

# Black Hole Remnants and Classical vs. Quantum Gravity

Forthcoming in *Philosophy of Science*

Peter Bokulich<sup>†‡</sup>  
University of Notre Dame

---

Belot, Earman, and Ruetsche (1999) dismiss the black hole remnant proposal as an inadequate response to the Hawking information loss paradox. I argue that their criticisms are misplaced and that, properly understood, remnants do offer a substantial reply to the argument against the possibility of unitary evolution in spacetimes that contain evaporating black holes. The key to understanding these proposals lies in recognizing that the question of where and how our current theories break down is at the heart of these debates in quantum gravity. I also argue that the controversial nature of assessing the limits of general relativity and quantum field theory illustrates the significance of attempts to establish the proper borders of our effective theories.

---

**1. Introduction.** Any attempt to formulate a full theory of quantum gravity faces the difficulty of establishing the dynamical behavior of the theory when spacetime itself is subject to quantum uncertainties. Typically, our only justification for offering any picture of truly quantum gravitational processes comes from arguing that in certain limits we can trust the models

---

<sup>†</sup>Department of Philosophy, 336 O'Shaughnessy Hall, University of Notre Dame, Notre Dame, IN 46556; e-mail: Bokulich.1@nd.edu

<sup>‡</sup>The content of this paper constitutes a substantial portion of Chapter IV of my forthcoming dissertation, "Horizons of Description: Black Holes and Complementarity." Thanks to Jim Cushing, Don Howard, Laura Ruetsche, and the Chicago Area Philosophy of Physics Group for helpful comments and discussions.

provided by classical general relativity (GR), low energy quantum field theory (QFT), and their hybrid offspring, semi-classical gravity. However, we recognize that these models will break down in certain regions, specifically in regions where interactions on the scale of the Planck energy are likely to occur. We then face the challenge of using our effective models to the extent that we can, marking off the areas where these models can no longer be trusted, and making reasonable guesses at what takes place beyond those borders.

A rich variety of responses to this challenge can be found in the debate surrounding the Hawking information loss paradox. Both the argument supporting the paradox (Hawking 1976) and the replies to it rely on semi-classical methods and guesses about the behavior of systems in the quantum gravitational regime, but there is considerable controversy over the success of these various arguments. Belot, Earman, and Ruetsche (BER) argue in their 1999 overview of responses to Hawking's argument that a class of proposed solutions to the paradox, namely, those that postulate the existence of black hole remnants, do not provide an adequate answer to Hawking's challenge. These proposed solutions, argue BER, either fail to rebut the argument driving the paradox or they only escape by denying the existence of the explanandum, that is, by denying the existence of black holes.

Here I argue that this criticism of the remnant proposals is misguided on three important counts. First, BER either fail to consider, or misrepresent, remnant scenarios that postulate the existence of residual black holes. This is a prominent response to Hawking's paradox that clearly escapes their argument. Second, they neglect or misunderstand the limits of the classical description of the spacetime postulated by the remnant theorists, and thereby miss the substance of the remnant proposal. Third, the definition of black holes they appeal to is overly restrictive

in a quantum gravitational context and leads them to misjudge a legitimate reply to Hawking's argument.

My purpose in rebutting BER's arguments is not to champion the remnant proposal, which faces serious physical challenges unrelated to BER's charges, but is instead to clarify the nature of the debate in which this scenario has been proposed. It seems that the root of BER's misreading of the proposal lies in their overreliance on theoretical descriptions that are called into question by this debate. Because the question of precisely how and where these descriptions fail is at the heart of this controversy, one cannot hope to understand force of Hawking's argument, or the responses to it, unless one recognizes the significance of arguments over the correct limitations of our classical and semi-classical theories.

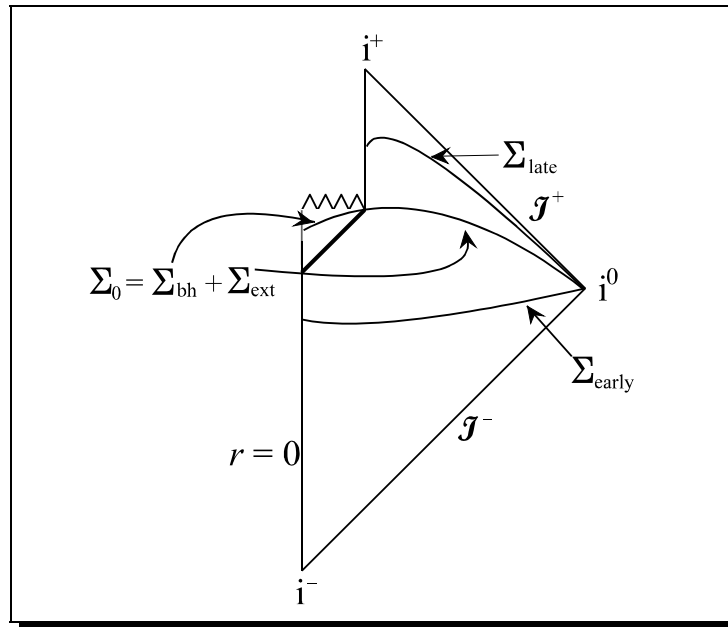
**2. The Information Loss Paradox.** The basis of the information loss paradox is Hawking's 1974 discovery that the theory of quantum fields in curved spacetime implies that black holes will give off thermal radiation at a temperature inversely proportional to their mass. Conservation of energy implies that the black hole will lose mass through this process, and if nothing halts the evaporation, the black hole will eventually disappear completely. This premise of complete evaporation is essential to Hawking's argument, and is denied by some of the remnant proposals that we will be considering in the next section.

