The Fate of Quantum Information in a Black Hole

Daniel Carmody*

Quantum Computation and Quantum Information (33-658)

(Dated: May 12, 2008)

I. INTRODUCTION

This past semester our class has spent a great deal of time discussing the principles of quantum mechanics as they relate to the new fields of quantum computation and quantum information. Most of the emphasis of this discussion was placed on quantum computation, the practical idea that the interesting properties of quantum mechanics may provide for a new approach to computing, and with this approach come advances in computing power and reductions in computing time. This idea was the focus of this class and has been the focus of much research.

The second part of our course title, quantum information, is a much more abstract idea, and as a result is a little more difficult to grasp. It will be expanded on in greater detail in a later section, but in brief it is the idea that a quantum system possesses a property, called information, that can be thought of as something real and intrinsic to a system. Often quantum information is mentioned in reference to the processing that occurs within a quantum computer, but it has broader applications than just this.

* dcarmody@andrew.cmu.edu

One of the more interesting circumstances in which the importance of quantum information arises is when its principles are applied to a different, seemingly unrelated, field of physics: general relativity. In particular, it is in black holes, where space-time is highly warped and our knowledge of the physical environment and processes is highly lacking, that both of these concepts are relevant. In this situation the physics of the very large meets the physics of the very small, and we encounter the problem that theoretical physicists have called the black hole information loss paradox.

This problem is concerned with the fate of quantum information in a black hole system. It was first introduced by Stephen Hawking in 1975 [1], and it has been argued about ever since. Before Hawking's work, it was assumed that mass would fall into a black hole, and the size of the black hole would grow monotonically, with no mechanism by which the size could decrease. Hawking found, though, that this was not the case, and that it was possible for a black hole to evaporate, eventually disappearing entirely out of existence. This seemingly benign process implied that black holes were capable of destroying quantum information, a statement that contradicted one of the central tenets of quantum mechanics. This was an unsettling idea, as it would require the entire field, a field that had proven itself to be very effective thus far, to be reformulated. This was such an unsavory idea that there were those who believed that there must be an alternate explanation. One of these was John Preskill, who believed so strongly in quantum mechanics that he was willing to bet on its veracity. In 1997, he proposed a wager to two of his colleagues - Stephen Hawking and Kip Thorne - and it is in this wager that the problem is probably most clearly expressed. The principle that Preskill hoped was true, the principle on which he was betting, was this:

"When an initial pure quantum state undergoes gravitational collapse to form a black hole, the final state at the end of black hole evaporation will always be a pure quantum state." [12]

The alternative to this, that championed by Thorne and Hawking, is that after collapse and evaporation a pure state may evolve into a mixed state. This terminology of pure versus mixed will be expanded on later, but the essential idea is that if a pure state evolves into a mixed one, then information is lost: the black hole has destroyed it.

Preskill thought that this destruction was impossible, and he felt that once the correct theory of quantum gravity is found, it will become clear that there is a method for the information to be recovered fully. Several suggestions have been proposed, but for one reason or another they have all fallen short.

So what is this paradox that has caused such prominent physicists to come to such different conclusions? And what is it about the information loss that causes such consternation? To address these issues, it is first necessary to introduce the idea of quantum information.

II. PRINCIPLES OF QUANTUM INFORMATION

Any discussion of quantum information naturally starts with the concept of a quantum bit, the qubit. Whereas a classical bit can be in one of two states, 0 or 1, a qubit can be in any of an infinite number of states. Often it is expressed in the Dirac notation,

$$|\psi\rangle = a\,|0\rangle + b\,|1\rangle \tag{1}$$

where $|0\rangle$ and $|1\rangle$ can be thought of as basis states, and a and b are chosen so that the magnitude of the total state is one ($|a|^2 + |b|^2 = 1$, for instance, if $a = b = \frac{1}{\sqrt{2}}$). This normalization is chosen to reflect the idea that the total probablity to be in *some* state is one ($\langle\psi|\psi\rangle = 1$). An example of a possible physical meaning of a qubit, and one commonly used to illustrate the concept, is the spin of an electron, where $|0\rangle$ means spin $+\frac{1}{2}$ and $|1\rangle$ means spin $-\frac{1}{2}$ (along some chosen direction). This is the state vector representation of the system.

