‘Information Loss’ Is Said in Many Ways, the Responses Legion, Not All Cogent

SIGRAV International School
“Quantum Effects in Curved Spacetimes and the Thermodynamics of Horizons”
19. Feb 2025

Erik Curiel

Lichtenberg Group for History and Philosophy of Physics, Bonn Universität

Black Hole Initiative (BHI), Harvard University

Quantum Information Structure of Spacetime Consortium (QISS)

`erik@strangebeautiful.com`
`http://strangebeautiful.com`

February 20, 2025

The subtlety of Nature far exceeds the subtlety of sense and intellect: so that these fine meditations, and speculations, and reasonings of men are a sort of insanity, only there is no one at hand to remark it.

– Francis Bacon
Novum Organum, Book I, Aphorism X

Ἀπολογία

(Apology)

– in physics we feel, I think justifiably, that we have come to learn something about the world, often something concrete, sometimes something deep

– this comes out most clearly when we recall that physics does more than predict experimental outcomes based on clearly formulated mathematical models

– it also teaches us about qualitative features of the world that we do not know how to model in anything like an adequate quantitative sense (turbulence, *e.g.*)

– and it teaches us about broad and global features of the world, about its *possible* behaviors, that, it seems, we need general theorems to characterize (the relationships among topology, causal and affine structures in GR, *e.g.*, captured by the classical singularity theorems)

– in philosophy, we try to understand what it is we've learned, and we realize that learning is only the first step in coming to understand

– in reflecting on the state of our knowledge, we recognize that there always remain open questions about that knowledge. . .

– how to clarify, elaborate and enrich the concepts and the relations among them we (are trying to!) use to formulate and represent the knowledge physics has given us, to grasp what conceptual possibilities are opened up or closed off by that knowledge in the best of cases, this sets up a self-sustaining feedback loop, a virtuous epistemic circle physics provides philosophy the knowledge to reflect on:

> *[W]e are met as cultivators of mathematics and physics. In our daily work we are led up to questions the same in kind with those of metaphysics; and we approach them, not trusting to the native penetrating power of our own minds, but trained by a long-continued adjustment of our modes of thought to the facts of external nature.*
>
> – James Clerk Maxwell (1870)
> "Address to the Mathematical and Physical Sections of the British Association"

philosophy provides physics concept clarification, and the questions whose investigation may not lead to definitive answers to the questions themselves, but pleasantly often opens up new avenues of research that lead us to more and deeper learning about the world:

> *[W]e must bear in mind that the scientific or science-producing value of the efforts made to answer these old standing questions is not to be measured by the prospect they afford us of ultimately obtaining a solution, but by their effect in stimulating men to a thorough investigation of nature. To propose a scientific question presupposes scientific knowledge, and the questions which exercise men's minds in the present state of science may very likely be such that a little more knowledge would shew us that no answer is possible. The scientific value of the question, How do bodies act on one another at a distance? is to be found in the stimulus it has given to investigations into the properties of the intervening medium.*

– James Clerk Maxwell (1875a)
"Attraction" (*Encyclopædia Britannica*, edition IX)

– in a field such as black hole thermodynamics (BHT), and semi-classical gravity (SCG) more generally...

– where we have not only *no empirical experience* to test our theorizing...

– but, much more important (and worse), we have none to *guide* and *constrain* it...

– where we have *not* been "trained by a long-continued adjustment of our modes of thought to the facts of external nature"...

- $\Longrightarrow$ investigations necessarily speculative in a way unusual even in theoretical physics
- $\Longrightarrow$ technically sophisticated, conceptually deep physical questions...
- inextricable from subtle philosophical considerations spanning ontology, epistemology, and methodology...
- in a way unusual even in theoretical physics

in such a field...

I see no clear line to be drawn to demarcate physics from philosophy

And so my task here today, as this Socratic ἀπολογία suggests, is...

to play Socratic gad-fly

(which, recall, involves clarificatory exposition[1])

1. Also recall the charge that got Socrates executed was "corrupting the youth"—I hope to do that here as well.

Outline

SCG and BHT

- without a doubt the most widely accepted, most deeply trusted results in theoretical physics in which GR and QFT (and thermodynamics!) work together in seemingly fruitful harmony—
- especially remarkable when one reflects on the fact that we have *absolutely no experimental or observational evidence for any of it*, nor hope of gaining empirical access any time soon to the regimes where such effects may appreciably manifest themselves
- $\Longrightarrow$ investigations necessarily speculative in a way unusual even in theoretical physics
- $\Longrightarrow$ technically sophisticated, conceptually deep physical questions inextricable from subtle philosophical considerations spanning ontology, epistemology, and methodology, again in a way unusual even in theoretical physics
- $\Longrightarrow$ why do we trust it?

- in such a situation, what besides the intuitions of eminent physicists do we have to guide our theorizing?
- can we have any epistemic control over the situation at all?

$\Rightarrow$ THE FUNDAMENTAL QUESTION!

The Information-Loss Paradox (ILP) in a Nutshell

1. stuff collapses to a black hole, or is thrown into an existing one
2. No-Hair theorems: at the classical level, all information about it all is utterly effaced from the outside
3. when quantum effects taken into account, black holes emit Hawking radiation...
4. far away, ignoring grey-body factors, an essentially Planckian spectrum
5. carries positive mass-energy away ⇒ the black hole shrinks ("evaporates")
6. eventually, the black hole "completely evaporates"[2]
7. but Planckian radiation carries *no* information (except an encoded temperature), "perfectly random"
8. all "information" about what collapsed/fell in *irretrievably lost*!
9. ⇒ *violation of fundamental principle of QM*, unitary evolution

2. Or: gets small enough that it doesn't have enough mass-energy, micro-states, ..., to maintain a "record" of everything that collapsed/fell in.

two notions of evolution

GR (well defined)

1. initial data: (h_{ab}, π^{ab}) on slice Σ
2. evolution into domain of dependence governed by Cauchy development induced by EFE

QFT (aspirational)

1. initial data: quantum state ψ_t defined on a Cauchy slice Σ_t
2. evolution between surfaces governed by something like Schrödinger unitary evolution:

$$\text{“}|\psi_t\rangle = \hat{U}_t|\psi_0\rangle\text{”}$$

where “$\hat{U}_t = \exp(-it\hat{H})$”

one might have thought the central issue of the ILP would standardly be posed as something like:

are these 2 notions of evolution in an appropriate sense consistent?

