

Quantum Mechanics of Black Holes

Edward Witten

The popular conception of black holes reflects the behavior of the massive black holes found by astronomers and described by classical general relativity. These objects swallow up whatever comes near and emit nothing. Physicists who have tried to understand the behavior of black holes from a quantum mechanical point of view, however, have arrived at quite a different picture. The difference is analogous to the difference between thermodynamics and statistical mechanics. The thermodynamic description is a good approximation for a macroscopic system, but statistical mechanics describes what one will see if one looks more closely.

In quantum mechanics, if a time-dependent transition is possible from an initial state $|i\rangle$ to a final state $|f\rangle$, then it is also possible to have a transition in the opposite direction from $|f\rangle$ to $|i\rangle$. The most basic reason for this is that the sum of quantum mechanical probabilities must always equal 1. Starting from this fact, one can show that, on an atomic time scale, there are equal probabilities for a transition in one direction or the other (1).

This seems, at first, to contradict the whole idea of a black hole. Let B denote a macroscopic black hole—perhaps the one in the center of the Milky Way—and let A be some other macro-

scopic body, perhaps a rock or an astronaut. Finally, let B^* be a heavier black hole that can be made by combining A and B . General relativity tells us that the reaction $A + B \rightarrow B^*$ will occur whenever A and B get close enough. Quantum mechanics tells us, then, that the reverse reaction $B^* \rightarrow A + B$ can also happen, with an equivalent amplitude.

The reverse reaction, though, is one in which the heavier black hole B^* spontaneously emits the body A , leaving behind a lighter black hole B . That reverse reaction is exactly what does not happen, according to classical general relativity. Indeed, the nonoccurrence of the reverse reaction, in which a black hole re-emits whatever it has absorbed, is often stated as the defining property of a black hole.

It seems, then, that black holes are impossible in light of quantum mechanics. To learn more, let us consider another physical principle that is also seemingly violated by the existence of a black hole. This is time-reversal symmetry, which says that if a physical process is possible, then the time-reversed process is also possible. Clearly, black holes seem to violate this as well.

Time reversal is a subtle concept, and elementary particle physicists have made some unexpected discoveries about it (2). However, for applications to black holes, the important problem with time reversal is that in everyday life, it simply does not appear to be valid. We can spill a cup of water onto the ground, but the water never spontaneously jumps up into the cup.

The explanation has to do with randomness at the atomic level, usually called entropy. Spilling the cup of water is an irreversible operation in practice, because it greatly increases the number of states available to the system at the atomic level, even after one specifies all of the variables—such as temperature, pressure, the amount of water, the height of the water above the ground, and so on—that are visible macroscopically. The water could jump back up into the cup if the initial conditions are just right at the atomic level, but this is prohibitively unlikely.

The second law of thermodynamics says that in a macroscopic system, like a cup of water, a process that reduces the randomness or entropy in the universe can never happen. Now suppose that we consider what is sometimes called a

School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA. E-mail: witten@ias.edu

mesoscopic system—much larger than an atom, but not really macroscopic. For example, we could consider 100 water molecules instead of a whole cupful. Then we should use statistical mechanics, which tells us that a rare fluctuation in which randomness appears to diminish can happen, but very rarely. Finally, at the level of a single particle or a handful of particles, we should focus on the fundamental dynamical equations: Newton's laws and their modern refinements. These fundamental laws are completely reversible.

Since the late 19th century, physicists have understood that thermodynamic irreversibility arises spontaneously by applying reversible equations to a macroscopic system, but it has always been vexingly hard to make this concrete.

Black Hole Entropy and Hawking Radiation

Now let us go back to the conflict between black holes and quantum mechanics. What is really wrong with the reverse reaction $B^* \rightarrow A + B$, wherein a heavier black hole B^* decays to a lighter black hole B plus some other system A ?

A great insight of the 1970s [originating from a suggestion by Bekenstein (3)] is that what is wrong with the reverse reaction involving black holes is just like what is wrong with a time-reversed movie in which a puddle of water flies off the wet ground and into the cup. A black hole should be understood as a complex system with an entropy that increases as it grows.

