On Information-Loss Paradoxes, Holographic Proofs of Unitarity, Causal Structure and Non-Locality

— or —

Why Marolf's Argument is Ingenious and Elegant but Probably Wrong

(in which I shall attempt to anger Wald at least once)

Erik Curiel

Munich Center For Mathematical Philosophy (LMU Munich)

Black Hole Initiative (Harvard University)

Quantum Information Structure of Spacetime Consortium

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

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 312032894, CU 338/1-2.
(Funded by the German Research Foundation (DFG) – project number 312032894, CU 338/1-2.)

Apology

Ἀπολογία

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
> "Address to the Mathematical and Physical Sections of the British Association"

philosophy provides physics 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
"Attraction" (*Encyclopædia Britannica*, edition IX)

in a field such as black hole thermodynamics, and semi-classical gravity more generally—where we have not only no empirical experience to test our theorizing, but, much more importantly (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, again 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

Outline

‘Information Loss’ Is Said in Many Ways

Marolf’s Boundary Unitarity Argument

Some Causality Conditions and a Theorem on Causal Structure

Marolf’s Dilemma

two notions of evolution

QFT

1. initial data: quantum state ψ_t defined on a Cauchy slice Σ_t
2. evolution between surfaces governed by standard Schrödinger unitary evolution:
$$|\psi_t\rangle = \hat{U}_t|\psi_0\rangle$$
where "$\hat{U}_t = \exp(-it\hat{H})$"

GR

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

one might have thought the central issue of the Information-Loss Paradox 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 the Schrödinger equation

(the semi-classical Einstein field equation (SCEFE) doesn't do the job, since it deals only with $\langle \hat{T}_{ab} \rangle$)

in the context of evaporating black hole spacetimes, there are 4 questions commonly posed, not always clearly distinguished from each other, and none clearly equivalent to—or even clearly related to—each other:

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

Page-curve problem does the entropy of Hawking radiation decrease at late times during evaporation?

final-state problem is Hawking radiation (*i.e.*, the quantum state of the field) in a pure state after evaporation ends?

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)?

- they all have more or less vaguely to do with the idea of unitarity—whether unitary evolution consistently holds in QFT for an evaporating black hole
- all complicated by fact that in scattering theory ‘unitarity’ has 2 completely different meanings:
  (1) conservation of probability
  (2) evolution from pure states to pure states
- neither straightforwardly related to standard definition of ‘unitary’ for self-adjoint operator on Hilbert space

this ambiguity comes out most clearly in the “standard” argument about the Hawking problem, and its relative the final-state problem

let’s focus for the moment on the final-state problem

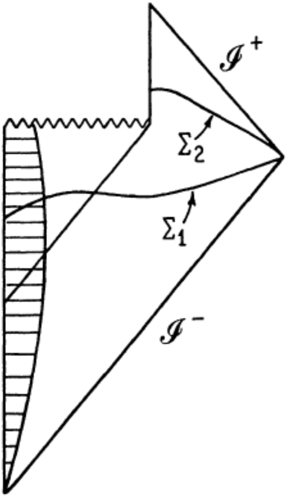


Figure 7.3. A conformal diagram illustrating the phenomenon of loss of quantum coherence in a spacetime in which black hole evaporation occurs.

(taken from Wald 1994)

summary

1. by Cauchy evolution, a pure state on Σ_1 develops into a mixed state on Σ_2
2. by Schrödinger evolution, 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
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

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

—addressing some combination, I think, of the final-state problem and the recovery problem, using 'unitarity' in the sense of the definition 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

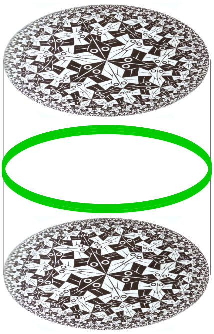
and it does so in a way that brings out the problem with clarity (because the argument itself is so crisp and clear)

—or, from a different perspective, it makes perspicuous some fundamental, otherwise hidden and severe seeming 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

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:

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

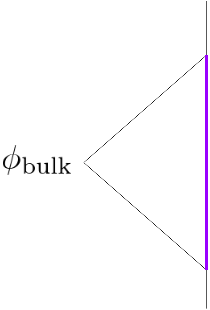
anything that can be measured at t_1 can also be measured at t_2,

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 $\phi(t) \in \mathcal{A}_{t,\Delta t}$



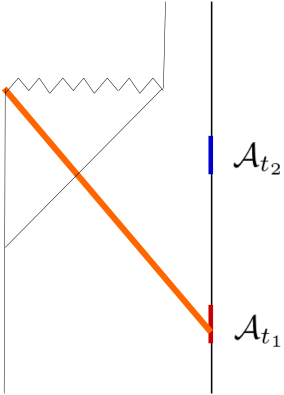
it is known how to solve for a field (outside of any horizons) in terms of integrals of boundary limiting values ϕ_{bulk}

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

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]