in fact, it is not, in large part because no one knows how to combine the 3+1 EFE with anything like the Schrödinger equation, whether appropriately or not

(the semi-classical Einstein field equation (SCEFE)

$$G_{ab} = 8\pi \langle \hat{T}_{ab} \rangle$$

doesn't do the job, since it deals only with $\langle \hat{T}_{ab} \rangle$)

- rather, in the context of evaporating BH spacetimes, there are (at least) 5 questions commonly posed...
- sometimes only implicitly
- none always clearly distinguished from others
- and none clearly equivalent to—or even clearly related to—others
- although some *prima facie* more intimately entangled than others

Hawking problem is there a unitary scattering matrix from $\mathscr{I}^-$ to $\mathscr{I}^+$? (Hawking 1975, 1976a)

final-state problem is Hawking radiation (*i.e.*, the quantum state of the field on some slice, preferably Cauchy) in a pure state after evaporation ends? (Wald 1994; Jacobson 2005)

recovery problem can the "information" encoded in a physical system be recovered after it enters a black hole (whether part of initial collapse or later addition)? (Hayden and Preskill 2007; Harlow and Hayden 2013)

causal-structure problem what is the global causal structure of a spacetime with an evaporating black hole, and is it compatible with unitary evolution? (Kodama 1979; Wald 1984)

Page-curve problem does the entropy of Hawking radiation decrease at late times during evaporation? (Page 1993; Akers et al. 2020)

- they all have more or less vaguely to do with the idea of unitarity...
- variously glossed as the possibility of retaining or maintaining
  1. "determinism"
  2. or "predictability"
  3. or "conservation of probability"
  4. or "evolution of pure states into pure states"
  5. or "entanglement between `in` and `out` modes of Hawking radiation"
  6. or ...
- $\Rightarrow$ can quantum fields unitarily evolve in a spacetime with an evaporating black hole, in any of these senses?

all complicated by fact that there are ***MANY***, for the most part radically different, forms of derivation of the Hawking effect, depending on:

spacetime (GR) choices

1. shape and character of spacetime:
   1.1 exact solution: Schwarzschild, Kerr, Reissner-Nordström, Kerr-Newman, dS-Schwarzschild, AdS-Schwarzschild, . . .
   1.2 *OR* abstract characterization: type of horizon, topology, symmetries, asymptotic structure, . . .
2. region of analysis (local, global, near-horizon, asymptotic)
3. some form of cosmic censorship (complete $\mathscr{I}^+$, non-singular event horizon, . . .)
4. causality conditions (*e.g.*, chronology)
5. stability assumptions ("small perturbations do not destroy the event horizon")
6. . . .

matter (QFT) choices

1. QFT formulation (S-matrix, algebraic, canonical based on Lagrangian, covariant phase space, low-energy perturbative quantum gravity, . . .)
2. flavor of QFT (scalar, vector, bosonic, fermionic, . . .)
3. imposition of (averaged) energy conditions or quantum energy inequalities

4. choice of state needed? if so, which? or generic conditions imposed?
5. choice of state eigenbasis needed? if so, which? or generic conditions imposed?
6. boundary conditions, *e.g.*, incoming Minkowski vacuum
7. approximations, *e.g.*, eikonal or WKB
8. constructibility of stress-energy tensor operator
9. niceness of state (*e.g.*, Hadamard)
10. adiabaticity conditions
11. insensitivity to trans-Planckian phenomena
12. various forms of locality and causality
13. cluster decomposition
14. ...

joint spacetime-matter choices

1. backreaction? (almost never!)
2. stationary, quasi-static, quasi-steady or dynamic matter+spacetime geometry
3. entropy conditions (*e.g.*, satisfaction of the GSL)

there are now, at a conservative estimate,

2,069,547,534

possible derivations[3]

3. with a tip of the hat to I. J. Good (1971)

most popular forms:

1. S-matrix *à la* Hawking's (1975) original
2. past-boundary *à la* Unruh's (1976) original
3. algebraic
4. canonical, based on Lagrangian
5. canonical, based on Cauchy development
6. canonical, based on "Schrödinger evolution"
7. tunneling
8. anomaly canceling
9. more general stress-energy tensor analysis
10. near-horizon symmetries
11. thermal atmosphere
12. holographic
13. renormalization group
14. analytic continuation
15. Euclidean path-integral
16. Lorentzian path-integral
17. low-energy perturbative quantum gravity (EFT)
18. perturbative canonical QG
19. perturbative LQG
20. perturbative string theory
21. . . .

this all leaves us with (at least) 2 *very serious problems* to consider:

1. how do all the radically different possible derivations square with each other (if at all)?
2. would an answer to ILP formulated with respect to one hold for any of the others?

especially in light of:

> $\Rightarrow$ many suggest radically different physical interpretations

example: spacetime picture of dynamics

S-matrix

- eikonal approximation gives well defined dynamics formulated using spacetime geodesics
- "particle flux" at $\mathscr{I}^+$ (relatively) well defined (asymptotic symmetries)

tunneling

- "positive energy particle tunneling out of black hole interior"
- ***cannot*** be given cogent interpretation based on propagation guided by (classical!) spacetime geometry (here, the affine structure)...
- without calling into question the classical spacetime causal structure

algebraic

1. gives well defined state
2. no picture of particle production at all—consistent with both (either?) of two previous?

what is the physical content of “the Hawking effect”?!

does it even deserve the honorific of the definite article?

and one final problem. . .

– the entire edifice of BHT, including outhouses like ILP, depend on interpreting BHs as true thermodynamical systems. . .

– best argument for that is Hawking radiation. . .

– *BUT* unlike the incandescence of a glowing lump of hot iron (real blackbody radiation!). . .

– Hawking radiation, in the semi-classical picture, *is not generated by micro-degrees of freedom of the event horizon* (which is treated as simple, classical geometrical structure)

– back-reaction is essentially never (~99.9%of the time) included

– and even then never the SCEFE. . .

– and even if SCEFE *were* used, that still doesn't include micro-degrees of freedom of the horizon

Hawking radiation is not blackbody radiation!

The Temperature Decoupling Problem

Why should we believe the Planckian temperature of Hawking radiation is the temperature of the black hole, when it is not generated by micro-degrees of freedom of the event horizon?

