

The Black Hole Information Paradox and Quantum Information Theory

Tracey E. Tessier
Department of Physics and Astronomy
University of New Mexico
Albuquerque, NM 87131-1156, USA

April 19, 2000

Abstract

Stephen Hawking's theory of black hole evaporation predicts the evolution of an initially pure quantum state to a mixed state. This result contradicts the fundamental assumption of unitary evolution in quantum mechanics. Further, in light of the information theoretic interpretation of entropy as a measure of missing or hidden information about a system, Hawking evaporation may be viewed as an information-destroying process. This paper reviews the black hole information paradox, and analyzes the more popular 'solutions' from the perspective afforded by quantum information theory. The resolution of the information loss problem promises, at the very least, to provide a unified treatment of thermodynamics and black hole mechanics, and may very well yield revolutionary insights into fundamental quantum processes.

1 Introduction

Black holes are some of the most enigmatic objects in the universe. They are the stuff of science fiction, where brave explorers venture, never to return to the universe that was once their home. They are also the source of one of the most paradoxical predictions in modern physics, the black hole information loss problem. The mere fact that the word paradox appears in the description of the problem implies that this is fertile ground for fundamental discovery. This word is used only when there is disagreement between well worn theory and experiment (although in this case there is no paradox, nature is always right), or between two or more accepted theories. In the current situation the discrepancy is between quantum mechanics, which predicts the unitary evolution of an initial pure state to a final state that is also pure, and Stephen Hawking's theory of black hole evaporation, which purports to evolve an initial pure state to a mixed state.

The revealed inconsistency between these two theories provides us with an excellent opportunity to further our understanding of natural processes. It seems obvious that any resolution of the paradox will, at the very least, result in a unified treatment of thermodynamics and black hole mechanics. There is also the possibility of gaining revolutionary insight into fundamental quantum physics. Finally, since gravitational and quantum effects are undoubtedly both important in the study of Planck-sized black holes, in the words of Ulf Danielsson and Marcelo Schiffer[1], "It might even be that the information paradox is our best clue to the elusive quantum gravity theory."

The purpose of this paper is to provide an overview of the black hole information paradox from the point of view afforded by quantum information theory. We will proceed by first introducing the basic properties of black holes, stressing the remarkable analogy between the known laws of thermodynamics and those of black hole mechanics. The essential ideas of quantum information theory needed for an analysis of this problem will then be presented, and applied to quantum mechanical black holes. Next, the popular 'solutions' to the paradox will be evaluated using quantum information theoretical concepts. As we will see, there is no definitive resolution. However, the application of quantum information theory does help to demonstrate the seemingly paradoxical nature of the problem, and makes predictions about certain properties that any resolution must satisfy. Finally, we suggest certain paths for further research which look especially promising.

2 Black Hole Fundamentals

The simplest definition of a *black hole* is a region of space-time from which all signals, including those propagating at the speed of light, cannot escape. They result from the gravitational collapse of massive bodies to extremely high densities. Space-time in the surrounding region is warped intensely enough to effectively 'pinch' off from the rest of the universe, becoming causally discon-

nected. This collapsing process continues, and results in the formation of a *singularity*, a point at which the constituent matter is squeezed to zero volume and infinite density. According to the cosmic censorship theorem, these exotic objects do not appear by themselves, *i.e.*, there are no “naked” singularities. Instead, they are always found to be shrouded by an *event horizon*.

The event horizon may be thought of as the boundary of the black hole inside which nothing can escape to the external universe. However, this boundary is geometrical rather than physical in character since, for massive enough black holes, there is minimal curvature at the event horizon. An astronomer, upon crossing the event horizon of an extremely massive black hole, would experience nothing out of the ordinary since the space-time at that point would appear locally flat[2].

The existence of black holes, although this term had not yet been introduced, was first contemplated by John Michell and Pierre-Simon de Laplace in the late 18th century[3, 4]. They considered astronomical bodies with densities so great that their escape velocities exceeded the speed of light. The critical radius (R_s) required for the formation of the black hole, known as the *Schwarzschild radius*, is found by these considerations to be:

$$R_s = \frac{2GM}{c^2}. \quad (1)$$

It follows that any spherical body of mass M confined within the Schwarzschild radius is a black hole.

If we consider the area of the event horizon of a Schwarzschild black hole, we find that it is proportional to the square of the mass of the black hole, *i.e.*, in units in which $G = c = 1$,

$$A = 4\pi R_s^2 = 16\pi M^2. \quad (2)$$

Accordingly, if two Schwarzschild black holes were to merge, the area of the resulting black hole would exceed the sum of the areas of the initial black holes. This is a special case of the more general result, discovered by Stephen Hawking, that the horizon area of an isolated black hole never decreases in any *classical* transformation[3].

The most general black hole predicted by Einstein’s general relativity is given by the *Kerr-Newman solution* discovered in 1965[3, 4]. Remarkably, this solution depends on only three parameters: the mass M , the electric charge Q , and the angular momentum \vec{L} , of the collapsing body. This observation is the source of the famous statement, due to John Wheeler, that “a black hole has no hair” in which to encode information about the matter that comprises it[3]. For example, all details of the shape of an initially non-spherical collapsing body will be quickly dissipated in the form of gravitational radiation as a black hole is created. While this information is not lost, since it is contained in the outgoing gravitational waves, other properties of black hole evolution do seem to destroy information, or at least store it and manipulate it according to fundamentally new physical laws.

3 The ‘Thermodynamics’ of Black Holes

A striking analogy between the properties of classical¹ black holes and the laws of thermodynamics has been developed by Stephen Hawking and Jacob Bekenstein[3]. This section serves to introduce these concepts, which will be extended in section 5 to include the effects of quantum mechanical processes. The following description of the four laws of black hole mechanics, corresponding to the known laws of thermodynamics, is based on the treatments given by J. P. Luminet[3] and P. Majumdar[4].