A quantum system may consist of multiple objects (for example, multiple electrons, each with an individual spin state). In this case the total system is taken to be the tensor product of the two separate states.

$$|\psi\rangle_{total} = |\psi\rangle_1 \otimes |\psi\rangle_2 \tag{2}$$

For instance, taking $|\psi\rangle_1 = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|\psi\rangle_2 = |0\rangle$,

$$|\psi\rangle_{total} = \frac{1}{\sqrt{2}}(|0\rangle\,|0\rangle + |1\rangle\,|0\rangle) = \frac{1}{\sqrt{2}}(|00\rangle + |10\rangle) \tag{3}$$

A state is said to be entangled if it *cannot* be written as the tensor product of two substates, for example,

$$|\psi\rangle_{ent} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{4}$$

An alternative formulation to state vectors are density matrices (also called density operators),

$$\rho = |\psi\rangle\,\langle\psi| \tag{5}$$

where $|\psi\rangle$ is a state as given by Eq. 1. An important concept for the purpose of understanding the black hole problem is the differentiation between a "pure" quantum state and a "mixed" state. When discussing this concept, it is necessary to use the density matrix description of a quantum system. In this formulism, a pure state is one in which the state is known exactly, i.e., there is only one $|\psi\rangle$ that makes up ρ.

A mixed state, on the other hand, is one in which the quantum state is *not* known exactly; there is only a probability to be in any number of states. An example of this is

$$\rho = p\,|\psi\rangle\,\langle\psi| + q\,|\phi\rangle\,\langle\phi| \tag{6}$$

where $|\psi\rangle$ and $|\phi\rangle$ are two separate quantum states, and p and q are the probabilities of the quantum system to be in either of these two states ($p + q = 1$).

One of the important and fundamental characteristics of quantum mechanics is that all evolution of a quantum system must be unitary. This condition is necessary in order to preserve probability. So a state vector $|\psi\rangle$, initially in the state $|\psi(0)\rangle$, acted on by a unitary time evolution operator $\hat{U}(t)$, will change as such:

$$|\psi(t)\rangle = \hat{U}(t)\,|\psi(0)\rangle \tag{7}$$

Knowing this, it is easily seen that a pure quantum system, as described by Eq. 5, will remain a pure system under unitary evolution. A density operator under unitary evolution will evolve like

$$\rho' = U\rho U^{\dagger} \tag{8}$$

Now that we have these tools, we can begin to talk about the information component of a quantum system. To do this, we introduce the definition of entropy of a quantum system.

When talking about information, entropy is interpreted as the amount of information *missing*. In general, entropy may be defined for any system that has a probability distribution associated with it. For a quantum system described by ρ, the entropy (as defined by Von Neumann [10]) is given by

$$S(\rho) = -Tr(\rho \, log_2 \rho) \tag{9}$$

which for a density matrix described by

$$\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{10}$$

is simply

$$S(\rho) = -\sum_i p_i \, log_2(p_i) \tag{11}$$

If one has a pure state, say $\rho = \left|0\right\rangle \left\langle 0\right|$, then it may be seen that the entropy will be zero,

$$S = -1 \; log_2 \, 1 \; (-\; 0 \; log_2 \, 0) = 0 \tag{12}$$

and it is clear that the entropy will be zero for any such pure state. An entropy of zero is interpreted as meaning that no information is missing from the system. For a mixed state, say $\rho = \frac{1}{2}\left|0\right\rangle \left\langle 0\right| + \frac{1}{2}\left|1\right\rangle \left\langle 1\right|$, the entropy will be some non-zero value.

$$S = -\frac{1}{2} \, log_2 \, \frac{1}{2} - \frac{1}{2} \, log_2 \, \frac{1}{2} = \frac{1}{2} + \frac{1}{2} = 1 \tag{13}$$

Likewise, all other mixed states will have entropies of greater than zero, and entropies of greater than zero are to be interpreted as indicating that information is missing from the system.

III. PROPERTIES OF BLACK HOLES

The second component of the problem that we must understand is the formation of black holes and the interesting properties that they exhibit. Though somewhat difficult to discuss formally, black holes are relatively easy to treat at a basic level.

A black hole is an object created through the gravitational collapse of a massive star. Most of the time the outward pressures of nuclear forces within stars are enough to overcome the inward gravitational pull. In some cases, however, as in dying stars when the nuclear fuel is burned off, the gravitational force wins out, and the mass is concentrated to such great density that, within a certain radius, even light is unable to escape its pull. For a given amount of mass M, the radius within which it must be concentrated is given by this equation:

$$R_S = \frac{2G_N M}{c^2} \tag{14}$$

This radius, called the Schwarzschild radius or event horizon, is the outside edge of the 'blackness' of a black hole. For an object with a mass on the order of the Sun, this radius turns out to be about 3km. Anything within this radius cannot escape to outside of it, so it may be considered off limits to an outside observer. An object, having fallen in past the event horizon, can have no causal connection with any objects or events at greater radii. The interior of a black hole is therefore a place about which we may really know nothing. Thus, despite its initial complexity, a black hole is a very simple object; it is almost completely defined by its size (its radius), which is in turn completely described by its mass.