- need further assumptions to bring two *prima facie* disparate phenomena—(presumed) micro-dynamics of horizon on one hand, and those of external quantum field on other—into explicit and harmonious relation with each other
- only then can conclude that temperature of quantum radiation with Planckian spectrum is sound proxy for temperature of black hole itself as determined by dynamics of (presumably) its very own micro-degrees of freedom
- but exactly lack of such bridging principles, as we will see, calls into question importance of the ILP
- so perhaps derivations of Hawking radiation themselves, which don't depend on coupling of quantum field with classical geometry, are already trying to tell us not to take any of this terribly seriously, with regard to fundamental issues. . .
- after all, we have *no evidence* that anything like "perturbative low-energy QG degrees of freedom" couple in the right way with degrees of freedom of *ordinary matter* (QFs) at energy and gravity scales *we can already probe!*
- *N.b.*: that last issue is particularly severe for the Page-Curve Problem, where assumed that "quantum-gravitational statistical-mechanical" degrees of freedom of black hole couple to Hawking radiation (ordinary QFs!)

our plaintive question is now sharpened:

$\Rightarrow$ can quantum fields unitarily evolve in a spacetime with an evaporating black hole,

– in any of the senses of "unitarity",

– with respect to (at least) the most popular, influential, physically perspicuous, mathematically rigorous, sexy, groovy, . . . , derivations,

– so that conclusions are insensitive to the Temperature Decoupling Problem,

– in a way that addresses at least some of the 5 formulations of the ILP?

and now the shrugging and the freaking out begin

(whether one shrugs or freaks out is strongly correlated with one's intellectual geneaology and rearing)

– all these issues come out with peculiar clarity in the "traditional" argument about the Hawking problem,

– and its close kin (kissing cousins?) the final-state and the causal-structure problems

let's focus for the moment on the final-state problem to get a feel for why the shrugging and the freaking out

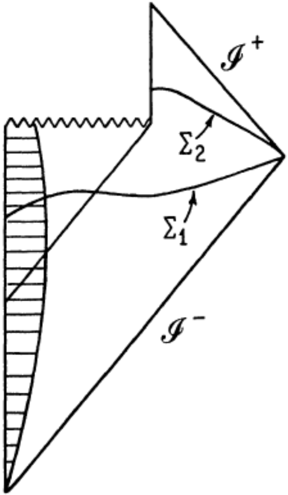


Figure 7.3. A conformal diagram illustrating the phenomenon of loss of quantum coherence in a spacetime in which black hole evaporation occurs.

(taken from Wald 1994)

summary

1. by Cauchy evolution (EFE – aspirational), a pure state on Σ_1 develops into a mixed state on Σ_2
2. by Schrödinger-type evolution (aspirational), this cannot happen
3. the root of the conflict lies in the pathology of the causal structure of the spacetime:
   i. Σ_1 is a Cauchy surface for that part of the spacetime lying beneath the null surface "connecting the evaporation region to $\mathscr{I}^+$"
   ii. Σ_2 is not (although it is a slice, a "partial Cauchy surface"[4])
4. $\Longrightarrow \Sigma_2$ can, at most, "know" about the information encoded in the mixed state defined on Σ_1 by tracing out that part of the state associated with the part of Σ_1 lying behind the event horizon
5. information is lost!

---

4. Technically: an achronal surface without edge, where an edge is all points on the surface having a point not on the surface to its chronological past and one to the future such that there is a timelike curve connecting the two not intersecting the surface

- arguments in favor of loss of unitarity (in basically all senses) are strong (Unruh and Wald 2017)
- why all the fuss in the case of black hole evaporation, but not in standard treatments of quantum fields in effective field theory (EFT) formalism, where unitarity is violated all the time...
- and, for that matter, in the case of measurements in standard quantum theory?

those who freak out have propounded a devils' legion of mechanisms, principles, frameworks, ..., and arguments based on them to save unitarity:

1. complementarity (Susskind et al. 1993; Nomura et al. 2013)
2. BH remnants (Giddings 1992; Chen et al. 2015)
3. fuzzballs (Mathur 2005; Mathur and Mehta 2024)
4. other kinds of holographic arguments (Marolf 2009; Chen et al. 2020)
5. firewalls (Almheiri et al. 2013; Bousso 2025)
6. manifold varieties of AdS/CFT stuff (Polchinski 2017; Akers et al. 2020)
7. "island" calculations (Almheiri et al. 2019; Penington 2019)
8. quantum error-correction codes (Akers and Penington 2022)
9. final information blasts (Bardeen 2014; Unruh and Wald 2017)
10. pervasive, promiscuously orgiastic nonlocality (Giddings 2006; Almheiri et al. 2021)
11. ...

in most cases, they destroy the village (QFT) to save the village[5]

5. Vietnam War, with regard to the Battle of Bên Tre, an unnamed American major: "It became necessary to destroy the village to save it."

all leaves us with a poignant question:

- why do many physicists feel confident that an effect predicted by what is manifestly an EFT calculation—
- evaporation due to emission of Hawking radiation—
- can be trusted to reveal features of an underlying fundamental theory of quantum gravity (Marolf 2017)?
- we do not try to glean such insight from any standard EFT as used, *e.g.*, in high-energy particle physics or condensed-matter physics
- why even demand unitary evolution for quantum fields in every crazy old spacetime one can contrive at the semi-classical level?

The Cascade Problem

Why assume failure of unitarity at semi-classical level automatically cascades down to failure of unitarity—or something else equally distasteful—at the level of a more fundamental theory of QG?

so let's look at a strong argument
in favor of unitarity

– Marolf (2009) proposed a novel form of “holographic” argument, which has become widely influential, that black hole evaporation is fundamentally a unitary process. . .

– addressing a combination of the final-state problem and the recovery problem

– using ‘unitarity’ in the sense applicable to self-adjoint operators and related to the evolution of pure states to pure states

– I think it is ingenious and elegant, and delivers real insight on a number of issues

– but I also think it begs a fundamental question, in the same way that all holographic arguments I know of for unitarity do (and island calculations, and many other types of argument as well)

– and it does so in a way that brings out the problem with clarity (because the argument itself is so crisp, clear and elegant)

– or, from a different perspective, it makes perspicuous some fundamental, otherwise hidden, and severe consequences of unitarity

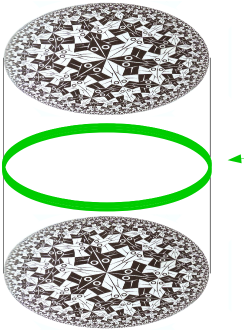
– to prove my *bona fides*, I will present the argument using the slides of a talk by Aron Wall from 2019, who vigorously defends the validity of the argument

– also because he provides a clear, elegant and physically intuitive exposition (with diagrams, which I suck at making)

Argument #2: Boundary Unitarity (Marolf)

assume asymptotically AdS, but do *NOT* assume AdS/CFT
(instead we will be *proving* that something like it must hold)

[a asymptotically flat argument exists, but is more subtle.]



argument concerns the set of all quantities that are measurable on the boundary at a given time t:

$$\mathcal{A}_{t,\Delta t}$$

allow a small "thickness" Δt to avoid worries about smearing operators in time...

basic principles of physics will now imply that the info that falls into a black hole remains accessible on the boundary...