Figure 1 is a Penrose conformal diagram representing the formation and complete evaporation of a black hole.<sup>1</sup> The shaded area represents the black hole, the region that is not in

---

1. The Penrose conformal diagrams that are used throughout this paper represent spherically symmetric spacetimes in which we have suppressed two angular dimensions. Thus each point of

the causal past of future null infinity  $\mathcal{J}^+$ . The causal curves that pass through the black hole cannot be extended to  $\mathcal{J}^+$  but instead are assumed to terminate in a curvature singularity indicated by the jagged line. Although the existence of a black hole classically implies that a spacetime is singular, we will see that many remnant theorists deny that black holes harbor curvature singularities.



**Figure 1**

the diagram represents a sphere. All of spatial infinity ( $r = \infty$ ) is represented by the single point  $i^0$ , at which all “time slices” (i.e., spacelike surfaces without edge) end. Past and future temporal infinity are represented by the points  $i^+$  and  $i^-$ , respectively. While spatial and temporal distances will be systematically distorted as one approaches these points at infinity, the null rays (light paths) of the spacetime are always represented as lines of  $45^\circ$ , regardless of the curvature of the spacetime. By clearly indicating the lightcone structure of a spacetime, these diagrams capture the possible causal (luminal or sub-luminal) paths of the spacetime. For example, they allow us to represent a black hole as a region from which nothing can escape to future null infinity,  $\mathcal{J}^+$ , without traveling faster than light, i.e., on a path more than  $45^\circ$  from vertical.

BER employ algebraic QFT to formulate their argument for information loss because of worries about non-unitarily equivalent representations of quantum fields in curved spacetimes. However, in the interests of brevity and clarity, and because nothing in my discussion will depend on this choice, I will use the more standard Hilbert space vocabulary. Consider the three spacelike hypersurfaces,  $\Sigma_{\text{early}}$ ,  $\Sigma_0$ , and  $\Sigma_{\text{late}}$ , on which we can define our global state at a “time.”  $\Sigma_0$  is the latest hypersurface that includes the black hole and is composed of the section inside the black hole, which we will call  $\Sigma_{\text{bh}}$ , and the section outside, labeled  $\Sigma_{\text{ext}}$ . We assume that our quantum field begins in a pure state,  $\rho_{\text{early}}$ , at an early time, that is, on our hypersurface  $\Sigma_{\text{early}}$ . Quantum mechanics tells us that this state will then unitarily evolve into another pure state,  $\rho_0$ , on  $\Sigma_0$ .

At this point in the argument we appeal to the validity of local QFT, which claims that  $\rho_0$  will be a state in  $\mathcal{H}_{\text{bh}} \otimes \mathcal{H}_{\text{ext}}$ , where these two Hilbert spaces will be defined by local field operators on  $\Sigma_{\text{bh}}$  and  $\Sigma_{\text{ext}}$  respectively. We can then define the component states  $\rho_{\text{bh}}$  and  $\rho_{\text{ext}}$  by tracing  $\rho_0$  over the degrees of freedom associated with  $\Sigma_{\text{ext}}$  and  $\Sigma_{\text{bh}}$ . These component states,  $\rho_{\text{bh}}$  and  $\rho_{\text{ext}}$ , will be pure states if and only if  $\rho_0$  is an uncorrelated, factorizable state, that is, if and only if

$$\rho_0 = \rho_{\text{bh}} \otimes \rho_{\text{ext}}.$$

However, it seems that we should expect there to be quantum correlations between the field on  $\Sigma_{\text{bh}}$  and the field on  $\Sigma_{\text{ext}}$ . Such correlations are quite easy to form – for example, the interaction of a pair of particles followed by one of them falling into the black hole will result in a correlated global state – and there seems to be no mechanism to prevent such correlations from forming

between the inside and outside of the black hole. Further, Hawking explicitly showed that the outgoing radiation will be correlated with the state of the quantum field inside the black hole. Given that  $\rho_0$  will be an entangled state, we must conclude that  $\rho_{\text{bh}}$  and  $\rho_{\text{ext}}$  will be mixed states.

A further requirement of locality, often called microcausality, requires all spacelike related observables to commute. This implies that no act represented by a local operator – for example, performing a measurement, or making local changes to the Hamiltonian density – can change the state of the fields outside of the causal future of the region associated with that local operator. BER appeal to microcausality and the fact  $\Sigma_{\text{late}}$  is completely spacelike related to the interior of the black hole to justify their “commutation condition.” This condition claims that any local observables associated with the black hole region, for example, any observable that can be constructed from the local field operators on  $\Sigma_{\text{bh}}$ , should commute with all local observables defined on the late time slice  $\Sigma_{\text{late}}$ . Given the validity of local QFT, we can then conclude that our late time state,  $\rho_{\text{late}}$ , will be independent of the local field values on  $\Sigma_{\text{bh}}$ , and that  $\rho_{\text{late}}$  will unitarily evolve from  $\rho_{\text{ext}}$ .