IV. BLACK HOLE EVAPORATION

In ‘classical’ general relativity, or that in which quantum effects are not considered, a black hole can only grow in size. Mass falls into the black hole, and its radius increases proportionately according to the formula (Eq. 14).

This classical conclusion is changed once the principles of quantum field theory are applied. In this theory, the vacuum is populated by virtual particle/anti-particle pairs, which normally recombine with each other and annihilate. At the surface (the event horizon) of a black hole, however, it is possible for this pair to become separated such that one lies outside the horizon and one within it. To an observer outside the horizon, being unable to see one of the pair, it would appear that there were particles being created and emitted from the surface of the black hole. These particles are radiated away (a phenomenom called Hawking radiation), giving the black hole a characteristic temperature,

$$T = \frac{\hbar c^3}{8\pi k_B G_N M} \tag{15}$$

as well as providing a scheme by which a black hole may lose energy and shrink. In this manner a black hole may radiate away its mass, eventually evaporating away entirely and shrinking (or exploding) out of existence. Hawking also found there to be an entropy associated with this radiation, proportional to the mass of the black hole squared,

$$S \propto M^2 \tag{16}$$

The fact that this radiation has entropy is an important concept, because it is in this entropy that it becomes apparent where the paradox arises.

V. THE PARADOX

Let us now consider a pure quantum system that falls into a black hole, or a black hole that was itself a pure quantum system before collapse. The black hole emits Hawking radiation, and the important thing to note about this radiation is that it is thermal, and thermal radiation is described as a mixed quantum state - it carries little, if any, information. One way to think about this is to consider that the radiation is emitted at the surface of the black hole, and can not be influenced in any way by anything within the Schwarzschild radius. Another way is to think of the radiation as an entangled quantum state. A simplified description of this is given by

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\bar{a}\rangle_i |a\rangle_o + |b\rangle_i |\bar{b}\rangle_o) \tag{17}$$

Here a and b represent two different sets of emitted particles (for instance, a photon of a certain wave number and polarization), with $\bar{a}$ and $\bar{b}$ being the anti-particles. The subscripts i and o indicate which particle is emitted outward from the horizon and which is trapped inside. The choice of particle or anti-particle as being emitted or trapped is arbitrary.

The total system $|\psi\rangle$ can be seen to be in a pure state, $\rho = |\psi\rangle\langle\psi|$, but what about the subsystems inside or outside the black hole? To describe these we need to make use of the reduced density operator, whose equation is given by

$$\rho^{i(o)} = Tr_{o(i)}(\rho) \tag{18}$$

The reduced density operator for outside the horizon would be given by

$$\begin{aligned} \rho^o = Tr_i(\rho) &= Tr_i(\frac{1}{2}(|\bar{a}\rangle_i |a\rangle_o + |b\rangle_i |\bar{b}\rangle_o)(\langle\bar{a}|_i \langle a|_o + \langle b|_i \langle\bar{b}|_o)) \\ &= \frac{1}{2}(\langle\bar{a}|\bar{a}\rangle_i |a\rangle_o \langle a|_o + \langle b|\bar{a}\rangle_i |a\rangle_o \langle\bar{b}|_o + \langle\bar{a}|b\rangle_i |\bar{b}\rangle_o \langle a|_o + \langle b|b\rangle_i |\bar{b}\rangle_o \langle\bar{b}|_o) \end{aligned}$$

$$= \frac{1}{2} \left|a\right\rangle_o \left\langle a\right|_o + \frac{1}{2} \left|\bar{b}\right\rangle_o \left\langle\bar{b}\right|_o \tag{19}$$

which is a mixed state. This might not initially appear to be an upsetting conclusion, but it hides a more sinister implication.

While the black hole exists there is no cause for alarm. The total system consists of the mixed state thermal radiation and the black hole, and the total system may still be a pure state - the information from the original system may still lie within the black hole, even if it is inaccessible behind the event horizon.