Axiom #1: $\mathcal{A}_{t,\Delta t}$ is an *algebra of operators* [QFT]

- vector subset of operators A, B... in global Hilbert space

- closed under addition (vector space): A + B

- closed under multiplication (algebra): AB

- and reasonable limits thereof (C* algebra)

These assumptions are totally standard in AQFT when describing the set of all measurable quantities in a region

Axiom #2: the Hamiltonian is measurable at boundary [GR]

$$H \in \mathcal{A}_{t,\Delta t}$$

In ANY **diffeomorphism-invariant** *theory of gravity (not just GR), the total energy is a pure boundary term (the ADM energy).*

Gauge symmetry implies that H = 0 locally, up to a total derivative that arises when the diffeo vector ξ *does not vanish on the boundary.*

The ADM energy is obtained from the case where ξ *limits to a time translation on the boundary.*

Axiom #3: The Hamiltonian generates time translations [QM]

a) H is a self adjoint operator

b) for all $\mathcal{O}$, $[H, \mathcal{O}] = i\frac{d\mathcal{O}}{dt}$ (Heisenberg picture)

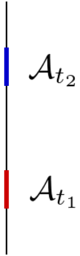
These rules are simply the **definition** *of the Hamiltonian in QM, which always exists if there is a time-translation symmetry acting on the complete Hilbert space.*

(the identification of this with the previous H is related to the exact equivalence of gravitational and inertial energy in GR.)

Axioms 1-3 imply that the boundary evolves unitarily!

If $\mathcal{O}(t)$ is a family of operators related by time translation symmetry, then you can solve for one time in terms of other times:

$$\mathcal{O}(t_1) = e^{iH(t_1-t_2)}\mathcal{O}(t_2)e^{-iH(t_1-t_2)}$$



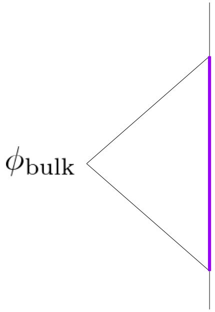
anything that can be measured at t_1
can also be measured at t_2,

because H and $\mathcal{O}(t_2)$ are in the algebra and the r.h.s. is just a limit of sums & products of those...

hence no info can be lost from the boundary (unless it was *never* on the boundary)

Axiom #4: There are other nontrivial operators in the algebra that can be excited to form a black hole [AdS QFT]

e.g. the boundary value of a scalar field $\phi(t) \in \mathcal{A}_{t,\Delta t}$



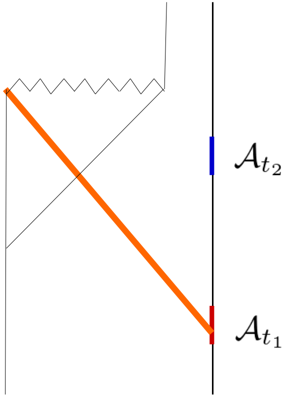
it is known how to solve for a field (outside of any horizons) in terms of integrals of boundary limiting values $\phi_{\rm bulk}$

sideways Cauchy problem subtle but basically OK

Hamilton-Kabat-Lifschytz-Lowe (free fields)
interacting case done perturbatively in 1/N
(should be good near infinity)

note we just need ***some*** *nontrivial field operator, (other than vacuum symmetry generators like H) on a small boundary interval*

Information is not lost into the black hole



excite fields at t_1 to form BH,
these fields carry info to the inside

at any later moment of time t_2
(even before the BH evaporates)
the information is still available
in principle, and can be measured
by a complicated experiment

Summary of Assumptions

#1: exists an *algebra of operators* $\mathcal{A}_{t,\Delta t}$. [QFT]

#2: $H \in \mathcal{A}_{t,\Delta t}$ [GR]

#3: a) H is a self adjoint operator

b) for all $\mathcal{O}$, $[H, \mathcal{O}] = i\frac{d\mathcal{O}}{dt}$ [QM]

#4: exist nontrivial operators in $\mathcal{A}_{t,\Delta t}$ that can be used to form black hole [AdS QFT]

I have many problems with the derivation—all identifiable because of beautiful clarity of argument!

I list a few:

1. no Bondi mass/news in asymptotically anti-de Sitter spacetimes
   1.1 how account for effects of gravitational radiation?
   1.2 boundary excitations creating BH in bulk will produce gravitational radiation, at least when BH is formed...
   1.3 so solving for later operator representing "information that went into BH" must account for "information in gravitational radiation"
2. identification of "H measurable at bdry", *viz.*, GR ADM mass, with "H generator of time translation symmetries", *viz.*, QFT Hamiltonian, is dubious at best:
   2.1 "exact equivalence of gravitational and inertial mass in GR" is neither exact nor an equivalence
   2.2 subtleties in identifying bulk timelike vector-fields asymptoting to boundary time-translations means can't rely on that unambiguously

3. almost every observable on boundary relevant to bulk physics, particulary H but also those creating bulk black holes, are *essentially and wildly non-local*
   3.1 we have no idea of what it means, how we would go about trying, to measure or observe such profoundly and promiscuously non-local phenomena, we know nothing physically about such observables. . .
   3.2 parlous to base arguments about fundamental matters on them![6]
4. BH formed from boundary excitation encloses small BH formed earlier, evaporates inside—guarantee that relevant boundary operators appropriately related?
5. final diagram essentially same as standard black-hole evaporation diagram (one from Wald 1994 earlier)
   5.1 strongly suggests bulk pure states *do not* evolve to pure states—how reflected in behavior of boundary operators?
   5.2 possible to encode bulk pure-to-mixed evolution to boundary pure-to-pure evolution?
   5.3 or all states on boundary mixed to start with?

---

6. Non-local phenomena in ordinary QM, such as the Aharonov-Bohm effect and the Berry phase, can give us no guidance here, for here the non-locality extends across *the entirety of a slice of asymptotic infinity*.

