A Primer on Black Hole Thermodynamics and the Hawking Effect

XXIV International Summer School in Philosophy of Physics
"Black Holes and the Information Loss Paradox"
7.–10. Jun 2021
Università degli Studi di Urbino "Carlo Bo"

Erik Curiel

Munich Center For Mathematical Philosophy
LMU Munich

Black Hole Initiative
Harvard University

erik@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.)

Outline

Black Holes As a Way of Life

Event Horizon Telescope: M87

why black holes matter

1. they're incredibly cool

2. they're incredibly weird: can teach us much about conceptual structure of general relativity, what is (in some sense) physically possible in the world

why black holes matter (cont.)

3. black hole thermodynamics: one of most important, central, and fruitful fields of study in theoretical physics, closely connecting disciplines once seen as largely independent:
   i. general relativity
   ii. quantum field theory
   iii. thermodynamics
   iv. cosmology
   v. statistical mechanics
   vi. particle physics
   vii. fluid dynamics
   viii. condensed matter
   ix. quantum optics
   x. . . .

why black holes matter (cont.)

4. widely considered best source of clues to a more fundamental theory of quantum gravity:
   i. simplest non-trivial "purely gravitational" system to try to quantize
   ii. their laws mix dependence on and independence from Einstein Field Equation (EFE) in funny way, so get both pure kinematics and dynamics
   iii. indirect access to micro-degrees of freedom and their dynamics by treating thermodynamics as arising from statistical mechanics

but what are they?

Field	Core Concepts
astrophysics	• compact object • region of no escape • engine for enormous power output
classical relativity	• causal boundary of the past of future null infinity (event horizon) • apparent horizon (all outgoing light rays "get turned around") • quasi-local horizon
mathematical relativity	• apparent horizon • singularity

Table: the core concepts common to different fields for characterizing black holes (from Curiel 2019)

Field	Core Concepts
semi-classical gravity	• same as classical relativity • thermodynamical system of maximal entropy
quantum gravity	• particular excitation of quantum field • ensemble or mixed state of maximal entropy • no good definition to be had
analogue gravity	• region of no escape for finite time ("long" compared to characteristic time) • same for low energy modes ("low" compared to characteristic energies)

Table: the core concepts common to different fields for characterizing black holes, cont. (from Curiel 2019) different properties one may demand of a "black hole"

- possesses a horizon that satisfies the four laws of black hole mechanics
- possesses a locally determinable horizon
- possesses a horizon that is, in a suitable sense, vacuum
- is vacuum with a suitable set of symmetries
- defines a region of no escape, in some suitable sense, for some minimum period of time
- defines a region of no escape for all time
- is embedded in an asymptotically flat spacetime
- is embedded in a topologically simple spacetime
- encompasses a singularity
- satisfies the No-Hair Theorem
- is the result of evolution from initial data satisfying an appropriate Hadamard condition (stability of evolution)

different properties one may demand of a "black hole" (cont.)

- allows one to predict that final, stable states upon settling down to equilibrium after a perturbation correspond, in some relevant sense, to the classical stationary black hole solutions (Schwarzschild, Kerr, Reissner-Nordström, Kerr-Newman)
- agrees with the classical stationary black hole solutions when evaluated in those spacetimes
- allows one to derive the existence of Hawking radiation from some set of independent principles of interest
- allows one to calculate in an appropriate limit, from some set of independent principles of interest, an entropy that accords with the Bekenstein entropy (*i.e.*, is proportional to the area of a relevant horizon, with corrections of the order of $\hbar$)
- possesses an entropy that is, in some relevant sense, maximal
- has a lower-bound on possible mass
- is relativistically compact
- formed by the gravitational collapse of matter

different subsets of these properties are used in different contexts in different investigations, often in the same field

but they are jointly inconsistent

$\Rightarrow$ no definition can accommodate all actual uses of the concept in contemporary physics serious methodological and epistemological (and ontological?) problems, a small taste:

1. event horizon is "global": makes implicit reference to "all future time"
2. that, and fact that nothing locally distinguished about event horizon $\Rightarrow$ *no local measurements can ever determine its location, much less whether there is one*
3. indeed, our universe is not "asymptotically flat", so it cannot have anything like an event horizon
4. $\Rightarrow$ the definitions used in classical general relativity *cannot* match those in astrophysics
5. *all* properties of black holes, all theorems, relied on by astrophysics assume event horizon—still applicable in real world?
6. so: how is SgrA* similar to and how different from a Kerr black hole?

even more deep and interesting epistemological and metaphysical problems in the offing, such as

- how empirical content of theories gets fixed
- what an appropriate semantics of theoretical terms can be
- whether there is a consistent ontology across theories
- ...

but, sadly, no time to go into any of it now

we will now try to be a bit more precise, laying down some definitions and stating some theorems, to give us a shared epistemic platform to hold onto when the physical winds start to blow rough and the metaphysical grounds in ceaseless turmoil seethe

a spacetime $\mathcal{M} = (M,\, g_{ab})$ is *temporally orientable* if there exists a continuous timelike vector field on it; equivalently, if one can choose one lobe of the null cone as "future" consistently over the entire spacetime; a *temporal orientation* is such a choice

from hereon, we assume every spacetime we deal with is temporally oriented, and a choice of orientation has been made

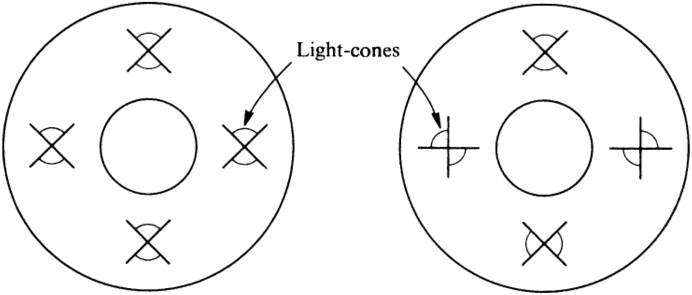


Figure 5.3. Two spacetimes based on the two-dimensional annulus. The first is time-orientable; the second is not.

