ETH zürich

Semester Project MSc Physics ETH

The Black Hole Information Loss Paradox in the Context of Analogue Gravity

Sébastien C. Garmier

sgarmier@student.ethz.ch

May 19, 2023

Supervisors:

Giulia Mazzola [1]

Prof. Dr. Renato Renner [1]

Abstract

We investigate the possibility of approaching the black hole information loss paradox from the point of view of analogue gravity models. More generally, we ask whether analogue models can at all make inferences about gravity. To this end we give an introduction to black holes, the black hole information loss paradox, and analogue gravity models, before attempting to formulate the information loss paradox in the context of analogue gravity. We find that crucially, the notion of black hole entropy is missing, placing a discussion of the paradox in that context out of reach. Simultaneously, we argue based on the ubiquity and generality of analogue models that they are unlikely to possess deep connections with gravity.

[1] Institute of Theoretical Physics, ETH Zürich, Switzerland

Contents

Introduction 2

Basic Concepts and Notation 3

1 Motivation: Spacetime as a Flowing Fluid 8

1.1 Schwarzschild Spacetime in Gullstrand-Painlevé Coordinates 8

1.2 Basic Notions of Analogue Gravity 12

1.3 Fluid-Flow Metrics 13

2 Black Holes 17

2.1 Black Holes and Event Horizons 17

2.2 Area Theorem 22

2.3 Four Laws of Black Hole (Thermo-) Dynamics 24

2.4 Bekenstein Entropy 28

2.5 Hawking Radiation 29

2.6 Back-Reaction 41

3 Black Hole Information Loss Paradox 44

3.1 Hawking's Version of the Paradox 44

3.2 Page Curve Version of the Paradox 45

3.3 A Recent Approach to a Solution: Replica Wormholes 49

3.4 Anatomy of the Paradox 52

4 Analogue Gravity 55

4.1 General Classical Models 55

4.2 Linear Sound in Irrotational, Barotropic, Perfect Fluids 60

4.3 Linear Sound in Bose-Einstein Condensates 65

4.4 Continuity and the Difficulty of Schwarzschild Geometry 68

5 Analogue Gravity and the Information Loss Paradox 71

5.1 The Missing Piece: Black Hole Entropy 72

5.2 Analogue Gravity Models are Simply Models 75

6 Conclusion and Outlook 77

A Details on the Zeroth and First Law of Black Hole Dynamics 78

B Details on the Derivation of Hawking Radiation 83

C Fluid Analogue Model for 1+1-Dimensional Schwarzschild Spacetime 89

Introduction

Analogue gravity is a research program aimed at modelling features of gravity using various physical systems [7]. Famously, it is possible to model massless field propagation in curved spacetimes with sound wave propagation in flowing fluids [74] [75]. The analogies between gravity and fluid dynamics go as far as predicting the existence of *Hawking radiation* in fluids with so-called *apparent horizons* [69] [76], surfaces which sound can only traverse in one direction due to sonic or supersonic fluid flow velocity. Interest in analogue gravity models has recently been renewed due to experimental confirmation of Hawking radiation within fluid analogue systems and further ongoing experimental efforts [37].

Hawking radiation is a crucial ingredient of the *black hole information loss paradox* [32] [55] [84]: the emission of Hawking radiation in a mixed quantum state during the *evaporation* of a black hole, predicted by quantum field theory in curved spacetime [31], is in conflict with the *thermodynamics of black holes*, which is derived from mostly geometric considerations [30] [8] [82]. Resolutions of the black hole information loss paradox might allow valuable insight into *quantum gravity* and is thus heavily studied; in particular, recent progress in the form of *replica wormholes* is promising for a potential resolution of the paradox [59] [1] [2].

One may thus wonder whether these two areas of research can be combined. Specifically: *can we learn about the black hole information loss paradox from analogue gravity models?* Related to this, and motivated by the recent surge in popularity of analogue models, we can ask the more general question: *can we infer anything at all about gravity from analogue models?*

It is the purpose of this work to first give an introduction to both the black hole information loss paradox and analogue gravity, and then to attempt answering these questions. We will argue that both questions are for now best answered with "no": the first one, because the crucial notion of *black hole entropy* is yet lacking in analogue gravity models and there is no clear way to define it; the second one in the sense that analogue models are *best used as models* and that there is no guarantee for phenomena occurring in gravity if the same phenomena occur in analogue models. Importantly, our conclusions *will not discredit the field of analogue gravity.*

The present work is structured in the following way: Section 1 motivates the field of analogue gravity by examining a specific form of metrics, so-called *fluid-flow metrics*. Section 2 treats the necessary background on black holes, including black hole thermodynamics and Hawking radiation. Section 3 then introduces the black hole information loss paradox, while Section 4 does the same for analogue gravity. With these preparations we tackle the two questions posed above in Section 5 before concluding and providing an outlook in Section 6. Many sections are designed with future sections in mind; this requires a certain amount of foresight but also prevents needless repetition of almost identical arguments in different circumstances. For instance, we already introduce the fundamentals of analogue gravity in Section 1, so that we can use fluid-flow metrics, which will become truly relevant in Section 4, already in Section 2 when treating black holes, especially when deriving Hawking radiation; thus we will not have to re-derive Hawking radiation for fluid-flow metrics later on.

Basic Concepts and Notation

Spacetime. The pair (M, g_{ab}) denotes a *spacetime* consisting of a smooth manifold M and a Lorentzian metric g_{ab}. If not stated otherwise, spacetime has one timelike dimension and three spacelike dimensions; occasionally we may explicitly generalize to d spacelike dimensions. At each $p \in M$ the metric $(g_{ab})_p$ locally defines two *light cones*. We will always assume (M, g_{ab}) to be *time-orientable*; that is, we assume it is possible to continuously designate one of these cones as the *future light cone* and the other as the *past light cone*. Intuitively speaking, we assume that globally the meaning of *future* and *past* is well-defined for each event.

We will assume that spacetimes do not contain closed timelike curves.

Definition 0.1: Hypersurfaces

An n-dimensional *hypersurface* $S \subset M$ is simply taken to mean an n-dimensional smooth submanifold of M.

Tensors and Signs. We use the metric sign convention $-{+}{+}{+}$ (more generally $-{+}\cdots{+}$) and borrow a version of the abstract index notation from [81]: tensor expressions such as g_{ab} or k^a are written with Roman indices at the beginning of the alphabet (a, b, c, d, e), taking values from 0 to 3 (or more generally to d), and are meant to be coordinate-independent. That is, any abstract index expression may be converted to an expression in any coordinate system simply by interpreting the indices as indices for this specific coordinate system, as in usual index notation. If an expression only holds in specific coordinates, we use Greek indices.

Sometimes, especially for explicit expressions of metrics, we use index-less notation such as $ds^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$. Einstein summation over repeated indices, once upstairs and once downstairs, is implied.

Roman indices in the middle of the alphabet (i, j, k, l) as in v^i, ${\delta^i}_j$ have a special role: they run from 1 to 3 (ore more generally to d) and *always label Cartesian coordinates in flat spacetime*, as is usually done in physics. Note that the position of these indices does not matter (*e.g.* $v^i = v_i$). To prevent confusion with the abstract indices mentioned above, we will always write sums over i, j, k and l explicitly.

A general metric tensor is written as g_{ab} and $\eta_{\mu\nu} = \mathrm{diag}(-1, 1, \ldots, 1)$ is the flat Minkowski metric in Minkowski coordinates. The *Riemann tensor* is denoted by R_{abcd}, the *Ricci tensor* by R_{ab}, and the *Ricci curvature* by R. T_{ab} is the *energy-momentum tensor*.

Theorem 0.2: Field Equations (Einstein)

The Einstein field equations read [23]

$$R_{ab} - \frac{1}{2} g_{ab} R = 8\pi T_{ab} \,. \tag{0.1}$$

Units. Whenever not specified, we use *Planck units*: $c = G = \hbar = k_B = 1$. Sometimes we explicitly use *geometric Units*: $c = G = 1$.

Causality. We will need some basic notions and results surrounding *causality*. See [81].

Definition 0.3: Causal Curves

A *future-directed, causal curve* $C : [a, b] \to M$ in a spacetime (M, g_{ab}) is a C^1 curve whose tangent vectors all lie on or within the future light cone; *i.e.* the tangent vectors are timelike or null. We say that a future-directed causal curve is *maximally extended*, if it is not part of a larger future-directed causal curve.

Intuitively, if causality is to be conserved, information may only be transported along future-directed causal curves. Fundamental physical field equations, including the Einstein field equations, tend to conserve causality, hence the importance of causal curves.

Definition 0.4: Causal Future and Past

The *causal future* $J^+(N) \subset M$ of a subset $N \subset M$ is the set of spacetime points which can be reached by future-directed causal curves originating in N. Similarly, the points p of the *causal past* $J^-(N) \subset M$ are such that there exists a future-directed causal curve originating at p and reaching some point in N.

Intuitively, $J^+(N)$ are the spacetime points *influencable* by N, and $J^-(N)$ are the points which *can influence* N. Note that $J^-(N)$ completely *determines* N in the sense that no points outside $J^-(N)$ can send information to N. Meanwhile, $J^+(N)$ is not necessarily completely determined by N, since one may have $J^-(J^+(N)) \supsetneq N$.

Definition 0.5: Cauchy Surface

A *Cauchy surface* Σ for a subset $N \subset M$ is a *spacelike hypersurface* (*i.e.* a hypersurface whose tangent vectors are all spacelike) contained in N such that every maximally extended, future-directed causal curve in N intersects Σ exactly once.

Note that every point p of N is then either in $J^+(\Sigma)$ or $J^-(\Sigma)$, and in both if and only if $p \in \Sigma$. Thus, $J^+(\Sigma) \cap N$ is completely determined by Σ.[1] Intuitively, Σ describes a generalization of the special-relativistic notion of 'space at one instant of time' on which initial-value data for the laws of physics may be provided. If N admits a Cauchy surface, we call N *globally hyperbolic*.

Quantum Physics. A quantum system is described by a *Hilbert space* $\mathcal{H}$, *i.e.* a complex vector space with sesquilinear scalar product which is complete with respect to the norm induced by the scalar product. An *operator* $\hat{A}$ on $\mathcal{H}$ is a linear map $\hat{A} : \mathcal{H} \to \mathcal{H}$; we denote operators with a caret. Physical *observables* are *Hermitian* operators, *i.e.* $\hat{O}^\dagger = \hat{O}$.[2] The (real) eigenvalues of an observable are the possible outcomes when measuring the observable.

A *state* of a quantum system described by $\mathcal{H}$ is a *density operator*:

Definition 0.6: Density Operators

A (linear) operator $\hat{\sigma} : \mathcal{H} \to \mathcal{H}$ with $\operatorname{tr} \hat{\sigma} = 1$, $\hat{\sigma} \geq 0$ (in the sense of eigenvalues) is a *density operator*. We also say that $\hat{\sigma}$ *is a state on* $\mathcal{H}$.

[1] And, if the laws of physics are time-reversible, $J^-(\Sigma) \cap N$ also.
[2] We will not need the technical distinction between *Hermitian* and *self-adjoint*.

Definition 0.7: Pure and Mixed States

A state of the form $\hat{\sigma} = |\psi\rangle\langle\psi|$, $|\psi\rangle \in \mathcal{H}$, is called *pure*. All other states are called *mixed*. If $\dim \mathcal{H} < \infty$, then the state

$$\hat{\pi} := \frac{\hat{\mathrm{id}}}{\dim \mathcal{H}} \tag{0.2}$$

is called the *completely mixed state* on $\mathcal{H}$.

One may transform density operators into new density operators by acting with a unitary transformation:

Proposition 0.8: Unitary Transformations

Let $\hat{U} : \mathcal{H} \to \mathcal{H}$ be *unitary*, (*i.e.* $\hat{U}$ is bijective and conserves the scalar product of $\mathcal{H}$), and let $\hat{\sigma}$ be a density operator. Then $\hat{U}\hat{\sigma}\hat{U}^{-1}$ is also a density operator.

In particular, *time evolution* of *closed* quantum systems is described by such a unitary transformation.

Definition 0.9: Mesurement

Assume that a quantum system is described by the state $\hat{\sigma}$, and that the observable $\hat{O}$ is now being measured. By diagonalization one can write $\hat{O} = \sum_o o\hat{P}_o$, where the sum runs over the eigenvalues o of $\hat{O}$ and $\hat{P}_o$ is the projector onto the eigenspace of o. The *possible outcomes* when measuring $\hat{O}$ are the eigenvalues of $\hat{O}$. The probability of obtaining the outcome o is

$$\mathrm{pr}[\hat{O} = o] = \mathrm{tr}(\hat{\sigma}\hat{P}_o). \tag{0.3}$$

The expectation value of this measurement is thus

$$\langle\hat{O}\rangle = \mathrm{tr}(\hat{\sigma}\hat{O}). \tag{0.4}$$

Assume that the outcome o has been achieved. The post-measurement state is then

$$\hat{\sigma}' = \frac{\hat{P}_o\hat{\sigma}\hat{P}_o}{\mathrm{pr}[\hat{O} = o]}. \tag{0.5}$$

We will often encounter *bipartite* states: $\hat{\sigma} : \mathcal{H} \to \mathcal{H}$, with $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ for two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. The state on $\mathcal{H}_1$ (on $\mathcal{H}_2$) is then obtained by *tracing out* $\mathcal{H}_2$ ($\mathcal{H}_1$):

Definition 0.10: Reduced States and Partial Trace

Given $\hat{\sigma}$ a state on $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ we define the *reduced states* on $\mathcal{H}_1$ and $\mathcal{H}_2$ as

$$\hat{\sigma}_1 := \mathrm{tr}_2(\hat{\sigma}), \qquad \hat{\sigma}_2 := \mathrm{tr}_1(\hat{\sigma}). \tag{0.6}$$

Here tr_1 is the *partial trace* on $\mathcal{H}_1$ defined as the linear extension of

$$\mathrm{tr}_1 : \hat{A}_1 \otimes \hat{A}_2 \mapsto \mathrm{tr}(\hat{A}_1) \cdot \hat{A}_2, \tag{0.7}$$

and similarly for tr_2.

One can show that this definition of reduced states is compatible with notions of measurements on subsystems and is thus sensible. For details, see *e.g.* [36].

Information Theory. Consider a *classical random variable* X, whose possible values x are distributed according to the probabilities $P(x)$. That is, $\sum_x P(x) = 1$, and $0 \le P(x) \le 1$ for all possible values x; we say that P is a *probability distribution*. One then defines [36] the *entropy* or *Shannon entropy* (after the mathematician Shannon):

Definition 0.11: Entropy (Shannon)

Given a random variable X distributed according to P, the *(Shannon) entropy* of X is defined to be

$$H(X)_P = H(P) := -\sum_x P(x) \log_2 P(x). \tag{0.8}$$

The entropy $H(P)$ characterizes *how much we do not know about a variable X distributed according to P*. We use the letter H to discern information-theoretical entropy from thermodynamic entropy S.

For a full discussion of entropy and its meaning, see [36].

Quantum Information Theory. For any state of a quantum system we can define the *(von Neumann) entropy* (after the physicist von Neumann), in some sense a generalization of Shannon entropy:

Definition 0.12: Entropy (von Neumann)

Given a quantum state $\hat{\sigma}$, the *(von Neumann) entropy* of σ is

$$H(\hat{\sigma}) := -\mathrm{tr}(\hat{\sigma} \log_2 \hat{\sigma}). \tag{0.9}$$

Note that for states of the form $\hat{\sigma} = \sum_x P(x) |\psi_x\rangle \langle\psi_x|$, with the states $|\psi_x\rangle$ orthonormal, and P a probability distribution, we find $H(\hat{\sigma}) = H(P)$. It holds that:

Proposition 0.13: Bounds on Entropies

We have

$$H(\hat{\sigma}) \ge 0, \tag{0.10}$$

with equality if and only if $\hat{\sigma}$ is pure.

With $\mathrm{supp}\,\hat{\sigma} := \{|\psi\rangle \in \mathcal{H} \mid \hat{\sigma}|\psi\rangle \neq 0\}$ the *support* of $\hat{\sigma}$ (a subspace of $\mathcal{H}$), and if $\dim \mathrm{supp}\,\hat{\sigma} < \infty$, it holds that

$$H(\hat{\sigma}) \le \log_2(\dim \mathrm{supp}\,\hat{\sigma}). \tag{0.11}$$

If $\dim \mathcal{H} < \infty$, then the maximal possible entropy is achieved by the completely mixed state:

$$H(\hat{\pi}) = \log_2(\dim \mathcal{H}). \tag{0.12}$$

Proposition 0.14: Entropy is Conserved in Unitary Evolution

Let $\hat{U} : \mathcal{H} \to \mathcal{H}$ be unitary. Then

$$H(\hat{U}\hat{\sigma}\hat{U}^{-1}) = H(\hat{\sigma}). \tag{0.13}$$

In particular, unitary time evolution conserves the entropy of states.

Finally, we state an important result concerning the entropies of reduced states:

Proposition 0.15: Entropies of Reduced States

For a state $\hat{\sigma}$ on $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ it holds that

$$H(\hat{\sigma}) \geq |H(\hat{\sigma}_1) - H(\hat{\sigma}_2)|. \tag{0.14}$$

In particular, $H(\hat{\sigma}_1) = H(\hat{\sigma}_2)$ if $\hat{\sigma}$ is pure. Consequently, if $\hat{\sigma}$ is pure then $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are either both pure or both mixed.

For example, the reduced states of an entangled pure state are both mixed. The entropies of the reduced states is in fact a measure of the *entanglement* between the two parts.

For a detailed treatment of quantum information theory, see [36].

1 Motivation: Spacetime as a Flowing Fluid

Let us begin with a seemingly unrelated issue: the coordinate singularity in Schwarzschild coordinates at the event horizon of Schwarzschild geometry. By removing this singularity with a clever choice of new coordinates, one may find the *Gullstrand-Painlevé* form of the Schwarzschild metric [28] [58]. Remarkably, it is possible to interpret the metric in these new coordinates as a *flowing fluid*, relative to which radial null geodesics propagate at a fixed speed; this resembles how in the high-frequency limit (in the *eikonal approximation* [42]) sound in a fluid propagates as *rays* with a fixed speed of sound relative to the flow of the fluid [39].

This invites the possibility of further analogies between flowing fluids and spacetime, and brings us directly to the field of *analogue gravity models*. We begin by deriving the above-mentioned result for Schwarzschild spacetime in Section 1.1. Section 1.2 then explains how this leads us to consider analogue gravity models; there we will see that so-called *fluid-flow metrics* are of great importance. We will discuss them in Section 1.3.