But we now recall that  $\rho_{\text{ext}}$  is a mixed state, which implies that  $\rho_{\text{late}}$  is also mixed (since unitary evolution preserves the purity or mixedness of states). It therefore appears that the formation and complete evaporation of a black hole cannot be described quantum mechanically. Quantum mechanics describes all evolution as unitary transformations of quantum states, but the evolution described above cannot be unitary since it begins with a pure state,  $\rho_{\text{early}}$ , and ends with a mixed state,  $\rho_{\text{late}}$ . This is Hawking's “paradox.”

**3. Black Hole Remnants.** Many physicists feel that Hawking's conclusion is unacceptable in light of the past success of describing time evolution by unitary transformations of quantum states. Further, there are concerns that the postulated non-unitary evolution may violate locality or energy conservation. Responses to the information loss paradox typically argue that some aspect of the above picture of evaporating black holes will be excluded by a full theory of quantum gravity, and thus we can expect such a theory to retain unitary quantum mechanical evolution. The remnant proposal, in particular, points to the fact that the semi-classical methods used to derive the Hawking effect are clearly invalid when the black hole reaches the Planck mass. This opens the possibility that quantum gravitational processes might safeguard, and perhaps even return, the information contained in the black hole, thus allowing for the unitary evolution of our global quantum state.

If Planckian physics were simply to shut down the Hawking radiation, then we would be left with a very small black hole that would be quantum entangled with the external state of the universe. We might call such an entity a "residual black hole" in recognition of the fact that it retains the essentially spatiotemporal character of a black hole. By contrast, we can consider a scenario that postulates the evolution of a Planck sized black hole into some new object: a "black hole remnant." Such a remnant is typically described as a fundamental particle, of an essentially quantum gravitational nature, that retains all information that falls into a black hole.

If the remnant, or the residual black hole, remains for all future time, then we call it a *stable* remnant. However, there is also the possibility that Planckian physics will allow the information stored in the remnant to return to the external universe. Once all the correlations are passed off to field operators localized outside the remnant, thus securing the purity of the

external state, the remnant can safely pass out of existence. These transient objects are referred to as *long-lived* remnants in recognition of the fact that their finite lifetime will nonetheless have to be quite lengthy, given the potentially immense amount of information that they will have to return, and the meager Planck-sized mass they have to return it with.

The amount of information that these remnants would have to store is the basis of the most substantial physical objection to this proposal. There would appear to be no bound on the number of internal degrees of freedom that a remnant would have to possess. Suppose, for example, that we form a black hole from a large amount of matter that is entangled with some system outside the black hole. We can then wait for the black hole to shrink by giving off thermal radiation that will be uncorrelated with any external system. We then feed in more correlated matter, let the hole shrink, and so on. There seems to be no bound on the number of correlations (the amount of information) that the black hole – and hence the remnant – would have to support. However, this seems to imply that a remnant should have an indefinitely large number of internal states, while only having Planckian mass. This in turn implies that such remnants should be pair produced without bound in background electromagnetic or gravitational fields, which is clearly false. (See, for example, Giddings 1995b, for more details of this argument.) While this is a very serious problem facing the remnant scenario, it is not the reason that BER offer for rejecting it.

**4. Belot, Earman, and Ruetsche's Assessment.** BER's paper first offers a derivation of pure to mixed state evolution, and then offers a taxonomy of proposed solutions to the Hawking paradox



based on which premise of the derivation is denied by a particular solution. My attempt to defend remnant proposals from BER's conclusions will require me to focus on the expected breakdown of local QFT and GR at Planck scales. However, BER explicitly assume the validity of local QFT, and of the classical general relativistic spacetime indicated in Figure 1. We therefore face the question of whether my criticisms actually confront BER's argument, or whether they simply point to a possibly fruitful extension of their project. I wish to set this question to one side here. My goal is not so much to decide whether BER's assessment of the remnant proposal is fair given the project they set out, as to clarify how this scenario should be understood in the context of the debate over information loss.

Given their aim of offering a taxonomy of responses to Hawking's Paradox based on how the response rejects a premise of their argument, what is BER's evaluation of the remnant proposal? They claim that the proposal faces a "fundamental difficulty" which they pose in the following dilemma.

Either remnants are remnants--that is, of black holes--in which case they do not provide for a satisfying resolution of the Hawking paradox, or they are not remnants--at least, not of black holes--in which case they can do nothing to address the problem of *black hole* evaporation. (BER 1999, 216)

The first horn of the dilemma, which will be discussed more fully in the next section, points out that the derivation of pure to mixed state evolution rests solely on the commutation condition, and the claim that there are quantum correlations between the interior and exterior of the black hole (BER's "correlation condition), where this black hole evaporates as indicated in Figure 1. The mere postulation of information retaining remnants, according to BER, does not appear to block any premise used in their derivation, and therefore this response is not "satisfying." On the

other hand, as we shall see in Section 6, one could see the remnant proposal as denying the existence of black holes. But this move, according to BER, merely changes the subject and does not address the concern at issue. Thus either remnants should be grouped with inadequate, confused responses to the paradox, or they should be seen as (apparently uninteresting) denials of the original assumption that black holes exist. I will argue below that the solution offered by remnants is both satisfying and completely relevant to the question of evaporating black holes, regardless of which horn of BER's dilemma we decide to face.