- Zeroth Law:

Thermodynamics: All parts of a system in thermodynamic equilibrium have the same temperature.

Black Hole Mechanics: The surface gravity (g) of a black hole is the same at all points along the event horizon. Surface gravity plays a role analogous to temperature in black hole mechanics.

- First Law:

Thermodynamics: The change in the internal energy of a system is given by

$$dU = TdS + PdV, \quad (3)$$

where the second term denotes work done on the system.

Black Hole Mechanics: The following differential formula may be obtained for a Kerr-Newman black hole:

$$dM = \frac{g}{8\pi}dA + \vec{\Omega} \cdot d\vec{L} + \Phi dQ, \quad (4)$$

where $\vec{\Omega} = 4\pi\vec{L}/MA$ represents the angular velocity of a point on the event horizon, and $\Phi = 4\pi Qr_+/A$ is the electrostatic potential. Here, r_+ denotes the distance to the event horizon, and is given by

$$r_+ \equiv M + \sqrt{M^2 - Q^2 - L^2/M^2}. \quad (5)$$

The second two terms in equation 4 represent the work done on a black hole. Making the identification $S = \alpha A$, where α is some constant completes the analogy. The validity of this identification is supported by the second law of black hole mechanics which we now state.

¹The constructs of general relativity are necessary for any discussion of black holes. Here, and throughout the remainder of this paper, we use the word classical to denote the absence of quantum mechanical processes.

- Second Law:

Thermodynamics: The entropy of a closed system can never decrease, *i.e.*, $dS \geq 0$.

Black Hole Mechanics: The area of the event horizon of a black hole can never decrease, *i.e.*, $dA \geq 0$.

The association of event horizon area with the concept of entropy is by far the most important of any of the relationships presented in this section. It is therefore worthwhile to look more closely at how this association comes about.

The total *mass-energy of a Kerr-Newman black hole* is given by[3]

$$M^2 = \frac{L^2}{4M_{ir}^2} + \left(\frac{Q^2}{4M_{ir}^2} + M_{ir} \right)^2, \quad (6)$$

where the *irreducible mass* (M_{ir}) is defined as

$$M_{ir} = \frac{1}{2} \sqrt{\left(M + \sqrt{M^2 - Q^2 - L^2/M^2} \right)^2 + L^2/M^2}. \quad (7)$$

The first term in equation 6 corresponds to the rotational kinetic energy of the black hole, while the second term represents the electrical potential energy. Finally, there is a contribution to the total mass-energy due to the 'irreducible' mass term. While it is possible, in principle, to extract the rotational and coulomb energies of a black hole to do useful work, the irreducible energy remains inaccessible to all classical physical processes. In fact, the mass-energy of a Schwarzschild black hole is due solely to the irreducible mass, *i.e.*, the mass that appears in equation 2 may be replaced by M_{ir} . This yields the relationship,

$$M_{ir} = \sqrt{\frac{A}{16\pi}}, \quad (8)$$

which, it turns out, holds in general for all black holes. The irreducible mass, and thus the area of the event horizon, cannot by definition be decreased by any classical process.

This result is just a restatement of Hawking's area theorem presented in section 2. However, this law only holds when considering classical black holes, and will have to be modified slightly when we consider black holes undergoing quantum processes.

- Third Law:

Thermodynamics: The temperature of a system can never be reduced to absolute zero by any finite number of operations.

Black Hole Mechanics: The surface gravity of a system can never be reduced to zero by any finite number of operations.

The implications of the third law of black hole mechanics can be better understood by considering equation 5 for the distance to the event horizon (r_+). If we require real solutions, then the values that the mass, electric charge, and angular momentum may take are constrained by the relation

$$Q^2 + L^2/M^2 \leq M^2. \quad (9)$$

If we now let $Q = 0$, which holds in general for any astronomical-sized black hole,² the requirement becomes

$$L \leq M^2.$$

The extremal solution $L = M^2$ corresponds to the vanishing of the surface gravity on the event horizon due to ‘centrifugal’ forces[3]. The requirement specified by equation 9 is simply another form of the cosmic censorship theorem presented in section 2.

4 Introduction to Quantum Information Theory

Before embarking on a systematic discussion of the black hole information paradox, we need a solid understanding of what we mean by information. Specifically, we need to identify how information is encoded in physical systems, and what it means for information to be destroyed. The scientific study of information is an active field of research, and is intimately connected with the thermodynamic ideas presented in the last section. It is to this discipline that we now turn in the hopes of gaining insight into these questions.

4.1 What is Information?

The purpose of this section is to address two fundamental questions: what is information, and why should physicists be interested in it. The answer is simple: information is physical. It is something (quantifiable) that is encoded in the physical state of a system. It is therefore expected that information and its processes conform to certain ‘laws’ analogous to, and ultimately determined by, the physical laws governing the system in which the information is encoded.

