The pursuit failed, as recalled in Penrose's book *Fashion Faith and Fantasy* [6]:

I had wanted to see whether an apparent contradiction between the steady-state model and standard general relativity ... might be averted by the presence of deviations from the complete symmetry that is employed in the usual steady-state picture. By the use of a geometrical/topological argument, I had convinced myself that such deviations from symmetry could not remove this contradiction.

But the good thing about scientific research is: the value of it is not always measured by the success or failure of its direct target. In many cases, valuable lessons that led to great success came out of research for which the direct target failed. Penrose's study of the steady-state model turned out to be one of them, out of which two lessons had been learned: one lesson is generic and strategic, namely abandoning symmetry may not always make a difference dramatic enough to save the day; the other lesson is more specific and tactical and showed that the geometrical/topological argument that was rather novel in general relativity research at the time had the power to achieve something traditional methods could hardly reach. These lessons paved a road for Penrose when his interest switched to black hole research, and the road drastically differed from that of Lifshitz and Khalatnikov's in both its goal and method.

What attracted Penrose into the area of black hole research, however, was not his research on the steady-state model, but related to "quasars" (quasi-stellar objects) which were discovered in 1963. These novel astronomical objects shine a hundred times brighter than a typical galaxy but their size is only a millionth of that of the latter (which makes it "quasi-stellar", hence its name), and therefore must be very compact. Preliminary analysis indicated that a giant black hole swallowing matter (including stars) was the most probable mechanism that could fuel so compact and energetic an object. This therefore provided indirect but strong support for the existence of black holes. Since everyone knows that symmetry cannot be exact, therefore if black holes exist, we must show they exist under generic conditions. This, together with the lessons learned from the study of the steady-state model, attracted Penrose into the area of black hole research with the goal of exploring the formation of singularity under generic conditions.

This goal is opposite to that of Lifshitz and Khalatnikov, and the best method to pursue it is to use the "geometrical/topological argument" which was the second lesson Penrose learned from the study of the steady-state model. The reason for this is simple: since the goal is to explore the formation of singularity under generic conditions — especially when there is no symmetry, so that properties such as shape and size become pretty much irrelevant. Furthermore, since general relativity is a highly geometric theory, the formation of singularity is a highly geometric problem about spacetime structure, and we all know that if properties such as shape and size are irrelevant in a geometric problem, what remains to be relevant will be topological properties, which thus justifies the use of the "geometrical/topological argument".

But even with the goal set, and the method well within the technical expertise of Penrose, whose Ph.D. research was in the area of geometry, reaching the goal is still highly challenging, and requires certain inspiration. Although the coming and going of inspiration is often difficult to trace, in the particular case of Penrose's black hole research, we are lucky enough to have the master's own reminiscences which he described in his popular book *The Emperor's New Mind* and on other occasions.

According to these reminiscences, Penrose got his inspiration late in the autumn of 1964, shortly after he started his black hole research. One day in that autumn, Penrose and mathematical physicist Ivor Robinson were walking along the street, chatting about something completely unrelated to black hole research, and were stopped by a red signal while crossing a side road. It was at that moment that an idea occurred to Penrose. Later that day, after Robinson left, Penrose returned to his office and (to quote his own words) "finally brought to mind the thought that I had had while crossing the street — a thought which had momentarily elated me by providing the solution to the problem that had been milling around at the back of my head!"

The thought that he so elaborately brought to mind and provided "the solution to the problem" is related to a concept called "closed trapped surface" whose fundamental property is: all light-like geodesics orthogonal to it — regardless of inward or outward propagation — are converging. To put it in layman's terms, it is a two dimensional closed surface from which light cannot escape. Equipped with this inspiration, and after several months' diligent work, Penrose proved an important result in 1965 that we will call the Penrose singularity theorem, and which is the earliest version of a class of theorems now called "singularity theorems".

SINGULARITY THEOREMS

What are singularity theorems? Or, to be more specific, what is the Penrose singularity theorem? Briefly speaking, the Penrose singularity theorem is a theorem that leads to the formation of a singularity by assuming three types of premises — a logical structure shared by all singularity theorems. Among the three types of premises, the first asserts certain generic properties of matter; the second imposes certain restrictions on spacetime itself; and the third assumes certain conditions of matter distribution. With these premises, Penrose proved that the formation of a singularity is generic and inevitable, and does not rely on symmetry.

In that same year (i.e. 1965) when Penrose proved his singularity theorem, the world's leading general relativity experts — including Lifshitz and Khalatnikov for whom permission to travel outside the Soviet Union was obtained not without trouble — gathered in London for the Third Conference on General Relativity and Gravitation. This is the stage on which Penrose's affirmative result and Lifshitz and Khalatnikov's negative result on the formation of singularities collided for the first time.

The "collision" didn't yield any immediate outcome, but Penrose's novel "geometrical/topological argument" attracted several young researchers who — in a manner similar to Penrose — also had technical strength in geometry and topology. Among them were the American theoretical physicist Robert Geroch and Sciama's graduate student Hawking, who we mentioned before; both were only twenty-three years old at the time. In the next several years, Penrose, Hawking, Geroch and others proposed and proved more versions of the singularity theorem, which differed from each other mainly in the details of the premises. Through the blooming of these singularity theorems, the existence of singularities and black holes gained more and more theoretical acceptance.