First, however, let us consider the case of residual black holes, that is, Planck scale black holes that no longer Hawking radiate. It is not clear whether BER take their criticisms to apply to this remnant scenario as well, or whether they consider it to be outside the scope of their discussion, as it denies the premise of complete evaporation. This is nevertheless a prominent remnant scenario that clearly deserves our attention.

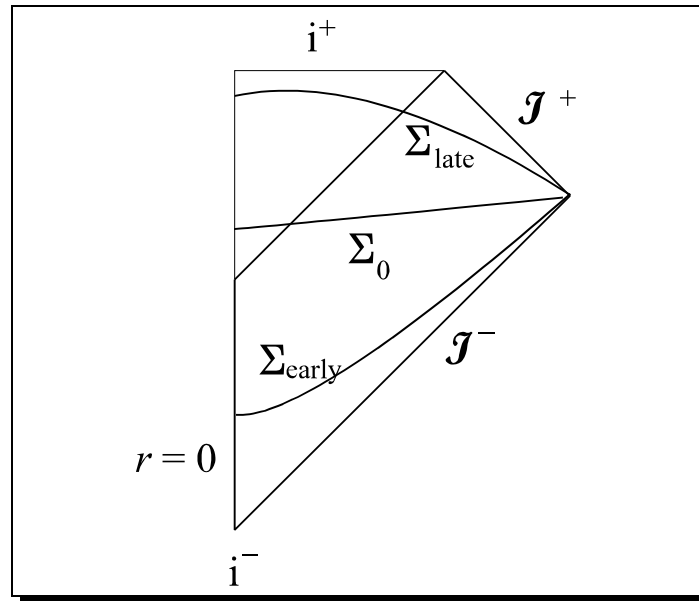


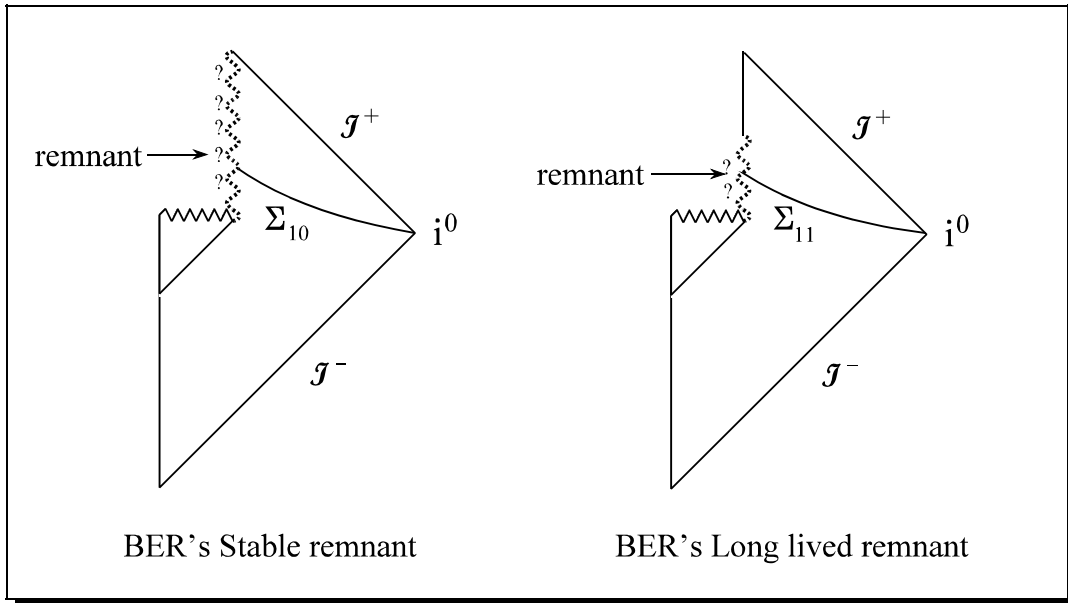
Figure 2

Such a scenario has been advocated by Banks (1995), who argues that Planck scale physics might shut down the Hawking radiation and prevent the formation of a singularity in the center of the black hole. As we can see in Figure 2 (copied from Banks 1995), appropriate late time slices will include all the matter that fell into the black hole, and thus our global state can remain pure. Although at any time the partial state associated with the region that is accessible to an external observer will be mixed, the evolution of the total state will be unitary, and thus we have a response to Hawking’s argument. This proposal violates BER’s commutation condition because there will be no time slice,  $\Sigma_{\text{late}}$ , whose observables will all commute with the black hole region. This is a simple consequence of the fact that the black hole continues to exist for all time. While there may be physical reasons to reject this proposal – based, for example, on the aforementioned worries about black hole pair production – it is nonetheless a conceptually adequate response to Hawking’s argument.

**5. Remnants vs. The Commutation Condition.** Banks’ proposal explicitly retains the spatiotemporal nature of the end product of black hole evaporation, which is therefore a residual black hole or a “remnant black hole” as BER also refer to it. The main target of their argument, however, is the black hole remnant proposal, which claims that Planck scale processes will *replace* the black hole with a new sort of fundamental particle, namely a black hole remnant.