Claude Shannon, working at the Bell Telephone Laboratory in the late 1940’s, laid the foundation of classical information theory while attempting to optimize the rate of transmission of coded messages across certain communication channels[5]. As a result of his efforts, Shannon was able to demonstrate the existence of a profound analogy between the thermodynamic entropy of a

²Due to the relative strength of the electromagnetic interaction as compared to the gravitational interaction, an astronomical black hole (which realistically does not form in a vacuum, but rather in a region occupied by charged particles), will neutralize itself quite rapidly[3].

system, and the information encoded in the system. Specifically, he showed that the average information content of a message ensemble (X) containing n different messages with associated probabilities of occurrence (p_i) is given by[5]:

$$H(X) = - \sum_{i=1}^n p_i \log_2 p_i \text{ bits.} \quad (10)$$

We refer to $H(X)$ as the *Shannon entropy* of the ensemble X in view of the fact that it is identical in form to the well known equation for the Gibbs entropy of a thermodynamic system. This result leads to the interpretation of the thermodynamic entropy of a system as a measure of missing or hidden information, since it represents the number of microstates that the system may inhabit after all macroscopic data has been taken into account[6]. Accordingly, an increase in entropy corresponds to a decrease in our state of knowledge about a system. Mathematically, this relationship has the form

$$\Delta I = -\Delta H. \quad (11)$$

Consider, for example, a situation in which we have an air-tight bottle, filled with smoke, and sitting in a corner of a room. If we remove the cap at time $t = 0$, the smoke will begin to exit the bottle and dissipate into the surrounding air. Accordingly, the thermodynamic entropy of the system increases until equilibrium is again established. This corresponds to a loss of information regarding the positions of the smoke particles. Before the cap was removed we could make the definite statement that all of the smoke particles were confined to the volume of the bottle (say 1 litre). Now the most we can say is that the smoke particles exist somewhere in the much larger volume of the room (obviously a weaker statement). The increase in the phase space available to the smoke particles is accompanied by an increase in entropy, and a corresponding decrease in our state of knowledge about the system.

While these ideas have proven themselves to be extremely powerful in many areas of research, we come to a dead end when discussing the black hole information paradox solely in terms of classical information theory. As described in section 3, the matter that comprises the irreducible mass of a black hole, and thus the information encoded in it, is inaccessible to any classical process. There is no paradox in the classical situation. The information encoded in the irreducible mass simply remains stored in the black hole. Although information is not destroyed by a classical black hole, it does remain forever inaccessible.

Of course black holes, like every other physical system, are fundamentally quantum mechanical in nature. A proper treatment of the information loss problem can only be performed within the context of quantum mechanics, requiring a quantum theory of information. The remainder of this section will be devoted to introducing the basic ideas of this theory, including: quantum superpositions, the concept of entanglement, the use of density matrices to represent incomplete knowledge, and the reversible, unitary evolution of a closed quantum system. These ideas will then be applied to quantum black holes in section 5.

4.2 Quantum Superpositions

The fundamental unit of classical information is the *bit*, which may be represented by a physical system that can be in either of two states, encoding a zero or a one respectively. Similarly, the fundamental unit of quantum information is the qubit, which may again be encoded in a two-level system, *e.g.*, a spin 1/2 particle, where a zero is represented by spin down along the z-axis $|\downarrow\rangle$, and a one by spin up along the z-axis $|\uparrow\rangle$. However, the states of a qubit are not limited only to the basis states. Rather, they include a continuously infinite number of superpositions of the two basis states. The most general state of a qubit is given by

$$|\psi\rangle = c_0 |\downarrow\rangle + c_1 |\uparrow\rangle, \quad (12)$$

where c_0 and c_1 are complex in general.

Although an arbitrary qubit may exist in any of an infinite number of superposition states, these states are extremely fragile since the result of any measurement is limited to the eigenstates of the up-down basis (when measuring spin along z). Obtaining the result $|\uparrow\rangle$, for instance, reveals only that the initial superposition state possessed some nonzero amplitude along $|\uparrow\rangle$. It yields no other information about the magnitude of this amplitude. Hence, a great deal of information encoded in the state of an arbitrary qubit is inaccessible in the sense that it cannot all be extracted by any conceivable measurement.

4.3 The No-Cloning Theorem

One immediate consequence of the fragile nature of quantum states, and the resulting inaccessibility of quantum information, is the *no-cloning theorem* which states that it is impossible to copy the state of an unknown quantum system. Following T. Spiller[7], the validity of this statement may be shown as follows.³

Consider a unitary operator U that successfully copies the basis states $|\downarrow\rangle$ and $|\uparrow\rangle$ onto some ancillary qubit initially prepared in the state $|\alpha\rangle$, *i.e.*,

$$U |\downarrow\rangle |\alpha\rangle = |\downarrow\rangle |\downarrow\rangle \text{ and } U |\uparrow\rangle |\alpha\rangle = |\uparrow\rangle |\uparrow\rangle.$$

If we now apply U to a superposition state of the form of equation 12, we obtain

$$U |\psi\rangle |\alpha\rangle = c_0 |\downarrow\rangle |\downarrow\rangle + c_1 |\uparrow\rangle |\uparrow\rangle. \quad (13)$$

However, the desired state in this situation is not given by equation 13, but by

$$|\psi\rangle |\psi\rangle = c_0^2 |\downarrow\rangle |\downarrow\rangle + c_0 c_1 (|\downarrow\rangle |\uparrow\rangle + |\uparrow\rangle |\downarrow\rangle) + c_1^2 |\uparrow\rangle |\uparrow\rangle. \quad (14)$$

It is apparent that there is no way to obtain equation 14 using the unitary operator (U), since it is unable to generate cross-terms of the form $|\downarrow\rangle |\uparrow\rangle$

³For simplicity, we neglect the internal state of the physical system that is actually performing the copying function in this proof. The argument may be generalized to include the state of the copier with no change in the results.

and $|\uparrow\rangle|\downarrow\rangle$. Thus, even if a unitary operation exists which will copy certain states with perfect accuracy, the same operation will not work in general for an arbitrary state. This is just another example of how the unavoidable disturbance of fragile quantum states during measurement renders a great deal of the information encoded in a quantum system inaccessible.

4.4 Entanglement of Quantum States

The quantum superpositions discussed in section 4.2 require only one qubit to manifest themselves. When we consider systems of more than one qubit, a new phenomenon, known as *entanglement*, becomes possible. Entanglement occurs when two initially separate quantum systems interact, and effectively lose their individual identities. This is evidenced by the fact that knowledge of the states of the component subsystems is insufficient to determine the state of the resulting composite system. Accordingly, we find that at least some of the information that used to reside in either of the two subsystems is now stored in correlations between the subsystems.

