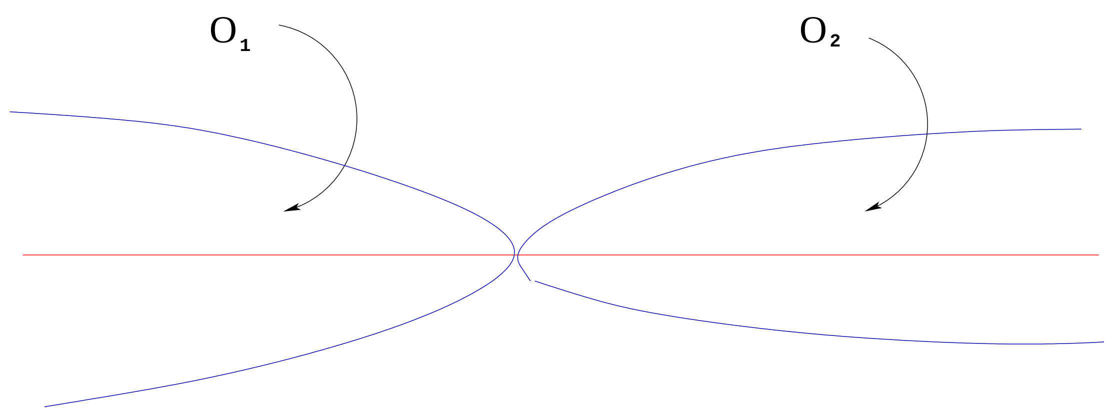
Information Loss

Robert M. Wald

Entanglement in Quantum Field Theory

Entanglement is a ubiquitous feature of quantum mechanics, but it is an essential feature of quantum field theory. Consider any two globally hyperbolic regions, O_1 and O_2, of spacetime that are causal complements of each other, as shown:



Let system 1 be the quantum field observables in O_1 and let system 2 be the quantum field observables in O_2.

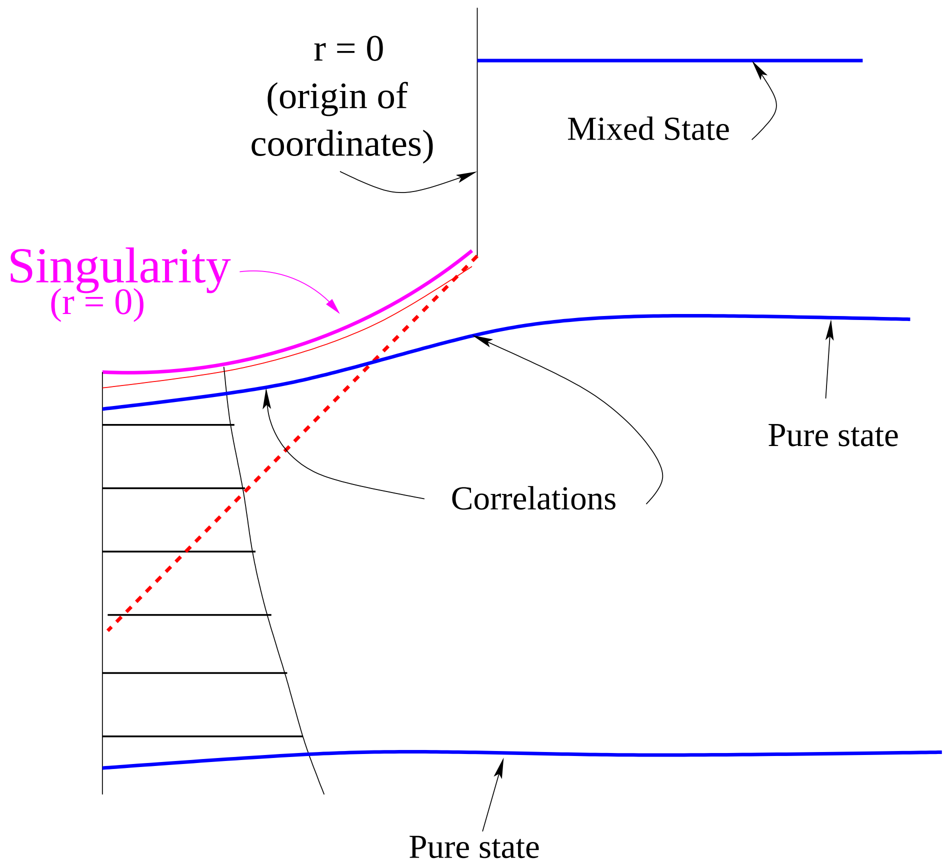
Then all physically reasonable states of the joint system will be strongly (in fact, infinitely) entangled. In particular, all physically reasonable states exhibit strong correlations at spacelike separations on small scales.