(taken from Geroch and Horowitz 1979)

fix spacetime $\mathcal{M} = (M,\, g_{ab})$, $p, q \in M$, $S \subset M$

chronological future $I^+(p) := \{q : \exists$ future-directed timelike curve from p to $q\}$

causal future $J^+(p) := \{q : \exists$ future-directed causal curve from p to $q\}$

timelike related p and q are timelike related if $p \in I^+(q)$ or $q \in I^+(p)$

causally related p and q are causally related if $p \in J^+(q)$ or $q \in J^+(p)$

closed timelike curve (CTC) a timelike curve that intersects itself

closed causal curve (CCC) a causal curve that intersects itself

chronology $\mathcal{M}$ satisfies chronology (or is chronological) if it has no CTCs

causality $\mathcal{M}$ satisfies causality (or is causal) if it has no CCCs

all relevant definitions can, *mutatis mutandis*, be formulated for pasts, using the indicator '-' rather than '+' (*e.g.*, the chronological past $I^-(p)$)

fundemental properties and results

1. $I^+(p)$ is open
2. $J^+(p)$ is not necessarily closed (if it is closed for all $p \in M$, then $\mathcal{M}$ is *causally simple*)
3. causality implies chronology
4. chronology does not imply causality

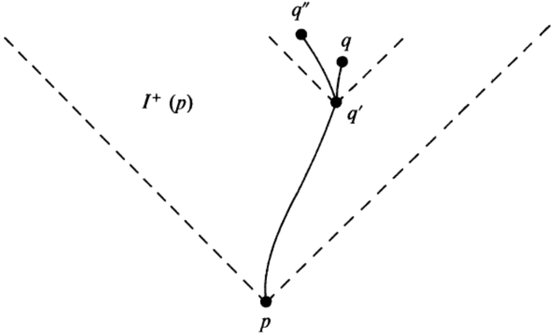


Figure 5.5. $I^+(p)$ is an open set. The set of all points q'' reached by future-directed timelike curves from q' form a neighborhood of q which lies inside $I^+(p)$.

(taken from Geroch and Horowitz 1979)

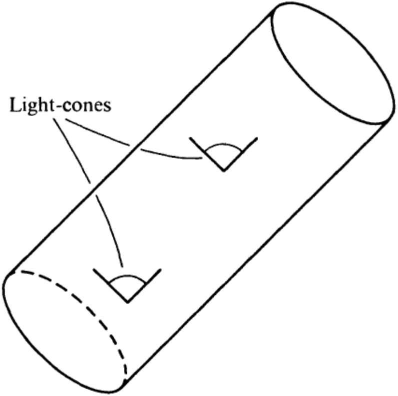


Figure 5.13. Strip of two-dimensional Minkowski spacetime between $t = 0$ and $t = 1$, with the point $(0, x)$ identified with $(1, x + 1)$. This spacetime has no closed timelike curves, but there are closed null curves.

(taken from Geroch and Horowitz 1979)

achronal S is achronal if no two of its points are timelike related

future domain of dependence for closed achronal S, $D^+(S) := \{p :$ every past-inextendible causal curve through p intersects $S\}$

domain of dependence $D(S) := D^+(S) \cup D^-(S)$

Cauchy surface/slice a closed, achronal S such that $D(S) = M$

globally hyperbolic a spacetime containing a Cauchy slice

Cauchy time function $t : M \to \mathbb{R}$ such that its levels $S_\tau = t^{-1}(\tau)$ are Cauchy slices

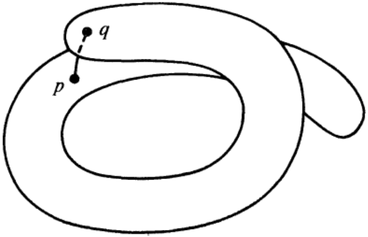


Figure 5.18. Spacelike submanifold of Minkowski spacetime (one spatial dimension suppressed) which is not achronal, e.g. because p precedes q.

(taken from Geroch and Horowitz 1979)

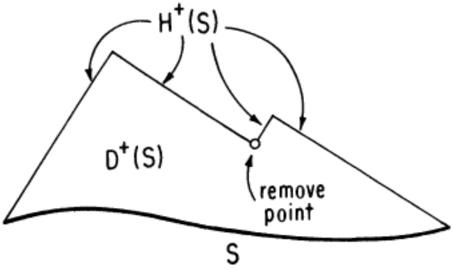


Fig. 8.11. A spacetime diagram showing the future domain of dependence, $D^+(S)$, and Cauchy horizon $H^+(S)$ of a particular closed achronal set S in Minkowski spacetime with a point removed.

(taken from Wald 1984b)

fundamental results about globally hyperbolic spacetimes

Theorem (Geroch 1970)

$\mathcal{M} = (M, g_{ab})$ *with Cauchy slice* S *is such that:*

1. S *is an embedded topological submanifold of* M *(from hereon, without loss of generality, we assume it to be a differential submanifold)*
2. $M = S \times \mathbb{R}$
3. M *is foliated by leaves diffeomorphic to* S*, defining a Cauchy time function on* $\mathcal{M}$
4. *every inextendible timelike curve intersects* S *exactly once*
5. $\mathcal{M}$ *is causal (indeed, it is stably causal—wait for Wednesday)*
6. $\mathcal{M}$ *is causally simple*
7. M *is the disjoint union* $I^-(S) \cup S \cup I^+(S)$
8. $\forall p, q \in M$, $J^-(q) \cap J^+(p)$ *is compact*