1.1 Schwarzschild Spacetime in Gullstrand-Painlevé Coordinates

Let us begin with the following well-known result:

Theorem 1.1: Schwarzschild Spacetime in Schwarzschild Coordinates

A solution to the vacuum ($T_{ab} = 0$) Einstein field equations (0.1) is the *Schwarzschild metric*, which in so-called *Schwarzschild coordinates* $t \in \mathbb{R}$, $r \in (0, r_s) \cup (r_s, \infty)$, $\theta \in [0, \pi)$, $\phi \in [0, 2\pi)$ takes the form

$$ds^2 = -\left(1 - \frac{r_s}{r}\right) \mathrm{d}t^2 + \left(1 - \frac{r_s}{r}\right)^{-1} \mathrm{d}r^2 + r^2 \,\mathrm{d}\Omega^2, \quad \mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta \,\mathrm{d}\phi^2. \quad (1.1)$$

$r_s := 2M$ is the *Schwarzschild radius* and $M \geq 0$ is the *mass* parameter.

See for instance the original derivation of the metric [70] due to Schwarzschild in 1916, or any textbook treatment such as [49, Chapter 23] or [81, Chapter 6]. Once *black holes* have been rigorously introduced in Section 2, we will see that Schwarzschild spacetime describes a black hole with mass M (hence the name for the parameter M). The apparent singularity at the horizon $r = r_s$ is a mere coordinate singularity and can be removed by a suitable choice of coordinates; see *e.g.* [81, Section 6.4]. In particular, observers can reach the *interior* $r < r_s$ from the *exterior* $r > r_s$ within a finite amount of proper time without experiencing infinite tidal forces; see *e.g.* [49, Section 25.5].

A clever and commonly employed choice of coordinates to get rid of the apparent singularity at $r = r_s$ are for instance *Kruskal-Szekeres coordinates* [38] [71]. We will instead make use of the fact that observers can cross the horizon inwards and replace the Schwarzschild time coordinate t by the proper time τ of one specific infalling observer. This will lead us to the Gullstrand-Painlevé form of the metric. Note that derivations of this form similar to our derivation below exist; see for instance [48].

Free-Falling Observers. A natural choice of special observer is one radially and freely falling from rest at infinity. Let us consider a slightly more general observer freely and radially falling, starting at rest relative to the radial coordinate r at $r = R > r_s$.

Due to the timelike Killing vector field $(\partial_t)^a$, the energy

$$E := -g_{ab}(\partial_t)^a u^b = (1 - r_s/r)\,\mathrm{d}t/\mathrm{d}\tau \quad (1.2)$$

is conserved along the worldline.[3] Due to the normalization $u_a u^a = -1$ of the four-velocity u^a of the falling observer, we have $-(1 - r_s/r)\,(\mathrm{d}t/\mathrm{d}\tau)^2 + (1 - r_s/r)^{-1}(\mathrm{d}r/\mathrm{d}\tau)^2 = -1$, *i.e.*

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 = E^2 - \left(1 - \frac{r_s}{r}\right). \tag{1.3}$$

Now $\mathrm{d}r/\mathrm{d}\tau = 0$ at $r = R$ and thus $E = \sqrt{1 - r_s/R}$. Hence,

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \sqrt{1 - \frac{r_s}{R}} \cdot \left(1 - \frac{r_s}{r}\right)^{-1}, \tag{1.4}$$

and

$$\frac{\mathrm{d}r}{\mathrm{d}\tau} = -\sqrt{\frac{r_s}{r} - \frac{r_s}{R}}. \tag{1.5}$$

We took the negative square root since the observer is infalling.

In the case of our spacial observer ($R \to \infty$) we get

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \left(1 - \frac{r_s}{r}\right)^{-1}, \qquad \frac{\mathrm{d}r}{\mathrm{d}\tau} = -\sqrt{\frac{r_s}{r}}, \qquad \frac{\mathrm{d}r}{\mathrm{d}t} = -\sqrt{\frac{r_s}{r}} \cdot \left(1 - \frac{r_s}{r}\right). \tag{1.6}$$

Conditions for the New Time Coordinate T. We now wish to define a new coordinate $T = T(t, r)$ that, when constrained to the worldline of the special observer, agrees with their proper time up to a constant. Since the special observer can reach the region $r < r_s$, we hope to thus define a time coordinate which is regular at $r = r_s$.

Since T should match τ up to a constant, we require

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\partial T}{\partial t} + \frac{\partial T}{\partial r}\frac{\mathrm{d}r}{\mathrm{d}t} \tag{1.7}$$

along the special observer's worldline, *i.e.*

$$1 - \frac{r_s}{r} = \frac{\partial T}{\partial t} - \frac{\partial T}{\partial r}\sqrt{\frac{r_s}{r}} \cdot \left(1 - \frac{r_s}{r}\right). \tag{1.8}$$

A priori there are many possible choices for T satisfying this condition. Intuitively speaking, T will only be fully determined by a partial differential equation *defined for all* (t, r) as well as boundary conditions for T. For now, we have no boundary conditions and a partial differential equation holding only on a one-dimensional subset of the (t, r)-plane (the observer's worldline).

We can thus simplify our search for a solution by requiring the condition (1.8) *everywhere*, not just on the special observer's worldline. This generalization even has a physical motivation: there is an infinitude of observers radially infalling after starting from infinity, distinguished by the coordinate time t at which they reach a specific waypoint $r_0 > r_s$ in their fall, that would qualify as special observers; one can show (*e.g.* [49, Chapter 31]) that the worldlines of all these observers fill the entire t-r-plane (with some technicalities at $r = r_s$, but otherwise even the black hole interior). Requiring (1.8) everywhere thus amounts to *treating all these observers equally* in the sense that T matches (up to a constant) the proper time τ of *any* special observer, when restricted to the worldline of said special observer. The constant may depend on the worldline.

---

[3] See [81] or [49] for discussions of spacetime symmetries and conserved quantities along geodesics using the language of Killing vector fields. See [41] for the basics of analytical mechanics underlying these discussions.

General Coordinates. Equation (1.8) leaves us with the freedom to choose $\partial T/\partial t =: A$, where A is a function of t and r. However, as we shall see, not every choice of A yields a meaningful coordinate system.

Given A, we find

$$\frac{\partial T}{\partial r} = \left[A - 1 + \frac{r_s}{r}\right] \cdot \sqrt{\frac{r}{r_s}} \cdot \left(1 - \frac{r_s}{r}\right)^{-1} \tag{1.9}$$

and with

$$\mathrm{d}T = A\mathrm{d}t + \frac{\partial T}{\partial r}\mathrm{d}r, \tag{1.10}$$

it follows that

$$\mathrm{d}t = \frac{1}{A}\mathrm{d}T - \left[1 - \frac{1}{A} + \frac{r_s}{rA}\right] \sqrt{\frac{r}{r_s}} \cdot \left(1 - \frac{r_s}{r}\right)^{-1} \mathrm{d}r. \tag{1.11}$$

With the definition

$$B := 1 - 1/A + r_s/rA, \tag{1.12}$$

we then get

$$-\left(1 - \frac{r_s}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{r_s}{r}\right)^{-1}\mathrm{d}r^2 \tag{1.13}$$

$$= -\left(1 - \frac{r_s}{r}\right)\frac{1}{A^2}\mathrm{d}T^2 - \left(B^2\frac{r}{r_s} - 1\right) \cdot \left(1 - \frac{r_s}{r}\right)^{-1}\mathrm{d}r^2 + 2\frac{B}{A}\sqrt{\frac{r}{r_s}}\,\mathrm{d}T\mathrm{d}r, \tag{1.14}$$

where $\mathrm{d}T\mathrm{d}r = (\mathrm{d}T \otimes \mathrm{d}r + \mathrm{d}r \otimes \mathrm{d}T)/2$ is the symmetric product.

We thus find that the Schwarzschild metric (1.1) can be written as

$$\mathrm{d}s^2 = -\frac{1}{A}\mathrm{d}T^2 + \frac{B}{A}\mathrm{d}T^2 - \left(B^2\frac{r}{r_s} - 1\right) \cdot \left(1 - \frac{r_s}{r}\right)^{-1}\mathrm{d}r^2 + 2\frac{B}{A}\sqrt{\frac{r}{r_s}}\,\mathrm{d}T\mathrm{d}r + r^2\mathrm{d}\Omega^2. \tag{1.15}$$

Finally, the new coordinate T is related to t and r by

$$T(t,r) = \int^t \mathrm{d}t'\, A(t',r) + \int^r \mathrm{d}r'\, A(t,r')B(t,r')\sqrt{\frac{r'}{r_s}} \cdot \left(1 - \frac{r_s}{r'}\right)^{-1}. \tag{1.16}$$

The lower integral bounds are unimportant as they only change T by an additive constant, which does not change whether T solves (1.8). Note that neither of the integrands can be identically zero, for this would restrict the new coordinate to a one- or zero-dimensional submanifold of the t-r plane. Thus, we must have that A and B are not identically zero.

Gullstrand-Painlevé Coordinates. By choosing $B = r_s/r$ (*i.e.* $A = 1$), the metric (1.15) becomes

$$\mathrm{d}s^2 = -\mathrm{d}T^2 + \left(\mathrm{d}r + \sqrt{\frac{r_s}{r}}\mathrm{d}T\right)^2 + r^2\mathrm{d}\Omega^2, \tag{1.17}$$

which is flat on any slice of constant T; we say it is *spatially flat*. Note that the overall metric is still curved, but curvature is restricted to terms involving the time coordinate T.

With $A = 1$ we have $T(r,t) = t + \tilde{T}(r)$, and

$$\tilde{T}(r) = \int^r \mathrm{d}r'\,\sqrt{\frac{r_s}{r'}} \cdot \left(1 - \frac{r_s}{r'}\right)^{-1} = -2r_s\int^{u(r)} \mathrm{d}u'\,\frac{1}{u'^2 - u'^4}. \tag{1.18}$$

In the second step we have substituted $u(r') = (r_s/r')^{1/2}$, $\mathrm{d}r' = -2r_s u'^{-3}\mathrm{d}u'$. We get (using partial fraction decomposition for instance)

$$\tilde{T}(r) = C + 2\sqrt{r_s r} + r_s \ln\left|\frac{\sqrt{r/r_s} - 1}{\sqrt{r/r_s} + 1}\right|, \tag{1.19}$$

where C is a real constant.

Let us summarize:

Proposition 1.2: Schwarzschild Metric in Gullstrand-Painlevé Coordinates

By replacing the Schwarzschild time coordinate t in the Schwarzschild metric (1.1) with a new time coordinate T, equal up to a constant to the proper time of observers radially and freely falling from infinity, it is possible to obtain the so-called *Gullstrand-Painlevé coordinate system*, in which the Schwarzschild metric takes the spatially flat form

$$\mathrm{d}s^2 = -\mathrm{d}T^2 + \left(\mathrm{d}r + \sqrt{\frac{r_s}{r}}\mathrm{d}T\right)^2 + r^2\mathrm{d}\Omega^2. \tag{1.20}$$

The new time-coordinate T is related to Schwarzschild time t according to

$$T = t + C + 2\sqrt{r_s r} + r_s \ln\left|\frac{\sqrt{r/r_s} - 1}{\sqrt{r/r_s} + 1}\right|, \tag{1.21}$$

where C is a real constant. One typically chooses $C := 0$.

This form of the metric was found independently by GULLSTRAND [28] and PAINLEVÉ [58] both in 1921, and not starting from the Schwarzschild metric. In fact, it was only in 1933 when LEMAÎTRE showed [44] that the metric expression found by GULLSTRAND and PAINLEVÉ described Schwarzschild spacetime.

Interpretation as a Flowing Fluid. Consider a radial curve with affine parameter λ:

$$u_\mu u^\mu = \varepsilon = -\left(\frac{\mathrm{d}T}{\mathrm{d}\lambda}\right)^2 + \left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + 2\sqrt{\frac{r_s}{r}}\frac{\mathrm{d}T}{\mathrm{d}\lambda}\frac{\mathrm{d}r}{\mathrm{d}\lambda} + \frac{r_s}{r}\left(\frac{\mathrm{d}T}{\mathrm{d}\lambda}\right)^2, \tag{1.22}$$

where $\varepsilon = -1$ for timelike curves and $\varepsilon = 0$ for lightlike ones. Solving for $\mathrm{d}r/\mathrm{d}T = (\mathrm{d}r/\mathrm{d}\lambda)\cdot(\mathrm{d}T/\mathrm{d}\lambda)^{-1}$, we find

$$\frac{\mathrm{d}r}{\mathrm{d}T} = -\sqrt{\frac{r_s}{r}} \pm \sqrt{1 + \varepsilon\left(\frac{\mathrm{d}T}{\mathrm{d}\lambda}\right)^{-2}}. \tag{1.23}$$

The curve is outgoing for $+$ and incoming for $-$.

For light rays, we thus have

$$\frac{\mathrm{d}r}{\mathrm{d}T} = -\sqrt{\frac{r_s}{r}} \pm 1 =: V(r) \pm c_s, \tag{1.24}$$

while radially moving observers satisfy

$$\frac{\mathrm{d}r}{\mathrm{d}T} = -\sqrt{\frac{r_s}{r}} \pm \sqrt{1 - \left(\frac{\mathrm{d}T}{\mathrm{d}\tau}\right)^{-2}}. \tag{1.25}$$

From this we conclude:

Proposition 1.3: Schwarzschild Spacetime as a Flowing Fluid

In Gullstrand-Painlevé coordinates, radial propagation of light rays is mathematically analogous to radial propagation of sound rays in a fluid, moving radially with speed $-(r_s/r)^{1/2}$, and speed of sound equal to 1 relative to the fluid. Furthermore, radial motion of observers corresponds to observers "swimming" in the fluid with less than the speed of sound. In particular, the special observers used in the definition of the coordinates are co-moving with the fluid, since for them $\mathrm{d}T/\mathrm{d}\tau = 1$.

It is in this sense that we can see Schwarzschild spacetime as a flowing medium. In this analogy, neither observers nor light (sound) can escape the region $r < r_s$, because the flow of the fluid becomes faster than the speed of light (sound).

It will become clear later that similar statements hold about timelike and lightlike curves through Schwarzschild spacetime which are not necessarily radial; the analogy will thus become even stronger. It is however useful to first introduce some terminology in the next section.

1.2 Basic Notions of Analogue Gravity

Building on the interpretation of Schwarzschild spacetime as a flowing fluid, we introduce here the different types of *analogue gravity models* (or "*analogue models*" for short) relevant for this work. We do this now, without a treatment of analogue models beyond the basic notions, since it will be beneficial to already have made contact to analogue gravity in some of the following sections, but most of them will not actually require a full-fledged discussion of analogue models. Once truly needed, we will come back to analogue models and discuss them in much more detail in Section 4.

Note that the term "analogue gravity model" is often used quite loosely (see [7] for an overview); to prevent confusion, we will stick to a stricter terminology explained here.

Classical Fluid-Flow Analogue Models. We begin by noting that in the eikonal limit, the (*wave*) propagation of a *massless Klein-Gordon field* (also called *massless scalar field*) can be seen as *rays* propagating along null geodesics [49]. Similarly, sound *ray* propagation is the eikonal limit of sound *wave* propagation. The analogy found in the example of Schwarzschild geometry above thus holds in the eikonal limit, and one may try extending the analogy between fluids and spacetime outside the eikonal limit, to *waves*.

That is, we want to find an analogy between *wave propagation of sound* in a fluid and *wave propagation of a massless scalar field* in curved spacetime. Because sound in a fluid only has one polarization, the scalar field should be a *real* scalar field. An explicit such analogy was first demonstrated by Unruh in 1981 [74] and later independently by Visser [75]. A fluid or general medium with sound propagation of this type is a so-called *classical fluid-flow analogue model of gravity*. The model in [74] and [75] is still arguably the most important classical fluid-flow analogue model, due to its wide applicability on the one hand and its simplicity on the other; we will derive it in Section 4.2.

Taking the eikonal approximation on both the classical fluid-flow analogue model and the scalar field in curved spacetime immediately yields an analogy of ray propagation. As in Schwarzschild spacetime above, we can thus interpret spacetime as a flowing fluid relative to which null geodesics (and even null curves) propagate at a fixed speed. This kind of interpretation is a general feature of the metrics encountered when discussing classical fluid-flow analogue models of gravity.

A metric *in special coordinates* $(t, \mathbf{x})$ making this interpretation explicit, with velocity defined relative to time t and position $\mathbf{x}$, is a so-called *fluid-flow metric*; they are the metrics relevant when discussing classical fluid-flow analogue gravity models. We will make these metrics precise in the next section and see that the Gullstrand-Painlevé form of Schwarzschild spacetime is (unsurprisingly) such a metric.

Quantum Fluid-Flow Analogue Models. Quantizing the sound waves of a classical fluid-flow analogue model is equivalent to quantizing the massless scalar field on spacetime; we thus obtain what we call a *quantum fluid-flow analogue model*: an analogy between quantum sound propagation in a fluid and the propagation of a real quantum Klein-Gordon field in curved spacetime described by a fluid-flow metric. We will use such models in Section 2.5 when deriving *Hawking radiation*; we will see that Hawking radiation crucially depends on the notion of an *apparent horizon*, which is easily defined within fluid-flow metrics.

More General Analogue Models. Despite the importance of fluid-flow models, one may consider more general analogue models. For classical models, we can consider ones which are not based on sound propagation in a fluid, but come from the propagation of any scalar field in an arbitrary system, resembling Klein-Gordon field propagation in curved spacetime; we will treat such models in Section 4.1 before specializing to fluid-flow models in Section 4.2. Finally, one may consider general quantum models, which do not come from quantizing a classical fluid-flow model. We will see an example for such a model in Section 4.3.

1.3 Fluid-Flow Metrics

As argued above, the metrics encountered in classical fluid-flow analogue models are the *fluid-flow metrics*. We construct these metrics here and introduce the important notion of an *apparent horizon*. As mentioned above, apparent horizons will be relevant for the existence of Hawking radiation.

General Flows. Consider a fluid flow with flow velocity $\mathbf{V}(t, \mathbf{x})$, and local speed of sound $c_s(t, \mathbf{x})$. In the eikonal approximation sound then follows curves which satisfy

$$\left(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - \mathbf{V}(t, \mathbf{x})\right)^2 = c_s(t, \mathbf{x})^2. \tag{1.26}$$

These are precisely the null curves of the metric (see also [7, Section 2.2])

$$\mathrm{d}s^2 = -c_s(t, \mathbf{x})^2 \,\mathrm{d}t^2 + \sum_{j=1}^{3} (\mathrm{d}x^j - V^j(t, \mathbf{x})\,\mathrm{d}t)^2. \tag{1.27}$$

A metric of this form is what we mean with a *fluid-flow metric*. But (1.27) is not yet the most general fluid-flow metric, since multiplying the metric by any regular conformal factor $\Theta(t, \mathbf{x}) > 0$ does not change the form of null curves.