A way of picturing remnants which does not involve a residual black hole is given in [Figure 3]. The spacetime in question still has the event horizon structure constitutive of a black hole, so while the remnant (the ??? of [Figure 3]) is not a remnant black hole, it is [a] remnant of a black hole, and so confronts the dilemma’s first horn. In this situation one can proclaim as loudly as one wants that information is stored in the remnant. Be that as it may, observables in the algebra associated with post-evaporation slice  $\Sigma_{10}$

...(stable remnant) or  $\Sigma_{11}$ ...(long-lived remnant) ought to commute with observables associated with the black hole interior. (BER 1999, 216)



**Figure 3**

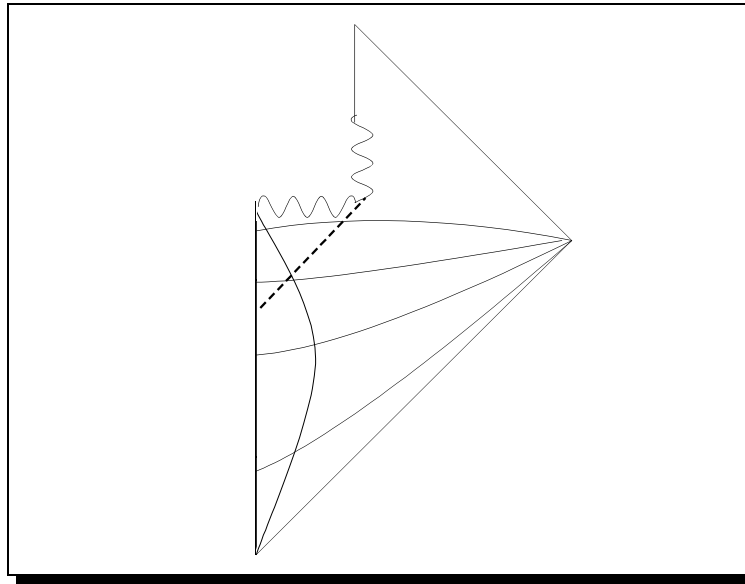
They consider one possibility for fleshing out the remnant story. The remnant (the ??? of their diagrams) could be a placeholder for a set of boundary conditions imposed at the singularity that would preserve unitarity.

But until remnant enthusiasts produce the new physics that incorporates the boundary conditions in a natural way, the present proposal 'solves' the information loss paradox only by inserting the missing information by hand, and 'remnant' is just a name that does nothing to justify the sleight of hand. (BER 1999, 218)

Leaving aside the question of whether this is a fair appraisal of the attempt to save unitarity through the imposition of boundary conditions, we should recognize that remnant theorists have other resources available.

The most common direction for the remnant proposal to take, and I think the most promising, is to deny that there is a true singularity in the center of black holes. (Whether these

objects then deserve the title “black hole” is a question I address in the next section.) Indeed, I believe that this how Giddings (1995a) intended the reader to interpret his figure that served as the model for BER's diagrams reproduced above.

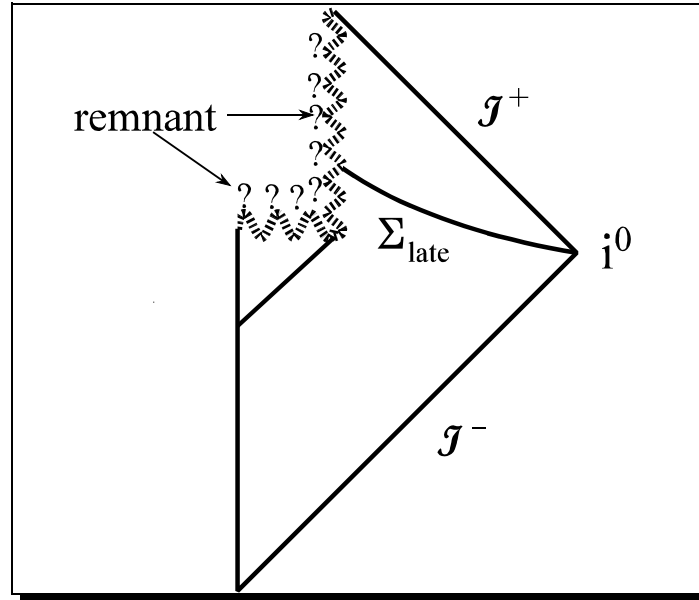


**Figure 4.** Giddings: “A Penrose diagram appropriate to a long-lived remnant scenario. The singularity is replaced by a planckian region near  $r = 0$ .”

While Giddings does not explicitly label the remnant in Figure 4, it seems clear that he intends the remnant, and not a curvature singularity, to be the terminus of all causal paths in the black hole. Why should this be a remnant rather than a singularity? Consider the claim that the remnant theorist is making. A large amount of matter undergoes gravitational collapse. This matter steadily loses mass through Hawking radiation, but it remains entangled with the external world. Eventually the black hole shrinks to Planck size and evolves from a black hole into a Planckian remnant. In the absence of evidence for a contrary interpretation, it would seem that we should take the claim that the information of the infalling matter is transferred to the remnant

to *imply* that the remnant is in the causal future of this matter.

While Figure 4 represents a long-lived remnant, we can also correct BER's representation of a stable remnant by replacing the spacelike singularity with a remnant, as indicated in Figure 5.