What are we to think, though, once the black hole has evaporated away entirely? All we are left with then is the outside state: thermal radiation in a completely mixed state carrying no information. Information has been destroyed, a pure state has evolved into a mixed state, unitary evolution has been proven false, probabilities will no longer be conserved, and quantum mechanics as we know it has failed us.

VI. POSSIBLE RECONCILIATIONS

Seeing as quantum mechanics has served us so well so far, some have felt it unnecessary to have to alter it, especially when there is so much that is unknown about the physics of black holes. Instead, they have searched for solutions elsewhere.

Perhaps a black hole does not evaporate or disappear completely, but instead leaves some kind of quantum scale remnant [8]. Knowledge of general relativity at this scale is incomplete, so it is possible that in the final moments of a black hole's life some new effect takes place that results in an object that contains all the information heretofore thought destroyed. The problem with this is that since there is an infinite number of ways to create a black hole (associated with an infinite number of quantum states that may go into its formation), and the final remnant must be unique and distinguishable for each one of these, there must be an infinite number of possible remnants. Accounting for this variability in an object that is of Planck scale is difficult to do, though it is not entirely out of the question.

Another approach is to consider that perhaps the surface of a black hole is more complicated than we give it credit. Current theory suggests that, though to an outside observer the event horizon is the limit of observation, to a freely falling observer the horizon is no place special. They fall past it without any momentous change. What if, though, there was some process by which an in-falling state interacted and became correlated with the horizon and the radiation being emitted there. Then there might be a way by which the Hawking radiation could carry information, and the paradox would be resolved. String theory has been considered as a likely candidate by which this may be done, and some attempts have been made along these lines [7].

Similarly, Hawking himself has introduced new ideas about the behavior of black holes. In 2005, he published a paper describing a method by which the surface of a black hole could fluctuate and information could escape [9]. Having reached this conclusion, Hawking conceded the bet to Preskill, and awarded him that which had been guaranteed to the winner of the wager: an encyclopedia "from which information can be recovered at will" [12]. Though Hawking has conceded, Kip Thorne remains unconvinced, and has not yet accepted a final resolution.

In the coming months there's a chance that some experimental evidence for black hole decay may be on the horizon (pardon the pun). Some theorists have predicted that the energies to be achieved by the Large Hadron Collider (LHC) will be sufficient to compactify particles into quantum scale black holes [13]. If this is the case, then perhaps some light may be shed on this paradox, and important knowledge gained into the realms of information, general relativity, and quantum gravity.

[1] S. Hawking, "Particle Creation by Black Holes", *Comm. Math. Phys.* 43 (1975) 199.

[2] J. Preskill, "Do Black Holes Destroy Information", *Prepared for the international Symposium on Black holes, Membranes, Wormholes, and Superstrings, The Woodlands, Texas, 16-18 January, 1992*

[3] J. Preskill, "Black holes and information: A crisis in quantum physics", *Caltech Theory Seminar, 21 October 1994*, arXiv:hep-th/9209058v1

[4] F. Bais and J. Farmer, "The Physics of Information", ITFA-2007-37

[5] S. Giddings, "The Black Hole Information Paradox," Proc. PASCOS symposium/Johns Hopkins Workshop, Baltimore, MD, 22-25 March, 1995, arXiv:hep-th/9508151v1

[6] S. Giddings, "Black holes, information, and locality", arXiv:0705.2197v2

[7] S. Mathur, "What Exactly is the Information Paradox", Proc. 4th Aegean Summer School on Black Holes, Mytilene, Greece, September 2007, arXiv:0803.2030v1

[8] S. Giddings, et. al., "Quantum Aspects of Gravity", Proc. APS Summer Study on Particle and Nuclear Astrophysics and Cosmology in the Next Millenium, Snowmass, Colorado, June 29 - July 14, 1994, arXiv:astro-ph/9412046v1

[9] S. Hawking "Information Loss in Black Holes", arXiv:hep-th/0507171v2

[10] M. Nielsen and I. Chuang, “Quantum Computation and Quantum Information”, Cambridge, England: University of Cambridge Press, 2000

[11] J. Hartle, “Gravity: And Introduction to Einstein’s General Relativity”, San Francisco: Addison Wesley, 2003

[12] *Black Hole Information Bet.* `http://www.theory.caltech.edu/~preskill/info_bet.html`

[13] “Microscopic Black Hole Production in TeV-Scale Gravity”, *Talk presented at XXX ITEP Winter School of Physics, Moscow, Russia, February 2002*, arXiv:hep-ph/0205024v3