- the future domain of dependence is the set of all points p such that a complete specification of physical data on S determines the state of spacetime at p—"every possible causal influence on p registered on S"
- a Cauchy slice can be thought of as "all of space at a moment of time"
- thus, globally hyperbolic spacetimes are appropriate settings for the initial-value formulation of GR: specify appropriate initial data on a Cauchy slice, and that determines the state of the entire spacetime
- Minguzzi and Sánchez (2008, p. 43, arXiv version):
  The compactness of the diamonds $J^-(q) \cap J^+(p)$ can be interpreted as "there are no losses of information/energy in the spacetime". In fact, otherwise one can find a sequence $\{r_n\} \subset J^-(q) \cap J^+(p)$ with no converging subsequence. Taking a sequence of causal curves $\{\gamma_n\}$, each one joining p, r_n, q, the limit curve γ_p starting at p cannot reach q. This can be interpreted as something which is suddenly lost or created in the boundary of the spacetime. That is, a singularity (this sudden loss/creation) is visible from q—there are "naked singularities".

Geroch and Horowitz (1979, p. 235):

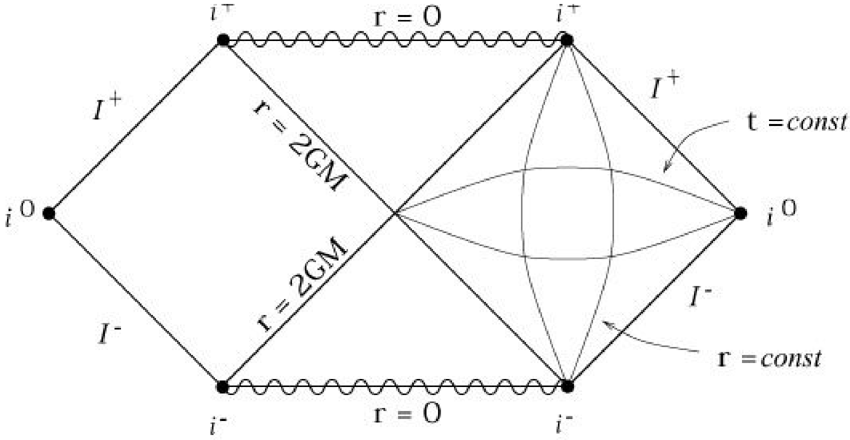
> *The subject of causal structure has the curious feature that there is a seemingly endless stream of plausible-sounding statements within its framework which are actually false. Frequently, these are exactly the statements one would like to have true in order to prove something interesting.*

for many problems in GR (*e.g.*, defining an "isolated system"), one wants to know how spacetime geometry behaves "infinitely far away—often, one wants to know that, in some appropriate sense, it "looks like Minkowski spacetime" (*i.e.*, is flat, geodesically complete and completely symmetric)

coordinate-based formulations, are unsatisfying and inadequate for many reasons, as is so often the case

as is so often the case, Penrose solved the problem for us (Penrose 1964) roughly speaking, a spacetime is *asymptotically flat* if an appropriate boundary, representing “points at infinity”, can be attached to it in a suitable way a little more precisely (but not much—the details are too technical), and with grievous abuse of notation and concepts, spacetime $\mathcal{M} = (M, g_{ab})$ is *asymptotically flat* if $\exists \bar{\mathcal{M}} = (\bar{M}, \bar{g}_{ab})$ with boundary such that:

1. $\exists$ conformal isometry $\psi : M \to \psi[M] \subset \bar{M}$, *i.e.*, ψ is a diffeomorphism on its image and $\bar{g}_{ab} = \Omega^2 g_{ab}$
2. $\dot{M} = \bar{M} - M$
3. $\exists \imath^0 \in \dot{M}$ such that $\dot{M} = \overline{J^-(\imath^0)} \cup \overline{J^+(\imath^0)}$ ("spatial infinity" is spacelike related to all points in M)
4. $\mathscr{I}^+ := \overline{J^+(\imath^0)}$ ("future null infinity" is generated by null geodesics emanating from $\imath^0$, and can be reached by complete null geodesics in M)
5. there exists an open neighborhood U of $\dot{M}$ such that the spacetime $(U, \bar{g}_{ab}|_U)$ is globally hyperbolic ("spacetime is well behaved far away from stuff")
6. complicated technical conditions on Ω and the space of null directions emanating from $\imath^0$



Penrose diagram of maximally extended Schwarzschild spacetime

energy conditions are a way of attempting to control the "unphysical" parts of the mathematics we use to represent the physical world—impose phenomenological criteria for deciding what counts as "physically reasonable" forms of matter

reflect anything about "fundamental" physics?

the peculiarity of energy conditions

1. of fundamental importance for the derivation of deep results
2. lack clear physical significance and interpretation
3. uncertainty about theoretical and epistemic character

postulate or derived consequence?

almost all fundamental propositions in general relativity admit interpretation as one or the other

not energy conditions

almost all not derivable from any other fundamental proposition

epistemic status?

- not “laws” on any standard construal
- nor empirical or inductive generalizations
- yet seem to represent structure at a deep level of our understanding of spacetime and quantum field theories

two kinds of formulation

Geometrical constraints on purely geometrical quantities

Physical constraints on T_{ab} itself

(we ignore a 3rd, the effective—see Curiel 2017a)

Null Energy Condition (NEC)

geometric for any null vector k^a, $R_{mn}k^m k^n \geq 0$: "gravity is attractive" (average radial acceleration of geodesic congruences)

physical for any null vector k^a, $T_{mn}k^m k^n \geq 0$: **?**

Weak Energy Condition (WEC)

geometric for any timelike vector ξ^a, $G_{mn}\xi^m \xi^n \geq 0$: **?**

physical for any timelike vector ξ^a, $T_{mn}\xi^m \xi^n \geq 0$: "ordinary energy density cannot be negative"