In a sense, this entropy measures the ignorance of an outside observer about what there is inside the black hole. When a black hole B absorbs some other system A in the process $A + B \rightarrow B^*$, its entropy increases, along with its mass, in keeping with the second law of thermodynamics. The reverse reaction $B^* \rightarrow A + B$ diminishes the entropy of the black hole as well as its mass, so it violates the second law.

How can one test this idea? If the irreversibility found in black hole physics is really the sort of irreversibility found in thermodynamics, then it should break down if A is not a macroscopic system but a single elementary particle. Although a whole cupful of water never jumps off the floor and into the cup, a single water molecule certainly might do this as a result of a lucky fluctuation.

This is what Hawking found in a celebrated calculation (4). A black hole spontaneously emits elementary particles. The typical energy of these particles is proportional to Planck's constant, so the effect is purely quantum mechanical in nature, and the rate of particle emission by a black hole of astronomical size is extraordinarily small, far too small to be detected. Still, Hawking's insight means that a black hole is potentially no different from any other quantum system, with reactions $A + B \rightarrow B^*$ and $B^* \rightarrow A + B$ occurring in both directions at the microscopic level. The irreversibility of classical black hole physics is just like the familiar irreversibility of thermodynamics, valid when what is absorbed by the black hole is a macroscopic system.

Although it was a shock at the time, perhaps in hindsight we should not be surprised that classical general relativity does not describe properly the emission of an atom or elementary particle from a black hole. After all, classical general relativity is not a useful theory of atoms and individual elementary particles, or quantum mechanics would never have been needed.

However, general relativity is a good theory of macroscopic bodies, and when it tells us that a black hole can absorb a macroscopic body but cannot emit one, we should listen.

Black Holes and the Rest of Physics

Is the quantum theory of black holes just a theoretical construct, or can we test it? Unfortunately, the usual astrophysical black holes, formed from stellar collapse or in the centers of galaxies, are much too big and too far away for their microscopic details to be relevant. However, one of the cornerstones of modern cosmology is the study of the cosmic microwave radiation that was created in the Big Bang. It has slightly different temperatures in different directions. The theory of how this came about is in close parallel with the theory of Hawking radiation from black holes, and its success adds to our confidence that the Hawking theory is correct.

Surprisingly, in the past 15 years, the theory of Hawking radiation and related ideas about quantum black holes have turned out to be useful for theoretical physicists working on a variety of more down-to-earth problems. To understand how this happened, we need one more idea from the early period.

The Membrane Paradigm for Black Holes

The entropy of an ordinary body like Earth or the Sun is basically a volume integral; to compute the entropy, one computes the entropy density and integrates it over the interior of Earth or the Sun.

Black holes seem to be different. According to Bekenstein and Hawking, the entropy of a black hole is proportional to the surface area of the black hole, not to its volume. This observation led in the early days to the membrane paradigm for black holes (5). The idea of the membrane paradigm is that the interactions of a black hole with particles and fields outside the hole can be modeled by treating the surface or horizon of the black hole as a macroscopic membrane. The membrane associated with a black hole horizon is characterized by macroscopic properties rather similar to those that one would use to characterize any ordinary membrane. For example, the black hole membrane has temperature, entropy density, viscosity, and electrical conductivity.

In short, there was a satisfactory thermodynamic theory of black hole membranes, but can one go farther and make a microscopic theory that describes these membranes? An optimist, given the ideas of the 1980s, might hope that some sort of quantum field theory would describe the horizon

of a black hole. What sort of theory would this be, and how could one possibly find it?

Gauge-Gravity Duality

The known forces in nature other than gravity are all well described in the standard model of particle physics in terms of quantum field theories that are known as gauge theories. The prototype is Maxwell's theory of electromagnetism, interpreted in modern times as a gauge theory.

Quantum gauge theory is a subtle yet relatively well-understood and well-established subject. The principles are known, but the equations are hard to solve. On the other hand, quantum gravity is much more mysterious. String theory has given some insight, but the foundations are still largely unknown.

In the 1990s, string theorists began to discover that aspects of black hole physics can be modeled by gauge theory (6, 7). Such insights led to a remarkable new way to use gauge theory to study black holes and other problems of quantum gravity (8).

This relied on the fact that string theory has extended objects known as branes (9), which are rather like membranes except that in general they are not two-dimensional. In fact, the word "brane" is a riff on "membrane."

