Black Holes and Information

Ronaldo Ortez
Physics
raortez@ucdavis.edu

Abstract

The information paradox was first introduced in 1974 by Stephen Hawking. Since then, it has fostered really fruitful work in hopes of being able to preserve the conservation of information which when thought about appropriately underlies all of quantum mechanics. In this paper, I go through many of the key points which highlight why this has been such a point of contention and some of the remarkable ways it fundamentally challenged how we think of information from a physical perspective.

0. Introduction

In 1800s during the peak of classical physics, physicists would likely have thought of the universe as a giant mechanical clock. This was a deterministic universe that operated by some arrangement of gears that stepped the universe forward in time. The goal of the physicist was to then describe the different gears and how they all fit together. This notion doesn't seem to be that different from the one in information theory where the universe is said to perform a computation to iterate into subsequent observable states. It seems a mere relabeling has taken place, so naturally one wonders what insights are added by this new framework. The 1900s brought with it quantum mechanics and the fall of the old mechanical universe because now the universe was seen to be inherently probabilistic. One was forced to abandon the notion of known states evolving deterministically to other known states. We would need a machine that could have multiple potential states governed by some probability distribution of states. What the physicist must then content himself by simply uncovering the appropriate probability distribution for the system and rules which produce them. It is here where the information theory paradigm is much better suited to handle a fundamentally probabilistic description. The machines of IT are characterized by the transition matrix that describes the probabilities a state transitions from one moment in time to the next. Additionally, a system can be characterized by an intrinsic randomness that will prevent the observer from ever predicting outcomes better than what the asymptotic entropy allows. One must then find a machine synchronizes to this value, and with a little more work one can in principal find the optimal predictive machine. The idea that such a machine uniquely exists provides a nice argument that this is fundamentally how the universe works, and this is how I understand the idea that the universe computes and how it surpasses the mechanical clockwork framework.

Physicists were able to take inspiration from this to suggest that information should be the fundamental book keeping device, and the dynamics are in the details of how the optimal machine reproduces the observed probability distribution. However, because probabilities arise not due to our ignorance but because of inherent uncertainty of the system, meanings had to be reworked a bit. the process described by the optimal Still, physicist use information slightly differently. If we use the heuristic notion that information is a measure of correlation between states that encodes the dynamics (the optimal machine) rather than the randomness (our ignorance) then this more closely approximates how physicist use the notion of information. We can relate Boltzmannian entropy to this notion by arguing entropy is a measure of our ignorance of a system. If we had perfect knowledge of the system then there would only be 1 possible microstate, and thus no entropy, but since our ignorance grows with the number of possible microstates, so does entropy. However, since we know quantum mechanics demands an inherently uncertain description such that even in the case of perfect knowledge we are left with probabilistic descriptions, physicist have instead opted to use entanglement entropy which is a measure of degree of an entanglement between states of a system. The Von Neuman entropy describes the "entanglement entropy" of the system, which is different from the classical entropy of the system. If there is no entanglement in a system, then it reduces to the classical case, and if the entanglement entropy is zero, we are in a maximally entangled state. Further the entanglement entropy is conserved under unitary transformation which will give us a nice way to talk about how information in quantum systems.

This line of thinking motivates a conservation law for information. And it was Hawking in 1974-1975 who first proposed that such a conservation principle could not hold in all the places of the universe. This sparked a debate that continues to even today. This paper is primarily aimed at following how this debate. First, I outline the thermodynamics of black holes which tells us fundamental things about information. Then I introduce the naïve version of the information paradox which states that information shouldn't be conserved. I then motivate a stronger version of this paradox known as the entropy subadditivity paradox. I then quickly discuss possible ways out of the paradox concluding with a quick description for the best framework that exists for resolving this paradox.

1. Black Hole Thermodynamics

In the early 70's it was worked out that the most general black hole (BH) solution to stationary, asymptotically flat, nonsingular regions outside a BH were described by a "no hair" theorem.