For example, all physically reasonable states in Minkowski spacetime display strong entanglement between the field observables in the left and right Rindler wedges. This accounts for why observers in a Rindler wedge see a (mixed) thermal state when the quantum field is in the (pure) Minkowski vacuum.

Entanglement with Black Holes

In a spacetime in which a black hole forms, there will be entanglement between the state of quantum field observables inside and outside of the back hole. This entanglement is intimately related to the Hawking radiation emitted by the black hole. In addition to the strong quantum field entanglement arising on small scales near the horizon associated with Hawking radiation, there may also be considerable additional entanglement because the matter that forms (or later falls into) the black hole may be highly entangled with matter that remains outside of the black hole.

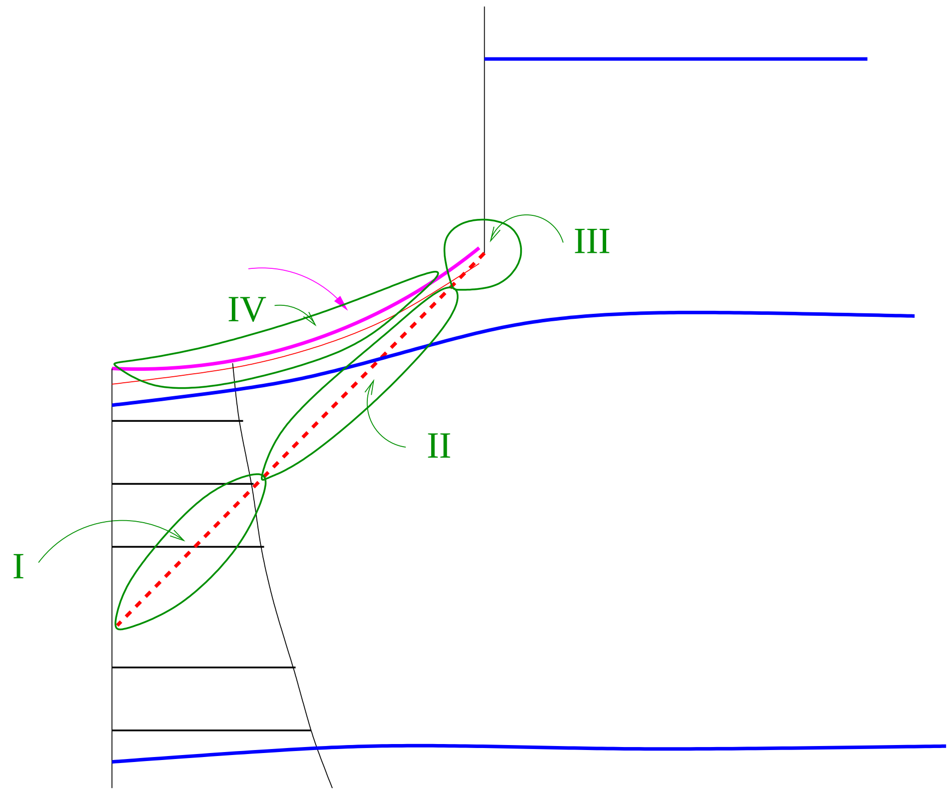
The Hawking effect and its back reaction effects give rise to the following semiclassical picture of black hole evaporation:



In a semiclassical treatment, if the black hole evaporates completely, the final state will be mixed, i.e., one will have dynamical evolution from a pure state to a mixed state.