Therefore, the general fluid-flow metric is

$$\mathrm{d}s^2 = \Theta(t, \mathbf{x}) \cdot \left[- c_s(t, \mathbf{x})^2 \,\mathrm{d}t^2 + \sum_{j=1}^{3} \left(\mathrm{d}x^j - V^j(t, \mathbf{x})\,\mathrm{d}t\right)^2 \right], \tag{1.28}$$

Notice that along timelike curves (such as the worldlines of observers) with proper time τ, we must have

$$\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 \geq \frac{1}{\Theta \cdot c_s^2} > 0, \tag{1.29}$$

because four-velocities have norm -1. In particular, the sign of $\mathrm{d}t/\mathrm{d}\tau$ may not change along the worldline. From (1.28) and again using the normalization of four-velocities, we find that timelike curves must obey

$$\left(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - \mathbf{V}(t,\mathbf{x})\right)^2 < c_s(t,\mathbf{x})^2. \tag{1.30}$$

Thus, timelike curves in fluid-flow metrics correspond to motion relative to the fluid with less than the speed of sound.

Radial Flows. Most of the time however we will be interested in *radial flows*: $\mathbf{V} = V(t,r)\,\mathbf{e_r}$, $c_s = c_s(t,r)$ and $\Theta = \Theta(t,r)$, where $r = |\mathbf{x}|$. The metric then becomes:

Proposition 1.4: Radial Fluid-Flow Metrics

The general fluid-flow metric with radial flow velocity profile $V(t,r)$, speed of sound $c_s(t,r)$ and overall conformal factor $\Theta(t,r)$ is

$$\mathrm{d}s^2 = \Theta(t,r)\cdot\left[-c_s(t,r)^2\,\mathrm{d}t^2 + \left(\mathrm{d}r - V(t,r)\,\mathrm{d}t\right)^2 + r^2\mathrm{d}\Omega^2\right]. \tag{1.31}$$

Since radial fluid-flow metrics (1.31) are specializations of general fluid-flow metrics (1.28), the conclusions about timelike and lightlike curves above also hold: lightlike curves move at the speed of sound, timelike curves with less than that.

We recognize that the Schwarzschild metric in Gullstrand-Painlevé coordinates (1.20) is a radial fluid-flow metric, with $\Theta = 1$, $c_s = 1$ and $V(r) = -\sqrt{r_s/r}$. So the analogy for Schwarzschild spacetime discussed previously really extends to all timelike and spacelike curves.

Apparent Horizons.[4] A spacetime-point $(t_0,\mathbf{x}_0)$ is called a *sonic point*, if the flow of the fluid at $(t_0,\mathbf{x}_0)$ reaches the speed of sound, *i.e.* if $|\mathbf{V}(t_0,\mathbf{x}_0)| = c_s(t_0,\mathbf{x}_0)$. More generally, we speak of *subsonic points* if $|\mathbf{V}| < c_s$ and *supersonic points* if $|\mathbf{V}| > c_s$. A region of supersonic flow at some fixed time t_0 is also called an *ergo-region* at t_0.

Consider a two-dimensional surface ∂U enclosing a volume U in space at a fixed time t_0. ∂U is called an (*outer-*) *trapped surface*, if the normal component of the fluid flow is inward-pointing and supersonic everywhere on ∂U.[5] Intuitively, any observer or light ray on ∂U is doomed to fall into U, *i.e.* to be "trapped" by the flow, at least for a short time after t_0.[6]

The limiting concept of a trapped surface is the *apparent horizon*:

Definition 1.5: Apparent Horizon

A surface ∂U enclosing a volume U in space at time t_0 such that the flow is inwards-pointing and sonic everywhere on ∂U is called an *apparent horizon.*

Intuitively, if an apparent horizon does not change in time, it is possible for light rays to stay stationary at an apparent horizon but observers on it are doomed to traverse it inwards. Typically, we will encounter apparent horizons as boundaries of regions containing trapped surfaces, such that the flow outside the horizon (at least for some finite distance) is subsonic and inside supersonic (again for at least some finite distance). For time-independent

[4] The following definitions are based on [7, Section 2.4.1].
[5] An *inner-trapped surface* is defined identically, except that the flow is everywhere outwards-pointing.
[6] For long times, the time-dependence of $\mathbf{V}$ may be such as to make surfaces which were previously trapped to be no longer trapped later.

apparent horizons of this type, observers can get arbitrarily close to the apparent horizon from the outside, but once they touch it, must traverse it. Note that the location of apparent horizons remain unchanged between conformally equivalent metrics. Figure 1 visualizes the idea of an apparent horizon in a fluid-flow metric.

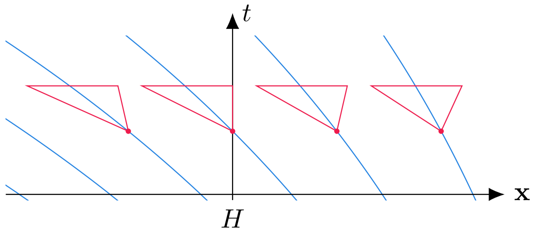


Figure 1: Apparent horizon H in a fluid flow metric. The flow (blue lines) is subsonic on the right and supersonic on the left of the horizon, flowing towards the left in both regions. Light cones (red) are shown at various positions. Inside the horizon, no matter whether a light ray is moving left or right relative to the fluid, it will always end up effectively moving to the left. On the horizon, it is possible for a light ray to stay in place. In this case, the flow and thus the horizon must not be changing with time.

The Painlevé-Gullstrand form of Schwarzschild spacetime discussed above clearly has a time-independent apparent horizon at $r = r_s$; moreover, the fluid flow is always inward-flowing and supersonic in the region $r < r_s$. Therefore, any light ray or observer which enter the region $r < r_s$, will never be able to escape to $r > r_s$. The region $r < r_s$ for all times t is an example of a *black hole*.

Black Holes. Intuitively speaking, a *black hole* is a region of spacetime from which no future-directed causal curve can escape "to infinity" [81]. In our example of Schwarzschild spacetime, "infinity" is at $r \to \infty$, and since the fluid flow is subsonic everywhere outside $r > r_s$, the region $r < r_s$ is indeed a black hole. The boundary of the black hole is called the *event horizon*. We will make both the notion of black hole and event horizon more precise in Section 2.

If the fluid flow is time-independent, then usually an apparent horizon is an event horizon;[7] otherwise, this may not be the case [7]. One can for instance consider an apparent horizon which is dissolving over time, as the fluid flow slows down; this would not be an event horizon, since after dissolution observers can again move freely in the fluid and escape to infinity.

Coordinate-Dependence. It is important to stress that a fluid-flow metric is a *metric*, a coordinate-independent object, *together with a specific choice of coordinates*, making the whole concept coordinate-dependent. This makes sonic points, trapped surfaces and apparent horizons equally coordinate-dependent. There are however choices of coordinates yielding fluid-flow metrics which are particularly natural; specifically, we will often be concerned with fluid flow metrics whose flow vanishes at infinity in an appropriate sense (see Section 2.5).

[7] This is for instance the case for Schwarzschild spacetime in Gullstrand-Painlevé coordinates. It is however not the case in the *Rindler wedge* of flat spacetime: the Rindler horizon can be seen as an apparent horizon, but it is not an event horizon. The difficulty here is that the fluid-flow metric (the Rindler wedge in Rindler coordinates) cannot describe the spacetime behind the horizon. As a rule of thumb, apparent horizons are event horizons in time-independent fluid-flow metrics, only if the horizon is not the boundary of spacetime.

In contrast, black holes and event horizons are coordinate-independent. More precisely, they are defined relative to *future null infinity*, a feature of *asymptotically flat spacetimes* (we will see these concepts in Section 2.1). Now the process of defining infinities in asymptotically flat spacetimes, and even defining asymptotic flatness itself, resembles choosing a particularly natural coordinate system. So even event horizons can be thought of as coordinate-dependent, although the choice of coordinates is heavily restricted and natural and thus practically forced upon us.[8]

In light of this it is less striking that features of analogue gravity such as apparent horizons are coordinate-dependent. Still, the distinction of apparent horizons and event horizons is an interesting question; for instance, as already mentioned a couple of times, we will see in Section 2.5 that Hawking radiation seems to require only an apparent horizon (although apparent and event horizon will be very similar in the considered case). Beyond this, we will not delve much deeper into the differences between the two types of horizon; for an overview, see *e.g.* [7] and sources therein.

ADM Metrics. Metrics of the form (1.28) or (1.31) are important besides their significance in the context of classical fluid-flow analogue gravity models, because they naturally arise (as special cases) when treating general relativity in the *initial-value formalism*, the *ADM-formalism* due to Arnowitt, Deser and Misner [3].

[8]There is however some freedom left, leading to *asymptotic symmetries* (see *e.g.* [81]). Roughly speaking, asymptotic flatness defines a standard of rest for large r, but this standard of rest is ambiguous (at least) up to Lorentz transformations. This makes sense, since in the asymptotically flat region inertial observers are precisely differentiated by Lorentz transformations, and the principle of relativity must hold.

2 Black Holes

After the perhaps uncommon introduction to black holes and horizons from the point of view of "flowing spacetimes" in the last section, we now turn to a more classical treatment. We begin by making the intuitive notion of a black hole as a spacetime region from which no causal curve can escape to infinity precise (Section 2.1). We continue with a discussion of aspects later needed for the black hole information loss paradox, namely the *area theorem* (Section 2.2), the *four laws of black hole (thermo-) dynamics* (Section 2.3), *Bekenstein entropy* (Section 2.4), *Hawking radiation* (Section 2.5), and the *back-reaction* of Hawking radiation on spacetime (Section 2.6).

2.1 Black Holes and Event Horizons

To precisely define a black hole, one needs a notion of infinity, which is conveniently found in *asymptotically flat* spacetimes. We proceed by defining first *asymptotic flatness* and then black holes. Furthermore, we state some of the most immediate properties of black holes and define the concept of *event horizon area*, a crucial ingredient for discussing the dynamics and thermodynamics of black holes later on.

Asymptotic Flatness, Penrose Diagrams. Roughly speaking, (M, g_{ab}) is asymptotically flat, if the metric g_{ab} "becomes flat" as one "goes far away". More precisely [81]:

Definition 2.1: Asymptotic Flatness, Simple Coordinate-Dependent Version

A spacetime (M, g_{ab}) is said to be *asymptotically flat*, if there exist coordinates x^μ with the property that

$$g_{\mu\nu} \to \eta_{\mu\nu} + \mathcal{O}(1/r) \qquad \text{for} \qquad r := \left[\textstyle\sum_{i=1}^{3}(x^i)^2\right]^{1/2} \to \infty \quad \text{at fixed } x^0. \tag{2.1}$$

For example, Minkowski and Schwarzschild spacetime are both asymptotically flat.

We may criticize this definition in two ways: firstly, the provided notion of infinity is *a direction in which one may take a limit*, which is still quite cumbersome, and secondly, it is coordinate-dependent. Here we are not so much interested in tackling the second issue; as it were, coordinate-independent definitions of asymptotic flatness are far from trivial (see [33] and [81] for two coordinate-independent approaches). We will thus stick to Definition 2.1 and introduce some further machinery in order to deal with the first issue.

A useful way to deal with infinity is to mathematically *compactify* spacetime down to a finite size and to then artificially add *points at infinity*, a method invented largely by Penrose [62] [63]. This is achieved through an *immersion* $\psi : M \to \tilde{M}$ of our spacetime (M, g_{ab}) into a larger, *unphysical* spacetime $(\tilde{M}, \tilde{g}_{ab})$. We require the push-forward of the metric to fulfil $(\psi_* g)_{ab} = \Theta^2 \tilde{g}_{ab}$, where Θ^2 is a non-negative function on M. In other words, ψ should be a *conformal immersion*, allowing us to interpret $(\psi(M), \tilde{g}_{ab})$ as the original spacetime (M, g_{ab}), up to a conformal scaling. The scaling is required for the compactification. We choose a conformal scaling, since then a curve in (M, g_{ab}) is causal (or lightlike, timelike, *etc.*) if and only if the corresponding curve in $(\psi(M), \tilde{g}_{ab})$ is causal (or lightlike, timelike, *etc.*).

More concretely: For large values of r, we may choose a time coordinate t and take r as a radial coordinate, supplemented by two angles θ and ϕ, such that $g_{\mu\nu}$ deviates from the Minkowski metric in spherical coordinates by at most $\mathcal{O}(1/r)$:

$$ds^2 = -\mathrm{d}t^2 + \mathrm{d}r^2 + r^2 \mathrm{d}\Omega^2 + \mathcal{O}(1/r). \tag{2.2}$$

Let us define new coordinates (T, R) to replace (t, r) such that the limit $r \to \infty$ becomes a finite limit $R \to R_*$. We can then artificially add limit points and thus construct ψ as well as $\tilde{M}$.

A simple approach would be to set $R = \arctan(r)$. This has the downside that light rays in the t-R-plane "slow down" for $R \to \infty$; we would like to keep the property that light rays travel "diagonally" in this plane. To solve this, we could additionally set $T = \arctan(t)$. The limit $t \to \infty$ at fixed r is however not necessarily well-defined (it is not mentioned in Definition 2.1); it thus makes little sense to also define a new coordinate for that limit. But note that the definition guarantees arbitrarily high values of t, and even asymptotic flatness, if r is sufficiently high; in other words, the limit $v := t+r \to \infty$ is well-defined and spacetime becomes flat in this limit. The same is true for the limit $u := t-r \to -\infty$. If we compactify along v and u, we keep light rays on diagonals and define useful coordinates. We thus set

$$T := \arctan(v) + \arctan(u), \quad R := \arctan(v) - \arctan(u), \quad v := t + r, \quad u := t - r. \quad (2.3)$$

A short computation shows that the metric in these coordinates is $ds^2 = \frac{1}{4}(1 + v^2)(1 + u^2)(-\mathrm{d}T^2 + \mathrm{d}R^2) + r^2 \mathrm{d}\Omega^2 + \mathcal{O}(1/r)$, which diverges as $\mathcal{O}(r^2)$ when $r \to \infty$ along either $v \to \infty$ or $u \to -\infty$. We can remedy this by multiplying with a conformal factor:

$$\Theta^2 := \frac{4}{(1 + v^2)(1 + u^2)}. \quad (2.4)$$

This defines the unphysical metric (for large R) as

$$\widetilde{ds}^2 = -\mathrm{d}T^2 + \mathrm{d}R^2 + \sin^2(R)\,\mathrm{d}\Omega^2 + \mathcal{O}(1/r^3) = \Theta^2\,\mathrm{d}s^2. \quad (2.5)$$

The coordinate transformation $(t, r) \rightsquigarrow (T, R)$ and the factor Θ are defined for large values of r, they thus only define the immersion ψ for large r. But we can always extend Θ and ψ to all M to complete the definition. It remains to define $\tilde{M} - \psi(M)$, at least to add the points at infinity.

For this we first consider the limit $r \to \infty$ at fixed t, corresponding to $R \to \pi$ and $T \to 0$. Since $\tilde{g}$ is well-defined at $R = \pi$, $T = 0$, we may extend $\tilde{M}$ to these points, the so-called points at *spacelike infinity*, denoted by i^0. Spacelike infinity is topologically equivalent to S^2 and not $\mathbb{R} \times S^2$ (the extent in t has disappeared by the two compactifications), but it geometrically corresponds to *a single point*, because the spherical part of the metric vanishes.

Taking the limit $v \to \infty$, *i.e.* $T + R \to \pi$, allows us to add the points at $T + R = \pi$, $0 < R < \pi$. Since these are points that can be reached by null curves (and even by timelike curves, which approach null curves, such as accelerated worldlines), we call these points *future null infinity*, denoted by $\mathcal{I}^+$. Note that the topology of $\mathcal{I}^+$ is $(0, \pi) \times S^2$, a fact reflected by its geometry. This is good news, since $\mathcal{I}^+$ is much more relevant than i^0, by virtue of it being reachable. With the limit $u \to -\infty$ we can analogously introduce the points at *past null infinity*, which we denote by $\mathcal{I}^-$.

Finally, we could attempt the limits $t \to \pm\infty$ at fixed r, corresponding to $T \to \pm\pi$, $R \to 0$. Similarly to i^0 we would call points in this limit the points of *future timelike infinity* and *past timelike infinity*, denoted by i^+ and i^- respectively. As for i^0, they each geometrically correspond to a single point. Causal and timelike curves can reach to i^- and i^+. These points are however not guaranteed to exist by the definition of asymptotic flatness; they thus will not be very important for future discussion.

The fact that i^0 (and $i^\pm$) are single points instead of $\mathbb{R} \times S^2$ (or S^2) is a consequence of our compactification along the directions u and v; but this will not bother us very much, as we will be most interested in $\mathcal{I}^\pm$.

Figure 2 summarizes these ideas in a so-called *Penrose diagram* for the special case of flat spacetime.

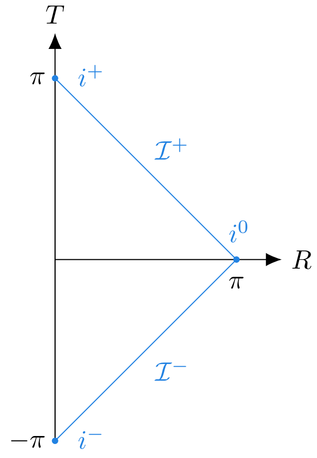


Figure 2: Penrose diagram of flat spacetime in spherical coordinates (angular part is suppressed). The infinities are indicated in blue. Penrose diagrams are always drawn such as to make null curves diagonals.

Black Holes and Event Horizons. With a working definition of asymptotic flatness in place, we can make the "escape to infinity" more precise. "Infinity" is most reasonably identified with future null infinity $\mathcal{I}^+$ of the unphysical spacetime, since spacelike infinity i^0 cannot be reached by future-directed causal curves and future timelike infinity i^+ may not exist. The causal past $J^-(\mathcal{I}^+)$ of future null infinity is then those points (in unphysical spacetime) from which *future-directed causal curves can escape to infinity.* The corresponding points in M are $\psi^{-1}(J^-(\mathcal{I}^+))$; it is physically sensible to pull back $J^-(\mathcal{I}^+)$ from $(\tilde{M}, \tilde{g}_{ab})$ onto (M, g_{ab}), because ψ is conformal and hence maps causal curves onto causal curves. Therefore, $B := M - \psi^{-1}(J^-(\mathcal{I}^+))$ is a candidate definition of the union of all black holes within M; $B = \emptyset$ indicates that no black hole is present.

We would also like to prevent *naked singularities* in our definition, that is singularities outside or on the boundary of B which can be probed by nearby observers. Singularities are not points of M and might not even be points of $\tilde{M}$. Their presence is however seen by the fact that some causal curves are inextendible; *i.e.* some causal curves end or even begin at singularities. A singularity can in principle be detected by an observer if causal curves both end and begin at it. We note that the existence of such singularities forbids the existence of a Cauchy surface, since the ending causal curves and the beginning ones cannot both intersect any candidate spacelike hypersurface. To get rid of naked singularities, we thus demand *strong hyperbolic predictability* for (M, g_{ab}): we require that there exist an open neighbourhood of $\overline{\psi(M - B)} = \overline{\psi(M) \cap J^-(\mathcal{I}^+)}$ which is globally hyperbolic.