> In standard gravity theory the most general stationary black hole exterior is described by the KN solutions with M, Q and J as its only parameters -Wheeler

Though there are other physically interesting cases that for example don't have spherical symmetry the main point stands: for all their potential complexity within the horizon, only a few parameters are required to describe the dynamics exterior to a BH. Or, to understand the "no hair" terminology if black holes had hair, then in principle one could say more about what is going in the interior, and there are varying versions of the "no hair" theorem. The "no hair" theorem is reminiscent of thermodynamics which can describe a large amount of microscopic complexity in equilibrium with only a few variables. So, Beckenstein began to think about the thermodynamics of a black hole and in particular he reasoned that not only should a BH have entropy, it should have very large entropy. It should have entropy to satisfy the 2nd Law, since entropy of the universe can't decrease even when an object with a given entropy crosses the event horizon. This lead Beckenstein to propose the generalized second law (GSL).

(1)

If we understand entropy as a measure of our ignorance of a system, then our ignorance is as large as possible because the inner part of the horizon is causally disconnected from us, outside observers. We cannot distinguish between various microstate arrangements of a BH.

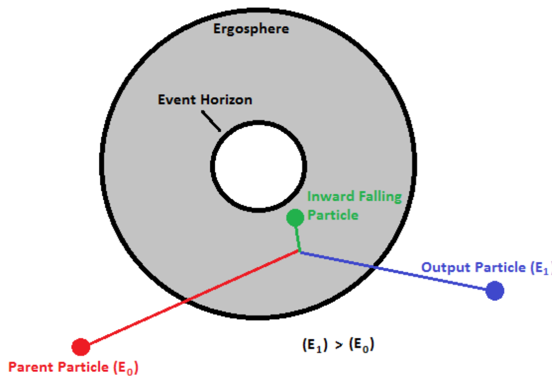
Next it was shown by Penrose and Hawking and others that area of a BH should never decrease in what is called area censorship. Since this is the same functional form for entropy, Beckenstein postulated that entropy should be proportional to the surface area of the BH as follows[6]:

(2)

Where is the Compton wavelength of the Plank mass.

Hawking would later calculate the temperature of a BH and was therefore able to work out the constant of proportionality. But we can do a better job of defining the thermodynamics of a BH by deriving an analogous 1st law that applies to BH's. However, rather than going through the full form involving all the possible BH parameters I will describe a slightly simpler BH that consists only mass and its angular momentum and motivate how one can up with the desire expression. We consider a Penrose process which describes how the mass of rotating black hole can vary (both increase and decrease). With this we can think about varying the mass parameter (equivalently energy) parameters and follow the typical approach of variational thermodynamics and consider how other variables would vary in response.

Penrose Process



To begin, consider the following scenario: an observer is falling into a rotating black hole with an object in his hand. Before reaching the event horizon he will cross into the ergosphere which is a region extending beyond the event horizon responsible for producing the frame-dragging effect. Both the observer and the object start out with some positive energy. Once inside the ergosphere, the observer can aim his projectile in such a way that the observer will be flung from the interior of the ergosphere to the outside with more energy than both he and the object entered with. This must mean that for energy to be conserved the still infalling object must now have negative energy. This is possible because objects can go from behaving timelike near the outer boundary of the ergosphere to behaving spacelike near the event horizon which results in the energy changing sign (from positive to negative). By this process, one can rob the BH of its rotational energy and therefore rob it of some mass. We can work out the limits to how much the angular momentum should change for a given mass change (all done in natural units). [10]

(3)

Where characterizes the ergosphere by being the rotational frequency of the ergosphere at the horizon boundary. To get a dependence on the area of the BH, we use the notion of the irreducible mass defined as:

(4)

which is nothing more than the mass of the Swarzschild BH in the limit where all the angular momentum goes to zero. We can now differentiate to work out how M_{irr} is affected by changes in the BH's mass or angular momentum to give the following:

(5)

Where is the angular momentum per unit mass. Of course, the is related to the area so we can rewrite this in terms of area and clean things up a bit to give the 1st law:

(6)

Where is called the surface gravity and is a measure of the curvature near the event horizon. We can see that in the limit of large M, the curvature goes to zero. Thus, one can cross the event horizon without having any local means of knowing.