Dominant Energy Condition (DEC)

geometric

1. for any timelike vector ξ^a, $G_{mn}\xi^m \xi^n \geq 0$, and $G^a{}_n \xi^n$ is causal: **?**
2. for any co-oriented timelike vectors ξ^a and η^a, $G_{mn}\xi^m \eta^n \geq 0$: **?**

physical

1. for any timelike vector ξ^a, $T_{mn}\xi^m \xi^n \geq 0$, and $T^a{}_n \xi^n$ is causal: "WEC + no superluminal propagation of energy-momentum"
2. for any co-oriented timelike vectors ξ^a and η^a, $T_{mn}\xi^m \eta^n \geq 0$: "every observer measures total energy-momentum flux to be causal and co-oriented with her arrow of time"

What "Depends" on the Einstein Field Equation?

commonly said:

- geodesic theorems
- causal hierarchy
- singularity theorems
- most black hole theorems
- positive energy theorems (ADM and Bondi)
- "conservation of matter" (T_{ab} vanishes on closed, achronal set $\Rightarrow$ vanishes in domain of dependence)
- gravitational entropy bounds
- Topological Censorship
- Cosmic Censorship Hypothesis
- Lorentzian splitting theorems
- a given globally hyperbolic extension of a spacetime is maximal
- CMC foliations for spatially compact spacetimes
- cosmological "No Hair" theorems
- ...

I claim:

none of them do!

in very strong sense...

assume logical negation of the Einstein field equation and theorems still hold

Actual Role of Einstein Field Equation:

provides physical interpretation

gives physical interpretation to conditions on "purely geometrical" object

logical form of arguments:

1. assume condition on some geometrical object (Riemann, Weyl, Ricci, Einstein, . . .)—almost always an energy condition
2. derive theorem in purely mathematical way
3. invoke EFE to give physical interpretation of condition assumed in step 1 and result derived in step 2

Geroch and Horowitz (1979):

> *One would of course have to impose some restriction on the stress-energy of matter in order to obtain any singularity theorems, for with no restrictions Einstein's equation has no content. One might have thought, however, that only a detailed specification of the stress-energy at each point would suffice, e.g. that one might have to prove a separate theorem for each combination of the innumerable substances which could be introduced into spacetime. It is the energy condition which intervenes to make this subject simple. On the one hand it seems to be a physically reasonable condition on all types of classical matter, while on the other it is precisely the condition on the matter one needs for the singularity theorem.*

methodological and epistemological problem

given the fact that so many of the results relied on in black hole physics do not depend on the EFE but rather only on energy conditions, how much do the problems we will consider bear on the validity of GR, if at all, rather than solely on that of QFT and particular theories of matter?

BUT: much more makes GR a physical theory (as opposed to the mathematical theory of Lorentzian 4-geometries) than the EFE, matters whose interpretation, meaning and physical significance don't depend on the EFE:

1. the geodesic principle
2. tidal forces
3. gravitational waves
4. the principle of equivalence
5. diffeomorphism invariance
6. Killing fields
7. causal structure
8. affine structure
9. global topology
10. spacetime dimension
11. the manifold's differential structure
12. the distinction among timelike, spacelike and null vectors
13. Lie derivative
14. volume elements
15. spinor structure
16. and on and on

so there is in fact plenty of reason to think that we may still need to modify some of GR in trying to find an adequate theory of QG even if what we need to modify has nothing to do with the EFE

(thanks to Yichen Luo for asking me a question that pushed me to clarify this)

general relativity

- valid from middling to largest spatial and temporal scales, and from near-perfect vacuum to super-dense matter
- dynamical spacetime geometry
- no quantum effects
- extreme causal and topological weirdness, seemingly inconsistent with quantum mechanics

quantum field theory

- valid for meso-molecular (and smaller) matter at smallest spatial and temporal scales, and for highest and lowest energies
- static, flat spacetime geometry
- no gravity (*n.b.*: this doesn't follow from previous, *e.g.*, "soft gravitons")
- extreme quantum weirdness, seemingly inconsistent with general relativity (*n.b.*: really superposition, *not* necessarily entanglement)

thermodynamics

- valid for all matter at "ordinary" spatial and temporal scales, and for all masses and energies
- no gravity, no quantum effects
- temporal asymmetry and irreversibility, seemingly inconsistent with general relativity and quantum field theory

but how do they fit together?

- each conceptually and physically independent of the other two
- each finds application in regimes well separated from those of the others
- the most characteristic features of QFT (superposition, entanglement, uncertainty principle, non-locality) are in manifest tension if not outright contradiction with those of general relativity (dynamical spacetime geometry, causal weirdness, non-trivial topology)
- they both are in manifest tension if not outright contradiction with the most characteristic feature of thermodynamics, *viz.*, irreversibility and temporal asymmetry

let's face it: we do not know how they fit together
(or even whether they do[1])

1. at root of many problems: quantum theory is fundamentally linear, general relativity is non-linear

with one major exception. . .

when the effects of quantum fields are taken into account, black holes in general relativity, even though they are nothing more than regions of empty spacetime, appear to become true thermodynamical objects, with an associated physical temperature and entropy

$\Rightarrow$ black hole thermodynamics, a deep and hitherto unsuspected connection among our three best, most fundamental theories

$\Rightarrow$ raising many deep philosophical problems and questions, most of which philosophers have not yet begun to grapple with

traditional philosophical puzzles

general relativity

- the nature of spacetime (substantivalism versus relationalism)
- deterministic or indeterministic? (the Hole Argument, naked singularities)

quantum field theory

- the Measurement Problem
- non-locality (Bell's Theorem, "action at a distance")
- coherent ontology?

thermodynamics

- the status of the Second Law (empirical generalization? law of nature?)
- temporal asymmetry (arrow of time)
- reduction of thermodynamics to statistical mechanics

The Central Problem of Black Hole Thermodynamics

What does it mean to conceive of and treat black holes, in the presence of quantum fields, as thermodynamical systems?