Thus, we define, in accordance with [81]:

Definition 2.2: Black Hole

Let (M, g_{ab}) be a strongly hyperbolically predictable spacetime; in particular, it is asymptotically flat with immersion $\psi : M \to \tilde{M}$ into a larger unphysical spacetime $(\tilde{M}, \tilde{g}_{ab})$. The region

$$B := M - \psi^{-1}(J^-(\mathcal{I}^+)) \subset M$$

is called the *black hole region* of M. The connected components of B are called *black holes.* If $B \neq \emptyset$, then M is said to be a *black hole* spacetime.

Let $B_0 \subset B$ be a black hole. We call the boundary $H_0 = \partial B_0$ the *event horizon* of the black hole B_0.

Figure 3 shows the Penrose diagrams of two important black hole spacetimes: that of a collapsing start forming a black hole, and that of Schwarzschild spacetime.

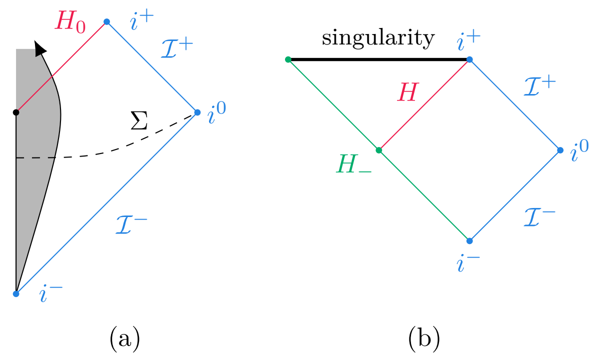


Figure 3: (a) Penrose diagram for a collapsing star forming a black hole (angular dimensions suppressed). The star's interior is shaded in grey, and the trajectory of its surface is indicated by an arrow. The event horizon H_0 does not always exist: for instance, it does not at the "time" of the spacelike slice Σ. The black hole interior lies to the top left of H_0, and we have left out details inside the black hole region. See *e.g.* [81, Section 12.1]

(b) Penrose diagram for Schwarzschild spacetime. The black hole region lies to the top left of the horizon H, and contains a singularity. H_- is the *anti-horizon* which becomes apparent when performing the conformal immersion of the spacetime; it is possible to analytically extend Schwarzschild spacetime through the anti-horizon, thus obtaining also a *white hole region* and an *anti-universe*. But since Schwarzschild spacetime is an idealization of an eternal black hole, and thus unphysical, these extensions are usually also thought of as unphysical. See *e.g.* [81, Sections 6.4 & 12.3].

We will only consider spacetimes for which all horizons are three-dimensional, piecewise smooth submanifolds. Note that our definition of black holes is slightly less general than the one given in [81], since we chose a less general (but simpler) version of asymptotic flatness; this will however not limit us. Let now H_0 be the (piecewise smooth) horizon of a black hole.

First Properties of Horizons. Being the boundary of the region $J^-(\mathcal{I}^+)$, which by definition is generated by causal curves, one might expect a black hole horizon to be generated by limits of those curves. This is indeed true, the limiting curves being null geodesics [60]:

Theorem 2.3: Horizon Generators (Penrose)

A *null generator* of H_0 is a future-directed null geodesic, which cannot be extended towards the future, and once it enters H_0 remains contained in H_0. There can thus be horizon generators which are always contained in H_0, and others which enter H_0 for the first and last time at some spacetime point, called a *caustic point*.

For every non-caustic point $p \in H_0$ there exists a *unique* (up to reparametrisation) null generator γ_p passing through p. For every caustic point $c \in H_0$ there exists at least one null generator passing through c, but there may be multiple. Two null generators can only intersect at some point p if they both enter H_0 at p, *i.e.* once within H_0, null generators do not cross.

See [49] for an illustrated proof. Importantly, the proof does *not* make use of the Einstein field equations (0.1). In fact, our discussion of black holes so far *did not depend on the dynamics* of the gravitational field at all, but only on the properties of causality as well as on the assumptions of asymptotic flatness and non-existence of naked singularities. If naked singularities were allowed, generators could potentially end at such naked singularities. This theorem is the cornerstone for much of the discussion on black hole dynamics and later thermodynamics.

One can show the following useful lemma along the way:

Lemma 2.4: Event Horizons are Null Hypersurfaces

A *null hypersurface* is a hypersurface $S \subset M$ such that at each $p \in S$ there exists a tangent vector $0 \neq k^a \in T_p S$ which is also a normal vector, *i.e.* $\forall v^a \in T_p S : k_a v^a = 0$.

An event horizon H_0 is a null hypersurface (whenever sufficiently differentiable), with normals provided by the tangent vectors of the null generators.

Proof. Assume towards contradiction that there exists a $v^a \in T_p H_0$ with $k_a v^a \neq 0$, where k^a is the null generator through p. Then there exists a suitable linear combination of v^a and k^a (hence lying in $T_p H_0$) which is timelike. From this, it is possible to construct a locally timelike causal curve in H_0, which could then be deformed, keeping it locally timelike and overall causal, to lead from within the black hole to future null infinity, a contradiction. □

Event Horizon Area. Let $B_0 \subset B$ be a black hole. One way to further investigate B_0 is to consider the black hole "at one moment in time". Take for this a Cauchy surface $\Sigma \subset \tilde{M}$ of an open neighbourhood $\tilde{V} \supset \overline{\psi(M - B)}$ in the unphysical spacetime. That such a neighbourhood exists is guaranteed by the assumption of strong asymptotic predictability. Note that the image of the event horizon $\psi(H_0)$ is contained in $\tilde{V}$. Therefore, $\psi^{-1}(\Sigma)$ may be seen as a "moment in time" of the spacetime M (potentially excluding some regions in the interior of B), and $H_{0,\Sigma} := H_0 \cap \psi^{-1}(\Sigma)$ is the *event horizon of* B_0 *at that time.*

$\psi^{-1}(\Sigma)$ can nowhere be tangential to H_0, since the tangent vectors of $\psi^{-1}(\Sigma)$ are all spacelike everywhere, but H_0 has at least one lightlike tangent vector everywhere, the null generator. $H_{0,\Sigma} = H_0 \cap \psi^{-1}(\Sigma)$ is thus either empty, or a two-dimensional, spacelike submanifold; we focus on the latter case. We may integrate over $H_{0,\Sigma}$:

Definition 2.5: Event Horizon Area

Let H_0 be a piecewise smooth event horizon of some black hole and let Σ be a moment in time in the above sense, *i.e.* a Cauchy surface of $\tilde{V} \subset \tilde{M}$. The *area of the event horizon* H_0 *at time* Σ, that is the *area of* $H_{0,\Sigma}$, is then defined as

$$A(H_{0,\Sigma}) := \mathrm{vol}_2(H_{0,\Sigma}) = \int_{H_{0,\Sigma}} \epsilon_{ab}, \qquad \epsilon := \sqrt{|\iota^* g|}\, \mathrm{d}x^1 \wedge \mathrm{d}x^2. \tag{2.6}$$

In other words, $A(H_{0,\Sigma})$ is obtained by integrating the two-dimensional volume form ϵ_{ab} obtained through pull-back of the 4D metric g_{ab} onto $H_{0,\Sigma}$ along the inclusion map $\iota : H_{0,\Sigma} \to M$.

For Schwarzschild spacetime, we have a single event horizon H with area

$$A(H_\Sigma) = A(H) = 4\pi(2M)^2 = 16\pi M^2, \tag{2.7}$$

independent of Cauchy surface Σ (see the discussion of Kerr-Newman spacetimes, of which Schwarzschild is a special case, in Appendix A).

2.2 Area Theorem

With black holes, event horizons and their area defined, we can investigate their dynamics. One of the most important results in this avenue is the *area theorem* due to HAWKING [30], found in 1971. It states that the area of an event horizon never decreases with time. We discuss it here in appropriate detail.

Cross-Sections of Null Congruences. Consider two instants in time Σ_1 and Σ_2 in the sense above (Cauchy surfaces of $\tilde{V} \subset \tilde{M}$), with Σ_2 in the future of Σ_1, *i.e.* $\Sigma_2 \subset J^+(\Sigma_1)$. We begin by noting that H_{0,Σ_1} is in a sense linked to H_{0,Σ_2} through null geodesics: for each $p \in H_{0,\Sigma_1}$ we can find a null generator passing through p which later also crosses H_{0,Σ_2}. The curves form a so-called *congruence of null curves*. By Lemma 2.4, H_{0,Σ_i}, $i = 1, 2$, is orthogonal to all those curves. Thus, $A(H_{0,\Sigma_i})$ may be understood as the *cross-sectional area* of the null congruence, with the cross-section taken at Σ_i.

Every point $p \in H_0$ can be labelled by two numbers x_1, x_2 determining the null generator on which p lies as well as by the affine parameter λ along the curve; this defines a coordinate system on H_0; at caustics, the coordinates are singular. The coordinate vector fields $(\partial_1)^a$ and $(\partial_2)^a$ describe the separation of geodesics in the congruence. Hence, their projections onto the surfaces H_{0,Σ_i} can be used to compute physical separations, and thus area elements of the cross-section. But since the H_{0,Σ_i} are perpendicular to the tangents k^a of the congruence, and because the k^a are null, the scalar products $(\partial_i)_a(\partial_j)^a$, $i, j = 1, 2$, are identical to the corresponding scalar products between the *projected counterparts of the coordinate vector fields.*[9]

Raychaudhuri Equation. Such scalar products evolve according to $\frac{\mathrm{d}}{\mathrm{d}\lambda}[(\partial_i)_a(\partial_j)^a] = (\partial_i)_a k^b (\partial_j)^a{}_{;b} + k^b (\partial_i)_{a;b}(\partial_j)^a = [(\partial_i)^a(\partial_j)^b + (\partial_i)^b(\partial_j)^a] k_{a;b}$. The last equality holds because $[\partial_i, k]^a = 0$, since they are coordinate vector fields. We decompose $k_{a;b} =: \frac{1}{2}\theta h_{ab} + \sigma_{ab} + \omega_{ab}$ into an isotropic part θ, a trace-free symmetric part σ_{ab}, and an antisymmetric part ω_{ab}, where the trace is taken with respect to the metric on any two-dimensional spacelike subspace perpendicular to k^a.[10]

The quantity θ describes isotropic expansion or contraction of the geodesic congruence's cross-sectional area element, σ_{ab} describes area-preserving shear and ω_{ab} describes area-preserving twist. Thus: *the sign of θ controls whether locally the horizon area, considered as cross-sectional area elements of the generator congruence, grows (+) or shrinks (−) over time.*

The evolution of θ is provided by the *Raychaudhuri equation* (after the Indian physicist RAYCHAUDHURI), which can be derived from the relative geodesic acceleration equation (see *e.g.* [81]):

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{1}{2}\theta^2 - \sigma_{ab}\sigma^{ab} + \omega_{ab}\omega^{ab} - R_{ab}k^a k^b. \tag{2.8}$$

It follows from an application of the *Frobenius theorem* that $\omega_{ab} = 0$ (see [81] for details).

Area Theorem. We thus find that if $R_{ab}k^a k^b \geq 0$ everywhere and $\theta < 0$ at some point, then θ will collapse to $-\infty$ within a finite amount of affine parameter. Hence, multiple nearby null generators will cross. This is impossible according to Theorem 2.3; thus, $\theta \geq 0$ if $R_{ab}k^a k^b \geq 0$. Since the null generators provide an injective map from H_{0,Σ_1} to H_{0,Σ_2}, we find:

[9] The same reasoning leads to the perhaps surprising conclusion that the cross-sectional area of thin light beams are Lorentz-invariant, a fact often met in relativistic geometric optics.

[10] Any choice is possible due to independence on projection of the relevant scalar products, as discussed above. For concreteness, we may for instance choose the tangent space of H_{0,Σ_1} or H_{0,Σ_2}.

Theorem 2.6: Second Law of Black Hole Dynamics: Area Theorem (Hawking)

Let H_0 be the event horizon of a black hole in a spacetime which satisfies

$$R_{ab}k^a k^b \geq 0 \tag{2.9}$$

everywhere and for every null vector k^a. Let Σ_1 and Σ_2 be Cauchy surfaces of $\tilde{V} \subset \tilde{M}$ with $\Sigma_2 \subset J^+(\Sigma_1)$. For the areas of H_0 as seen by Σ_1 and Σ_2 it then holds that

$$A(H_{0,\Sigma_2}) \geq A(H_{0,\Sigma_1}). \tag{2.10}$$

Let us remark some points:

1. The theorem as stated in 2.6 applies to connected components of B, so it must also apply to all of B.
2. $H_{0,\Sigma}$ can have multiple connected components, *i.e.* multiple black hole horizons at time Σ, even if B_0 has a single connected component. For instance, what we intuitively view as a merger of "two black holes" is all contained in a single connected component of B.

   If the horizon generators traversing a connected component $H_1' \subset H_{0,\Sigma_1}$ (*i.e.* a black hole horizon at time Σ_1) all traverse a connected component of $H_2' \subset H_{0,\Sigma_2}$ later on, then the area theorem even holds for $A(H_1')$ and $A(H_2')$. This can be seen by carefully inspecting the above proof outline.

   So it is not possible to locally decrease the horizon area of a "single black hole", even if in total the area increases.
3. So far, we have not used the Einstein field equations. In this sense, the dynamics of black hole event horizons described by the area theorem are *not* a consequence of the dynamics of spacetime. Rather, it is based on the purely geometric properties of null geodesics, and the assumptions we made.
4. Amid the assumptions, the condition (2.9) stands out. We can better understand it by finally invoking the Einstein field equations; we quickly find that (2.9) is equivalent to $T_{ab}k^a k^b \geq 0$ for all null vectors k^a. But this in turn follows by continuity from the so-called *weak energy condition*:

   $$T_{ab}v^a v^b \geq 0, \qquad \text{for all timelike vectors } v^a. \tag{2.11}$$

   The weak energy condition states that energy density measured by local observers may never be negative, a condition that is usually sensible to assume. Note however that (2.11) and even (2.9) are violated in the process of Hawking radiation, so the area theorem must not necessarily hold in that case [81, Section 14.4]; we will come back to this in Section 2.6.

   Starting with the assumption (2.11), which is much more natural to assume than (2.9), the Einstein field equations thus nevertheless enter the picture. In this sense, the dynamics of event horizons, as described by Theorem 2.6, are a consequence of the dynamics of spacetime, and of the weak energy condition.
5. The area theorem 2.6 is similar in form to the *second law of thermodynamics*, which states that *entropy may never decrease*; see *e.g.* [40]. We will see that the other three laws also resemble thermodynamic laws, leading to the thermodynamic interpretation of black holes.

   As with all second-law-type theorems, one may wonder where the time-asymmetry is coming from. We can trace it back to Definition 2.2 of a black hole, which was heavily time-asymmetric.

2.3 Four Laws of Black Hole (Thermo-) Dynamics

The area theorem invites the possibility of a thermodynamic interpretation of black holes. This possibility is further strengthened by the *four laws of black hole dynamics*, due to Bardeen, Carter and Hawking [8] in 1973. They take the area theorem as second law and provide a zeroth, first and third law, all to some extent similar to thermodynamic laws. Since we will not need the remaining three laws as much as the area theorem, we will only present them briefly here; a treatment of the zeroth and first law more akin in depth to that of the area theorem in the previous section can be found in Appendix A. We close the section by discussing how the four laws of black hole dynamics can be interpreted as thermodynamic laws. Here we will always consider a single black hole and unambiguously denote its horizon by H.

Stationary Electrovac Black Holes. The zeroth, first and third law of thermodynamics typically concern systems in *equilibrium* [40]. The black-hole-notion corresponding to a system in equilibrium are *stationary*,[11] *electrovac* black hole spacetimes. Those are spacetimes that intuitively "do not change with time" as seen by observers at infinity, and are thought to describe spacetimes after gravitational perturbations have died down and all energy-momentum has been swallowed by the black hole, with the possible exception of a remaining electromagnetic field. Recent progress [27] shows that at least slowly rotating black holes are stable to perturbations, providing evidence that stationary spacetimes really are equilibrium states.

Stationary electrovac black holes are surprisingly simple: they are completely described by the *Kerr-Newman metric* [52], which is parametrized by the three parameters *mass* $M \geq 0$, *angular momentum* $J \in \mathbb{R}$ and *electric charge* $Q \in \mathbb{R}$. In particular, all stationary electrovac black hole spacetimes are also *axisymmetric*. The simplicity of stationary electrovac black holes is captured by the *no-hair-theorem*; see [81, Section 12.3] for a discussion thereof. This is also very much in line with a possible thermodynamic interpretation of black holes, since thermodynamic systems in equilibrium are typically described by few, macroscopic quantities [40]. The case $J = 0$, $Q = 0$ is the familiar *Schwarzschild spacetime*, which we are primarily interested in; recall that we discussed the Schwarzschild metric in Gullstrand-Painlevé coordinates at the beginning in Section 1.1.

There exists a family of so-called *static* observers in the asymptotic region with the properties that they perceive the black hole as unchanging in time and that their worldlines are orthogonal to some spacelike hypersurface; the latter property ensures that they are not rotating around the axis of the black hole; see [81]. The parameters M, J and Q are the mass, angular momentum and electric charge static observers attribute to the black hole, as follows: In the weak-field, linear regime of general relativity (*i.e.* $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ and considering only linear terms in $h_{\mu\nu}$) it is possible to determine moments of the energy-momentum tensor $T_{\mu\nu}$ such as energy, angular momentum, *etc.* enclosed in a volume of space using suitable integrals of $h_{\mu\nu}$ over the surface of the volume; for instance, one may compute the total mass using Gauss' law of gravity, which holds in the linear regime. Since asymptotically flat spacetimes enter the linear regime towards infinity, we can *define* energy, angular momentum, *etc.* by the corresponding surface integrals even in asymptotically flat spacetimes. However, since asymptotically flat spacetimes do not everywhere fit into the linear regime, the quantities thus obtained are not necessarily moments of $T_{\mu\nu}$ but contain contributions from curvature; *e.g.* black holes with vanishing energy-momentum tensor can have a mass. In order to perform the split $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ we need a notion of reference frame, which is supplied by the static observers. This is the sense in which M, J and Q

[11] An asymptotically flat spacetime is said to be stationary, if it admits an asymptotically timelike Killing vector field ξ^a. The vector field provides a time standard at least at infinity.

correspond to quantities the static observers measure.[12] One also identifies $E = M$ as the *energy* of the black hole. See [49, Chapter 19] and [81] for details.

Finally, we mention that it is possible to define the *angular velocity* Ω_H *of the horizon* for Kerr-Newman black holes (see *e.g.* [33]). For us, it is enough to know that Ω_H carries the same sign as J and in particular vanishes if and only if J vanishes.