2. Information Bounds

Its worth pausing for a moment to highlight some of the conclusions we can draw from the BH 1^{st} Law. First there is the following correspondence between the thermodynamic 1^{st} law and the BH 1^{st} law.

(7)

Second the entropy of a BH is really large. To first order of magnitude the entropy of a million solar mass BH is larger than the entropy of all the matter content of the universe. So, a maximally entropic universe is a universe filled just with BH's. This large quantity of entropy suggest that this is as compact as we could store information. Any further and the spacetime will collapse to form a BH.

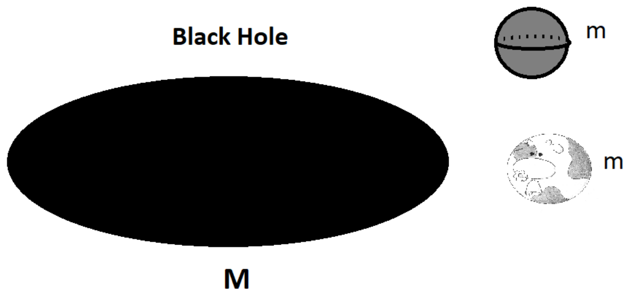
Another puzzling observation is that this information limit is proportional to the area of the BH and not the volume. If we take this seriously it seems that in 3 dimensions information cannot scale faster than the area enclosing the volume. This is very counterintuitive as we can imagine making up 3-D space by a lattice with Plank length spacing (Compton wavelength of Plank mass) then we certainly expect the entropy to scale with the number of sites which in turn scales with the volume. It must then be the case that many of these degrees of freedom are redundant and one needs to only consider the bounding surface area to entirely describe the information content. This goes by the name of the holographic information bound.

However, the holographic information bound is very large, and perhaps there are better more realistic bounds that have more to do with what we might actually hope to achieve. Again for context:

> "An object the size of a music compact disk would be allowed by the bound an information capacity of up to 10^{68} bits. Present technology can only store 10^{10} bits on it, and is expected to improve only by a few orders of magnitude" [6]

Instead of a deriving a bound by allowing the object to collapse into a BH, we can think of the following scenario. Consider a composite object very far away from a BH(Schwarzschild) that is allowed to slowly and allow it to slowly fall slowly towards the BH. We let the time it takes to fall be such that the mass lost by the BH through Hawking radiation is the same as the mass of the infalling object. This ensures that there is no difference between the BH mass M at its initial and final times. Because the entropy of the BH only depends on M, this means that the BH entropy didn't change. The entropy of the object must then be less than the entropy of the Hawking radiation. In order to satisfy the 2^{nd} law. One then finds a more reasonable bound on the information able to be stored on a compact disc to be 10^{40} bits. Still a very large bound.

3. Naïve Information Paradox



We are now able to pose the naïve version of the information paradox. Suppose we have 2 spherical objects, one is a plane sphere with mass m and the other is a spherical with patterns etc. similarly with mass m. Now we allow both to cross the event horizon. Because of the no-hair theorem as external observers the only information we have access to is increased mass of the black hole by m. Of course, this means we can no longer distinguish between the cow and all its features and our featureless spherical test mass.