– I now address what I see as the most fundamental and exemplary problem with Marolf's argument, which illustrates the conflict I believe lies at the heart of all related ways of posing the issue

– the nub of my problem with Marolf's argument is that it implicitly assumes the interior of spacetime is causally well behaved (all required actions on the boundary can propagate in a determinable way anywhere into the interior of spacetime I want)—

– but the final-state problem strongly suggests that that is exactly what is up for grabs

– to make the claim precise, I introduce a few causality conditions and a theorem by Lesourd (2019), to show where in Marolf's argument the implicit assumption is made without explicit justification[7]

---

7. I am *not* claiming that no justification is possible, only that no compelling one has been given by the community which advocates for such arguments and conclusions as Marolf's.

for $\mathcal{M} = (M,\, g_{ab})$:

distinguishing $\forall p, q \in M$, $I^-(p) = I^-(q)$ or $I^+(p) = I^+(q)$ implies $p = q$

reflecting $\forall p, q \in M$, $I^-(p) \subseteq I^-(q)$ iff $I^+(q) \subseteq I^+(p)$

causally continuous $\mathcal{M}$ is both distinguishing and reflecting

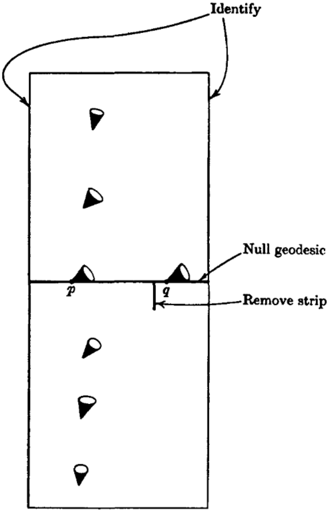


FIGURE 37. A space in which the causality and past distinguishing conditions hold everywhere, but the future distinguishing condition does not hold at p or q (in fact, $I^+(p) = I^+(q)$). The light cones on the cylinder tip over until one null direction is horizontal, and then tip back up; a strip has been removed, thus breaking the closed null geodesic that would otherwise occur.

(taken from Hawking and Ellis 1973)

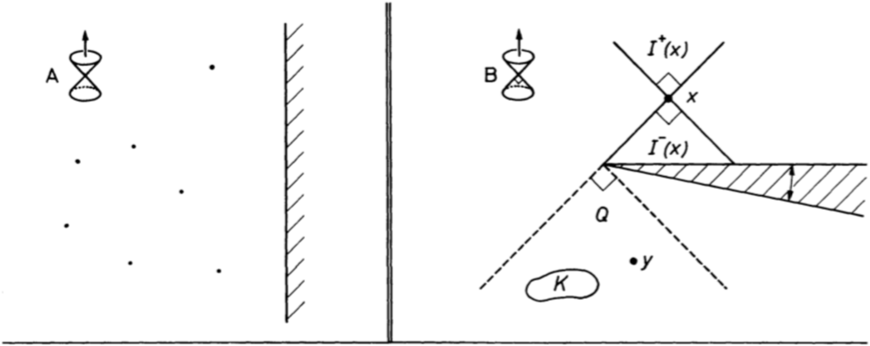


Fig. 1.2. The spacetime A is reflecting; the spacetime B is not

(taken from Hawking and Sachs 1974)

a causally discontinuous spacetime

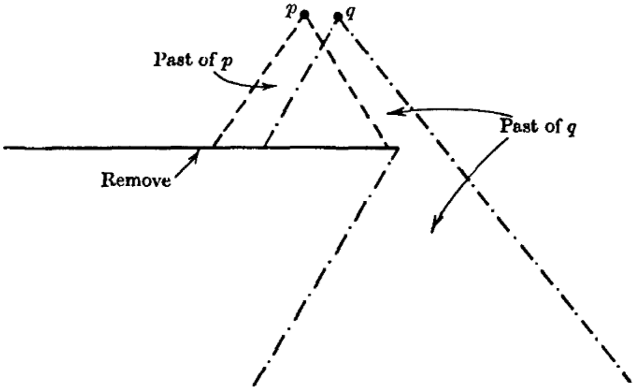


FIGURE 42. A small displacement of a point from p to q results in a large change in the volume of the past of the point. Light cones are at $\pm 45°$ and a strip has been removed as shown.

(taken from Hawking and Ellis 1973)

Theorem (Lesourd 2019)

Let $\mathcal{M} = (M, g_{ab})$ be a chronological spacetime with timelike asymptotic boundary $\mathscr{I}^+$ as in Marolf's argument, having topology $V \times \mathbb{R}$, such that:

1. *there is a non-trivial black hole region and event horizon;*
2. $\dot{I}^-(\mathscr{I}^+) \subset \overline{I^-(\Sigma)}$*, where Σ is a complete cross-section of $\mathscr{I}^+$, i.e., a spacelike submanifold of $\mathscr{I}^+$ with topology V.*

Then $\mathcal{M}$ is causally discontinuous.

Condition 2 captures the idea that the event horizon persists only up to a finite "moment of time" in the interior of the spacetime, *i.e.*, that the black hole evaporates.

(I slightly simplify (!) the formulation, but the original statement of the theorem implies this one.)

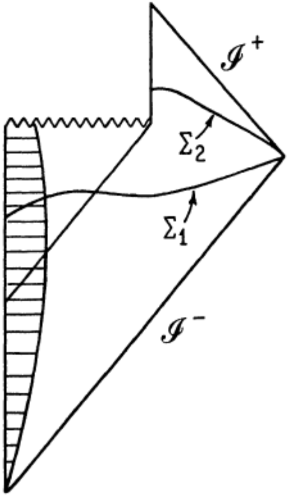


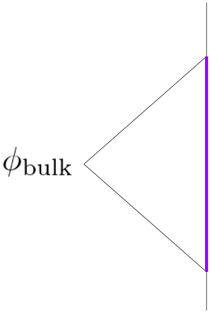
Figure 7.3. A conformal diagram illustrating the phenomenon of loss of quantum coherence in a spacetime in which black hole evaporation occurs.