**Figure 5**

Both the long-lived and the stable variants of this remnant proposal escape BER's argument by denying their commutation condition. The late time slice,  $\Sigma_{\text{late}}$ , includes the remnant, which in turn is in the causal future of all observables localized inside the black hole. Thus these observables will generally not commute with all observables on  $\Sigma_{\text{late}}$ . In the case of stable remnants only the observables associated with the remnant itself will fail to commute with observables localized inside the black hole. (Here we simply assume that it is legitimate to refer to the remnant degrees of freedom as "observables," even though they cannot even in principle be measured by a macroscopic observer.) If we have a long-lived remnant, on the other hand, field observables in the causal future of the remnant will also fail to commute with the black hole

observables.

We now face the question of whether this denial of the commutation condition entails a violation of microcausality. If our spacetime were adequately represented by Figure 1, then the black hole region and  $\Sigma_{\text{late}}$  would clearly be spacelike related, because there is no causal curve connecting these two regions. However, now that we have *replaced* the singularity with a quantum gravitational remnant, what can we say about the existence, or lack of existence, of causal paths out of the black hole?

We seem to have two options. The first is to agree that there are no causal (i.e., luminal or subluminal) curves out of the black hole, but claim that in or around the remnant spacelike related observables will fail to commute. Our second option is to deny that all of  $\Sigma_{\text{late}}$  is spacelike related to the black hole region, either because there are causal paths *through* the remnant to  $\Sigma_{\text{late}}$ , or because the underlying quantum nature of the spacetime does not allow the unambiguous specification of whether these regions are spacelike or timelike related. While remnant theorists often do not distinguish between these two ways of fleshing out their position, their rejection of the commutation condition at the Planck scale is typically quite explicit. For example, immediately before his diagram of a long-lived remnant (Figure 4 above) Giddings writes:

Another possibility is that the information is radiated after the black hole reaches  $M \sim m_{\text{pl}}$  and the semiclassical approximation fails. Here ordinary causality no longer applies to the interior of the black hole, and it's quite plausible that the information escapes. (Giddings 1995a, 551)

But is this failure of “ordinary causality” due to superluminal interactions across Planckian distances, or a breakdown in the causal structure of spacetime? Or is this distinction somehow ill

posed in a truly quantum theory of spacetime?

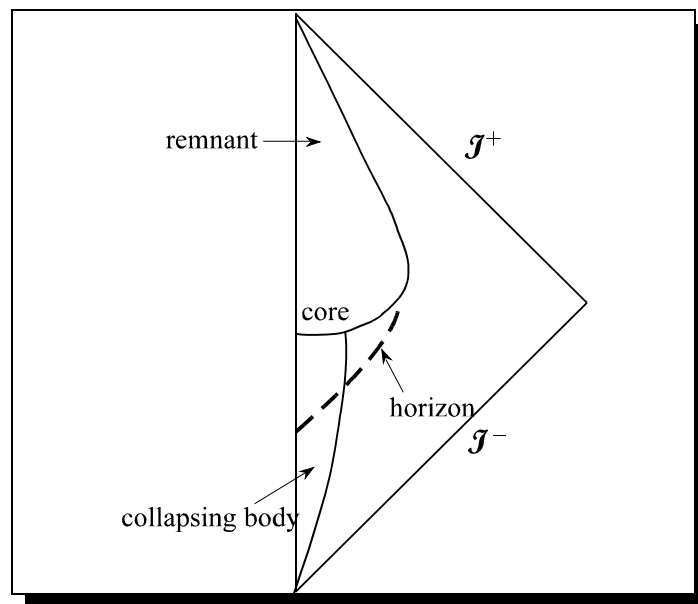
In support of reading the proposal as postulating superluminal information transfer, we can point out that remnant theorists typically claim that remnants are fundamental Planck-sized particles. The claim that they are fundamental presumably implies that they have no internal degrees of freedom of a spatial nature, i.e, they will have no spatially related dynamically distinguishable parts. But if they have a finite (Planckian) size, we can apparently no longer insist that all spacelike related observables commute. If a portion of an extended remnant is in the causal future of a black hole, and another portion of the remnant is in the causal past of a region of  $\Sigma_{\text{late}}$ , then an observable associated with the black hole will generally not commute with an observable defined on that region of  $\Sigma_{\text{late}}$ . The late time observable need only be in the causal future of *some* portion of the spatially extended remnant, not necessarily the portion that lies inside the lightcone of the black hole observable.

If we take the second alternative and argue that the black hole region is not (definitely) spacelike related to  $\Sigma_{\text{late}}$ , then we face both the question of the accuracy of Figure 5, and the second horn of BER's dilemma; for it is not clear that such a suggestion is compatible with the claim (depicted in the diagrams of the remnant spacetimes) that these regions are not in the causal past of  $\mathcal{J}^+$ , which is the *definition* of a black hole. One response to this challenge is to claim that the diagram is *limited*: the classical regions have been accurately portrayed, but not all of spacetime is amenable to this classical description. We misunderstand the proposal if we argue that because there are two regions of spacetime that *would* be spacelike related if the spacetime were completely classical, therefore operators associated with these regions "ought to



commute.” Unlike the stable residual black hole scenario postulated by Banks, an essential part of this remnant scenario is a specification of regions where our classical approximations break down, and it would be a mistake to insist on the commutation condition in this regime.