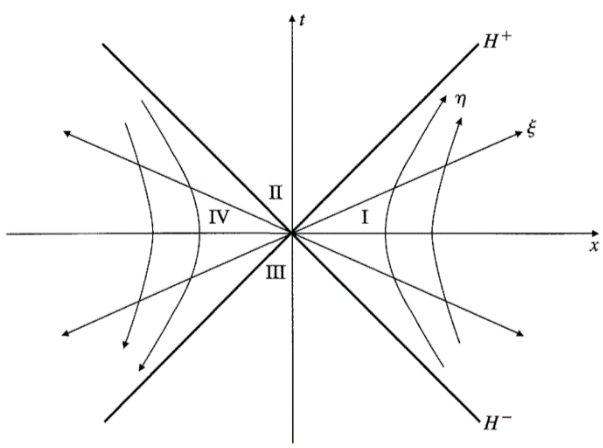
This alone doesn't constitute any information paradox as we would just have to accept that there are places in the universe that information can go which we won't have access to. Not to different from the idea that as the universe continues to expand, the observable universe is gradually shrinking. Nothing prohibits this. The paradox results from there being a temperature associated with the BH which suggests that it is in fact radiating particles at a rate inversely proportional to the mass, so for stellar black holes, the initial temperature is low resulting in a low power radiated by Stefan Boltzmann law, but as the BH shrinks the power radiated increases until the BH can evaporate entirely. Now the question becomes, what happened to all the information. The Black hole radiation obeys a thermal plank spectrum meaning that the photons are entirely uncorrelated with one another. It will be impossible to tell if the BH was filled with mostly spherical cows or from nice spherical test masses. This may not initially seem problematic if we take a classical perspective. Certainly, there are many chaotic systems where reversibility is impossible, or stated otherwise the idea that multiple states lead to same final state. In fact, the second law notoriously prohibits reversibility. This is a pill forced upon us by the nonzero ignorance we bring to every macro system.

However, if we take a quantum perspective, and we remember that unitarity is one of fundamental requirements of all quantum theories which means that we should always be able to achieve reversibility. To speak more precisely about this, it is often described as the phenomenon of pure states evolving to mixed states. The result of such an evolution would violate unitarity because certain entangled states must necessarily cease to exist, and the probability distribution would no longer sum to one. Similarly, it would require that multiple quantum states evolve to the same next state. Something that is easy to show is forbidden by quantum mechanics.

4. Black Hole Radiation

To get a better grasp on the precise nature of the paradox it is instructive to see the source of Hawking's radiation. The idea is that if we apply quantum field theory to a spacetime like that near a BH's horizon, we get interesting effects where particles seem to emerge from the vacuum state. To do this properly, we do need a quantum theory of gravity because if we were to restore the units of the Hawking temperature, we would notice the dependence on which tells us this is a quantum effect. Absent such a tool, Hawking applied QFT to a smooth background spacetime, which in principle should be a reasonable thing. Frequently, Hawking radiation is explained by arguing that entangled pairs of particles will randomly come into existence near the horizon of a BH. One imagines that one of the particles randomly ventures across the horizon while the other stays outside. Since nothing can ever escape back through the horizon, there will appear to be a net flux of particles that weren't able to annihilate with their antimatter pair. This is an entirely heuristic explanation that isn't borne out by calculations. There are several problems with calculations most notably being that this is a highly localized description, whereas the event horizon is a highly nonlocal quantity. Though Hawking introduced this explanation as a means of perhaps getting a more intuitive understanding of what might be going on, he cautioned against taking it too seriously. Rather than going through Hawking's original ray tracing argument, which took the form of a scattering effect, I will very briefly explain how the same thermal radiation results from a more universal effect called the Unruh effect.

The essential point is that in a curved background, the positive frequency modes of your field theory cannot be uniquely specified, so in general when we define the vacuum using an annihilation operator on empty Fock space, the annihilation operator will be symmetric under Bagliobov transforms which corresponds to a linear combination of creation and annihilation operators in another frame, and so, when we calculate the vacuum expectation value we will instead see that nonzero number of particles will exist in the vacuum state. To illustrate this, we will approximate a very smooth curvature by a uniformly accelerated reference frame (valid by the equivalence principle).