Following S. Popescu and D. Rohrlich[8], we consider a bipartite system consisting of two spin 1/2 particles A and B . If we take as our basis:

$$\left\{ |\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, |\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\},$$

then the basis of the combined system (AB) may be written as

$$\{ |\downarrow\rangle_A |\downarrow\rangle_B, |\downarrow\rangle_A |\uparrow\rangle_B, |\uparrow\rangle_A |\downarrow\rangle_B, |\uparrow\rangle_A |\uparrow\rangle_B \},$$

where we have used the notation, $|\downarrow\rangle_A |\downarrow\rangle_B = |\downarrow\rangle_A \otimes |\downarrow\rangle_B$.

If we now let A and B interact, then the most general state of our bipartite system is given by

$$|\psi\rangle_{AB} = \sum_{i,j} c_{ij} |i\rangle_A |j\rangle_B, \quad (15)$$

where $i, j \in \{\downarrow, \uparrow\}$, and the c_{ij} 's encode the joint probability amplitudes for simultaneously finding particle A in state $|i\rangle$ and particle B in state $|j\rangle$. A special case of equation 15 occurs when the probability amplitudes for A and B are independent of each other, *i.e.*, $c_{ij} = c_i^{(A)} c_j^{(B)}$ for all i and j . States of this form are known as product states since they may be written in the factored form:

$$|\psi\rangle_{AB} = \sum_{i,j} c_i^{(A)} c_j^{(B)} |i\rangle_A |j\rangle_B. \quad (16)$$

A state that cannot be written in the form of equation 16 is called *entangled*, *e.g.*,

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{2}} (|\downarrow\rangle_A |\downarrow\rangle_B + |\uparrow\rangle_A |\uparrow\rangle_B) \quad (17)$$

is a maximally entangled state. The effect of the interaction resulting in equation 17 is to entangle the particles A and B so that they do not possess well-defined states independent of each other. We can no longer treat A and B as two independent objects. Rather, we must think only in terms of the joint system AB .

Entangled systems are characterized by *correlations* that exist between their component subsystems. The state described by equation 17 for example, represents a perfect correlation between the subsystems A and B , *i.e.*, if A is measured to be in the state $|\downarrow\rangle$ ($|\uparrow\rangle$), then we know with certainty that B will also be found in the state $|\downarrow\rangle$ ($|\uparrow\rangle$).

The entanglement of physical systems is a physical process. Thus, we might expect information theoretic consequences, and this is indeed the case. Information that used to be associated with the individual particles is now encoded in the correlations between A and B . This is verified by the fact that local operations on A (or B) alone are unable to extract very much of the information that is encoded in the joint system.

If we consider separately, any subsystem of an entangled system, we find that the subsystem is described by a mixed state, even though it is part of a larger pure state. It is the mathematical representation of these mixed states to which we now turn. As we will see, the density matrix formalism is capable of characterizing the result of any conceivable local measurement on a subsystem of a larger entangled system.

4.5 Mixed States and Density Matrices

All of the systems that we have considered so far have been described by pure states. Indeed, it may be that nature always allows us to find a pure state description of a composite system if we look hard enough, *i.e.*, if we identify all relevant parts of the system, and all interaction terms of the appropriate Hamiltonian. Nevertheless, in situations where we have incomplete information about a composite system, we still require a mathematical formalism that will capture our state of knowledge regarding the part of the system to which we have access. The density matrix formalism used to describe *mixed states* is well suited for this task.

The concept of a density matrix is usually introduced as a mathematical construct that represents an ensemble of possible pure states with associated probabilities of occurrence (p_i). In this picture, the *density matrix*, defined as

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|, \quad (18)$$

may be interpreted as the average projection operator of all constituent states $|\psi_i\rangle$. However, the concept of a density matrix is also extremely useful in representing subsystems of a larger, pure entangled system. As we saw in the previous section, information may be stored not only in the physical components of a system, but also in the correlations between two or more components. When one

of these component subsystems is considered apart from the rest of the system, we have incomplete knowledge of the overall pure state. The local object is described by a mixed state which is well-represented by the density matrix formalism.

The *reduced density matrix* (ρ_A) for a subsystem A of a larger entangled pure state system (AB), is obtained by 'tracing out' the remainder of the system, *i.e.*,

$$\rho_A = \text{tr}_B |\psi\rangle_{AB} \langle\psi|_{AB}. \quad (19)$$

Accordingly, ρ_A fully characterizes the result of any conceivable local measurement performed solely on A . However, it does not represent the fact that a measurement of the state of subsystem A affects, in a nonlocal way, the result of a subsequent local measurement performed on B , and vice-versa.

Consider, for example, a system of two spin 1/2 particles described by the pure entangled state given by equation 17,

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{2}} (|\downarrow\rangle_A |\downarrow\rangle_B + |\uparrow\rangle_A |\uparrow\rangle_B),$$

where we are allowed to make local measurements on particle A , but particle B is locked in a vault, precluding us from making any measurements on it. (This thought experiment has obvious implications for the problem of black hole information loss, and will be discussed further in section 5).

Calculating the reduced density matrix of A from equation 19, we obtain

$$\rho_A = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

The effect of our ignorance regarding half of a maximally entangled state is that the half that we do have access to is described by a maximally mixed state, even though it is a component of a larger pure state. Thus, ρ_A yields the probabilities to find A in various states, but it says nothing about the fact that measuring particle A to be in the state $|\downarrow\rangle$ ($|\uparrow\rangle$) ensures that particle B will also be found in the state $|\downarrow\rangle$ ($|\uparrow\rangle$). The reduced density matrix contains no information about the correlations that exist between A and B . It is in this sense that entanglement enables a system to encode more information than just that stored locally in the component subsystems.