(taken from Wald 1994)

now, recall this crucial step in Marolf's argument:

Axiom #4: There are other nontrivial operators in the algebra that can be excited to form a black hole [AdS QFT]

e.g. the boundary value of a scalar field $\phi(t) \in \mathcal{A}_{t,\Delta t}$



it is known how to solve for a field (outside of any horizons) in terms of integrals of boundary limiting values ϕ_{bulk}

sideways Cauchy problem subtle but basically OK

Hamilton-Kabat-Lifschytz-Lowe (free fields)
interacting case done perturbatively in 1/N
(should be good near infinity)

note we just need ***some*** *nontrivial field operator,*
(other than vacuum symmetry generators like H)
on a small boundary interval

⇒ "sideways Cauchy problem subtle but basically OK"—*only assuming that the interior of the spacetime is causally well behaved!!!*

– that is to say, assuming that the interior is not, *e.g.*, causally discontinuous because a black hole has badly evaporated

- the boundary theory must capture the phenomena of the entire interior, otherwise there is the possibility that a black hole evaporates non-unitarily in a way that does not register on the boundary
- thus, in order for the boundary theory to capture the phenomena of the entire interior, there must be observables capable of affecting every region of the interior
- otherwise, information about what happens in that region is not necessarily recoverable at the boundary
- *but that is what causal discontinuity calls into question*

cannot reject problem by claiming we need full theory of QG to know what happens around "the evaporation point":

- Lesourd's theorem shows causal discontinuity occurs *arbitrarily far* from *any* neighborhood of "the evaporation region"
- in particular, the spacetime is causally discontinuous in regions asymptotically far from any neighborhood of "the evaporation region"...
- in a way completely independent of details of classical geometry or QG weirdness in neighborhood of evaporation region

Conjecture

Unitary ("Schrödinger/Heisenberg-like") evolution is impossible in a causally discontinuous spacetime.

"some reasonable, and reasonably comprehensive, class of parabolic and hyperbolic PDEs (including unitary evolution of QFs for reasonable Hamiltonians) do not have well-posed initial-value formulations in causally discontinuous spacetimes" there is, of course, a way around the problem: one accepts the argument by noting that

1. if an adequate underlying theory of QG is pervasively promiscuously non-local
2. so that information characterizing *any* "small region around the evaporation region" can be (at least in principle) recovered from information characterizing *any* small region asymptotically far away
3. then there can be no true causal discontinuity

this, I take it, is the pill those who like the argument indeed swallow—whether bitter or not a matter of personal taste (and *de gustibus non disputandum est*[8])

8. Except that my taste is better than everyone's else, as an objective fact. Obvs.

I balk, for three reasons:

1. it is not clear to me how pervasively promiscuous non-locality at a fundamental QG level can efface manifest causal discontinuity at the level of classical spacetime geometry in regions where curvature can be arbitrarily small
2. and, again, they've destroyed the village (QFT) to save the village...
3. failure of unitarity for a theory of QG seems to me a less radical departure from well established physics than a non-locality that has more or less every region of spacetime encoding information about every other other region, no matter whether to the past or future, or spacelike-related, no matter how distant...
4. and no matter how weak gravitational effects may be
5. but most importantly—this would be a *profoundly radical conclusion* to draw about fundamental theory from what is, at bottom, only a semi-classical, effective field-theoretic description

but revolutions in physics *are* radical!

that's life in the big QG city

deal with it

– *BUT* this would be more radical than past revolutions, I claim,

– because in the past, when profoundly radical new ideas were introduced (Newton's Second Law and universal gravity, the electromagnetic field, relativity of simultaneity, non-commutativity of operators, *etc.*)...

– they were *inspired by and in response to* empirical data that could not otherwise be explained,

– and, more important, their theoretical development was *constrained and guided by* that empirical data

$\Rightarrow$ this is most assuredly not the case here

– in the end, I see three viable alternatives:

1. accept non-unitarity at in the semi-classical regime, reject the Cascade Problem
2. accept non-unitarity at in the semi-classical regime, accept the challenge of the Cascade Problem, try to explain how a deeper theory does or does not maintain unitarity
3. require or argue of unitarity in the semi-classical regime, accept radical consequences for our understanding of currently entrenched physics (in particular, QFT)

– I don't think there is a principled, knockdown argument for any of them

– how one moves forward largely dictated by one's *epistemic style*:

- try to be as conservative and epistemically secure as possible in speculative theorizing, retain as much as possible of existing understanding of entrenched physics
- embrace radical novelty in speculative theorizing as the way to move forward

each has its virtues and demerits

can we now understand the shrugging and freaking out better?

one plausible answer for why to freak out:

- in black-hole evaporation there is an articulated dynamical—physical!—mechanism that directly yields violation of unitarity
- in clear opposition to standard EFT calculations. . . [9]
- in which violation of unitarity is manifestly an artifact of the approximations and idealizations involved in the mathematical construction of the EFT formalism. . .
- not reflecting or representing anything physical
- also, it happens at arbitrarily low energies (large black holes), where one *expects* the "high-energy quantum gravitational modes" one ignores in SCG (whatever they may be) to be essentially non-existent

9. Also: in clear opposition to "measurement collapse"!

moreover...

- any possible failure of unitarity follows from geometry of classical spacetime background...
- *NOT* the dynamics of the quantum field itself
- in particular, *global* features of the large-scale structure of such spacetimes responsible for failure of unitarity
- one can plausibly expect, moreover, such global features insensitive to any QG effects near evaporation region...
- since those effects are appropiately localized in a small spacetime region (as Lesourd's theorem makes precise)—
- on pain of truly mind-boggling non-locality in QG

– THUS: onset of “new physics” here (as one expects from a breakdown of an EFT) heralded by *complexity in the causal structure* of spacetime (possibly underwritten by extreme non-locality in the QG regime)...

– by high *entropy*, *not* by high *energy/momentum* of “gravitational phenomena”, as for loss of unitarity in pedestrian EFTs in QFT

$$\Downarrow$$

– attempting to draw lessons about underlying QG from our understanding of SCG as an EFT is parlous at best...

$\Rightarrow$ in so far as its structure as an EFT seems radically different from EFTs we know and love in QFT

So, again: SCG is just an EFT—why get fussed about problems in an EFT?