It is no exaggeration to say that the Central Problem affects essentially every traditional philosophical problem of all three theories:

- restricting and refining how they can be cogently formulated
- changing the criteria for what may and may not count as satisfying answers
- suggesting new avenues of attack

AND it results in several entirely new, deep problems, independent of the traditional ones

importance of philosophical study

- without a doubt the most widely accepted, most deeply trusted results in theoretical physics in which those theories work together—but those theories are *prima facie* inconsistent
- AND no empirical access to those regimes
- $\Longrightarrow$ absolutely no experimental or observational evidence for any of it—why do we trust it?
- $\Longrightarrow$ investigations necessarily speculative in a way unusual even in theoretical physics
- $\Longrightarrow$ technically sophisticated physical questions inextricable from subtle philosophical considerations spanning ontology, epistemology, and methodology, again in a way unusual even in theoretical physics

Black hole thermodynamics and results concerning quantum fields in the presence of strong gravitational fields more generally are without a doubt the most widely accepted, most deeply trusted set of conclusions in theoretical physics in which general relativity and quantum field theory work together in seemingly fruitful harmony.

This is 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.

for the remainder of this talk, unless explicitly stated otherwise, we will use the classic definition of a black hole, based on the idea of an isolated system that is a “region of no escape” (in future lectures, we will expand our repertoire)

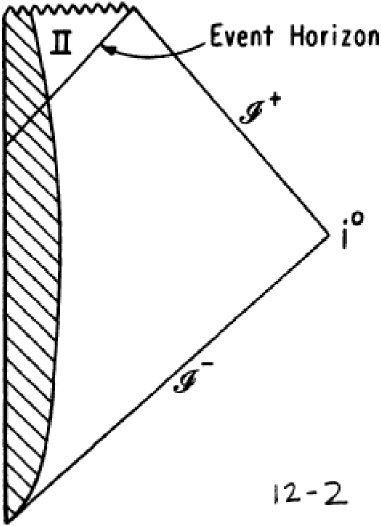
for this to make sense, we need a region such that, having reached it, one has indubitably “escaped”—gotten arbitrarily far away—from every other region in spacetime; but that is what the idea of asymptotic infinity captures

so the black hole region in a spacetime is the set of all points such that no future-directed causal curve from one can reach $\mathscr{I}^+$

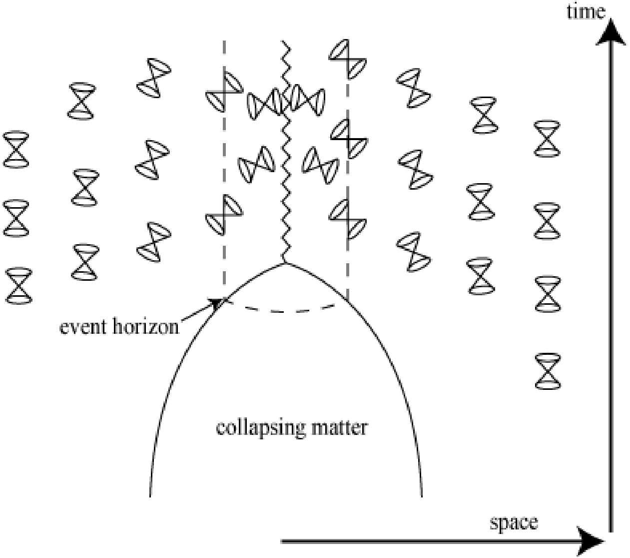
for an asymptotically flat spacetime $\mathcal{M} = (M,\, g_{ab})$:

black hole $B := M - J^-(\mathscr{I}^+)$

event horizon $H := \dot{B} = M \cap \dot{J}^-(\mathscr{I}^+)$



(taken from Wald 1984b)


time
event horizon
collapsing matter
space

beware a common misconception!

black holes aren't always super dense, super compact, super strong gravity:

- if all the stars in the Milky Way gradually aggregate towards the galactic center while keeping their proportionate distances from each other, they will all fall within their joint Schwarzschild radius long before they are forced to collide
- the average density of matter in that system would be infinitesimal by Terrestrial standards
- there is, in other words, nothing necessarily unphysical or mysterious or exotic about the interior of an event horizon formed from the aggregation of matter

first hint at black hole thermodynamics

No-Hair Theorem

A stationary electrovac black hole is completely characterized by three numbers, its mass, angular momentum and electric charge.

$\Rightarrow$ black hole "macrostate" completely independent of how it formed (compatible with *any* "microstructure")!

just like ordinary thermodynamical systems

Zeroth Law

Thermodynamics

The temperature T is constant throughout a body in thermal equilibrium.

Black Holes

The surface gravity κ is constant over the event horizon of a stationary black hole.

First Law (Energy Conservation)

Thermodynamics

change in energy = (temperature × change in entropy) + work done

$$(\mathsf{d}E = T\mathsf{d}S + p\mathsf{d}V + \Omega\mathsf{d}J)$$

Black Holes

change in mass = (surface gravity × change in area) + "rotational work"

$$(\delta M = \frac{1}{8\pi}\kappa\delta A + \Omega_{\mathsf{H}}\delta J)$$

Second Law

Thermodynamics

entropy never decreases ($\delta S \geq 0$) in any physical process

Black Holes

area of event horizon never decreases ($\delta A \geq 0$) in any physical process

Third Law (Nernst Theorem)

Thermodynamics

$T = 0$ is not achievable by any physical process

Black Holes

$\kappa = 0$ is not achievable by any physical process

“Minus First Law” (Brown and Uffink)

Thermodynamics

isolated thermodynamical systems tend to approach a unique equilibrium state

Black Holes

isolated, non-stationary black holes tend to settle down to a unique stationary state (Kerr-Newman spacetime)

quantum field theory on curved spacetime

perhaps surprising to learn that there is a consistent, rigorous theory of quantum fields on relativistic spacetimes (algebraic or axiomatic):