Branes can be described by gauge theory; on the other hand, because black holes can be made out of anything at all, they can be made out of branes. When one does this, one finds that the membrane that describes the horizon of the black hole is the string theory brane.

Of course, we are cutting corners with this very simple explanation. One has to construct a string theory model with a relatively large negative cosmological constant (in contrast with the very small positive one of the real world), and then, under appropriate conditions, one gets a gauge theory description of the black hole horizon.

Solving the Equations of Gauge Theory

Gauge-gravity duality was discovered with the aim of learning about quantum gravity and black holes. One can turn the relationship between these two subjects around and ask whether it can help us better understand gauge theory.

Even though gauge theory is the well-established framework for our understanding of much of physics, this does not mean that it is always well understood. Often, even if one asks a relatively simple question, the equations turn out to be intractable.

In the past decade, the gauge theory description of black holes has been useful in at least two areas of theoretical physics. One involves heavy ion collisions, studied at the Relativistic Heavy Ion Collider at Brookhaven. The expanding fireball created in a collision of two heavy nuclei turns out to be a droplet of nearly ideal fluid. In principle, this should all be described by known equations of gauge theory—quantum chromodynamics, to be precise—but the equations are intractable. It turns out that by interpreting the gauge theory as a description of a black hole horizon, and using the

Black Holes

Einstein equations to describe the black hole, one can get striking insight about a quantum almost-ideal fluid (10). This has become an important technique in modeling heavy ion collisions.

Condensed matter physics is described in principle by the Schrödinger equation of electrons and nuclei, but for most systems, a full understanding based on the Schrödinger equation is way out of reach. Nowadays, there is great interest in understanding quantum critical behavior in quasi-two-dimensional systems such as high temperature superconductors. These systems are studied by a wide variety of methods, and no one approach is likely to be a panacea. Still, it has turned out to be very interesting to study two-dimensional quantum critical systems by mapping them to the horizon of a black hole (11). With this approach, one can perform calculations that are usually out of reach.

Among other things, this method has been used to analyze the crossover from quantum to

dissipative behavior in model systems with a degree of detail that is not usually possible. In a sense, this brings our story full circle. The story began nearly 40 years ago with the initial insight that the irreversibility of black hole physics is analogous to the irreversibility described by the second law of thermodynamics. In general, to reconcile this irreversibility with the reversible nature of the fundamental equations is tricky, and explicit calculations are not easy to come by. The link between ordinary physics and black hole physics that is given by gauge-gravity duality has given physicists a powerful way to do precisely this. This gives us confidence that we are on the right track in understanding quantum black holes, and it also exhibits the unity of physics in a most pleasing way.

References and Notes

1. The precise mathematical argument uses the fact that the Hamiltonian operator H is hermitian, so that the transition amplitude $\langle f|H|i\rangle$ is the complex conjugate of the transition amplitude $\langle i|H|f\rangle$ in the opposite direction.
2. The precise time-reversal symmetry of nature also includes reflection symmetry and charge conjugation.
3. J. Bekenstein, *Phys. Rev. D Part. Fields* **7**, 2333 (1973).
4. S. W. Hawking, *Nature* **248**, 30 (1974).
5. K. S. Thorne, D. A. MacDonald, R. H. Price, Eds., *Black Holes: The Membrane Paradigm* (Yale Univ. Press, New Haven, CT 1986).
6. A. Strominger, C. Vaia, *Phys. Lett. B* **379**, 99 (1996).
7. S. S. Gubser, I. R. Klebanov, A. W. Peet, *Phys. Rev. D Part. Fields* **54**, 3915 (1996).
8. J. M. Maldacena, *Adv. Theor. Math. Phys.* **2**, 231 (1998).
9. M. J. Duff, R. R. Khuri, J.-X. Lu, *Phys. Rep.* **259**, 213 (1995).
10. P. K. Kovtun, D. T. Son, A. O. Starinets, *Phys. Rev. Lett.* **94**, 111601 (2005).
11. S. Sachdev, *Annu. Rev. Cond. Matt. Phys.* **3**, 9 (2012).

Acknowledgments: This research was supported in part by NSF grant PHY-096944.
10.1126/science.1221693