The nonlocal character of the correlations described above might appear to provide a method of superluminal, even instantaneous, communication through the use of a space-like separated, entangled system. However, this is not a possibility. Quantum mechanics remains compatible with the predictions of special relativity since no information can be transmitted using only local operations on a subsystem of a larger entangled system. Although we will not formally prove this, the essential argument as outlined by Popescu and Rohrlich[8], goes as follows.

Consider once again the entangled bipartite system AB described by equation 17, this time with a space-like separation between particle A and particle

B. Suppose our friends, Alice and Bob, have access to particles *A* and *B* respectively. Then the outcome of any local measurement that Alice(Bob) can perform on *A*(*B*) will depend only on the reduced density matrix ρ_A (ρ_B), which contains no information about the system-wide correlations. The reduced density matrix of system *B* is therefore independent of (and unaffected by) any local measurement performed on *A* and vice-versa. Since Alice can only measure particle *A* and Bob can only measure particle *B*, superluminal communication is not possible.

4.6 The von Neumann Entropy of a Quantum System

In 1935, even before the work of Shannon, John von Neumann defined the concept of the entropy (*S*) of a mixture of quantum states. Following R. Jozsa[9], we recreate the derivation from an information theoretic point of view.

Assume in analogy with the classical case, which yields the Shannon entropy of a system, that we have an ensemble of *n* different quantum states $|\psi_i\rangle$, each of which has an associated probability of occurrence (p_i). By equation 18, this ensemble is completely represented by

$$\rho = \sum_{i=1}^n p_i |\psi_i\rangle \langle \psi_i|. \quad (20)$$

Now, if $\lambda_1, \dots, \lambda_n$ are the eigenvalues of ρ , and $|\lambda_1\rangle, \dots, |\lambda_n\rangle$ are the corresponding eigenvectors, then given that $\{\lambda_1, \dots, \lambda_n\}$ forms a classical probability distribution, we may write equation 20 as[9]

$$\rho = \sum_{i=1}^n \lambda_i |\lambda_i\rangle \langle \lambda_i|. \quad (21)$$

Assuming that we have chosen the $|\lambda_i\rangle$'s in equation 21 to be mutually orthogonal, which we may always do, this situation is analogous to the classical case in that the states are all perfectly distinguishable. Thus, we are justified in calculating the Shannon entropy of this ensemble using equation 10 (where we use $S(\rho)$ in place of $H(X)$ to denote the von Neumann entropy):

$$S(\rho) = - \sum_{i=1}^n \lambda_i \log_2 \lambda_i. \quad (22)$$

Finally, since $\lambda_i \log_2 \lambda_i$ are the eigenvalues of $\rho \log_2 \rho$, and we are summing these eigenvalues, equation 22 may be recast in the form[9]

$$S(\rho) = -\text{tr}(\rho \log_2 \rho). \quad (23)$$

Equation 23 gives the von Neumann entropy for any mixture of quantum states described by the density matrix ρ . In analogy with the information theoretic interpretation of the Shannon entropy, $S(\rho)$ can also be taken to be a

measure of missing information, this time in quantum systems.⁴ This association is borne out, for example, in the observation that the von Neumann entropy for a pure state vanishes while mixed states have finite entropies. Finally, we note that the von Neumann entropy obeys

$$0 \leq S(\rho) \leq \log_2 n, \quad (24)$$

with the maximum value of $S(\rho)$ occurring for a uniform distribution of quantum states where all of the p_i 's are equal to $1/n$.

4.7 Unitary Evolution

Quantum mechanics predicts that all closed systems undergo unitary evolution described by the *Schrodinger equation*

$$H |\psi\rangle = i\hbar \frac{\partial |\psi\rangle}{\partial t}, \quad (25)$$

where H is the *Hamiltonian* of the system. Integration of equation 25 yields an equation for the time evolved state $|\psi(t)\rangle$:

$$|\psi(t)\rangle = U |\psi(0)\rangle, \quad (26)$$

where $U = e^{-\frac{i}{\hbar} \int_0^t H dt'}$ is the unitary time evolution operator[7]. According to equation 26, quantum mechanics predicts the evolution of a closed system, represented by a pure initial state, to a pure final state. It is vital that all of the interactions that our physical system undergoes be represented in H if the above considerations are to hold. If these conditions are met, it follows that $\frac{\partial S}{\partial t} = 0$ as we let the system evolve in time, since $S(\rho)$ for any pure state density matrix is zero. This is equivalent to the statement that the information we have about the system in question is conserved, in particular, it does not decrease with time.

This concludes our foray into quantum information theory. We now return to a discussion of black holes, which we view from this point on as quantum mechanical objects. This approach results in the prediction at the heart of the information paradox, Stephen Hawking's theory of black hole evaporation.

5 Quantum Black Holes

Stephen Hawking showed in the early 1970's that black holes emit thermal radiation in a quantum evaporation process[4]. The details of this process are unimportant for our purposes. Suffice it to say that it is somewhat analogous to pair-production of elementary particles in a polarized vacuum, where one of

⁴This interpretation is not on quite the same footing as in the classical case since, according to Holevo's theorem, $S(\rho)$ only gives an upper bound on the information content of a system[9]. However, this distinction is unimportant for our purposes.

the created particles escapes to infinity, while the other is captured by the black hole[3].