1. fixed background classical spacetime geometry
2. non-interacting quantum fields propagating as "test matter"—"free fields"

(also standard canonical and Lagrangian formulations, but they're *messy* and raise yet more mathematical and physical problems; some results known in axiomatic framework for interacting quantum fields in lower dimensions, but nothing of interest)

but freedom is scary...

semi-classical gravity (SCG)

shackle the fields to curvature with "back-reaction" $\Rightarrow$ semi-classical Einstein field equation (SCEFE)

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

classical Einstein tensor =
expectation value of stress-energy tensor as quantum operator

no completely rigorous mathematical theory, only standard formulations (Lagrangian, S-matrix, path integral, canonical, Hamiltonian-Jacobi, ...), perturbative warts, non-renormalizability, and all

fundamental question:

How does the sober, rigorous and precise Apollonian convocation of classical Lorentzian geometry and the exuberantly inexact and informal Dionysian fandango of quantum field theory come into mutually fruitful contact, so as to give the joy of material content to the former and the restrained discipline of consistent structure to the latter?

MANY severe technical, physical and conceptual problems for QFT-CST and SCG, but we must, regretfully, put them aside (I daren't even try to list them all)

Laws of Black Holes Mechanics. . .

or Thermodynamics?

classically, the analogy is purely formal (says orthodoxy[2]), *i.e.*, there is no consistent thermodynamics of classical black holes, because:

- classical black holes are perfect absorbers, emitting nothing $\Rightarrow$ temperature absolute zero irrespective of value of κ
- Geroch's infamous thought-experiment: use a classical black hole as heat sink to transform heat into work with 100% $\Rightarrow$ temperature absolute zero
- area nothing like entropy
- surface gravity nothing like temperature

(everything and its mother has EoM of SHO, but an alternating-current circuit is not physically a pendulum)

(ask me some time about Alain Connes's "Laws of Asshole Mechanics")

2. But see Curiel (2017b)—although Curiel himself believes it only about half the time.

but take quantum effects into account...

Hawking discovered that a black hole behaves like a perfect black body in the sense of ordinary statistical thermodynamics: in the presence of a quantum field in its vacuum state, thermal radiation with the Planckian power spectrum characteristic of a perfect blackbody at a fixed temperature equal to κ is generated in the neighborhood of the horizon of a stationary black hole. It glows like a lump of smoldering coal even though nothing, not even light, should be able to escape from it!

$\Rightarrow$ now we can take the formal analogy to be a truly physical one

Generalized Second Law

- seems easy to violate standard Second Law when black holes are around:
    1. throw favorite highly entropic system into black hole
    2. the entropy of the world outside the event horizon—a causally isolated system—spontaneously decreases
- $\Rightarrow$ Bekenstein proposed Generalized Second Law: *total* entropy, black hole (area) + ordinary matter outside, never decreases
- many powerful (purely theoretical!) arguments supporting it
- its validity is best argument we have for attributing a *physical* entropy to black holes—all other attempts I know of to argue for the attribution begs questions left and right

overwhelming consensus today is that black holes truly are thermodynamical objects, and the laws of black hole mechanics just are the laws of ordinary thermodynamics extended into a new regime, to cover a new type of physical system

(compare extension of thermodynamics to encompass electromagnetic blackbody radiation at end of 19th Century) but what on Earth can that mean?!

how can an empty, locally undistinguished region of spacetime have thermodynamical properties?

- difficult to think of two more different quantities than entropy and spatial area...
- unless it be temperature and surface gravity
- how can these possibly be *identical*?
- $\Rightarrow$ *deep* problem for conceptual understanding of inter-theory relations: "same" quantity as represented in different theories

also:

- laws of ordinary thermodynamics are empirical generalizations
- but laws of black hole mechanics are theorems of differential geometry
- $\Rightarrow$ how can they possibly be "the same"?

let's sketch a picture of Hawking radiation

I won't describe any particular derivation, as the rough, intuitive ones tend to be badly misleading, and the precise, rigorous ones are too technically demanding given the constraints of this lecture

I'll just sketch the basic ingredients any derivation requires, and state the generic conclusion

the setting: QFT-CST

1. a quantum field in an appropriate vacuum state treated as "test matter": doesn't contribute to curvature, propagates freely against the classic spacetime background
2. nonetheless, the quantum field and the spacetime geometry are not wholly independent, as the conformal and affine structures of the spacetime geometry guide the propagation and Cauchy development of the quantum field ("geodesic principle")
3. the spacetime geometry is a black hole, *i.e.*, an event horizon in the classic sense

the basic ingredients:

1. topological assumptions (*e.g.*, domain of outer communication is simply connected)
2. causal assumptions (*e.g.*, chronology)
3. stability assumptions ("small perturbations do not destroy the event horizon")
4. assumptions about asymptotic behavior and structure (instantaneous symmetries, "almost-conserved" quantities, energy fluxes that don't contribute to curvature)
5. arguments *cum* assumptions about the singularity structure (correlation functions) of the vacuum state for a given kind of quantum field near the event horizon (it should "look like" the singularity structure associated with the Minkowski vacuum state)
6. arguments that transforming the vacuum state to present it as it would appear to a static observer has a particular effect on the singularity structure, so that a characteristic pattern of mode-splitting occurs at the horizon, negative-energy modes falling in, positive-energy modes radiating away
7. the exponential schism between the motion of static and inertial observers then automatically translates into exponential scaling of the radiating modes in a Planckian thermal spectrum

conclusion:

- "at infinity", late-time static/inertial observers (they're the same thing there) measure a flux of excited modes in the ambient quantum field with a Planckian thermal distribution, encoding a temperature $= \dfrac{\kappa}{2\pi}$—"Hawking radiation"
- Hawking radiation is in fact thermalized throughout almost all the interior of spacetime, redshifted to higher temperature values as seen by static observers as one approaches closer to the event horizon

enormous conceptual problem with Hawking radiation (which is, oddly, almost never talked about): *it is not standard blackbody radiation!*

it is not generated by micro-degrees of freedom of the event horizon, like electromagnetic blackbody radiation of hot iron is caused by jiggling of its atoms and electrons

rather Hawking radiation is excited modes of *external* quantum field (the quantum field is not even *coupled* to the spacetime geometry, because we're treating it as test-matter!)

so why should we take it as a proxy for the *black hole's* temperature?

perhaps most puzzling of all

How did a theoretically predicted phenomenon (Hawking radiation), derived by combining seemingly incompatible theories in a novel way so as to extend their reach into regimes that we have no way of testing in the foreseeable future, constrained only by principles based on physical intuition not honed in those regimes, become the most important touchstone for testing novel ideas in theoretical physics? Can it play that role? What epistemic warrant do we or can we have for it in the end?

Hawking radiation by itself, derived in QFT-CST, does not lead to black hole evaporation; one must "couple" the Hawking radiation to the spacetime geometry by the SCEFE ("backreaction")

there are no exact solutions of the SCEFE known for a spacetime representing an evaporating black hole (in 4-d)

and the approximate ones we have are severely limited in their scope and in their potential to yield general insight, by dint of various kinds of almost absurdly idealizing assumptions

so let's take a few minutes to discuss the complex and difficult methodological and epistemological situation we find ourselves in—it turns out to be, from the point of view of a philosopher interested in the structure of our knowledge in physics, extraordinarily rich and deserving of investigation what to do, then, when you can't solve equations, and:

1. either other mathematical difficulties prohibit approximative methods and numerical simulation (as in SCG);
2. or one does not want detailed information about individual models or individual possible solutions (because, *e.g.*, the only ones we can derive are trivial in one way or another—as in SCG),
3. and narrow classes of solutions (depending on absurd idealizations, *e.g.*) just won't do the job?

especially when it is difficult to experimentally access regimes in which one expects the behavior of interest to manifest itself—as in SCG

Then one can try to prove or derive general results; but what exactly is it that one is then doing?

theorems

(1) either assume something of a suitably generic character (*e.g.*, an energy condition), and try to do derive or argue for something of a suitably generic character

1. non-existence claims (non-singular cosmological models)
2. rigidity/stability claims (topology—past-singular closed cosmological models satisfying the strong energy condition)
3. scarcity/genericity claims (measure—cosmological models with a Killing field)
4. non-constructive existence claims (classic singularity theorems)
5. non-constructive behavioral claims (formation of closed trapped surfaces under gravitational collapse)

schematic models

(2) or construct models based on general principles—what I will call *schematic models*:

1. fix general principles and generic conditions (the framework of SCG)
2. characterize general structures that embody the principles and conform to the conditions (a quantum field satisfying the semi-classical Einstein field equation on an asymptotically flat black hole spacetime)
3. derive a statement about generic features of the character and behavior of the general structures (Hawking effect: black holes evaporate)

what exactly is it that one is doing with a schematic model?

- these models often involve behavioral claims (in the sense of a general result)
- they are not individual solutions to equations of motion or field equations
- usually, no exact individual solutions are known that represent any thing like such systems, especially not in the generality postulated
- they do not otherwise represent individual systems in any straightforward sense
- sometimes, they are not grounded in exact, rigorous general results
- they are not approximations to or idealizations of solutions
- there are no clearly specifiable families of solutions they correspond to

- they rather represent general features that we expect certain (often loosely characterized, merely postulated) families of solutions to have

- almost always, one can (loosely) characterize such (postulated) families in several different, often mutually contradictory ways

- thus, they have an interpretive looseness and a flexibility of consequence to them not characteristic of exact or approximative solutions so, to return to black hole evaporation:

in addition to the assumptions used to derive Hawking radiation, we now also need:

1. imposition of the SCEFE
2. behaviorial assumptions about the coupling of the metric and the matter field ("almost-stationary", adiabatic, ...)

finally, the “derivation”, *i.e.*, construction of the schematic model

complex, largely heuristic arguments based on subtle assumptions relating quasi-local intra-spacetime behavior to global, asymptotic behavior

> “local negative energy fluxes at the horizon associated with the positive thermal Hawking flux decrease the global ADM mass, which in turn decreases the local area of the event horizon—radiating black holes evaporate”

Hawking (1975):

> *It should be emphasized that these pictures of the mechanism responsible for the thermal emission and area decrease are heuristic only and should not be taken too literally. It should not be thought unreasonable that a black hole, which is an excited state of the gravitational field, should decay quantum mechanically and that, because of quantum fluctuation of the metric, energy should be able to tunnel out of the potential well of a black hole.*

interpretive looseness and flexibility of consequence as a general feature of schematic models: not a unique behavior, nor system, nor even kind of system is picked out

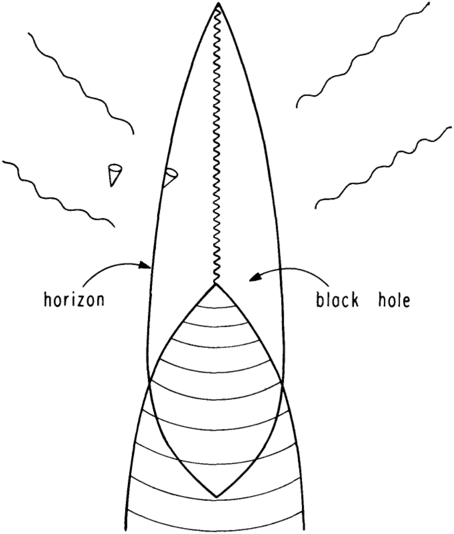


Figure 1 A space–time diagram showing the formation of a black hole by the collapse of matter and its subsequent evaporation by emission of Hawking radiation.