'Information Loss' Is Said in Many Ways

Marolf's Boundary Unitarity Argument

Some Causality Conditions and a Theorem on Causal Structure

Marolf's Dilemma

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 will introduce a few causality conditions and a theorem by Lesourd (2019), and show where in Marolf's argument the implicit assumption is made without justification 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

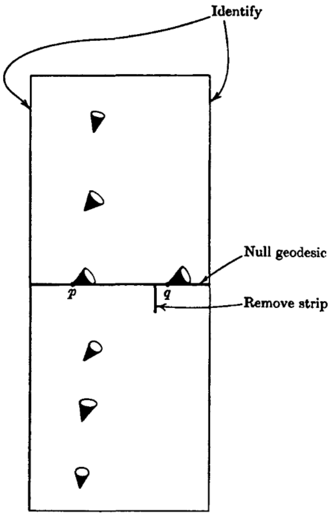


FIGURE 37. A space in which the causality and past distinguishing conditions hold everywhere, but the future distinguishing condition does not hold at p or q (in fact, $I^+(p) = I^+(q)$). The light cones on the cylinder tip over until one null direction is horizontal, and then tip back up; a strip has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

(taken from Hawking and Ellis 1973; the spacetime is reflecting)

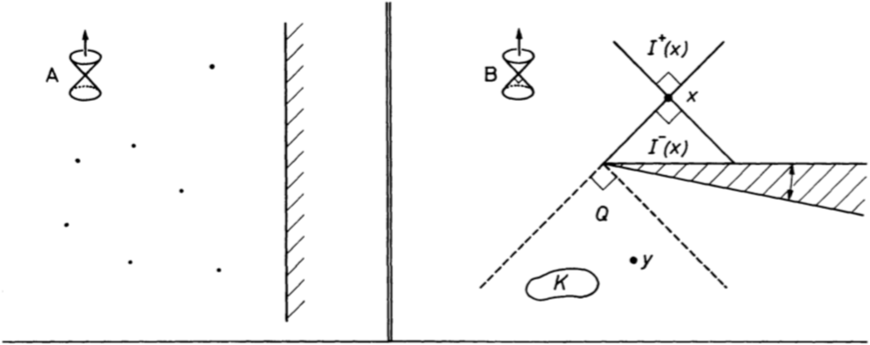


Fig. 1.2. The spacetime A is reflecting; the spacetime B is not

(taken from Hawking and Sachs 1974; both spacetimes are distinguishing)

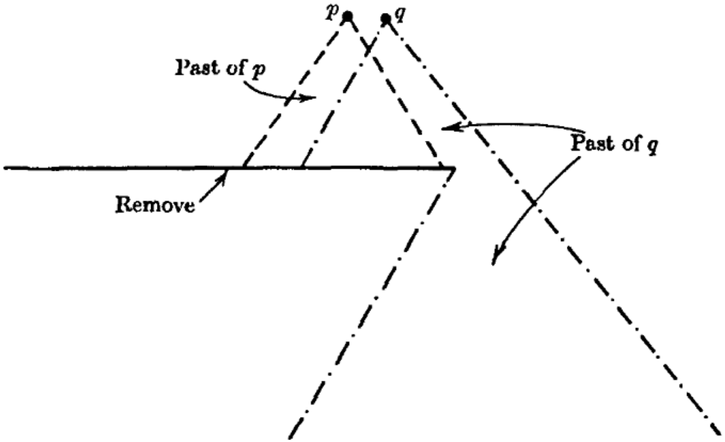


Figure 42. A small displacement of a point from p to q results in a large ohange in the volume of the past of the point. Light cones are at $\pm 45°$ and a strip has been removed as shown.

(taken from Hawking and Ellis 1973; the spacetime is causally discontinuous)

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.)

an evaporating black hole spacetime is causally discontinuous

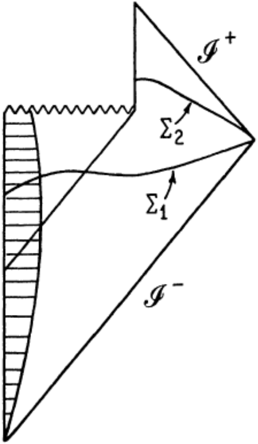


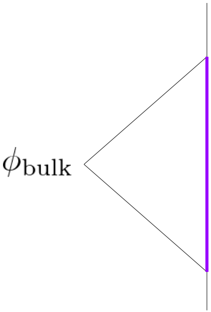
Figure 7.3. A conformal diagram illustrating the phenomenon of loss of quantum coherence in a spacetime in which black hole evaporation occurs.

(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 $\phi_{\rm bulk}$

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

"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

one cannot reject this problem by claiming that we need a full theory of QG to know what happens around "the evaporation region":

- Lesourd's theorem shows the causal discontinuity to occur 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 the details of the geometry in the neighborhood of the evaporation region

Conjecture

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

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*)

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. 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, no matter how distant
3. 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

now, this is 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), it was inspired by and in response to empirical data that could not otherwise be explained

this is most assuredly not the case here

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.

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.

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.

Wald, Robert M. 1994. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics.* Chicago: University of Chicago Press.