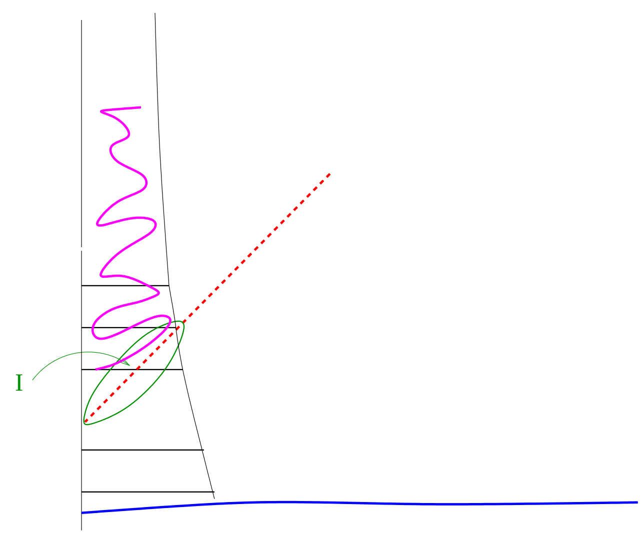
What's Wrong With This Picture?

If the semiclassical picture is wrong, there are basically 4 places where it could be wrong in such a way as to modify the conclusion of information loss:



Possibility I: No Black Hole Ever Forms (Fuzzballs)

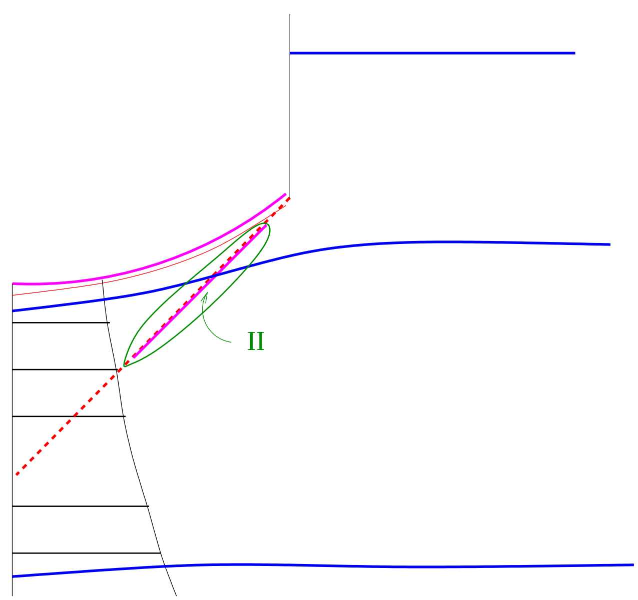
In my view, this is the most radical alternative. Both (semi-)classical general relativity and quantum field theory would have to break down in an arbitrarily low curvature/low energy regime.



Note that if the fuzzball or other structure doesn't form at just the right moment, it will be “too late” to do anything without a major violation of causality/locality in a low curvature regime as well.