First Law of Black Hole Dynamics One can obtain an equation similar to the first law of thermodynamics (see *e.g.* [40]) when considering infinitesimal changes $\delta M = \delta E$, δJ and δQ, while requiring for the black hole to remain a Kerr-Newman black hole, essentially considering all potentially possible infinitesimal processes in the family of equilibrated black holes. Namely:

Theorem 2.7: First Law of Black Hole Dynamics, Kerr-Newman version

Consider the variation of parameters (M, J, Q) of Kerr-Newman spacetimes. It holds that

$$\delta M = \frac{\kappa}{8\pi}\delta A(H) + \Omega_H \delta J + \Phi\delta Q, \tag{2.12}$$

where

$$\Phi = \frac{Q}{A(H)} \cdot \left(M + \sqrt{M^2 - a^2 - Q^2}\right) \tag{2.13}$$

is the electrostatic potential on the event horizon and κ is the *surface gravity* (see below).

Bardeen, Carter and Hawking [8] actually obtained a more general version of this theorem, also allowing for certain non-Kerr-Newman spacetimes and the potential presence of matter in the form of a perfect fluid.

The first law of black hole dynamics indeed looks like the first law of thermodynamics, $\delta E = T\delta S +$ (other contributions), with M identified as energy E, if we also interpret $A(H)$ as entropy S and $\kappa/8\pi$ as temperature T. We know from the area theorem (second law), theorem 2.6, that the horizon area $A(H)$ could potentially be seen as at least monotonously increasing with entropy, so this idea could indeed work.

Zeroth Law of Black Hole Dynamics. The *surface gravity* κ appearing in the first law of black hole dynamics is a function defined on the horizon of stationary black holes (not necessarily electrovac) with a rather technical definition which we will not need here; see *e.g.* [81, Section 12.5].

The name "surface gravity" derives from special cases (for instance Schwarzschild), where one can show that κ is the acceleration felt by a stationary observer at infinity holding a mass on a string just above the horizon [81, Section 12.5]. In the Schwarzschild case we have (see Appendix A and sources therein)

$$\kappa = \frac{1}{4M}. \tag{2.14}$$

The zeroth law of black hole dynamics roughly states [8] that the surface gravity is *constant* everywhere on the horizon for certain types of black holes, including all Kerr-Newman black holes. Thus, it makes sense that κ could be included in the first law of black hole dynamics.

The zeroth law of thermodynamics states that *equilibrium is an equivalence relation* (*i.e.*, two systems in equilibrium with a third are also in equilibrium with each other). A consequence

[12] The charge Q is computed not from $h_{\mu\nu}$, but from the electromagnetic field strength tensor $F_{\mu\nu}$, albeit in a very similar way.

of the zeroth and first laws of thermodynamics is that *temperature is well-defined* for systems in equilibrium, in the sense that the equivalence classes of equilibrium can be labelled by a single number, temperature T. See [40].

Compare this to the zeroth law of black hole dynamics which states that surface gravity is well-defined for "equilibrium" black holes; in this sense, surface gravity is analogous to temperature. Note however that the equilibrium describing stationary black holes is a slightly different notion than the equilibrium between thermodynamic systems: for black holes, we can only say whether a *single* black hole spacetime is in equilibrium (by checking whether it is stationarity and electrovac), not whether two black hole spacetimes are "in equilibrium with each other". Hence, the analogy between the zeroth law of black hole dynamics and the zeroth law of thermodynamics is much weaker than for the other two laws.

Third Law of Black Hole Dynamics. One finds that $\kappa = 0$ is only possible for so-called *extremal* black holes, that is, they are on the verge of losing their event horizon and becoming naked singularities; see *e.g.* [81, Section 12.3].

One might thus hope that nature prevents us from reaching $\kappa = 0$. In light of possible parallels between black hole dynamics and thermodynamics, this would correspond to the fact that absolute zero ($T = 0$) cannot be reached by a finite sequence of processes, a fact that follows from the *third law of thermodynamics* (version formulated by Planck): the entropy S approaches a value independent of the state of the system, if $T \to 0$ [67].

Bardeen, Carter and Hawking hypothesized [8] that a law for the unreachability of $\kappa = 0$ exists, and Wald later showed [79] that $\kappa = 0$ was indeed unreachable for a specific family of processed applied to Kerr-Newman black holes. Finally, Israel stated and proved a powerful version of the law [35]. Since it is rather technical and because we will not need the third law rigorously, we do not state it here but refer the reader to literature.

The third law of black hole dynamics is analogous to the unreachability of absolute zero temperature in equilibrium thermodynamics, *but not* to the third law of thermodynamics. In fact, black holes with $\kappa = 0$ can have all possible values for their horizon area $A(H)$, and there cannot be a law for black holes perfectly analogous to the third law of thermodynamics, see [81, Section 12.5].

Reversible and Irreversible Processes. In equilibrium thermodynamics, the second law usually follows by considering *reversible* and *irreversible* processes [40]. One can similarly define such processes for black holes: Christodoulou [18] and later Ruffini [19] discussed such processes and concluded simultaneously with Hawking [30] that horizon area should play the role of entropy. Thus, the area theorem, when restricted to "equilibrium" black holes, can also be understood from the point of view of such transformations. Reversible processes are those that leave $A(H)$ invariant, irreversible ones are those that increase it. For an extensive overview of these processes, in particular of the *Penrose process* allowing for arbitrary, reversible changes of angular momentum J [61] [64], see [49, Section 33.8]

These processes allow us to understand the first law of black hole dynamics (2.12) not just as infinitesimal relations holding for black holes in the Kerr-Newman family, but also as a statement about the change of quantities brought about by infinitesimal processes on black holes followed by equilibration [82]. This much resembles how the first law is viewed in thermodynamics.

Interpretation of the Four Laws as Thermodynamic Laws. Interpreting the four laws of black hole dynamics as thermodynamic laws entails interpreting certain quantities as thermodynamic quantities. We reasonably keep the interpretation of E as the energy of the black hole. We take black holes to be *at equilibrium*, if they are of the Kerr-Newman family. The second law of black hole dynamics 2.6 then suggests that the *entropy* of a black hole is

$$S := f(A(H)), \tag{2.15}$$

where f is a monotonically increasing function yet to be determined. The first law 2.7 further tells us that for fixed J and Q

$$\delta E = \frac{\kappa}{8\pi}\delta A(H) = \frac{\kappa}{8\pi f'(A(H))}\delta S, \tag{2.16}$$

analogous to the first law of thermodynamics $\delta E = T\delta S$, which allows us to define the *temperature* of the black hole as

$$T := \frac{\kappa}{8\pi}\frac{1}{f'(A(H))}. \tag{2.17}$$

The function f cannot be determined at this stage; a very good guess come from BEKENSTEIN's information-theoretical approach [9] to black hole thermodynamics, which was then further cemented by HAKWING's discovery of Hawking radiation [31]. We will address both of these in the coming sections.

Problems. Despite a strong link between thermodynamics and black hole dynamics, there are still differences. For one, when discussing the zeroth law of thermodynamics, there is no readily available notion of black holes being in equilibrium *with each other*, since the considered systems are entire spacetimes which cannot interact. This is partly remedied by considering multiple black holes within the same spacetime, far separated from each other; then, exchange of gravitational waves and electromagnetic radiation (*e.g.* Hawking radiation) allows for interaction. Another difference is the fact that the area theorem 2.6 also applies to individual black holes and it is impossible to locally decrease horizon area while globally keeping or increasing horizon area (see above for detailed notions). This is in contrast to thermodynamics, where there is no reason why entropy might not decrease locally as long as it does not decrease globally in closed system. See [82] for a detailed discussion of the differences between thermodynamics and black hole dynamics.

Beyond Equilibrium and Phenomenological Thermodynamics. While the zeroth, first and third law only hold for equilibrium black holes and thus have their counterparts in *equilibrium* thermodynamics, the second law holds for *all* black holes and thus resembles a second law from *non-equilibrium thermodynamics*, such as the *H-theorem* from kinetic gas theory. Of course, this gives us no reason to assume that black holes are gases described by kinetic theory, but it invites the possibility that they are described by some yet unknown microscopic theory.

This possibility is further strengthened by the fact that the thermodynamic laws of virtually all known thermodynamic systems such as gases, solid-state systems and even quantum systems are *phenomenological rather than fundamental*: they derive from underlying microscopic laws such as kinetic theory, atomic theory or quantum (field) theory.

At the least we would expect gravity to emerge from some kind of quantum gravity, so this could be the source of the observed thermodynamic laws. But it is also possible [54] that another (perhaps classical) theory hides between general relativity and quantum gravity, similarly to how kinetic gas theory stands between the phenomenological thermodynamics of gases and quantum theory of the gas particles.

The question of the microscopic origin of black hole thermodynamics has no conclusive answer as of today [82]. Undoubtedly, an answer or even an attempt at an answer would have far-reaching consequences for the search of a theory of quantum gravity.

2.4 Bekenstein Entropy

The black hole entropy S can also be understood as an information-theoretical entropy, as was demonstrated by Bekenstein in 1973 [9]. With this, it is possible to estimate the monotonously increasing function f in (2.15). We give a brief account of the argument. We use geometric units here ($c = G = 1$, but $\hbar$ and k_B are left explicit).

Dropping Bits into Black Holes. Bekenstein [9] begins by considering Kerr black holes (Kerr-Newman with $Q = 0$), and builds on the area theorem, Theorem 2.6, as well as the black-hole processes described by Christodoulou (see previous section), to make the assumption that

$$H_B = f(A) \tag{2.18}$$

is the information-theoretical entropy of a black hole with event horizon area A (previously called $A(H)$, but now we leave out the explicit mention of the horizon H to prevent confusion), with f some monotonously increasing, but unknown function. Note that the paper by Bardeen, Carter and Hawking [8] on the four laws of black hole dynamics (see previous section) had only been published shortly after Bekenstein's paper.

He argues that dropping a fundamental particle with no internal structure into a black hole should result in the black hole's entropy increasing by *at least one bit*, reflecting the uncertainty about whether the particle still exists after it has entered the horizon. Indeed, a probability distribution $P(\text{yes}) = P(\text{no}) = 1/2$ has an entropy of $H(P) = 1$ according to (0.8). The plan is then to compute the minimal possible increase in area ΔA_{min} one can obtain by dropping the particle into the black hole (different ways of dropping the particle may result in different area increases, we want the smallest possible area increase). This increase should then correspond to an increase of entropy $\Delta H_B = 1$ by one bit. From (2.18) it then follows that for any area A,

$$1 = \Delta H_B = f(A + \Delta A_{\text{min}}) - f(A), \tag{2.19}$$

allowing us to find f.

Minimal Area Increase. The most complicated part of Bekenstein's argument is the computation of the minimal horizon area increase ΔA_{min}. From earlier computations of processes where point particles are lowered into black holes [18], one obtains that $\Delta A_{\text{min}} = 0$, leading to f being ill-defined. Under the more reasonable assumptions that the dropped particle cannot be localized below a certain length scale, Bekenstein extended this result [9] to:

Lemma 2.8: Minimal Horizon Area Increase (Bekenstein, Christodoulou)

If a particle of rest mass m and proper radius b is dropped into a Kerr black hole (Kerr-Nemwan with $Q = 0$), the minimal possible horizon area increase is

$$\Delta A_{\text{min}} = 8\pi m b. \tag{2.20}$$

For the detailed computations, see the original paper [9]. We only note that they make use of the Einstein field equations (0.1) in a subtle way: it is assumed that the *energy* and *angular momentum* of the particle's geodesic (defined by the two conserved quantities associated with the time-translational and axial Killing vector fields present in Kerr spacetime, see *e.g.* [81, Section 12.3]) are added to the black hole's energy (total mass M) and angular momentum J. This is indeed true if the Einstein field equations hold.

We note that Christodoulou's result $\Delta A_{\text{min}} = 0$ occurs because for particles with $b = 0$ there exist geodesics which just graze the horizon, and it is possible that although the mass of the black hole increases due to the energy of the geodesic, its angular momentum decreases by just the right amount, such that the area is not affected overall.

Two natural choices for b are present: the *Compton wavelength* $\hbar/m$ and the *Schwarzschild radius* $2m$. The Compton wavelength is larger for $m^2 < \hbar/2$, giving $\Delta A_{\min} = 8\pi\hbar$. The Schwarzschild radius is larger for $m^2 > \hbar/2$, giving $\Delta A_{\min} = 16\pi m^2 > 8\pi\hbar$. Therefore, the smallest possible increase is $\Delta A_{\min} = 8\pi\hbar$.

Note that the lower bounds for the proper radius b of the particle are mere estimates. Thus, the smallest increase of area is really only roughly $8\pi\hbar$, up to a prefactor of order one: $\Delta A_{\min} \sim \hbar$.

Bekenstein Entropy. We can now solve (2.19) and find that f is a linear function:

Theorem 2.9: Bekenstein Entropy (Bekenstein)

The information-theoretic entropy of a black hole is the *Bekenstein entropy*

$$H_B = f(A) \sim \frac{A}{\hbar}. \tag{2.21}$$

As we will see in the next section, the missing factor in Bekenstein's formula (2.21) required to make H_B a thermodynamic entropy is $k_B/4$. The factor of k_B is to be expected when transitioning from an information-theoretic entropy to a thermodynamic one; this transition also entails a factor of $\ln 2$. The remaining factor can be ascribed to the uncertainty of the approximation $\Delta A_{\min} \sim \hbar$; it is fixed by the temperature of Hawking radiation, as we will see.

Further Results. With Bekenstein entropy in place, it is possible to prove a *generalized second law*, holding for a region of spacetime containing a black hole and regular matter [10]: the total matter entropy plus black hole entropy may never decrease.

Furthermore, one can derive an upper bound on the entropy-to-energy ratio of any system with some given radius [11]. This so-called *Bekenstein bound* is saturated by a black hole.

2.5 Hawking Radiation

Hawking's 1975 discovery [31] of the thermal radiation emitted by black holes, now called *Hawking radiation*, introduced a radical shift away from the classical idea of black holes being "completely black" towards objects which emit quantum radiation. Hawking radiation occurs in any quantum field on spacetime, but is most prominent in massless fields such as the electromagnetic field. As mentioned above, this discovery would allow for a better thermodynamic interpretation of black holes: the emitted radiation turns out to have a blackbody spectrum with temperature $T_H = \kappa/2\pi$, thus allowing us to interpret T_H as the temperature of the black hole, in accordance with the suspicions raised in the previous section. Perhaps most important for us is the fact that the emitted Hawking radiation is in a *mixed* quantum state, implying entanglement between the radiation and some other system, if the unitarity of quantum mechanics is to be kept intact. As we will see, this will be an important ingredient for the black hole information loss paradox.

We derive Hawking radiation in this section for general radial fluid-flow metrics of the form (1.31), with some slight restrictions; as we have argued in Section 1.3, these are the metrics with appropriate coordinates relevant for classical fluid-flow analogue models, with the Schwarzschild metric in Gullstrand-Painlevé coordinates being one of them. We employ a generalization of the derivation in [69] (they use simpler metrics) and sometimes use tools from the derivation in [76] (they forego much of the canonical quantization step, yielding a very compact derivation); in particular, we will use the idea of [76] that Hawking radiation does not necessarily need an event horizon, but only an apparent horizon which may even be slowly evolving in time.

We will thus see that Hawking radiation is a very general phenomenon which has not much to do with gravity; in fact, the Einstein field equations (0.1) will never enter the derivation. Instead, we will arrive at the conclusion of [76]: that Hawking radiation at temperature $T_H = \kappa/2\pi$ is a generic feature of any system with a metric (importantly, gravitational *or* analogue) containing a slowly-evolving apparent horizon with well-defined *surface gravity* κ. We will complete the link to Section 2.3, by noting that this surface gravity is the same quantity in the context of this section as it was in Section 2.3.

Because of the generality of the argument will we not just derive Hawking radiation for the context of general relativity, as is usually done in derivations such as the one in [81], but we will also be able to conclude that Hawking radiation occurs in quantum fluid-flow analogue models if an apparent horizon is present.

To keep the present section as short as possible, we have moved some computations to Appendix B.

Preparations. Of all the metrics described by (1.31), we consider only those that are asymptotically flat for $r \to \infty$, and possess a single apparent horizon at some radius $r_H(t) > 0$, with the flow being supersonic inwards for $r < r_H(t)$ and subsonic for $r > r_H(t)$ at any time t. These assumptions make sense, since we wish to study an isolated black hole. Note that the apparent horizon must not necessarily be the event horizon, since $r_H(t)$ is allowed to change over time; roughly speaking, a piece of fluid once contained within but close to the horizon could find itself outside again the next moment. We will assume slow time dependence for the quantities V (and thus r_H), Θ and c_s, so the event horizon (if one exists) and the apparent horizon should be fairly close to each other.[13] We also assume V, Θ and c_s to be sufficiently smooth.

When quantizing a classical field, we typically have the choice over two methods [65]: *canonical quantization* and *path integral quantization*. We choose here to use canonical quantization, as was done in HAWKING's original derivation [31], is done in [69], and is strongly implied in [76]. We choose canonical quantization, since this most readily gives us access to the particles created in the Hawking radiation process. Note that path integral approaches exist, famously the one by HARTLE and HAWKING [29].

We could in principle consider any field theory on the given curved spacetime and investigate a potential Hawking radiation effect. As a toy model, one however often considers the *massless, real Klein-Gordon field*. It is known [31] that many of the features characteristic of Hawking radiation are exhibited already by the comparatively simple massless, real Klein-Gordon field. HAWKING [31] originally considered the Klein-Gordon field for simplicity. We will do so too.

Field Modes. As is usual with canonical quantization of fields [65], we begin by investigating the solutions of the classical field equation; we decompose these into linearly independent modes, which are then quantized.

In our case, the field equation is the massless Klein-Gordon equation in curved spacetime:

$$\partial_\mu \left(\sqrt{|g|}\, g^{\mu\nu} \partial_\nu \Phi \right) = 0, \tag{2.22}$$

where $g_{\mu\nu}$ is the metric (1.31) of Proposition 1.4, $\Phi = \Phi(t, r, \theta, \phi)$ is the scalar field, and we have divided by the usual prefactor $1/\sqrt{|g|}$.

[13] We will soon make precise what we mean by "slow time dependence".

Let us briefly assume that the metric components do not change at all in time, *i.e.* V, c_S and Θ are time-independent. Then it is useful to introduce the *out- and ingoing null coordinates* as

$$u := t - \int^r \frac{dr'}{c_s(r') + V(r')}, \qquad v := t + \int^r \frac{dr'}{c_s(r') - V(r')}, \tag{2.23}$$

since the metric (1.31) then takes a simpler form. See Appendix B for a motivation and derivation of these coordinates. At the apparent horizon we have $V = -c_s$ and u becomes undefined; or at least, special care is needed when considering the coordinate u on both sides of the horizon simultaneously. To prevent problems, we can define *two* variables $u_<$ and $u_>$ instead of one, on the regions $r < r_H$ and $r > r_H$ respectively, by choosing two different lower bounds $r_< < r_H$ and $r_> > r_H$ for the integral in the definition of u:

$$u_< := t - \int_{r_<}^r \frac{\mathrm{d}r'}{c_s(r') + V(r')}, \qquad u_> := t - \int_{r_>}^r \frac{\mathrm{d}r'}{c_s(r') + V(r')}. \tag{2.24}$$

We will later see that the integral in u can actually be extended through $r = r_H$ thanks to a natural choice of regularization; so after all, $u_<$ and $u_>$ will be related and unified in a single coordinate u. We will come back to this issue.