**6. Remnants vs. the Definition of Black Holes.** However, we still have not fully confronted the second horn of BER’s dilemma, for some versions of the remnant scenario imply that, strictly speaking, there will be no points that fail to be in the causal past of  $\mathcal{J}^+$ . BER criticize one such scenario, due to Giddings, on precisely these grounds: it denies the existence of a global event horizon, and thus denies the existence of a black hole. The spacetime of Giddings’ massive remnant proposal is sketched in Figure 6 (reproduced from Giddings 1992, Figure 3). Here the collapsing matter does not form a singularity, but rather a Planckian “core” that superluminally expands past the horizon, thus allowing the information to return to the external universe.



**Figure 6**

BER question whether it is appropriate to refer to this proposal as a solution of the problem of information loss in black holes, since there is no true black hole in Figure 6.

One disturbing feature of this proposal is that the core expands at a superluminal rate. Worse, [Figure 6] incorporates no singularity of gravitational collapse. ... Because the spacetime lacks a genuine black hole, the surface labeled 'horizon' in [Figure 6] must be an apparent horizon, an object locally delineated, rather than a true event horizon... And because the relevant event horizon structure is missing, the theorems that underwrite Hawking radiation do not apply. In short, the labels of 'black hole' and 'black hole evaporation' strike us as misnomers when applied to [Figure 6]. ... It is less a solution to the information loss paradox than a sweeping denial of the problem. But perhaps that is his point. (BER 1999, 219)

While I cannot fully address all of the objections BER raise here, a number of points should be made on behalf of remnant proposals that deny the existence of singularities.

The question of how we should use the term "black hole" is, of course, not the significant issue here. We can, if we like, identify black holes by the formation of apparent horizons. This was the concept that Wheeler first applied the term to in 1967, five years before Hawking introduced his account of global event horizons. Alternatively, we could, with very little violence to the term, identify a black hole with the region from which all causal curves necessarily terminate in a Planckian remnant. We might call the border of this region a "remnant horizon" if we wanted to distinguish it from an apparent horizon. Note that nothing could escape this region without passing through the Planckian remnant, a prospect not even a fundamental particle is likely to survive. But nothing of consequence hangs on this terminological issue.

The substantial question is whether apparent horizons or remnant horizons will be adequate for the derivation of Hawking radiation. Presumably it is this worry that prompts BER to charge Giddings with offering a "sweeping denial of the problem." If he were denying the

force of the argument for Hawking radiation, there might be something to this charge (although I would still worry about the distinction between solving a problem and denying that there is a problem). But, while Hawking's derivation of black hole radiation rests on the presence of a global event horizon (and hence the existence of a singularity), one can consistently accept the predictions of Hawking's semiclassical model while denying that the spacetime in question is singular. For example, 't Hooft (1985) has demonstrated with a simplistic "brick wall" model that in certain situations one can derive Hawking-like radiation from the explicitly unitary evolution of quantum fields. While 't Hooft is not a proponent of the remnant scenario, his model does indicate that nonsingular spacetimes can reproduce the effective descriptions offered by Hawking.

Granting that Hawking-like radiation is *possible* in the absence of true black holes, we still might wonder why a remnant theorist would go through the effort of trying to recover Hawking's prediction of black hole evaporation. One typically does not see explicit arguments for the claim that large black holes will give off Hawking radiation even if quantum gravitational processes prevent the formation of a true singularity and global event horizon. This is because all the participants in the debate generally agree that Hawking's prediction will be *effectively* accurate for large black holes. The debate centers around *how* accurate this effective description is likely to be (e.g., whether the radiation will be truly thermal as predicted by Hawking), and what happens when the effective description breaks down, for example, when the black hole shrinks to Planck size. The remnant theorist agrees with Hawking that the semi-classical description ought to be highly accurate on time-slices that stay well away from regions of strong gravitational curvature. This implies that a large black hole should behave *as if* it were in the

process of forming a true curvature singularity throughout most of its lifetime. Therefore it should be giving off thermal Hawking radiation – and any Planckian mechanism that might prevent the formation of a true singularity would seem to be unable to shut down this early time radiation.

The conviction that Hawking's prediction should be accurate for early times, along with the rejection of the claim that black holes completely evaporate away destroying quantum coherence, led remnant proponents to look for an explicit model, however idealized, of black holes evolving into remnants. The central project that occupied these theorists in the early 1990s was trying to find a simplified (usually two dimensional) unitary model of a black hole giving off radiation and shrinking, while still evolving unitarily. Although a plausible quantum gravitational model for a black hole remnant was never found, it should be clear that it would be illegitimate to dismiss such a project as not addressing the question of information loss in *black holes*.

There are serious problems facing the remnant proposal. Indeed, I know of no physicist who still defends this position. Giddings and Banks, for example, have both abandoned the remnant view in favor of something like black hole complementarity (private communication), a position we will briefly consider in the following section. However, these problems do not lie in a failure to address Hawking's argument adequately – if the scenario were plausible, it *would* resolve the information loss paradox – but rather in physical predictions that are difficult to reconcile with other aspects of physics (e.g., low mass objects having an exceedingly large number of internal degrees of freedom), and in an apparent inconsistency between the predicted behavior of remnants and the quantum gravitational theory that currently seems most promising to the majority of high energy physicists, namely string theory.