Possibility II: Major Departures from Semiclassical Theory Occur During Evaporation (Firewalls)

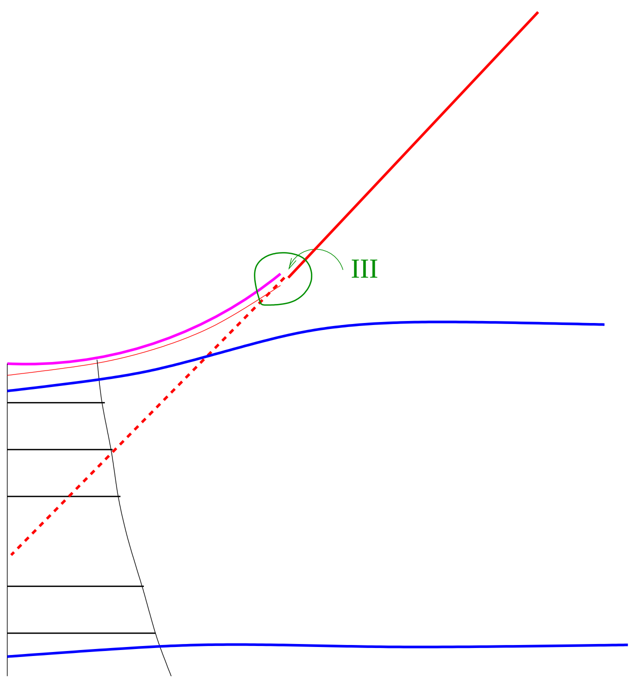
This is also a radical alternative, since the destruction of entanglement between the inside and outside of the black hole during evaporation requires a breakdown of quantum field theory in an arbitrarily low curvature regime.



A singular state at the horizon is clearly seems *necessary* to avoid quantum field entanglement with the black hole, but it is far from clear that it is *sufficient*, e.g., it would seem that one would also need violations of causality/locality to destroy the entanglement between matter that formed the black hole and matter that never fell in.

Possibility III: Remnants

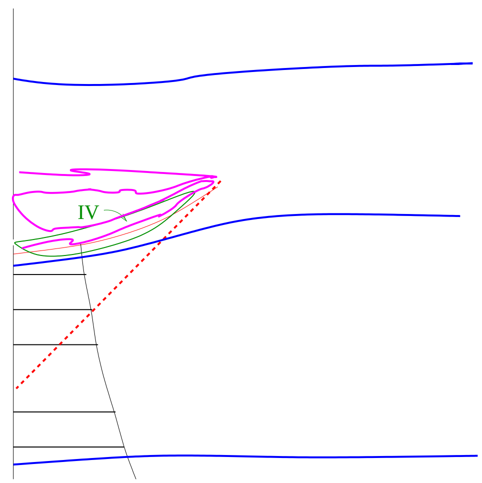
This is not a radical alternative, since the breakdown of the semi-classical picture occurs only near the Planck scale.



However, it is not clear what “good” the remnants do (since the "information," although still present, is inaccessible), and there are thermodynamic problems with them. I don't know of any present advocates of remnants.

Possibility IV: A Final Burst

This alternative requires an arbitrarily large amount of "information" to be released from an object of Planck mass and size.



The "burst" would clearly have to be highly non-classical until it reached a very large size. I do not know of any present advocates of bursts.

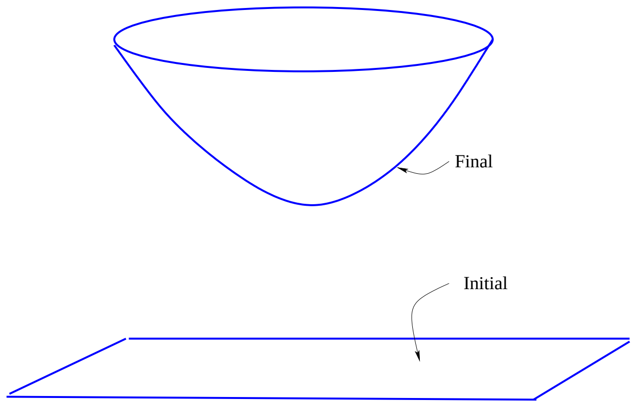
Arguments Against Information Loss:

Violation of Unitarity

In scattering theory, the word "unitarity" has 2 completely different meanings: (1) Conservation of probability; (2) Evolution from pure states to pure states.