Furthermore, we will decompose Φ into a radial and angular part, with the angular part further decomposed into spherical harmonics $Y_{l,m}$, $l = 0, 1, 2, \ldots$ and $m = -l, -l+1, \ldots, l$. This makes sense given the spherical symmetry of the metric (1.31).

One can now show (see Appendix B) that the general solution of (2.22) is

$$\Phi(t, r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{Y_{l,m}(\theta, \phi)}{r\sqrt{\Theta(r)}} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{\sqrt{\omega}} \sum_{\alpha = u_<, u_>, v} A_{\omega,\alpha}^{l,m} F_{\omega,\alpha}^{l}(r)\, \mathrm{e}^{-\mathrm{i}\omega\alpha}, \tag{2.25}$$

with the reality condition

$$A_{\omega,\alpha}^{l,m} = \bar{A}_{-\omega,\alpha}^{l,-m} \tag{2.26}$$

imposed, and $F_{\omega,\alpha}^{l}(r)$ are real-valued functions defined on the respective domains of $u_>$, $u_<$ and v. These functions encode the precise radial fall-off rate of modes; in flat spacetime, we would have $F_{\omega,\alpha}^{l} = \text{const}$. We can choose them such that $F_{\omega,u_>}^{l}, F_{\omega,v}^{l} \to 1 + \mathcal{O}(r^{-1})$ for $r \to \infty$.

If the metric is not time-independent, then this solution for Φ remains approximately correct for those frequencies ω compared to which the time-dependence of the metric is slow:

$$|\omega| \gg \max\{|\dot{V}/V|, |\dot{c}_s/c_s|, |\dot{\Theta}/\Theta|\}. \tag{2.27}$$

In those cases one can effectively ignore the derivatives $\dot{V}$, $\dot{c}_s$ and $\dot{\Theta}$. We will assume this for all frequencies which we are interested in; this is what we mean by "slow time dependence".

Spherically Symmetric Modes ($l = 0$). Let us restrict to the spherically symmetric modes $l = 0$ (and hence $m = 0$). We will later see that modes with higher angular momentum do not matter for the main features of Hawking radiation; for that reason, higher modes are often ignored in derivations of Hawking radiation (as for instance in [81]).

We now have (leaving out the indices $l = 0$, $m = 0$)

$$\Phi(t, r, \theta, \phi) = \frac{1}{r\sqrt{\Theta(r)}} \int_0^{\infty} \frac{\mathrm{d}\omega}{\sqrt{\omega}} \sum_{\alpha = u_<, u_>, v} F_{\omega,\alpha}(r) \cdot \left[A_{\omega,\alpha}\mathrm{e}^{-\mathrm{i}\omega\alpha} + \bar{A}_{\omega,\alpha}\mathrm{e}^{\mathrm{i}\omega\alpha}\right]. \tag{2.28}$$

Quantization. During canonical quantization, classical *real* observables are replaced by quantum observables, that is *Hermitian operators* [65]. The commutation relations between quantum observables are taken from the Poisson brackets of the corresponding classical observables: $[\hat{X}, \hat{Y}] = i\{X, Y\}$, where $\hat{X}$ and $\hat{Y}$ are the quantum Hermitian observables corresponding to the classical real observables X and Y.

In our case of a real classical scalar field theory awaiting quantization, the most important observables are the field Φ as well as its canonical conjugate. Note that the definition of a canonical conjugate requires the choice of a *time standard* (see *e.g.* [41] [42]). However, all the time standard is needed for, is to give a different description of the dynamics which happens to *look* highly non-covariant, although the dynamics *are* clearly covariant in the case of a Klein-Gordon field. So roughly speaking, the choice of time standard should not influence the physics of the field once quantized. We may thus take the time coordinate t as standard, which at infinity corresponds to the time of static observers. Another very useful time standard will be the proper time of observers falling into a horizon.

It is often more useful to deal with the modes $\propto A_{\omega,\alpha}\,\mathrm{e}^{-\mathrm{i}\omega\alpha}$ instead of the field Φ. But the amplitudes are *complex-valued* and thus the modes do not directly qualify as classical observables. To circumvent that we can split them into real and imaginary parts, which themselves are each real observables, and will be replaced by Hermitian operators during quantization. Accordingly, the amplitudes $A_{\omega,\alpha}$ will be replaced by *not necessarily Hermitian* operators $\hat{A}_{\omega,\alpha}$. Because in the classical field all modes are independent of each other (vanishing Poisson brackets, since we can view them as individual physical systems), so will they be in the quantum theory (vanishing commutators). From this we can immediately conclude that $[\hat{A}_{\omega,\alpha}, \hat{A}^\dagger_{\omega',\alpha'}] = \hat{C}_\alpha(\omega) \cdot \delta(\omega - \omega') \cdot \delta_{\alpha,\alpha'}$, for some operators $\hat{C}_\alpha(\omega)$, and all other commutators vanish. In fact, it even holds that

$$[\hat{A}_{\omega,\alpha}, \hat{A}^\dagger_{\omega',\alpha'}] = C_\alpha(\omega) \cdot \delta(\omega - \omega') \cdot \delta_{\alpha,\alpha'} \cdot \hat{\mathrm{id}} \tag{2.29}$$

for some real-valued functions $C_\alpha(\omega)$, $\alpha = u_<, u_>, v$.[14] Since our spacetime is asymptotically flat our $u_>$- and v-modes modes should behave as in flat spacetime in the asymptotic region, and thus it follows that $C_{u_>}(\omega) > 0$ and $C_v(\omega) > 0$ [65]. The precise values depend on the exact fall-off given by $F_{\omega,u_>}$ and $F_{\omega,v}$, which we already fixed earlier (see Appendix B for details). Furthermore, $C_{u_<}(\omega) < 0$;[15] We will come back to the precise value of $C_{u_<}(\omega)$ later.[16]

Particles at Infinity. The commutation relations (2.29) imply that for every ω and α *separately* the Hilbert space of field states (*i.e.* the space on which our operators act) contains a *ladder structure*: one of either $\hat{A}_{\omega,\alpha}$ or $\hat{A}^\dagger_{\omega,\alpha}$ serves as a *creation operator* and the other as the corresponding *annihilation operator*, allowing us to add and remove "things" *in discrete steps* to and from the field. These "things" provide the simplest notion of *particles*; and since every mode has its own ladder structure independent of all others, each particle has

[14] A complex-valued classical observable z which is a linear combination of the canonical variables (here $\Phi(x)$ and $\Pi(x)$) satisfies $\{z, \bar{z}\} \propto \mathrm{i}$; this is readily checked by explicitly writing this linear combination and evaluating the Poisson brackets. In our case, quantization gives $[\hat{A}_{\omega,\alpha}, \hat{A}^\dagger_{\omega',\alpha'}] \propto \hat{\mathrm{id}}$.

[15] See *e.g.* [69] for details. The negative sign has to do with the way modes are extracted from the field and its canonical conjugate via Fourier transforms with respect to well-defined coordinates such as r, and u changing its dependence on r when traversing the horizon; this r-dependence of u will come up again and play an important role for identifying particles seen by an infalling observer.

[16] We could rescale the operators $\hat{A}_{\omega,\alpha}$ by $|C_\alpha(\omega)|^{-1/2}$, in order to get $|C_\alpha(\omega)| \rightsquigarrow 1$. There is however also another issue to consider when choosing the normalization of modes: the relative normalization of modes in the Hamiltonian, or equivalently, the relative *norm* of modes. This ensures that the frequency of modes correctly translates to energy of particles, and is usually implemented without much additional comment in flat spacetime [65]. Choosing a reasonable normalization thus might leave us with non-trivial $C_\alpha(\omega)$. We will not discuss the Hamiltonian or norms of modes, as this requires more detailed knowledge of the functions $F_{\omega,\alpha}$, and thus keep the $C_\alpha(\omega)$ explicit.

also its own *mode*. Which operator takes the role of creation operator depends on the sign of $C_\alpha(\omega)$: if $C_\alpha(\omega) > 0$, then $\hat{A}^\dagger_{\omega,\alpha}$ is the creation operator. Thus, $\hat{A}^\dagger_{\omega,u_>}$, $\hat{A}^\dagger_{\omega,v}$ and $\hat{A}_{\omega,u_<}$ are creation operators. See [65] [15].

Due to asymptotic flatness, a static observer at infinity, *i.e.* at rest with respect to r, θ and ϕ, must observe the same local dynamics of the field as if they were an inertial observer in flat spacetime; in particular, their notion of particle must be the same as in flat spacetime. We know from quantum field theory in flat spacetime [65] that particles observed by a given inertial observer correspond to modes which are *stationary* with respect to the observer's proper time τ (*i.e.* the time dependence is of the form $\mathbf{e}^{\mathrm{i}\omega\tau}$ for some constant ω). In our case of static observers at infinity, the proper time is t. But the $u_>$- and v-modes are stationary with respect to t and extend into the asymptotic region. Therefore, they are precisely the modes of particles observed by static observers at infinity:

Lemma 2.10: Particles at Infinity

The particles of frequency ω and with direction of motion $u_>$ or v (outgoing or ingoing) seen by static observers at infinity are annihilated by

$$\hat{a}_{\omega,u_>} := \hat{A}_{\omega,u_>}, \quad \hat{a}_{\omega,v} := \hat{A}_{\omega,v} \tag{2.30}$$

and created by $\hat{a}^\dagger_{\omega,u_>}$, and $\hat{a}^\dagger_{\omega,v}$. It holds that

$$[\hat{a}_{\omega,\alpha}, \hat{a}^\dagger_{\omega',\alpha'}] = C_\alpha(\omega) \cdot \delta(\omega - \omega') \cdot \delta_{\alpha,\alpha'} \cdot \hat{\mathrm{id}} \tag{2.31}$$

for $\alpha = u_>, v$, and $C_\alpha(\omega) > 0$.

The *vacuum* $|0_a\rangle$ is a normalized state such that

$$\hat{a}_{\omega,\alpha} |0_a\rangle = 0, \qquad \forall\omega, \quad \alpha = u_>, v, \tag{2.32}$$

indicating the absence of any particles. It is the outgoing particles ($u_>$) that we mean when speaking of Hawking radiation. We will thus ultimately have to consider the expectation values of the *number operators*

$$n_{\omega,u_>} = C_{u_>}(\omega)^{-1}\, \hat{a}^\dagger_{\omega,u_>} \hat{a}_{\omega,u_>}. \tag{2.33}$$

The normalization $C_{u_>}(\omega)^{-1}$ is chosen such that one-particle states $\hat{a}^\dagger_{\omega,\alpha} |0\rangle$, $\alpha = u_>, v$, count as exactly one particle.

Not knowing $C_{u_>}(\omega)$ explicitly will introduce an unknown *grey-body factor* in the final result; we will see this towards the end of this section. If needed, $C_{u_>}(\omega)$ can be computed from the full mode expansion, which requires knowledge of the exact fall-off $F_{\omega,u_>}(r)$ of solutions.

Particles Seen by Other Observers. The decomposition (2.28) is not the only decomposition of the field providing a notion of particle [15]: Consider any decomposition of the general solution,

$$\hat{\Phi}(x) = \sum\!\!\!\!\!\!\int_j \left(\hat{q}_j f_j(x) + \hat{q}^\dagger_j \bar{f}_j(x)\right), \qquad [\hat{q}_j, \hat{q}^\dagger_{j'}] = c_j \cdot \delta_{j,j'} \cdot \hat{\mathrm{id}}, \qquad c_j > 0, \tag{2.34}$$

where the modes are indexed by j and all other commutators vanish. The sum-integral sign indicates that this expansion may contain sums and integrals, and $\delta_{j,j'}$ is accordingly a product of Dirac and Kronecker delta terms. With this decomposition, $\hat{q}_j$ become the annihilation operators and $\hat{q}^\dagger_j$ the creation operators, providing a notion of particle as described above. The decomposition (2.25) was of this type, with $j = (l, m, \omega, \alpha)$ and $f_j \propto F^{l,m}_{\omega,\alpha}(r)\mathrm{e}^{-\mathrm{i}\omega\alpha}$.

As in the special case above, the *vacuum* $|0\rangle$ is defined to be a normalized state annihilated by all annihilation operators [15]: $\hat{q}_j|0\rangle = 0, \forall j$, and the *number operators* $\hat{n}_j := c^{-1}\hat{q}_j^\dagger\hat{q}_j$ measure the number of particles in the mode j.

As with the special case of particles at infinity above, the modes f_j leading to the most natural interpretation of particles for any given observer with proper time τ are *stationary* with *constant, positive frequency* $\omega_j > 0$ along the worldline of the observer, *i.e.* $f_j \sim \mathrm{e}^{-\mathrm{i}\omega_j\tau}$ on the worldline.[17] Positive frequency ensure that the *energy* of the mode is also positive.

Note that if two complete sets of modes $\{f_j\}_j$ and $\{g_j\}_j$ are such that f_j is a linear combination of the g_j's but not the $\bar{g}_j$'s, and thus similarly for g_j being expressible through the f_j's alone, *i.e.* if both sets agree on the notion of positive frequency, then the vacua defined by the respective sets of modes are equal. This follows by writing the linear combinations explicitly and from it deducing linear relations between the corresponding operators [15]. Two observers surveying the same region of spacetime agree on the direction of time flow, and thus on the notion of positive frequency; they must therefore observe the same vacuum. So if we are only interested in the vacuum of a given observer, we do not strictly have to find the stationary modes for that observer.

As a consequence, the vacua of two observers can only be different, *if the observers survey two different regions of spacetime.* Importantly, the regions surveyed by two observers *may overlap* (but not coincide), in which case there can still be disagreement about which modes have positive frequencies (observers can disagree on the local concept of frequency in regions where only one of the observer has access), and so the notions of vacuum can still be different for these observers.

This case is precisely what will happen below: we will see that while an observer outside and very far from the horizon might share a region of spacetime with an observer falling through the horizon (*i.e.* the two observers both oversee the outside of the horizon), the outside observer manages to measure particles in the field while the infalling observer sees a vacuum; this is possible because the outside observer can never hope to access the region below the horizon, which the infalling observer of course can.

Infalling Observers. Computing the expectation values of the number operators (2.33) requires knowing the state of the field. Our strategy will be to find a different family of observers for which the field is in the vacuum state. By transforming the annihilation operators between families we will then be able to compute expectation values of the number operators (2.33).

As explained above, we will consider observers infalling through the horizon as our second family. Such observers are in a sense also the most natural ones, since they are inertial and do not see any horizons. The absence of horizons in particular means that there is no other family of observers with a more complete view of spacetime. We will thus assume that the infalling observers see a vacuum.[18] Note that shortly before they crossed the horizon, they should also observe a vacuum: for any sudden change in observed particles while crossing the horizon is forbidden by the *principle of equivalence.*

Extension of u-Coordinate Through the Horizon. Before finding the particles seen by infalling observers, there is a technical problem we need to solve: the separation between sets of u-modes ($u_>$ and $u_<$) exist due to u as defined in (2.23) not being regular at the

[17] Further dependencies of the mode on τ besides $\mathrm{e}^{-\mathrm{i}\omega_j\tau}$ do not change the meaning of particle much if they are slower than the exponential dependence; this is for instance the case here, since our modes are radial and will thus fall off at some rate, even in flat spacetime. Furthermore, it is often only possible to find perfectly stationary modes in spacetimes with high symmetry.

[18] The observers need to be inertial, since accelerated observers see *Rindler horizons* [49], which lead to *Unruh radiation* [26] [73].

horizon; but from the point of view of the infalling observer, the horizon is not special at all, revealing that the separation of modes must be a purely mathematical artefact due to our choice of coordinates. Note further that only the outgoing coordinates are affected, since v is everywhere regular. Therefore, it should be possible to replace $u_<$ and $u_>$ by a new, *single* coordinate.

Specifically, we will attempt to regularize the integral in the first definition of (2.23) to define a new coordinate, which we will also call u. One way to do this is to first set the lower integral bound to $r_>$, but extend its definition to *all values of* r, with the intent of regularizing the divergence later:

$$u := t - \int_{r_>}^{r} \frac{\mathrm{d}r'}{c_s(r') + V(r')}, \qquad \forall r > 0. \tag{2.35}$$

The divergence for $r < r_H$ occurs because close to the horizon,

$$c_s(r') + V(r') = \kappa \cdot (r' - r_H) + \mathcal{O}((r' - r_H)^2), \tag{2.36}$$

where

$$\kappa := \frac{\mathrm{d}}{\mathrm{d}r}\left(c_s + V\right)\Big|_{r=r_H} = \frac{1}{2c_s}\frac{\mathrm{d}}{\mathrm{d}r}\left(c_s^2 - V^2\right)\Big|_{r=r_H}. \tag{2.37}$$

Note that $\kappa > 0$. In the Schwarzschild case with Gullstrand-Painlevé coordinates, this becomes

$$\kappa = \frac{\mathrm{d}}{\mathrm{d}r}\frac{r_s}{2r}\Big|_{r=r_s} = \frac{1}{2r_s}, \tag{2.38}$$

which, upon comparison with (2.14), is precisely the *surface gravity* of the black hole (hence the choice of symbol).

The integral can be regularized by introducing a small imaginary part in the denominator. Luckily, a natural choice for such a regularization comes from elsewhere in the theory: from the iε*-prescription* due to FEYNMAN (see *e.g.* [65] for a detailed treatment). We find (see Appendix B for details) that

Lemma 2.11: Continuation of u Through the Horizon

u can be extended through the horizon. Choosing $u = u_>$ for $r > r_H$, we get

$$u = u_< - \frac{\mathrm{i}\pi}{\kappa} + \mathcal{O}(\varepsilon) \tag{2.39}$$

for $r < r_H$. u is still undefined at $r = r_H$ due to the pole, but u is analytical for r in the upper half-plane around the pole. κ is the surface gravity.

This shows that $u_<$ and $u_>$ are not really independent. Furthermore, we can now find two new sets of independent modes,

$$A_{\omega,+} \cdot \frac{F_\omega(r)}{r\sqrt{\Theta(r)}} \cdot \left[\theta(r - r_H)\mathrm{e}^{-\mathrm{i}\omega u_>} + \theta(r_H - r)\mathrm{e}^{-\mathrm{i}\omega u_<}\mathrm{e}^{-\pi\omega/\kappa}\right], \tag{2.40}$$

$$A_{\omega,-} \cdot \frac{F_\omega(r)}{r\sqrt{\Theta(r)}} \cdot \left[\theta(r - r_H)\mathrm{e}^{\mathrm{i}\omega u_>} + \theta(r_H - r)\mathrm{e}^{\mathrm{i}\omega u_<}\mathrm{e}^{\pi\omega/\kappa}\right], \tag{2.41}$$

which are analytical in r through the horizon (in the upper half-plane), thanks to the extension of u just described. See Appendix B for details (in particular, we could combine $F_{\omega,u_>}(r)$ and $F_{\omega,u_<}(r)$ in a single function). See [76] for a derivation of these modes with eikonal techniques. The new modes are better suited to describe the field seen by the infalling observer, since they do not suffer from some unnatural separation at the horizon.