Using Rindler coordinates we can describe what uniform acceleration looks like relative to inertial Minkowski coordinates (x,t) as shown in the diagram above. That is, relative to a stationary observer, an observer moving with constant acceleration will move along parabolas that asymptote to H^+ (corresponding to a null light ray). Again, uniform gravitational fields are nice approximations for the very smooth spacetime we expect to exist near the event horizon of large BH (recall dependence of surface gravity on mass)

In a quantum field theory, we find the modes that are allowed to exist by solving the equation of motion that results from a particular choice of Lagrangian. We consider a free field that satisfies the Klein-Gordon equation and as usual we get two solutions with coefficients that are set by the relevant boundary conditions. To quantize the field the coefficients become creation and annihilation operators that serve the function of creating or destroying modes defined by (plane waves for free fields) having a particular momentum. Thus, to define a field at a point we sum over all the relevant modes.

(8)

Now we will want to define these modes in a frame that is uniformly accelerating, that is, in terms of our Rindler coordinates. Thus in terms of Minkowski operators that produce an empty vacuum state in the inertial frame we can write the normalized Rindler operators as:

(9)

We see that the Rindler operators are linear combinations of positive and negative Minkowski frequency modes. The variable refers to the constant acceleration of the frame. An interesting thing to compute is the number operator, (number of particles in a given state) applied to the vacuum. This gives:

(10)

Here we see that a thermal plank spectrum arises from the vacuum expectation value of the number operator. Thus, an observer moving with constant acceleration relative to a flat Minkowski vacuum will see a thermal spectrum of particles. Using the equivalence principle, we can understand it more easily this way, suppose we are in a space ship and we are uniformly accelerating at a constant rate. Then our vacuum will contain thermal spectrum of particles, and if we accelerate to arbitrarily high then we along with our space ship will burn up. This same kind of acceleration is produced by the gravitational field near the event horizon of a black. The result is a thermal spectrum of particles that appear to radiate outward towards us, the stationary observers.

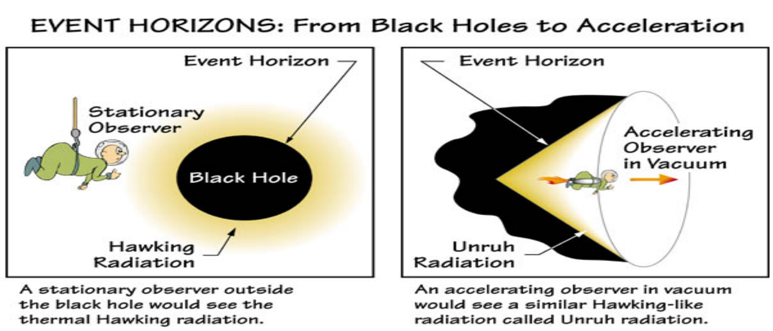


Fig. 1

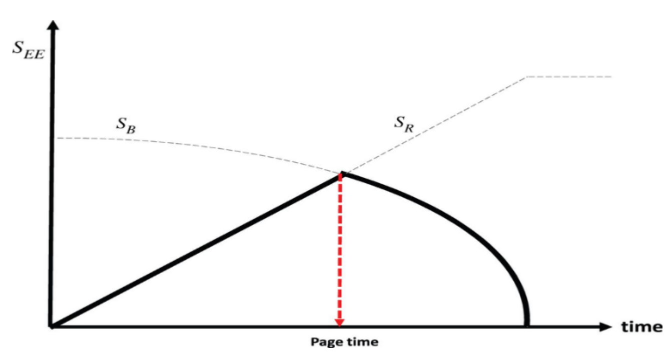
So, to summarize, because the mass of the a sufficiently large BH gently curves spacetime near the event horizon, the result of this curvature is to create a thermal spectrum of particles as seen by an outside observer. Again, what we would like to understand is how the information of the matter that falls into the black hole might be secretly encoding in this thermal spectrum of particles. Given what we've learned it's not at all obvious how any correlations can be established between the global curvature of the black hole and the local process of something falling into the BH. Although to be fair, one of the things we've failed to consider is the backreaction of the radiation on the curvature itself. It seems to be worthwhile to switch into thinking about how the spacetime itself might be correlated from point to point.