This radiation is predicted to occur at a temperature proportional to the surface gravity of the black hole, further strengthening the thermodynamic analogy. Specifically, the temperature (T) of the Hawking radiation is given by[3]

$$T = \frac{\hbar}{2\pi} g.$$

However, the process of Hawking evaporation violates the second law of black hole mechanics presented in section 3, which states that the event horizon area is a nondecreasing quantity. This observation leads us to define the total entropy of the universe to be

$$S_{universe} = S_{bh} + S_{ext}.$$

Here, S_{bh} is the entropy associated with the area of the event horizon, and S_{ext} is the entropy of the space-time exterior to the black hole. The generalized version of the second law

$$dS_{universe} \geq 0$$

appears to remain valid[3].

The effects of Hawking evaporation lead to paradoxical conclusions regarding the information encoded in the matter that goes into the creation of a black hole. Imagine for instance, that a black hole is created by the gravitational collapse of a massive body described by a pure quantum state. If Hawking's calculations (which predict that the outgoing radiation is thermal) are correct, and if the black hole evaporates completely, then the information originally encoded in the matter has been destroyed by the black hole evaporation process[1, 11]. This is because thermal radiation is described by a mixed state, which has a nonvanishing entropy. The entropy-information relationship,

$$\Delta I = -\Delta S,$$

implies that information is irretrievably lost. Considerations such as these lie at the heart of the black hole information paradox.

There are many difficulties with any of the myriad proposed solutions to the information loss problem, and we do not attempt to discuss them all. In particular, we neglect ideas such as the nucleation of baby universes to which the missing information escapes, and the existence of certain types of "quantum hair" in which information can be encoded in subtle ways. More information on these and other proposals can be found in [1, 11].

The remainder of this paper will be devoted to an analysis, using the tools of quantum information theory developed in section 4, of two main classes of proposed solutions: the existence of correlations in the emitted Hawking radiation, and the existence of black hole remnants. Of course, it may be that information is in fact destroyed by black holes, forcing us to give up the notion of unitary

evolution. This would be the most radical possible outcome, necessitating a fundamental reformulation of quantum mechanics. However, it is difficult to envision what form such a replacement theory might take without the guiding light of experimental data. Therefore, we do not pursue this idea any further.

5.1 Hawking Radiation

The most obvious place to begin our search for the information initially encoded in the matter comprising a black hole is the emitted Hawking radiation. Perhaps Hawking's calculations are an invalid approximation, and the outgoing radiation is actually in a pure (or nearly pure) state. This would imply that the radiation is uncorrelated (or very weakly correlated) with the internal state of the black hole.

Now consider a pure state system falling into a black hole. If we assume that the information encoded in the in-falling matter remains intact as it crosses the event horizon, which is locally flat for a large enough black hole, then some mechanism must exist to transfer this information to the Hawking radiation. If it were possible to copy the incoming information at the event horizon, it could be made available to the outgoing radiation, and the paradox averted. However, the no-cloning theorem excludes such a possibility.

The reason that the unitary operation, defined by equation 13, is unable to copy an unknown quantum state, is that it is actually too good at making copies. The resulting correlations are always perfect. Perhaps this is the mechanism by which to resolve the paradox. After all, if the correlations are always perfect between the radiation and the black hole, then knowledge of the emitted radiation yields the internal state of the black hole. There is then no need to trace out the black hole subsystem. Unfortunately, this mechanism does not work either, as demonstrated by Danielsson and Schiffer[1].

Recalling equation 13,

$$U |\psi\rangle |\alpha\rangle = c_0 |\downarrow\rangle |\downarrow\rangle + c_1 |\uparrow\rangle |\uparrow\rangle,$$

written in the basis

$$\left\{ |\downarrow\rangle = \frac{1}{\sqrt{2}} (|\rightarrow\rangle + |\leftarrow\rangle), |\uparrow\rangle = \frac{1}{\sqrt{2}} (|\rightarrow\rangle - |\leftarrow\rangle) \right\},$$

we have

$$c_0 |\downarrow\rangle |\downarrow\rangle + c_1 |\uparrow\rangle |\uparrow\rangle = \frac{c_0 + c_1}{2} (|\rightarrow\rangle |\rightarrow\rangle + |\leftarrow\rangle |\leftarrow\rangle) + \frac{c_0 - c_1}{2} (|\rightarrow\rangle |\leftarrow\rangle + |\leftarrow\rangle |\rightarrow\rangle). \quad (27)$$

It is apparent from equation 27 that the correlations are, in general, not perfect in all bases. Only when $(c_0 = -c_1)$ $c_0 = c_1$ is the (anti-)correlation perfect[1]. We must, in the general case, still trace over the unknown (and inaccessible) state of the black hole. Therefore, assuming that information survives its transit

across the event horizon, a violation of causality seems to be necessary in order to return the information stored in the in-falling matter to the external universe.

Perhaps then, the information encoded in the in-falling matter does not survive the event horizon crossing as one would naively expect. Let $\{|i\rangle\}$ denote a basis for the initial quantum state of the in-falling matter, and consider the situation in which the internal state of the black hole is completely uncorrelated with the outgoing radiation. Paraphrasing an argument due to Preskill[11], we require that each basis state $|i\rangle$ evolve according to

$$|i\rangle \longrightarrow |i\rangle_{bh} |j\rangle_{ext},$$

where $\{|i\rangle_{bh}\}$ and $\{|j\rangle_{ext}\}$ are the basis states of the black hole and the Hawking radiation, respectively. Next, considering a superposition of the basis states $|i\rangle$ which evolves according to

$$\sum_i c_i |i\rangle \longrightarrow \sum_{i,j} c_{ij} |i\rangle_{bh} |j\rangle_{ext}, \quad (28)$$

we find that correlations will, in general, exist between the internal state of the black hole and the outgoing radiation. It is only when the above may be written as a product state (see equation 16) that the two subsystems remain uncorrelated. This implies (in the case where we consider the Hawking radiation to be in a pure state), that the state $|i\rangle_{bh}$ must be the same in each term, *i.e.*, the black hole must exist in a unique state. More generally, for Hawking radiation that is nearly pure, the internal state of the black hole must be nearly unique[11].