The part of the field which until now was decomposed into $u_>$- and $u_<$-modes can thus be rewritten as a linear combination of the new modes given by (2.40) and (2.41) and their complex conjugates. Once quantized, the operators $\hat{A}_{\omega,\pm}$ satisfy commutation relations required of annihilation operators.

Particles seen by Infalling Observers. Consider a radially, freely falling observer in the process of crossing the horizon. Due to (1.30) It always holds that $\mathrm{d}r/\mathrm{d}t < V(r) + c_s(r)$, and $\mathrm{d}r/\mathrm{d}t < 0$. While still outside the horizon ($r > r_H$), we have $V(r) + c_s(r) > 0$ and thus

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\mathrm{d}u_>}{\mathrm{d}t} = 1 - \frac{\mathrm{d}r}{\mathrm{d}t}\frac{1}{c_s(r) + V(r)} > 0. \tag{2.42}$$

After the observer has passed the horizon ($r < r_H$), we have $V(r) + c_s(r) < 0$ and hence

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\mathrm{d}u_<}{\mathrm{d}t} = 1 - \frac{\mathrm{d}r}{\mathrm{d}t}\frac{1}{c_s(r) + V(r)} < 0, \tag{2.43}$$

because $\mathrm{d}r/\mathrm{d}t \cdot [V(r) + c_s(r)]^{-1} > 1$. Recall from (1.29) that $\mathrm{d}t/\mathrm{d}\tau > 0$, so that

$$\frac{\mathrm{d}u}{\mathrm{d}\tau} > 0 \quad \text{for } r > r_H, \qquad \frac{\mathrm{d}u}{\mathrm{d}\tau} < 0 \quad \text{for } r < r_H. \tag{2.44}$$

Remarkably, (the real part of) u first increases for our observer, but then decreases again after they crossed the horizon. This makes the identification of positive-frequency modes in the region around the horizon rather interesting. This complication does not arise with v-modes, since v is regular at the horizon and $\mathrm{d}v/\mathrm{d}\tau > 0$ always.

Thus, the (+)-mode (2.40) has a positive frequency from the point of view of the observer for $r > r_H$, and a negative one for $r < r_H$; we also note that the negative frequency part is suppressed by $\mathrm{e}^{-\pi\omega/\kappa} < 1$. Similarly, the (−)-mode (2.41) has a positive frequency for $r < r_H$, and a negative one for $r > r_H$; this mode is also suppressed by $\mathrm{e}^{-\pi\omega/\kappa} < 1$ in the latter region. Intuitively, since in both modes the negative-frequency regions are suppressed, we can still see them as "mostly" positive frequency modes for our observer and associate them with annihilation operators; this is what we will do (see Appendix B for more comments on this assumption).

It will be useful to normalize the modes seen by the infalling observer so that the commutators become the same as for particles at infinity, equation (2.31). We thus get (see Appendix B for details):

Lemma 2.12: Particles seen when Falling Through the Horizon

The particles seen by observers freely falling radially though the horizon, are annihilated by

$$\hat{b}_{\omega,+} := \frac{1}{\sqrt{2\sinh(\pi\omega/\kappa)}}\left[\mathrm{e}^{\pi\omega/2\kappa}\hat{A}_{\omega,u_>} - \mathrm{e}^{-\pi\omega/2\kappa}\hat{A}_{\omega,u_<}\right], \tag{2.45}$$

$$\hat{b}_{\omega,-} := \frac{1}{\sqrt{2\sinh(\pi\omega/\kappa)}}\left[\mathrm{e}^{\pi\omega/2\kappa}\hat{A}^\dagger_{\omega,u_<} - \mathrm{e}^{-\pi\omega/2\kappa}\hat{A}^\dagger_{\omega,u_>}\right], \tag{2.46}$$

$$\hat{b}_{\omega,v} := \hat{A}_{\omega,v} \tag{2.47}$$

and created by the Hermitian transposes of these operators. It holds that

$$[\hat{b}_{\omega,\pm}, \hat{b}^\dagger_{\omega',\pm'}] = C_{u_>}(\omega)\cdot\delta(\omega-\omega')\cdot\delta_{\pm,\pm'}\cdot\hat{\mathrm{id}}. \tag{2.48}$$

Accordingly, we define the vacuum $|0_b\rangle$ of the infalling observers as a normalized state such that

$$\hat{b}_{\omega,\alpha} |0_b\rangle = 0 \qquad \forall\omega, \quad \alpha = +, -, v. \tag{2.49}$$

The assumption that observers radially and freely falling through the horizon see a vacuum means that the field is in the state $|0_b\rangle$. Note that $|0_b\rangle$ and $|0_a\rangle$ *are not necessarily the same states*; we will see shortly that they are really not.

The equations (2.45) and (2.46) form a so-called *Bogoliubov transformation.* To compute expectation values $\langle 0_b| \hat{n}_{\omega,u_>} |0_b\rangle$ we will need $\hat{a}_{\omega,u_>} = \hat{A}_{\omega,u_>}$. Solving the Bogoliubov transformation for it yields (see Appendix B)

$$\hat{A}_{\omega,u_>} = \frac{1}{\sqrt{2\sinh(\pi\omega/\kappa)}} \left[e^{\pi\omega/2\kappa} \hat{b}_{\omega,+} + e^{-\pi\omega/2\kappa} \hat{b}^{\dagger}_{\omega,-} \right]. \tag{2.50}$$

Before computing the expectation values, let us summarize our progress so far in Figure 4.

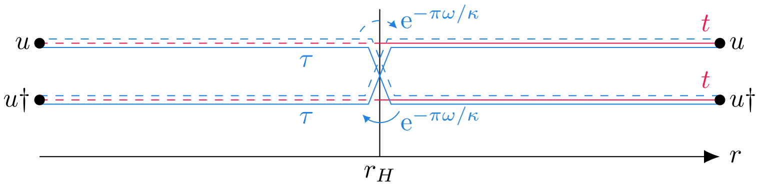


Figure 4: Outgoing modes as seen from infinity (solid red lines) use coordinate time t as time standard and *a priori* only exist outside the horizon. Inside the horizon, we find a separate set of outgoing modes (dashed red lines), implicitly also using t as time standard. The upper lines correspond to the annihilation parts of the modes, the lower lines are the creation parts of the modes.

The analytical extension of u across the horizon allows us to find two alternative outgoing modes, which use the proper time τ of an infalling observer as time standard. They can be thought of as analytical extensions of the original exterior mode to the interior (solid blue lines), and analytical extension of the original interior mode to the exterior (dashed blue lines) respectively. Again, the annihilation and creation parts are both shown, matching the parts of the original modes from which they were extended (*e.g.* for the solid blue lines, the labels on the right, the outside, determine annihilation and creation). The new modes are both linear combinations of the old exterior and interior modes, with an exponential relative scaling between terms in the combinations stemming from the analytical continuation (the arrows indicates which side receives the scaling). Importantly, the notion of "positive frequency" switches between u-modes for the infalling observer as they traverse the horizon, which means that the annihilation parts of the new modes are combinations of old annihilation *and* creation parts, and similarly for the creation parts (indicated by the crossing blue lines).

Hawking Radiation. Using the number operators at infinity (2.33), the Bogoliubov transformation (2.50), the commutation relations (2.48) and the definition of the vacuum (2.49), we can compute the expected number of outgoing particles observed by the static observers at infinity:

$$\begin{aligned} N_{\omega,u} &= \langle 0_b | \hat{n}_{\omega,u_>} |0_b\rangle = C_{u_>}(\omega)^{-1} \langle 0_b | \hat{a}^\dagger_{\omega,u_>} \hat{a}_{\omega,u_>} |0_b\rangle \\ &= \frac{C_{u_>}(\omega)^{-1}}{2\sinh(\pi\omega/\kappa)} \langle 0_b | \left(e^{\pi\omega/2\kappa} \hat{b}^\dagger_{\omega,+} + e^{-\pi\omega/2\kappa} \hat{b}_{\omega,-} \right) \left(e^{\pi\omega/2\kappa} \hat{b}_{\omega,+} + e^{-\pi\omega/2\kappa} \hat{b}^\dagger_{\omega,-} \right) |0_b\rangle \\ &= \frac{C_{u_>}(\omega)^{-1}\, e^{-\pi\omega/\kappa}}{e^{\pi\omega/\kappa} - e^{-\pi\omega/\kappa}} \langle 0_b | \hat{b}_{\omega,-} b^\dagger_{\omega,-} |0_b\rangle \propto \frac{C_{u_>}(\omega)^{-1}}{e^{2\pi\omega/\kappa} - 1}. \end{aligned} \tag{2.51}$$

We arrive at the remarkable result:

Theorem 2.13: (Hawking) Hawking Radiation

The outgoing, the spherically symmetric ($l = 0$) part of the radiation observed by static observers at infinity follows a *Planckian spectrum*, *i.e.* a *thermal spectrum*, at the so-called *Hawking temperature*

$$T_H = \frac{\hbar}{k_B} \frac{\kappa}{2\pi} \tag{2.52}$$

(with physical constants restored). This outgoing radiation is called *Hawking radiation*. In the case of a Schwarzschild black hole, κ is given by (2.14), and thus (in Planck units) $T_H = 1/8\pi M$.

The spectrum is Planckian up to grey-body factors ultimately coming from the exact fall-off of modes.

Note that no such radiation is observed in the v-modes, because the v-modes of the infalling observer agree with those of observers at infinity. Hawking radiation is however observed in the $u_<$-mode, and the same spectrum is observed; this can easily be seen by repeating the above argument for the $u_<$-modes instead of $u_>$ (but keeping in mind that $\hat{A}_{u_<}$ is a creation operator, not an annihilation operator).

Higher Modes ($l > 0$). For $l > 0$, we must work with the full mode expansion (2.25); in particular, the reality condition (B.2) means that amplitudes not just with $\pm\omega$, but with $\pm m$ are related. Apart from this, the above analysis can essentially be repeated for higher l. And except from different grey-body factors $C_{l,\alpha}(\omega)$, the result turns out the same.

See *e.g.* [76] for a simple argument why higher modes exhibit the same Planck spectrum, and even a mass term would not change it. See also the original paper [31] for a more in-depth argument.

Hawking Radiation is a Kinematic Effect. We would like to emphasize (see also the discussion in [76]) that the way we have derived Hawking radiation did not require any use of the Einstein field equations (0.1), but only the properties of the metric and some limited notions of observers and the equivalence principle. It is in this sense that Hawking radiation is merely a *kinematic effect*. In particular, it can occur in analogue models *that do not* exhibit the dynamics of the Einstein field equations.[19]

Back to Black Hole Thermodynamics. Recall from Section 2.3 that due to the laws of black hole dynamics, we can interpret black holes as thermodynamic objects. In particular,

[19] We will see in Section 4.1 that up to very rare, for us unimportant but still interesting exceptions, analogue models do not exhibit Einsteinian dynamics.

Kerr-Newman black holes are at equilibrium, with a temperature (2.17):

$$T = \frac{\kappa}{8\pi} \frac{1}{f'(A(H))}, \tag{2.53}$$

where $A(H)$ is the area of the horizon and f is a monotonically increasing, but yet unknown function. Furthermore, it was not clear how different black holes could be in thermal equilibrium with each other, or with traditional thermodynamic systems.

We now know that black holes emit thermal Hawking radiation at temperature $T_H = \kappa/2\pi$. Since radiation can travel from one black hole to another, it constitutes a way by which black holes can thermalize with each other or with other thermodynamic systems. This leads us to set $T = T_H$ [81, 14.4]. Therefore, we have (at least for Schwarzschild black holes[20]), that

$$f'(A(H)) = \frac{1}{4}, \tag{2.54}$$

and thus the entropy (2.15) of the black hole is

$$S = f(H(A)) = \frac{A(H)}{4}, \tag{2.55}$$

where one typically normalizes $S = 0$ in the absence of a horizon. In conventional units, we have

$$S = k_B \frac{A(H)}{4 l_P^2}, \tag{2.56}$$

with $l_P^2 := G\hbar/c^3$ the *Planck area*. This fixes the unknown prefactor present in the formula for Bekenstein entropy (2.21).

State of the Radiation. Theorem 2.13 characterizes the expectation values of number operators for outgoing particles observed at infinity. But often one is also interested in the actual quantum state of the radiation at infinity; this will in particular be important for the black hole information loss paradox.

For this we remark that because the horizon splits the u-modes into two parts, the Hilbert space $\mathcal{H}_u$ of outgoing (u-mode) radiation should factor into two parts:

$$\mathcal{H}_u = \mathcal{H}_{u_>} \otimes \mathcal{H}_{u_<}. \tag{2.57}$$

See for instance [80] or [81, Section 14.3].[21] We will be interested in the part of the state on $\mathcal{H}_{u_>}$: if $\hat{\sigma}_u : \mathcal{H}_u \to \mathcal{H}_u$ is the density operator of the total state, then the part of the state on $\mathcal{H}_{u_>}$ is obtained by tracing out the part on $\mathcal{H}_{u_<}$:

$$\hat{\sigma}_{u_>} := \mathrm{tr}_{u_<}(\hat{\sigma}_u). \tag{2.58}$$

To find the state of the radiation on $\mathcal{H}_u$ is not quite straightforward, and we give here only a very rudimentary and heuristic argument; see [80] for details.[22] The expectation values of particle numbers at infinity (2.51) together with the fact that the same expectation values hold for particle numbers in the $u_<$-modes (see note earlier), *suggests* that the state $|0_b\rangle$ of

[20] Only Schwarzschild, because we only checked the equivalence between surface gravity κ and the quantity κ occurring in the derivation of Hawking radiation for the Schwarzschild case. The same however holds for more general black holes [31].

[21] Note that these sources use $\mathscr{H}$ to signify a single-particle Hilbert space, and $\mathscr{F}$ for the Fock space constructed from $\mathscr{H}$. We use $\mathcal{H}$ for the Fock space.

[22] There, Schwarzschild spacetime is mainly used (with some comments on Kerr spacetime, *i.e.* Kerr-Newman with $Q = 0$). But the main ideas should transferrable to more general metrics.

the field is a superposition of particles *occurring in pairs*, with one inside and one outside the horizon. More precisely,

$$|0_b\rangle = \int \mathrm{d}\omega \sum_{n=0}^{\infty} \chi_\omega(n) \, |u_>, \omega, n\rangle \otimes |u_<, \omega, n\rangle \,, \tag{2.59}$$

with $|u_>, \omega, n\rangle \in \mathcal{H}_{u_>}$ the state of n particles with frequency ω in the $u_>$-mode, $|u_<, \omega, n\rangle \in \mathcal{H}_{u_<}$ is the state of n-particles of frequency ω in the $u_<$-mode, and $\chi_\omega(n) \in \mathbb{C}$ is an unknown function of n and ω containing the details of the state. We then have $\hat{\sigma}_u = |0_b\rangle \langle 0_b|$ and tracing out $u_<$ leaves us with

$$\hat{\sigma}_{u_>} = \int \mathrm{d}\omega \sum_{n=0}^{\infty} |\chi_\omega(n)|^2 \, |u_>, \omega, n\rangle \langle u_>, \omega, n| \,. \tag{2.60}$$

A similar result is obtained for modes with $l > 0$. The outgoing radiation is in a *mixed* state. Of course, $\chi_\omega(n)$ should be such as to reproduce the statistics of (2.51):

$$\sum_{n=0}^{\infty} n|\chi_\omega(n)|^2 \propto \frac{C_{u_>}(\omega)^{-1}}{\mathrm{e}^{2\pi\omega/\kappa} - 1}. \tag{2.61}$$

In conclusion (see the original paper [80] for an actual proof):

Theorem 2.14: State of Outgoing Hawking Radiation at Infinity (Wald)

The quantum state of outgoing Hawking radiation is a *mixed*, *thermal* state at the Hawking temperature T_H.

Furthermore, every escaping particle is created together with a particle captured by the black hole.

This implies that the state of Hawking radiation at infinity carries a non-zero von Neumann entropy; this will be a crucial ingredient for the black hole information loss paradox in Section 3.

Hawking Radiation in Analogue Models. We have shown that Hawking radiation occurs in the metrics we would expect to obtain from fluid-flow analogue models. But in doing so, we have used notions from general relativity: observers, inertial reference frames, and the principle of equivalence. These notions exist also mathematically in fluid-flow analogue models, but they might not have a direct physical meaning. We discuss them here and argue that the extent to which they occur in fluid-flow analogue models is enough to make the above argument for Hawking radiation also work in situations where the metric does not come from spacetime, but from some fluid-flow analogue model.

The metric provided by the analogue model carries a notion of timelike curves and even timelike geodesics. Therefore, the worldlines of observers, *in the sense of general relativity*, exists in analogue models; we call these *general-relativity-type observers*. They are purely mathematical ideas and do not necessarily correspond to actual observers present in the physical system of the analogue model; the latter we call *physical observers*.

Physical observers, no matter if they are moving or stationary in the fluid flow of the model, will always observe the time standard provided by the coordinate t, and therefore their four-velocity will not be on-shell necessarily.[23] But due to (1.29), $\mathrm{d}t/\mathrm{d}\tau$ never changes sign or becomes zero for any general-relativity-type observer with proper time τ. Furthermore,

[23] A four-velocity v^a is on-shell if $v_a v^a = -1$.

for $r \to \infty$, the observers even coincide since the fluid flow of the metric comes to a halt, showing that $\mathrm{d}t/\mathrm{d}\tau > 0$. Wherever we only need $\mathrm{d}t/\mathrm{d}\tau > 0$, we may substitute physical observers for general-relativity-type observers in the above argument.

Apart from $\mathrm{d}t/\mathrm{d}\tau > 0$ we have also used the principle of equivalence holding for inertial general-relativity-type observers: we argued that freely falling general-relativity-type observers who just crossed the horizon should observe a vacuum even shortly before infalling. This requires explanation when we try to use physical observers instead. A careful re-examination of the above steps shows that we can always use observers which are momentarily co-moving with the flow of the fluid-flow metric, since they can be momentarily inertial. For physical observers co-moving with the flow, a weaker version of the equivalence principle should still hold: from their perspective, the flow is at rest, but possibly departing from rest *to first order*; note that in general relativity, departure from local flatness happens *to second order* only. If the departure from rest is small, which can always be achieved for *e.g.* smooth flows, then the dynamics of the fluid, and thus the resulting Klein-Gordon equation for sound waves (see Section 4 below for details) become the equation in a fluid at rest. There should thus be no creation of sound waves in the vicinity of co-moving observers and the overall argument still holds.