5. Information Subadditivity Paradox

An instructive analogy is as follows:

> A cold piece of coal is illuminated by a laser beam. The system is in a pure state: coal in its ground state and beam in a coherent state (analogous to the sphere of superfluid). Experience will tell us the coal will heat up and radiate (black hole forming and radiating). The beam is interrupted (no matter is thrown into black hole after its formation) The coal cools while radiating thermally (Hawking radiation). The coal cools totally and returns to its ground state (black hole evaporates), which is, of course, pure. [6]

We can state the information paradox as a pure state evolving into a mixed state. This is true in both these cases, we begin with a pure states (coal in ground state) and we conclude with a pure state(coal back in its ground state) but also have a collection of thermal photons which seems to be mixed. No one suggest that information has been lost in this case. Rather the information is encoded in tiny correlations between the thermal photons at different times. This example also illustrates how difficult it can be to determine differences between mixed states and pure states because any course inspection of the photons would look mixed. More generally, it can be shown that even quantum states can appear thermal (eigenstate thermalization hypothesis, [7])



As a way out of the naïve IP, we can argue that despite Hawking radiation appearing thermal, it's possible that correlations still exist between the outgoing radiation and modes near the inner surface of the horizon. We can use entanglement entropy to quantify the correlations that must exist in the outgoing radiation. Further, it must be the case that the BH entropy must serve as an upper limit to the entanglement entropy. As the BH begins to radiate, entanglement entropy of Hawking radiation must increase, but the since Hawking radiation causes the BH to evaporate the BH entropy must decrease. Eventually the BH will evaporate which means the entanglement entropy must also be zero. Thus, it doesn't seem possible for correlations in the Hawking radiation to carry away the requisite correlations to preserve information with the BH.

More precisely, we can state the entropy subadditivity paradox as follows: We consider a large Schwarzchild BH that evaporates and so decreases in size. After emitting some radiation, we label the following 3 subsystems: A) the old thermal radiation B) the quanta being emitted just outside the horizon C) the quanta emitted just inside the horizon. Recall that entanglement entropy is S_{EE}=0 for a maximally entangled states, and becomes maximal when the system becomes nearly thermalized (very small entanglement between the states). We can deduce the following for the system:

1) $S_{AB}<S_A$, This says that the combined system AB is more entangled than the A by itself. That is, there is more entanglement between the quanta just being emitted and the old quanta than there is the old quanta itself. This is essentially the claim we are relying on to rescue the notion that information isn't lost.
2) $S_{BC}=0$, This says that the ingoing and outgoing quanta are maximally entangled, meaning by measuring one mode you immediately know the properties of the other set of modes.
3) $S_B=S_C>0$, this follows form the fact B and C are both individually thermal, so they should have nonzero entanglement entropy.

It is known that quantum entanglement entropy satisfies the strong subadditivity property given by:

(1)

Now lets put in what we know, by 2) we get:

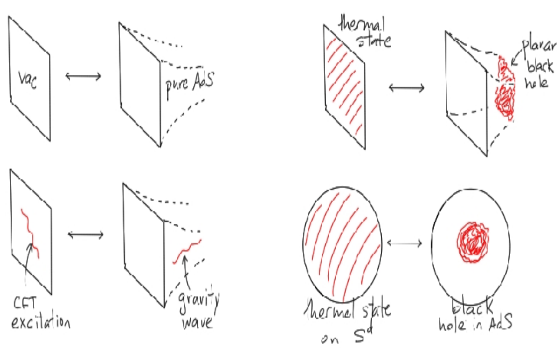
(2)

By looking at 1) we see that this relation cannot hold. Moreover, by using relation 3) we violate this condition to at least 1st order, that is, by a significant amount. Almheiri, Marolf, Polchinski, Sully (AMPS) have argued that to avoid this paradox, we must give up the S_{BC}=0 condition by positing the existence of a firewall that is able destroy the entanglement of modes near the horizon. The existence of this firewall would mean that the curvature near the horizon wouldn't be smooth contrary to what GR predicts and as a result none of our effective field theories used to probe these regions are viable.