This result is in agreement with the idea that “black holes have no hair”, and are therefore incapable of storing the information associated with the in-falling matter. One possibility is that some sort of ‘bleaching’ process occurs upon crossing the event horizon, which removes all information about the state of the in-falling matter, and makes it available to the outgoing radiation. The challenge facing proponents of this type of process as a possible resolution to the information loss problem is to conceive of a reasonable mechanism to accomplish this task.

It is also quite possible that Hawking’s semi-classical calculations are very nearly correct after all, and strong correlations exist between the internal state of the black hole and the outgoing radiation. This situation is somewhat analogous to that of the entangled system considered in section 4.5, where one particle is locked in a safe, preventing us from making any measurements on it. Since the internal state of the black hole is inaccessible to measurement, the remainder of the system (the Hawking radiation), is described by a mixed state. The joint system, black hole plus external universe, may still be described by an entangled pure state. However, if the black hole evaporates completely, then the radiation has become the entire system, and a pure state has evolved into a mixed state[11]. This is the source of the paradox.

Assuming Hawking’s calculations to be valid, it seems our only recourse is to consider what happens at the Planck scale, when it is certain that semi-classical approximations no longer hold. Two possible outcomes are the cessation of

black hole evaporation, and the slow leakage of information back to the universe, corresponding to stable and meta-stable black hole remnants respectively.

5.2 Black Hole Remnants

In this section we consider the possibility that some as yet unknown law of Planck scale physics halts the black hole evaporation process, resulting in the formation of miniature black hole remnants. The idea is that, as long as the black hole does not evaporate completely, unitarity may be saved. The information might remain forever stored in (but not be destroyed by) these remnants. Alternatively, physics at the Planck scale might allow remnants to emit their own form of radiation, returning the information to the universe. There are many quantum field theoretic arguments against these stable and meta-stable remnants on the grounds that they would experience infinite production rates[11, 12]. The problem is that, since the initial mass of (and, thus, the initial amount of encoded information in) a black hole is arbitrary, an infinite number of different remnant species are possible. Even if the actual production rate of each individual species is very small, the infinite number of different species still results in an overall infinite production rate. Leaving these objections aside, we consider what constraints quantum information theory applies to remnants if they do exist.

The following derivation, due originally to Danielsson and Schiffer[1], demonstrates that, if the Hawking radiation is very nearly thermal, then the black hole must possess a very large number of internal states. We begin by defining the information content (I) of a system as:

$$I = I_{max} - S(\rho),$$

where $I_{max} = S(\rho)_{max}$. This definition obviously has the right character at the extreme values, since $S(\rho) = S(\rho)_{max} \Rightarrow I = 0$, and $S(\rho) = 0 \Rightarrow I = I_{max}$. Using equation 24, which implies that $I_{max} = \log_2 N$, and the definition of von Neumann entropy (equation 23), yields

$$I = \log_2 N + tr(\rho \log_2 \rho). \quad (29)$$

Now, suppose that the bipartite system consisting of a black hole and the external universe is initially described by a pure state, *i.e.*,

$$|\psi\rangle = \sum_{i,j=1}^{N_{bh}, N_{ext}} c_{ij} |i\rangle_{bh} |j\rangle_{ext},$$

where N_{bh} and N_{ext} are the dimensions of the black hole and external universe Hilbert spaces respectively. Obviously, we expect $N_{ext} \geq N_{bh}$.

Applying equation 29, the information contents of the black hole and the external universe are given by

$$I_{bh} = \log_2 N_{bh} + tr(\rho_{bh} \log_2 \rho_{bh}),$$

and

$$I_{ext} = \log_2 N_{ext} + \text{tr}(\rho_{ext} \log_2 \rho_{ext}),$$

where the corresponding reduced density matrices are found from equation 18 to be

$$\rho_{bh} = \sum_{m,i,k}^{N_{bh}, N_{ext}} c_{im} c_{km}^* |i\rangle_{bh} \langle k|,$$

and,

$$\rho_{ext} = \sum_{m,j,l}^{N_{bh}, N_{ext}} c_{mj} c_{ml}^* |j\rangle_{ext} \langle l|.$$

The total information content of the entire system may then be defined as,

$$I_{tot} = I_{bh} + I_{ext} + I_{bh,ext}, \quad (30)$$

where $I_{bh,ext}$ represents the information encoded in the system correlations.

The above results, by themselves, do not provide very much in the way of constraints on the information content of a black hole remnant. However, Danielsson and Schiffer[1] have shown that a lower limit given by

$$I_{ext,min} = \log_2 N_{ext} - \log_2 N_{bh} \quad (31)$$

can be placed on the information content of the Hawking radiation. This lower limit corresponds to the case where the emitted radiation is most nearly thermal, and contains very little information.

Assuming Hawking's prediction to be correct, *i.e.*, $I_{ext} \approx 0$, at least until the black hole has evaporated to the Planck scale, equation 31 implies that N_{bh} must be of the same order as N_{ext} . The number of internal states of a black hole remnant must, therefore, be extremely large. Note that equation 30 does not necessarily require that all of the information actually be stored in the internal states of the remnant. Much of it may reside in correlations with the emitted radiation. However, even if the bulk of the information is stored in the correlations, the remnant must still have the capacity to encode half of this information[1]. We may therefore, in analogy with classical statistical mechanics, regard S_{bh} as the logarithm of the number of black hole microstates. However, in light of the conjecture that "black holes have no hair", the nature of these microstates is unclear.