(especially one whose character we clearly do not yet understand)

- we want to use SCG as basis for arguments whose conclusions we want to have confidence in—we want, essentially, to use it as part of an evidential network to buttress the assertability of claims in BHT (*inter alia*)
- but we have no entrenched empirical knowledge about SCG
- $\Rightarrow$ *it can't confer confirmation...*
- because one of essential elements of confirmation is that anything that gets it can then be used as evidence for other claims
- but that is exactly what Hawking radiation, BHT and SCG and such cannot do...
- at least not in our current epistemic state
- $\Longrightarrow$ we must be careful in trying to use SCG to draw fundamental lessons!

what is going wrong? my diagnosis:

we have no epistemic control over SCG and BHT...
and many physicists seem not to acknowlege it

epistemic control includes at a minimum:

1. that the theory must have an *empirical rendition*: some understanding of how the theory can make contact with empirical data that we have in hand, or that we can foresee how to acquire in a way compatible with the experimental component of the current epistemic state
2. to understand what will and will not count as evidence in favor and against, and why
3. to understand the physical theory's regime of applicability and so, *a fortiori*, its breakdown scales
4. to understand the asymptotics of physically important quantities and dynamics
5. to understand what approximations and idealizations are justified, and why
6. to understand what forms of argumentation (*e.g.*, heuristic, perturbative, *etc.*) are legitimate in different investigative contexts in the ambit of the theory, and why
7. to understand the conditions under which we do and do not not need explicit schematization of the observer

8. to understand as well what is the physical significance of the different and various kinds of the structures, entities, components, ..., of the theory's formalism
9. and to understand the kinds of physicality those parts all respectively can have, and the circumstances in which they respectively can have them
10. to understand the theory's relations to other valuable theories (theories precedent to the theory itself, and theories laterally and antecedently related in an epistemic, conceptual or practical sense)—not just by approximation and limiting relations and such, but also *relations among concepts*
11. when there is more than one formulation of the theory and they differ to a degree that their interrelations are not perspicuous (as in SCG and BHT!), to understand how they relate, formally, conceptually and interpretively, in such a way as to treat the same family of physical systems

perhaps the most important and severe problems here are:

1. approximations and idealizations:
    - we deal here with a *new type* of approximation/idealization, even at the purely classical level. . .
    - not values of quantities ($1 \sim 10^3$ for the astrophysicist), nor localized configurations (sphere for a chicken, for the condensed matter physicist), nor . . .
    - here, we substitute one *global spacetime structure* for another
    - we have *precious little* empirical experience doing this (EHT, LIGO/Virgo/KAGRA)
    - and there, we have physically compelling justifications ("Kerr is idealization of isolated system")
    - we have nothing like that for the Hawking effect
2. intimately related: little to no understanding of regime of applicability and breakdown scales
3. little to no understanding of physical significance of many of the most important entities and structures of the formalisms
4. little to no understanding of how all the different frameworks relate to each other

– there is an active danger in assuming that the semi-classical regime is now well understood

– and that its results may be used as the touchstone for testing programs of QG. . .

– that recovering those results acts as a minimal criterion of adequacy, if not something epistemically stronger. . .

– if it turns out not to have been right, all of QG has been wasted effort

it makes sense to hedge our bets

- I am not saying we shouldn't use SCG/BHT as grounds for *speculative* investigation
- I am rather saying that we should be more critical, more skeptical, and more modest in our understanding of the epistemic control we have for it. . .
- both with regard to our understanding of it. . .
- and with regard to our confidence in using it as the ground for further investigation
- $\Rightarrow$ we should be clearer on our epistemic state with regard to it

recall the quotation from Maxwell: what we are doing here is exactly *not* approaching the questions

> *not trusting to the native penetrating power of our own minds, but trained by a long-continued adjustment of our modes of thought to the facts of external nature.*

for our minds have not been trained
by external nature for studying
these phenomena!

A Knight of Faith? Or a Lost Soul?

- Is there a consistent picture of spacetime geometry, matter, and their interaction in the framework of SCG?
- Can such a conceptually and mathematically problematic framework give good results, and, if so, how?
- Why have faith in results from a framework with such manifest, serious, unresolved problems?
- How to make progress in important parts of theoretical physics if one doesn't have faith?

Make the Leap of the Absurd?
or
Remain a Skeptic and Be Damned?

I Recommend. . .

Socratic Irony

a dialectical process of moving from skepticism to faith, and back again, as our epistemic circumstances evolve under constant questioning, knowing that we do not know

Akers, Chris, Netta Engelhardt, and Daniel Harlow. 2020. "Simple Holographic Models of Black Hole Evaporation". *Journal of High Energy Physics* 2020:32. arXiv:1910.00972 [hep-th], doi:10.1007/JHEP08(2020)032.

Akers, Chris, and Geoff Penington. 2022. "Quantum minimal surfaces from quantum error correction". *SciPost Physics* 12 (5): 157. arXiv:2109.14618 [hep-th], doi:10.21468/SciPostPhys.12.5.157.

Almheiri, Ahmed, Netta Engelhardt, Donald Marolf, and Henry Maxfield. 2019. "The Entropy of Bulk Quantum Fields and the Entanglement Wedge of an Evaporating Black Hole". *Journal of High Energy Physics* 2019:63. arXiv:1905.08762 [hep-th], doi:10.1007/JHEP12(2019)063.

Almheiri, Ahmed, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini. 2021. "The Entropy of Hawking Radiation". *Review of Modern Physics* 93 (3, Jul–Sep): 035002. doi:10.1103/RevModPhys.93.035002.

Almheiri, Ahmed, Donald Marolf, Joseph Polchinski, and James Sully. 2013. "Black Holes: Complementarity or Firewalls?" *Journal of High Energy Physics* 2013 (02): 062. arXiv:1207.3123v4 [hep-th], doi:10.1007/JHEP02(2013)062.

Bardeen, James M. 2014. "Black Hole Evaporation without an Event Horizon". arXiv:1406.4098 [gr-qc].

Bousso, Raphael. 2025. "Firewalls From General Covariance". arXiv:2502.08724 [hep-th].

Chen, Hong Zhe, Zachary Fisher, Juan Hernandez, Robert C. Myers, and Shan-Ming Ruan. 2020. "Information Flow in Black Hole Evaporation". *Journal of High Energy Physics* 2020 (3): 152. arXiv:1911.03402 [hep-th], doi:10.1007/JHEP03(2020)152.

Chen, P., Y. Ong, and D.-H. Yeom. 2015. "Black Hole Remnants and the Information Loss Paradox". *Physics Reports* 603 (11): 1–45. arXiv:1412.8366 [gr-qv], doi:10.1016/j.physrep.2015.10.007.

Giddings, Steven B. 1992. "Black Holes and Massive Remnants". *Physical Review D* 46 (4, 15 Aug): 1347–1352. arXiv:hep-th/9203059, doi:10.1103/PhysRevD.46.1347.

———. 2006. "Black Hole Information, Unitarity, and Nonlocality". *Physical Review D* 74 (10, 17 Nov): 106005. doi:10.1103/PhysRevD.74.106005.