Failure of (1) would represent a serious breakdown of quantum theory (and, indeed, of elementary logic). However, that is not what is being proposed by the semiclassical picture.

Failure of (2) would be expected to occur in any situation where the final "time" is not a Cauchy surface, and it is entirely innocuous.



For example, we get "pure $\rightarrow$ mixed" for the evolution of a massless Klein-Gordon field in Minkowski spacetime if the final "time" is chosen to be a hyperboloid. This is a *prediction* of quantum theory, not a *violation* of quantum theory.

The "pure $\rightarrow$ mixed" evolution predicted by the semiclassical analysis of black hole evaporation is of an entirely similar character.

I find it ironic that some of the same people who consider "pure $\rightarrow$ mixed" to be a violation of quantum theory then endorse truly drastic alternatives that *really are* violations of quantum (field) theory in a regime where it should be valid. I have a deep and firm belief in the validity of the known laws of quantum theory (on length and time scales larger than the Planck scale), and I will continue to vigorously defend quantum theory against those who may have initially set out to try to save it but who somehow got diverted into trying to destroy it.

Arguments Against Information Loss: Failure of Energy and Momentum Conservation

Banks, Peskin, and Susskind argued that evolution laws taking "pure $\rightarrow$ mixed" would lead to violations of energy and momentum conservation. However, they considered only a "Markovian" type of evolution law (namely, the Lindblad equation). This would not be an appropriate model for black hole evaporation, as the black hole clearly should retain a "memory" of what energy it previously emitted.

There appears to be a widespread belief that any quantum mechanical decoherence process requires energy exchange and therefore a failure of conservation of energy for the system under consideration. This is true if the "environment system" is taken to be a thermal bath of oscillators. However, it is not true in the case where the "environment system" is a spin bath. In any case, Unruh has recently provided an example of a quantum mechanical system that interacts with a "hidden spin system" in such a way that "pure $\rightarrow$ mixed" for the quantum system but exact energy conservation holds.

Bottom line: There is no problem with maintaining exact energy and momentum conservation in quantum mechanics with an evolution wherein "pure $\rightarrow$ mixed".

Arguments Against Information Loss: AdS/CFT

The AdS/CFT argument against the semiclassical picture is simply that if gravity in asymptotically AdS spacetimes is dual to a conformal field theory, then since the conformal field theory does not admit "pure $\rightarrow$ mixed" evolution, such evolution must also not be possible in quantum gravity.

AdS/CFT is a conjecture. The problem with using AdS/CFT in an argument against information loss is not that this conjecture has not been *proven*, but rather that it has not been *formulated* with the degree of precision needed to use it reliably in such an argument: "Information loss" in black hole evaporation is the statement that the bulk observables at late times are not the complete set of bulk observables. AdS/CFT says that the complete set of bulk observables should be in 1-1 correspondence with the complete set of CFT observables. Ordinary Hamiltonian evolution of the CFT says that the CFT observables at late times are equivalent to the complete set of CFT observables. To complete this to make an argument against information loss, one needs to argue that the bulk observables at late times are in 1-1 correspondence with the CFT observables at late times. But the correspondence between bulk and CFT observables is nonlocal in spacetime, and very little of the "dictionary" is explicitly known. It seems to me that there is plenty of room for nuances in the dictionary when black holes are present that would allow the late time observable correspondence to fail. Put another way, why can't the CFT state at late times continue to encode the information that went into a black hole, even though that information is no longer accessible to late time bulk observers? In that case, there would still be an exact AdS/CFT correspondence, the late time CFT state would be pure, and yet the late time bulk observers will have experienced information loss.

So, I hope that the AdS/CFT ideas can be developed further so as to make a solid argument against (or for!) information loss. Such an argument would then undoubtedly provide some explanation of *how* information is regained—not just that it must happen somehow or other. Until then, I'm sticking with information loss!