Finally, we consider the existence of meta-stable remnants. At the Planck scale, where Hawking's calculations are surely invalid, it may become possible for information to be returned to the universe by allowing the black hole remnant to emit its own form of radiation. This would entail subtle correlations between quanta emitted earlier and that being emitted now. This is in some

ways analogous to burning a book and reconstructing it from a history of the emitted radiation[11].

One might think that allowing these meta-stable remnants to eventually evaporate in this way would solve the problem of infinite production rates. Unfortunately it does not. The problem is that an arbitrary black hole remnant can, in principle, contain an enormous amount of information. However, at the Planck scale, there is little remaining energy ($E \sim m_{pl}$, where m_{pl} is the Planck mass) with which to convey this information back to the external universe. Accordingly, a large number of low energy particles must be used. This turns out to be a very slow process. Simple estimates[11, 12] of the lifetime (τ) of a black hole remnant yield

$$\tau \sim \left(\frac{M}{m_{pl}} \right)^4 t_{pl}, \quad (32)$$

where M is the initial mass of the black hole, and t_{pl} is the Planck time. Equation 32 predicts that a remnant of a black hole formed from the mass of a small mountain would decay in a time larger than the current age of the universe[12]. Accordingly, these remnants are effectively stable as far as the production rate arguments go. Any resolution of the information paradox in terms of black hole remnants has to deal successfully with the problem of infinite remnant species.

6 Summary and Future Directions

No satisfactory resolution to the black hole information paradox has yet been found. In fact, any solution might well require us to give up long cherished ideas, or force a fundamental reformulation of physical laws. The quantum information theoretical concepts we have introduced here, while not providing any obvious answers, do help to demonstrate the paradoxical nature of the problem, and to highlight the difficulties that any of the proposed resolutions have to overcome.

The no-cloning theorem, for example, precludes a conservative solution to the problem in which the information is somehow copied at the event horizon, and then made available to the outgoing radiation. Assuming that information encoded in a physical system remains intact upon crossing the event horizon of a black hole, a non-local transfer of information, acting on a macroscopic scale, seems to be required to return the information to the external universe via the emitted Hawking radiation. Alternatively, if we give up this assumption, we are left with the possibility that some kind of ‘bleaching’ process strips away the information from a physical system at the event horizon, making it available to the outgoing radiation.

If Hawking’s calculations are correct, and there is very little information contained in the radiation, it seems that we must rely on the existence of fundamentally new physical laws, which manifest themselves at the Planck scale, if we are to save unitarity. Even if this occurs, it seems that completely halting the evaporation process, or letting the information eventually leak back slowly,

leads to theories of black hole remnants plagued by an infinite number of possible remnant species. Finally, we must consider the possibility that black holes really do destroy information, forcing the abandonment of unitary evolution in quantum mechanics.

The remarkable analogy between the laws of thermodynamics and those of black hole mechanics seems to offer the best clue to an eventual resolution of the information paradox. Thinking along these lines, one might ask if it is possible to understand the entropy of a black hole (S_{bh}), in terms of underlying microstates, in analogy with thermodynamics and statistical mechanics. In other words, do mass, electric charge, and angular momentum define the macrostate of a black hole in the same way that pressure, volume, and temperature define the macrostate of a thermodynamic system? If this interpretation is viable, then the black hole entropy provides a measure of our lack of knowledge of the underlying microstates that correspond to a given macrostate. However, if black holes really have no hair, *i.e.*, no degrees of freedom by which to distinguish various microstates, then the character of these microstates is unclear. In any event, it is difficult to anticipate the level of insight that would be gained if the two sets of laws, whose derivations could hardly be more different, could be shown to be different manifestations of the same underlying principle.

It is reasonable to expect that any resolution of the black hole information paradox will result in the unification described above. Further, there seems to be no question that the information loss problem offers a tremendous opportunity for fundamental discovery. In the words of John Preskill[11], "it seems increasingly likely to me that it is as hopeless to reconcile relativistic quantum mechanics with black hole evaporation as it would have been to understand the spectrum of black body radiation using classical physics. The information loss paradox may be a genuine failing of 20th century physics, and a signal that we must recast the foundations of our discipline."

References

- [1] U. H. Danielsson and M. Schiffer. *Quantum Mechanics, Common Sense, and the Black Hole Information Paradox*, gr-qc/9305012, 1993.
- [2] L. Thorlacius. *Black Hole Complementarity*, astro-ph/9412046, 1994, pp. 15-17.
- [3] J. P. Luminet. *Black Holes: A General Introduction*, astro-ph/9801252, 1998, pp. 1-16.
- [4] P. Majumdar. *Black Hole Entropy and Quantum Gravity*, gr-qc/9807045 v3, 1998, pp. 1-4.
- [5] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, 1962.

- [6] L. Boltzmann. *Theoretical Physics and Philosophical Problems*, Reidel Publishing Co., Boston, 1974.
- [7] T. Spiller. *Basic Elements of Quantum Information Technology*, printed in *Introduction to Quantum Computation and Information*, World Scientific Publishing Co., Singapore, 1998, pp. 1-28.
- [8] S. Popescu and D. Rohrlich. *The Joy of Entanglement*, printed in *Introduction to Quantum Computation and Information*, World Scientific Publishing Co., Singapore, 1998, pp. 29-48.
- [9] R. Jozsa. *Quantum Information and Its Properties*, printed in *Introduction to Quantum Computation and Information*, World Scientific Publishing Co., Singapore, 1998, pp. 49-69.
- [10] V. Erolov and L. Thorlacius. *Quantum Aspects of Gravity*, astro-ph/9412046, 1994, pp. 2-5.
- [11] J. Preskill. *Do Black Holes Destroy Information?*, hep-th/9209058, 1992.
- [12] S. B. Giddings and L. Thorlacius. *Introduction to the Information Problem*, astro-ph/9412046, 1994, pp. 6-8.