(taken from Wald 1984a)

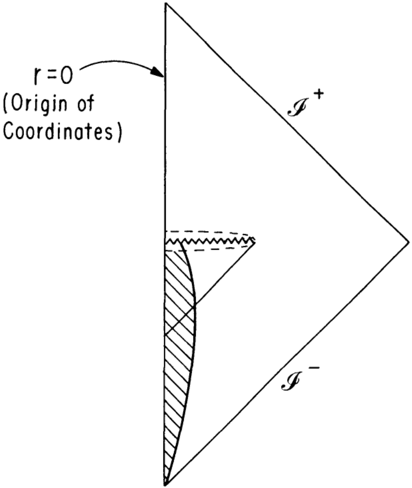
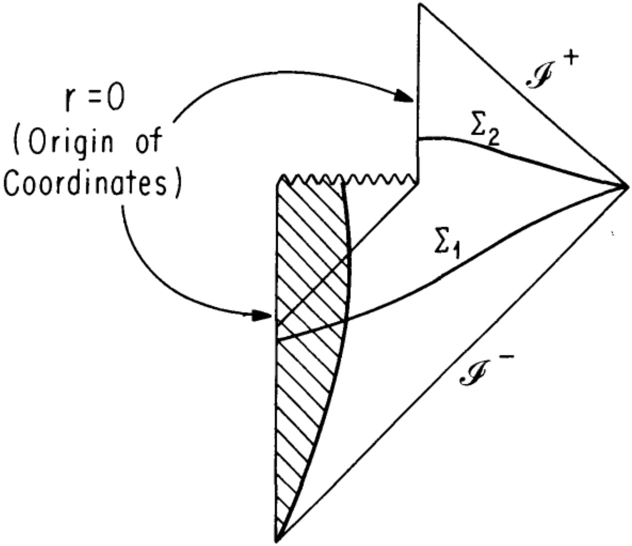


Figure 2 A conformal diagram showing the causal structure of a space–time where a black hole forms and evaporates in the manner of possibility (2) discussed in the text. Information cannot propagate outward through the black hole horizon without violating causality. The dotted lines near the classical singularity indicate a region where the classical description of space–time structure breaks down, thereby perhaps avoiding the presence of a singularity and allowing the information contained in the black hole to escape. However, this would require an enormous release of information with total energy only of order of the Planck energy. $\mathscr{I}^-$ and $\mathscr{I}^+$ are past null infinity and future null infinity.

(*ibid.*)



(*ibid.*)

it is exactly this interpretive looseness and flexibility of consequence of the schematic model of black hole evaporation that lends the Information-Loss Paradox it's open-ended character:

- there are many different formulations of the problem, none clearly equivalent to any of the others
- there are many different proposals in response to each of the different formulations, each of which some reasonable (and some unreasonable) people clearly see as a "solution" and other reasonable (and unreasonable) people don't

before moving on, however, let's briefly pause to contemplate how BHT and the Hawking effect may require new approaches to traditional philosophical problems, as I mentioned earlier would happen—in this case, radical changes to picture of ontology of spacetime and matter?

quantum field theory

- **matter** excitations in fields
- **spacetime** fixed, not affected by matter (Minkowski spacetime)

general relativity

- **matter** non-zero mass-energy ($T_{ab} \neq 0$, Ricci curvature)
- **spacetime** dynamical, affected by matter (zero mass-energy $\Rightarrow$ only Weyl curvature)

black hole thermodynamics spacetime curvature *transformed into* matter and matter *transformed into* spacetime curvature

Curiel, Erik. 2017a. "A Primer on Energy Conditions". Chapter 3 in *Towards a Theory of Spacetime Theories,* edited by D. Lehmkuhl, G. Schiemann, and E. Scholz, 13:43–104. Einstein Studies. New York: Birkhäuser. A slightly revised and updated version can be found at <http://strangebeautiful.com/papers/curiel-primer-energy-conds.pdf>., doi:10.1007/978-1-4939-3210-8$\underscore$3.

———. 2017b. "Are Classical Black Holes Hot or Cold?" Preprint: arXiv:1408.3691 [gr-qc]. Latest version available: <http://strangebeautiful.com/papers/curiel-class-bhs-hot-or-cold.pdf>.

———. 2019. "The Many Definitions of a Black Hole". *Nature Astronomy* 3:27–34. Free read-only SharedIt link: <https://rdcu.be/bfNpM>. Preprint: arXiv:1808.01507 [physics.hist-ph], doi:10.1038/s41550-018-0602-1.

Geroch, Robert. 1970. "Domain of Dependence". *Journal of Mathematical Physics* 11 (2): 437–449. doi:10.1063/1.1665157.

Geroch, Robert, and Gary T. Horowitz. 1979. "Global Structure of Spacetimes". Chapter 5 in *General Relativity: An Einstein Centenary Survey,* edited by S. Hawking and W. Israel, 212–293. Cambridge: Cambridge University Press.

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

Minguzzi, Ettore, and Miguel Sánchez. 2008. "The Causal Hierarchy of Spacetimes". In *Recent Developments in Pseudo-Riemannian Geometry,* edited by Dmitri V. Alekseevsky and Helga Baum, 299–358. ESI Lectures in Mathematics and Physics. Zurich: European Mathematical Society Publishing House. arXiv:gr-qc/0609119v3, doi:10.4171/051.

Penrose, R. 1964. "Conformal Treatment of Infinity". In *Relativité, Groupes et Topologie,* edited by C. DeWitt and B. DeWitt, 565–586. New York: Gordon / Breach.

Wald, Robert M. 1984a. "Black Holes, Singularities and Predictability". 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.

———. 1984b. *General Relativity.* Doi:10.7208/chicago/9780226870373.001.0001. Chicago: University of Chicago Press.