6. Ads/CFT

Here, I will briefly mention ways out of the information paradox (IP). The first proposition is that as the BH begins to evaporate the information left behind gets compactified. At some point our notions of a gently curved spacetime are no longer valid and maybe it is here when the information would eventually be released, or its even been suggested that quantum gravity would permit a plank sized object to hold 10^{40}bits of information which is comparable to a BH entropy. Its possible that the black hole is instead a wormhole and the information comes out the other end. There is of course the firewall proposal mentioned earlier or fuzzballs which suggest that vacuum regions form near the horizon of the infalling material which would produce sizeable corrections to the field theory. There are even suggestions that very long wavelength gravitons can produce an infinite number of conserved charges outside of the BH. There would then be an infinity of ways to preserved information. However, if we stick to the idea that the correlations are somehow preserved in the outgoing radiation, or rather that just because the outgoing radiation appears to be thermal it actually is thermal.

Indeed, this is a daunting prospect because we would be looking for very very tiny correlations. Let's go back to observation that we must somehow be able to describe a huge number of microstates within a BH. Where are these microstates? Well, Leonard Susskind has really been the advocate of the holographic principle which got its initial inspiration by noticing that the entropy bound was given by the area and not the volume. He argued that this should be a universal feature, that the total information within the system can be described by the degrees of freedom contained only the bounding region. This idea has found really sure footing since the AdS/CFT conjecture which allows one to describe a gravitational theory in antiDesitter space by a Conformal Field theory on the boundary. I shamelessly took these drawings from Raamsdonk's lecture notes. In AdS/CFT one establishes a dictionary between the relevant observables of one theory in the bulk and the other on the boundary. The key point is that the CFT theory is unitary and is a quantum theory, so it should show how the correlations are preserved, but being a tightly coupled theory, the details have not yet been worked out, but it seems the growing consensus is that we'll be able to recover the correlations albeit probably having to sacrifice locality to get there.



One of the really remarkable ideas that has come from this endeavor is that spacetime might arise from entanglement. In fact, one may be to connect spacetimes merely by entangling the degrees of freedom or tear them apart by disentangling them. This also addresses the puzzle of why you would only need the boundary degrees of freedom to describe the bulk. That is, if spacetime is maximally entangled then the bulk degrees of freedom are redundant and don't hold any new information.

References

[1] D. Harlow, “Jerusalem Lectures on Black Holes and Quantum Information,” arXiv: 1409.1231v4 [hep-th] 24 Sep 2015.

[2] Suvrat Raju, “Quantum Aspects of Black Holes and Information Paradox”, www.suvratraju.net/classes [Lecture notes]

[3] M. V. Raamsdonk, “Building up spacetime with quantum entanglement,” arXiv:1005.3035v1 [hep-th] 17 May 2010

[4] D. Marolf, “The Black Hole information problem: past, present, and future,” arXiv:1703.02143v2 [gr-qc] 25 May 2017

[5] S. D. Mathur, “The information paradox: A pedogogical introduction,” arXiv:0909.1038v2 [hep-th] 25 Jan 2011

[6] J. D. Beckenstein, “Black holes and information theory,” arXiv:quant-ph/0311049v1 9 Nov 2003

[7] M. Srednicki, “Chaos and Quantum Thermalization” Phys Rev E 50 (1994) 888, [arXiv:cond-mat/9403051]

[8] S. Carrol, “Spacetime and Geometry: An introduction to General Relativity,” 2014

[9] A. Aguirre, B. Foster, Z. Merali, “It From Bit or Bit From It?” 2015

[10] A. Almheiri, D. Marolf, J. Polchinski, J. Sully, “Black Holes: Complementarity or Firewalls?” arXiv:1207.3123[hep-th] 13 Apr 2013