This trend finally shook Lifshitz and Khalatnikov. In September 1969, American physicist Kip Thorne visited the Soviet Union [8]. Lifshitz took the opportunity to hand a manuscript to him, and asked him to submit it to *Physical Review Letters* which Soviet scientists could not do themselves at that time without first going through a lengthy security clearance process. In the manuscript, Lifshitz acknowledged the error of his and Khalatnikov's earlier work that let them to believe that singularities could not exist without symmetry.

This concession by Lifshitz and Khalatnikov removed the main objection among physicists regarding the existence of singularities and black holes in general relativity (careful readers might have noticed, in the domain of existence, we sometimes used singularity and black hole in a somewhat exchangeable way, as if a result about one can automatically extend to the other. The relation between the two is actually a quite subtle one; interested readers may consult, for instance, [3]). But the Penrose singularity theorem itself still has a weak point that needs to be — and can be — addressed. As we mentioned before, singularity theorems all assumed three types of premises. Among them, the generic properties of matter basically asserts that the energy density must be non-negative, which is widely considered valid in classical physics; the conditions of matter distribution are known to be realizable in physical processes such as the collapse of sufficiently large stars. But the restriction imposed on spacetime itself by the Penrose singularity theorem turns out to be too theoretical and too strong. In fact, this restriction, which requires the existence of a so-called Cauchy hypersurface, is so theoretical that it is hard to think of any observational evidence that can possibly verify it, and it is so strong that it was actually violated by a counter-example posted by Penrose himself in a paper published in the same year (i.e. 1965) as his singularity theorem paper.

This weak point in the Penrose singularity theorem was no secret to researchers in that area. Hawking, for instance, commented in his autobiography *My Brief History* [2] that it is possible that early singularity theorems "simply proved that the universe didn't have a Cauchy surface" rather than proved the existence of singularities and black holes. Penrose himself later, in a collaborative work with Hawking, also admitted that [1] "the assumption of the existence of a global Cauchy hypersurface is hard to justify from the standpoint of general relativity". Trying to improve on the Penrose singularity theorem by eliminating this weak point was actually a major motivation behind the multiple versions of singularity theorems proposed and proved in those years.

In the end, Penrose and Hawking collaborated on a paper titled "The Singularities of Gravitational Collapse and Cosmology", that laid down a theorem now called the Hawking-Penrose singularity theorem, published in 1970. In this theorem, which covers both black hole singularity and cosmological singularity, premises are made much more realizable. We all know from basic logic that for a simple logical deduction to draw a correct conclusion, not only the deduction itself must be valid, but the premises must also be valid. Similarly, for a theorem that was intended to describe the physical world, a physically relevant conclusion relies not only on the mathematical correctness of the theorem, but also the physical realizability of the premises. In this sense, the Hawking-Penrose singularity theorem is physically much more relevant than earlier versions of the singularity theorem due to its much more realizable premises. Among all singularity theorems, if we are to select one that best deserves the praise of "the discovery that black hole formation is a robust prediction of the general theory of relativity" which won half of the Nobel Prize in Physics this year, it should be the Hawking-Penrose singularity theorem. And it is a pity that Stephen Hawking has already passed away — a pity both for Hawking and for the history of the Nobel Prize.

EPILOGUE

Let me end this introduction by pointing out a very unique peculiarity that distinguishes Penrose's black hole research from not only the other achievements that have won the Nobel Prize in Physics, but most other physics research in general, namely the singularity theorems Penrose and others proved are very similar to pure mathematical theorems, except that these theorems are theorems *in the framework of general relativity* rather than in an ordinary mathematical axiomatic system. A singularity theorem merely draws "a robust prediction" based on general relativity, in the sense that even if its prediction is invalidated by observation, it is most likely general relativity rather than the singularity theorem that will be in trouble. A singularity theorem might become physically irrelevant in such cases, but its correctness as a mathematical theorem may well remain.

Black holes are astronomical objects, and for this reason many have considered this year's Nobel Prize in Physics as yet another case when astronomical research won a physics prize. But at least for the half prize that goes to Penrose, it is perhaps better considered as a case of *mathematical research* winning a physics prize, which is much rarer — perhaps completely unique so far, and therefore adds a far more colorful chapter in the history of Nobel Prizes.

References

[1] S. Hawking, Hawking on the Big Bang and Black Holes, (World Scientific Publishing Co. Pte. Ltd., 1993).
[2] S. Hawking, My Brief History, (Bantam Books, 2013).
[3] C. Lu, From Singularities to Wormholes: Selected Topics in General Relativity, (Tsinghua University Press, 2013).
[4] R. Penrose, The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics, (Penguin Books, 1991).
[5] R. Penrose, Cycles of Time: An Extraordinary New View of the Universe, (Vintage Books, 2012).
[6] R. Penrose, Fashion, Faith, and Fantasy in the New Physics of the Universe, (Princeton University Press, 2016).
[7] J. M. M. Senovilla and D. Garfinkle, arXiv:1410.5226 [gr-qc].
[8] K. S. Thorne, Black Holes and Time Warps: Einstein's Outrageous Legacy, (W. W. Norton & Company, 1995).

Changhai Lu is a science writer in the area of Physics, Astronomy and Mathematics. He received his Ph.D. in Physics from Columbia University in 2000. He has published more than ten books, two of which have won national level awards. He also published numerous science articles in magazines, newspapers and internet media.