**7. Quantum Gravity and Border Disputes.** I would like to conclude by suggesting that the foregoing discussion offers some lessons for exploring scenarios in which we expect our current theories to break down, as in the case of evaporating black holes. In these situations we can often do considerably more than simply guess at how a system will behave and report that our current theories are not completely trustworthy. We often have the resources to develop some picture of *where* and *how* the theories will break down. In the case of black hole evaporation, for example, we can appeal to the fact that local QFT begins to run into problematic infinities when we consider very short distances, and the fact that interactions at Planckian energies should themselves create black holes, to argue that our current theories are inadequate in such a regime. As the remnant theorists emphasize, however, these same arguments seem to indicate that the semi-classical approximation *should* be adequate when these energies are absent.

BER recognize that QFT and GR are only approximately true, but they claim that “we cannot know how good the approximation is, or even what ‘approximately’ means, until we know how to combine QM and GR in one theory” (BER 1999, 221). While there is a good deal of truth in this statement, it neglects the fact that we sometimes can establish some reasonable boundary lines for the validity of our existing theories. Further, we can hope to glimpse some features of a full theory of quantum gravity by working to establish the proper domain of the low energy theories we expect it to reduce to. For this reason we should see the ??? in Figure 5 not merely as a placeholder for some future physics, but also (and perhaps more substantially) as a proposal for the proper *border* of our semi-classical description of the situation. Of course, remnant theorists aimed at more than the demarcation of our current theories, but the question of

when we should expect the effects of Planckian physics to become manifest remains an important and controversial question.

To see this it will be helpful to introduce, briefly and superficially, one other response to the information loss paradox: black hole complementarity. Advocates of this view argue that Planckian physics will not be restricted to the center of a black hole, but will become apparent at the event horizon as well. They justify this assumption by pointing out that the modes of the quantum fields that support Hawking radiation become highly energetic when propagated back in time to the event horizon, and by arguing that these high energies indicate that our semiclassical theories are unreliable even at the event horizon of a large black hole. Superstring models provide some further support for this claim. While the details of this proposal are beyond the scope of this paper, we can still recognize that the key issue of debate between the remnant theorist and the black hole complementarian will be the question of whether classical GR and standard QFT adequately describe areas of low curvature far from the center of a black hole. Remnant theorists appeal to the equivalence principle to argue that Planckian effects will be negligible until one reaches the center of a black hole, black hole complementarians counter by pointing to the high energies involved in descriptions that external observers will offer of the region around the event horizon. The force of this debate is lost if we simply assume the validity of local QFT and GR, or if we fail to distinguish between proposals (such as some remnant scenarios) that postulate violations of the commutation condition over Planckian distances, and other proposals (such as black hole complementarity) that claim that this condition will break down over macroscopic, or even astronomical scales.

We should recognize that Hawking's argument for information loss also relies on

assumptions about how and where GR and QFT will break down. The claim that a black hole will continue to radiate when it has reached Planck size can only be justified by assumptions about how a full theory of quantum gravity is likely to behave. Even the claim that a singularity and a global event horizon will form from the gravitational collapse of a sufficiently massive body requires us to make assumptions that go well beyond our semi-classical models, for the singularity theorems rest on energy conditions that are generally violated by quantum fields in curved spacetimes. Indeed, it is the violation of these energy conditions that allows us to escape Hawking's area increase theorem and claim that black holes shrink as a result of giving off radiation.

An understanding of the proposed *limitations* of our physical descriptions is therefore necessary for an adequate account of this particular debate in quantum gravity. But I would suggest that the importance of finding the proper boundaries of merely "effective" theories extends well beyond this limited context. Presumably an adequate interpretation of either GR or QFT would also need to offer an account of when and where these theories are no longer reliable. And if this is true of our physical theories it is certainly also true in other sciences. Given that every scientific theory we have is (at best) an effective theory – able to provide an adequate description in certain situations, but *ineffective* in others – we should recognize that erecting signposts in borderlands will be an essential part of laying claim to the understanding of nature offered by our theories.

## REFERENCES

- Banks, T. (1995), "Lectures on Black Holes and Information Loss", *Nuclear Physics B Proceedings Supplement* 41: 21-65.
- Belot, Gordon, John Earman, and Laura Ruetsche (1999), "The Hawking Information Loss Paradox: The Anatomy of a Controversy", *British Journal for the Philosophy of Science* 50: 189-229.
- Giddings, Steven B. (1992), "Black Holes and Massive Remnants", *Physical Review D* 46: 1347-1352.
- \_\_\_\_\_ (1995a), "Quantum Mechanics of Black Holes", in E. Gava, A. Masiero, K. S. Narain, S. Randjbar-Daemi, and Q. Shafi (eds.), *1994 Summer School in High Energy Physics and Cosmology. ICTP Series in Theoretical Physics-Vol. 11*, Singapore: World Scientific, 530-574.
- \_\_\_\_\_ (1995b), "Why Aren't Black Holes Infinitely Produced?", *Physical Review D* 51: 6860-6869.
- Hawking, Stephen W. (1976), "Breakdown of Predictability in Gravitational Collapse", *Physical Review D* 14: 2460-2473.
- 't Hooft, Gerard (1985), "On the Quantum Structure of a Black Hole", *Nuclear Physics B* 256: 727-745.