Good, Irving J. 1971. "46656 Varieties of Bayesians (#765)". *American Statistician* 25 (5, January): 62–63. Reprinted in Good (1982), pp. 20-21.

———. 1982. *Good Thinking: The Foundations of Probability and Its Applications.* Minneapolis: University of Minnesota Press.

Harlow, Daniel, and Patrick Hayden. 2013. "Quantum Computation vs. Firewalls". *Journal of High Energy Physics* 2013 (06): 085. arXiv:1301.4504 [hep-th], doi:10.1007/JHEP06(2013)085.

Hawking, S. W. 1976a. "Breakdown of Predictability in Gravitational Collapse". *Physical Review D* 14 (10, Nov): 2460–2473. doi:10.1103/PhysRevD.14.2460.

———. 1976b. "Erratum: Particle Creation by Black Holes". *Communications in Mathematical Physics* 46 (Jun): 206. [Erratum to: Hawking (1975)], doi:10.1007/BF01608497.

Hawking, S. W., and G. F. R. Ellis. 1973. *The Large Scale Structure of Space-Time.* Cambridge: Cambridge University Press. doi:10.1017/CBO9780511524646.

Hawking, S. W., and R. K. Sachs. 1974. "Causally Continuous Spacetimes". *Communications in Mathematical Physics* 35 (4): 287–296. Freely available at http://projecteuclid.org/euclid.cmp/1103859625, doi:10.1007/BF01646350.

Hawking, S. W. 1975. "Particle Creation by Black Holes". *Communications in Mathematical Physics* 43 (Aug): 199–220. [Erratum: Hawking (1976b)], doi:10.1007/BF02345020.

Hayden, Patrick, and John Preskill. 2007. "Black Holes As Mirrors: Quantum Information in Random Subsystems". *Journal of High Energy Physics* 2007 (JHEP09(2007)): 120. doi:10.1088/1126-6708/2007/09/120.

Jacobson, Ted. 2005. "Introduction to Quantum Fields in Curved Spacetime and the Hawking Effect". Chapter 2 in *Lectures on Quantum Gravity,* edited by Andres Gomberoff and Donald Marolf, 39–89. Series of the Centro De Estudios Cienti'ficos. NY: Springer. arXiv:gr-qc/0308048. Based on lectures given at the CECS School on Quantum Gravity in Valdivia, Chile, Jan 2002, doi:10.1007/0-387-24992-3_2.

Kodama, Hideo. 1979. "Inevitability of a Naked Singularity Associated with the Black Hole Evaporation". *Progress of Theoretical Physics* 62 (5, Nov): 1434–1435. doi:10.1143/PTP.62.1434.

Lesourd, Martin. 2019. "Causal Structure of Evaporating Black Holes". *Classical and Quantum Gravity* 36 (2): 025007. arXiv:1808.07303 [gr-qc]., doi:10.1088/1361-6382/aaf5f8.

Marolf, D. 2009. "Unitarity and Holography in Gravitational Physics". *Physical Review D* 79 (4, 15 Feb): 044010. arXiv:0808.2842 [gr-qc], doi:10.1103/PhysRevD.79.044010.

Marolf, Donald. 2017. "The Black Hole Information Problem: Past, Present, and Future". *Reports on Progress in Physics* 80 (9, Jul): 092001. doi:10.1088/1361-6633/aa77cc.

Mathur, Samir D. 2005. "The Fuzzball Proposal for Black Holes: An Elementary Review". *Fortschritte der Physik* 53 (7–8): 793–827. arXiv: hep-th/0502050, doi:10.1002/prop.200410203.

Mathur, Samir D., and Madhur Mehta. 2024. "The Fuzzball Paradigm". arXiv:2412.09495 [hep-th].

Maxwell, James Clerk. 1870. "Address to the Mathematical and Physical Sections of the British Association". In Maxwell 1890, 215–229.

———. 1875b. "Attraction". In Maxwell 1890, 485–491.

———. 1875a. "Attraction: A Dictionary of Arts, Sciences, Literature and General Information". In *The Encyclopædia Britannica: A Dictionary of Arts, Sciences, Literature and General Information,* Ninth, edited by T. S. Baynes, volume III, 63–65. Edinburgh: Adam & Charles Black. Reprinted as Maxwell (1875b).

———. 1890. *The Scientific Papers of J. C. Maxwell.* Edited by W. D. Niven. Volume II. Cambridge: University of Cambridge Press.

Nomura, Yasunori, Jaime Varela, and Sean J. Weinberg. 2013. "Complementarity Endures: No Firewall for an Infalling Observer". *Journal of High Energy Physics* 2013 (03): 059. arXiv:1207.6626 [gr-qc], doi:10.1007/JHEP03(2013)059.

Page, Don N. 1993. "Information in Black Hole Radiation". *Physical Review Letters* 71 (23, 6 Dec): 3743–3746. doi:10.1103/PhysRevLett.71.3743.

Penington, Geoffrey. 2019. "Entanglement Wedge Reconstruction and the Information Paradox". arXiv:1905.08255 [hep-th].

Polchinski, Joseph. 2017. "The Black hole Information Problem". Chapter 6 in *New Frontiers in Fields and Strings,* edited by Joseph Polchinski, Pedro Vieira, and Oliver DeWolfe, 353–397. Singapore: World Scientific. Presented at the 2014–15 Jerusalem Winter School and the 2015 TASI. arXiv:1609.04036 [hep-th], doi:10.1142/9789813149441_0006.

Susskind, Leonard, Lárus Thorlacius, and John Uglum. 1993. "The Stretched Horizon and Black Hole Complementarity". *Physical Review D* 48 (8, 15 Oct): 3743–3761. arXiv:hep-th/9306069v2., doi:10.1103/PhysRevD.48.3743.

Unruh, W. G. 1976. "Notes on Black Hole Evaporation". *Physical Review D* 14 (4, 15 Aug): 870–892. doi:10.1103/PhysRevD.14.870.

Unruh, William G., and Robert M. Wald. 2017. "Information Loss". *Reports on Progress in Physics* 80 (9): 092002. arXiv:1703.02140 [hep-th], doi:10.1088/1361-6633/aa778e.

Wald, Robert M. 1984. "Black Holes, Singularities and Predictability: Essays in Honor of the 60th Birthday of Bryce S. DeWitt". In *Quantum Theory of Gravity: Essays in Honor of the 60th Birthday of Bryce S. DeWitt,* edited by S. Christensen, 160–168. Boca Raton, FL: CRC Press.

———. 1994. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics.* Chicago: University of Chicago Press.