In conclusion, we expect Hawking radiation to also occur in quantum fluid-flow analogue gravity models. In fact, Hawking radiation in analogue models has recently been experimentally verified; see *e.g.* [37] and the sources therein for an overview.

2.6 Back-Reaction

We have seen in Section 2.5 that the presence of an apparent horizon leads to a Planckian spectrum of outgoing particles observed at infinity. With this radiation, energy is transported away from the black hole towards infinity. We thus expect the black hole to lose some of its mass to Hawking radiation; we say that it is *evaporating*. For this to happen, there must be a so-called *back-reaction* effect of the Hawking radiation onto spacetime. We give here a short overview of back-reaction. If not stated otherwise, the contents of this section are a based on [81, Section 14.3]; see there and sources therein for details.

Semi-Classical Einstein Field Equations. To truly understand back-reaction would require a quantum version of the Einstein field equations, in turn requiring some notion of quantum gravity. Since such a theory is not available, one often considers the *semi-classical Einstein field equations*

$$R_{ab} - \frac{1}{2}Rg_{ab} = 8\pi \operatorname{tr}\big(\hat{T}_{ab}\,\hat{\sigma}\big), \tag{2.62}$$

where $\hat{\sigma}$ is the quantum state of matter (generally a density operator), and $\hat{T}_{ab}$ is the *operator-valued energy-momentum tensor* obtained from substituting classical observables in the energy-momentum tensor T_{ab} for their quantum observable counterparts.[24]

In the case of a massless Klein-Gordon field one finds

$$\hat{T}_{ab} = \nabla_a\hat{\Phi}\nabla_b\hat{\Phi} - \frac{1}{2}g_{ab}\nabla_c\hat{\Phi}\nabla^c\hat{\Phi}. \tag{2.63}$$

This expression contains infinities similar to the zero-point energies routinely encountered in quantum field theory even in flat spacetime. However, they cannot easily be removed in curved spacetime because there is no generally agreed-upon vacuum state (in flat spacetime, one could remove those with *normal ordering*, such that $\langle 0|\,\hat{T}_{ab}\,|0\rangle = 0$ for the vacuum $|0\rangle$).

[24] Notably this does not take into account the back-reaction of quantized gravitational waves (*i.e. gravitons*); but we will not need those here.

Assuming these issues have been removed, the computations involving the stress tensor are also more involved than in classical general relativity, even if only working with the expectation values $\mathrm{tr}(\hat{T}_{ab}\,\hat{\sigma})$, because these must not necessarily follow the typical energy conditions encountered in classical general relativity (recall Sections 2.2 and 2.3). In fact, negative expectation values for energy densities are possible with very reasonable states of the field.

This for instance happens with Hawking radiation, allowing the violation of the weak energy condition (2.11) and ultimately of the curvature condition (2.9), which implies that the area Theorem 2.6 may no longer hold. Indeed, as we will see, Hawking radiation should lead to a *decrease* in horizon area over time due to back-reaction. See [81, Section 14.4].

Qualitative Back-Reaction. Instead of using the semi-classical Einstein field equations (2.62), we can already learn much about back-reaction from the simple assumption that *the black hole looses mass at the same rate as energy is carried away by Hawking radiation.* An estimate for this rate of energy loss is given by the *Stefan-Boltzmann law*

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\sigma_{\mathrm{SB}} T^4 A, \tag{2.64}$$

where σ_{SB} is the *Stefan-Boltzmann constant* and A is the area of the emitting body. With this we obtain:

Proposition 2.15: Mass Loss due to Hawking Radiation

In the Schwarzschild case the mass loss due to Hawking radiation can be estimated as

$$\frac{\mathrm{d}M}{\mathrm{d}t} \sim -AT^4 \sim -\frac{1}{M^2}. \tag{2.65}$$

Here we have used that for Schwarzschild black holes, the horizon area is proportional to M^2, see (2.7), and also that the Hawking temperature T_H is proportional to $1/M$, see (2.52).

Note that even with this simplified assumption, we require some basic notion of dynamic coupling between spacetime and matter resembling the Einstein field equations: for in situations without black holes and with classical instead of quantum radiation, the Einstein field equations predict that the energy inside a region of space decreases as radiation escapes. See Appendix A for the definition of energy in curved spacetime.

Black Hole Evaporation. From (2.65) we learn that while the rate of mass loss is initially very small for massive black holes, it then increases without bound as the mass decreases. If the formula is to be believed, then black hole evaporation ends in a kind of explosion. Without a theory of quantum gravity, we however cannot be certain about the final stages of evaporation, when the black hole's mass reaches the Planck mass and its radius becomes comparable to the Planck length.

Figure 5 shows the formation and subsequent evaporation of a black hole in a Penrose diagram, assuming that the black hole evaporates completely. We will need this diagram again in the next section when discussing the black hole information loss paradox.

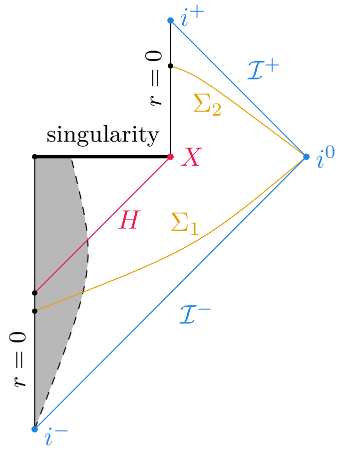


Figure 5: Penrose diagram of black hole formation due to a collapsing mass (grey) and subsequent evaporation. The point X at the top right end of the horizon H marks the point around which quantum gravity is expected to become important. It is easily seen to be a naked singularity (and thus Definition 2.2 has to be slightly adjusted, so we can still call this spacetime a black hole spacetime). Hawking radiation escapes towards $\mathcal{I}^+$. After the black hole has evaporated, the spacetime has again the same conformal structure as flat spacetime.

For later we have also included two spacelike hypersurfaces Σ_1 and Σ_2, one before the formation of the black hole, and one after it has evaporated. Note that neither of these are Cauchy surfaces for the entire exterior due to the presence of X: physics on Σ_2 is not completely determined by Σ_1.

3 Black Hole Information Loss Paradox

With black holes discussed in Section 2, we can now finally turn to one of the two main topics of this work: the *black hole information loss paradox*. One roughly distinguishes two versions of the paradox [84]: the original version due to HAWKING in 1976 [32], and a later, more striking and robust version due to PAGE in 1993 [55]. We will mostly work with the second version of the paradox.

We nevertheless begin with a short account of HAWKING's earlier version in Section 3.1. In Section 3.2 we then turn to the more modern version of the paradox. Section 3.3 addresses a recent approach toward a solution, involving so-called *replica wormholes*. Finally, we summarize the crucial ingredients of the paradox in Section 3.4, to facilitate the later discussion of the paradox in the context of analogue gravity models in Section 5.

3.1 Hawking's Version of the Paradox

We consider a black hole, evaporating due to Hawking radiation exhibited by some quantum field, as represented in Figure 5. The main feature of the paradox can already be seen by noting that the spacelike hypersurface Σ_2 is not completely determined by the spacelike hypersurface Σ_1: there are causal curves originating at the naked singularity X and reaching Σ_2. Thus, between before the formation of the black hole (at the time of Σ_1) and after its complete evaporation (at the time of Σ_2), *predictability of physical processes is not guaranteed*. HAWKING showed that under reasonable assumptions there is indeed a breakdown of predictability [32].

We will consider black holes with zero angular momentum ($J = 0$) and no electric charge ($Q = 0$). Our summary here is largely based on the account of the paradox given in [84].

Evolution of Quantum Fields. Assume that prior to the formation of the black hole (*i.e.* on Σ_1) the quantum field is in a given *pure* state.

According to the laws of quantum physics, we would expect for this state to evolve *unitarily*. As soon as the horizon forms, one may split the state into two subsystems, split by the horizon: the *exterior*, describing the field outside the horizon, and the *interior* of the black hole, describing the field inside the black hole.

Now the information about the interior state can never leave the black hole due to the horizon. This information is lost from the outside (at the latest once the black hole is evaporated, but really it is never accessible to the outside once the black hole has formed). To obtain the final state of the field outside we need to trace out the state which described the interior after the evaporation is complete (*i.e.* on Σ_2).

Entanglement and Non-Unitary Evolution. Unitary evolution allows for entanglement between the interior and exterior. And if entanglement has built up during the unitary evolution, then taking the trace implies that the initially pure state has evolved into a *mixed* state of the exterior, which after the black hole has disappeared describes the field on all of spacetime.

We know from Section 2.5 that Hawking radiation on the exterior is indeed entangled with the interior and thus itself in a mixed state. Thus, according to Proposition 0.13, the exterior state has a non-zero entropy once evaporation completes. Since the state of the field was originally pure, the entropy of the exterior state must have *increased* during black hole formation and subsequent evaporation; and because one can understand entropy as *missing information* [36], *information has been lost in the process.* Proposition 0.14 states that the entropy of a state is conserved if the evolution is *unitary*. Thus, we can conclude that black hole formation and evaporation would be a *non-unitary* process, contradicting the unitarity of quantum mechanics. We stress that the evolution of the entire state, *i.e.* interior plus exterior, is still unitary; at least until the singularity, when the interior state looses its meaning. But as seen from the outside, the process appears non-unitary.

Paradox. The paradox occurs as a conflict between the unitarity of quantum physics, generally believed to be a hallmark of any quantum theory [43], on the one hand, and the predictions of quantum field theory on a curved spacetime background in the form of Hawking radiation. The paradox inherits its name from the increase in total entropy and thus the loss of information in the process of black hole formation and subsequent evaporation. But the consequent non-unitarity of this process is what really leads to the paradox.

Note that the argument can be extended to also include the black hole in the quantum state, thus making the state describe the whole evaporation process. The outside-part of the state then also includes the properties of the black hole seen from outside the horizon (mass, angular momentum and charge, see Section 2.3). But these three classical pieces of information are in no way enough to store the information of the quantum field within the horizon, not to speak of whatever other information is needed to describe the inside of the black hole. The evaporation is still non-unitary and information loss occurs.[25]

There has been considerable criticism of Hawking's paradox, see *e.g.* [56] for a review of the discussion surrounding the paradox. Perhaps most strikingly [84], we do not know the physics around the naked singularity X, where quantum gravitational effects are expected to dominate. This is one reason why we will not continue investigating Hawking's version of the paradox and instead turn to the more modern version due to Page.

3.2 Page Curve Version of the Paradox

It was later found, following a paper by Page [55], that there is an even stronger version of the paradox, occurring already when the black hole is still macroscopic, far away from X and thus not needing any discussion of the potential naked singularity. We will cover this version of the paradox here, and it will be the version discussed in Section 5 in light of analogue gravity models.

Hilbert Space of an Evaporating Black Hole. Following [55], we assume that our system is at all times described by a bipartite Hilbert space

$$\mathcal{H} = \mathcal{H}_\mathrm{b} \otimes \mathcal{H}_\mathrm{r}, \tag{3.1}$$

where $\mathcal{H}_\mathrm{b}$ is the Hilbert space describing the black hole and its interior, and $\mathcal{H}_\mathrm{r}$ describes the radiation outside the black hole. Furthermore, we assume that if $\mathcal{H}_{\sim M} \subset \mathcal{H}_\mathrm{b}$ denotes the subspace of black holes with mass near M then $\dim \mathcal{H}_{\sim M}$ decreases as M decreases. Finally, assume that at any given time, the state $\hat{\sigma}_\mathrm{b}$ of the black hole will be confined to $\mathcal{H}_{\sim M}$ for some value of M, *i.e.* that it will be a density operator acting on $\mathcal{H}_\mathrm{b}$ with support contained in $\mathcal{H}_{\sim M}$.

The last assumption simply reflects the fact that we are considering almost entirely classical black holes which have very sharply but not perfectly defined mass M (hence the notation $\sim M$). And since mass loss due to Hawking radiation is very slow for most parts of the evaporation (see (2.65)), we can take M to be changing only very slowly, allowing the state to adapt adiabatically to the new mass. In fact, we can assume that our black hole is very well described by a Schwarzschild black hole of mass M.[26]

[25] This is how the paradox is for instance taken up in [55].

[26] Admittedly, the definition of $\mathcal{H}_{\sim M}$ is quite fuzzy: one could argue that a space much larger or smaller

Thermodynamic, Coarse-Grained Black Hole Entropy and the Central Dogma. Let us comment in detail why the second assumption above makes sense. Recall that

$$H_{\mathrm{TD},\sim M} := \log_2(\dim \mathcal{H}_{\sim M}) \tag{3.2}$$

(the meaning of "TD" will be clear soon) is the maximal entropy a state on $\mathcal{H}_{\sim M}$ can have, and that it is reached for the completely mixed state

$$\hat{\pi}_{\sim M} = \frac{\hat{\mathrm{id}}_{\mathcal{H}_{\sim M}}}{\dim \mathcal{H}_{\sim M}}. \tag{3.3}$$

$\hat{\pi}_{\sim M}$ describes a situation where every state of mass $\sim M$ is equally likely.

Turn to black hole thermodynamics for a moment. The black holes considered there are completely classical (or at least in a regime such that their quantum nature is not apparent). If the black hole and its interior really are described by a quantum state acting on a Hilbert space $\mathcal{H}_\mathrm{b}$, then the laws of black hole thermodynamics must be a *thermodynamic limit* of the microscopic, quantum laws describing the precise physics on $\mathcal{H}_\mathrm{b}$. More precisely, since the mass of classical (static) black holes is fixed, we expect the black hole be in a roughly *microcanonical* regime [40]: the actual state of the black hole is unknown, but all states with energy $\sim M$ are *equally likely*, and states of other energies are very unlikely. Thus, we expect a classical black hole, in the regime where black hole thermodynamics holds, to be roughly in the state $\hat{\pi}_{\sim M}$.

Consequently, the entropy $H_{TD,\sim M}$ should reflect (up to a constant) the thermodynamic entropy of a black hole (hence the choice of symbol). Now we expect thermodynamic entropies to be *extensive* quantities [40], increasing monotonously with energy; and since the black hole disappears at $M = 0$, we expect any thermodynamic entropy of the black hole to also vanish at $M = 0$. Thus, we see that $\dim \mathcal{H}_{\sim M}$ should decrease as M decreases, and $\mathcal{H}_{\sim M}$ should become nearly trivial at $M = 0$ (the exact behaviour there is not clear, due to the uncertainty in mass ΔM in $\mathcal{H}_{\sim M}$, which becomes comparable to M as the black hole disappears).

This argument is further strengthened by the fact that with the black hole entropy (2.56) discussed in the previous section we really have a thermodynamic entropy for our black hole; and indeed it has the property just described.

Let us make two remarks. Firstly, $H_{TD,\sim M}$ is a so-called *coarse-grained von Neumann entropy*: it is the von Neumann entropy, maximized over all quantum states which agree (perhaps up to some degree of accuracy) on a small set of observables (so-called *macroscopic observables*) [2]. Concretely, the states are black hole states $\hat{\sigma}_\mathrm{b} : \mathcal{H}_\mathrm{b} \to \mathcal{H}_\mathrm{b}$ and the only macroscopic observable is the mass of the black hole. The maximization naturally leads us to consider states on $\mathcal{H}_{\sim M}$, of which $\hat{\pi}_{\sim M}$ has the largest entropy. Ordinary von Neumann entropy is then called *fine-grained entropy*, to distinguish it from coarse-grained entropy.

Secondly, the presently discussed assumption (the second assumption above), refined by the black hole entropy expression (2.56), is very closely related to the so-called *central dogma of black hole thermodynamics*: *a black hole of mass M as seen from the outside is an ordinary quantum system with Hilbert space $\mathcal{H}_{\sim M}$ of dimension $\sim 2^S = 2^{A(H)/4}$, which evolves unitarily.* Besides black hole thermodynamics, the central dogma is also supported by results in certain approaches to quantum gravity (such as *string theory* with *AdS/CFT correspondence*). See [2] and sources therein for details.

Page Curve. Page [55] now considers the case where the entire evolution is unitary. The crucial difference to the situation in the previous section is that the dimension of the

than $\mathcal{H}_{\sim M}$ adequately describes a black hole of mass M. But as it turns out, this will not change the end result, since we only really require that $\dim \mathcal{H}_{\sim M}$ decreases as M decreases.

relevant Hilbert space describing the black hole, *i.e.* $\dim \mathcal{H}_{\sim M}$, becomes very small as the evaporation nears its end. Thus, there is simply not much possibility for entanglement between the black hole and the exterior radiation towards the end of evaporation. So even when we eventually trace out $\mathcal{H}_\mathrm{b}$, we lose virtually nothing. In this setup, *there is ideally no information loss.*[27] [28]

At any time, the state of the radiation $\hat{\sigma}_\mathrm{r} : \mathcal{H}_\mathrm{r} \to \mathcal{H}_\mathrm{r}$ can be obtained from the total state $\hat{\sigma} : \mathcal{H} \to \mathcal{H}$ by tracing out the black hole part:

$$\hat{\sigma}_\mathrm{r} = \mathrm{tr}_\mathrm{b}(\hat{\sigma}). \tag{3.4}$$

We assume that prior to the formation of the black hole the total state is pure, hence it will remain pure during evaporation.

We can now ask how the entropy $H(\hat{\sigma}_\mathrm{r})$ evolves. For this we note that

$$H(\hat{\sigma}_\mathrm{r}) = H(\hat{\sigma}_\mathrm{b}) \leq \log_2(\dim \operatorname{supp} \hat{\sigma}_\mathrm{b}) = \log_2(\dim \mathcal{H}_{\sim M}). \tag{3.5}$$

Here we have used Proposition 0.15 for the equality and then Proposition 0.13 for the inequality. The last equality follows from our assumptions about black hole evaporation above. Thus, $H(\hat{\sigma}_\mathrm{r})$ will eventually have to decrease, in line with entanglement decreasing towards the end of evaporation.

With some further assumptions on the state $\hat{\sigma}_\mathrm{r}$, one can find a plausible form for the evolution of $H(\hat{\sigma}_\mathrm{r})$. Namely, Page assumes that the total system is in a random pure state at any time; the result will thus be a *typicality result*; it is qualitatively drawn in Figure 6.[29] This so-called *Page curve* initially rises with an almost constant rate, before suddenly stalling at the *Page time*, and then falling as quickly as it rose. The initial rate is in fact maximal: at early times, the emitted radiation is close to a maximally mixed state, and consequently no information can escape the black hole.[30] Only after the turnaround will information come out, and very quickly at that [34]. In what follows, we will only need the qualitative shape of the curve; see [55] for details. Realistic page curves are computed in [57].

Hawking's Prediction. We saw in Section 2.5 that independently of the black hole's previous history the escaping Hawking radiation should be completely entangled with the inside of the black hole. Thus, at any time during evaporation, $H(\hat{\sigma}_\mathrm{r})$ should increase. Hawking's prediction is qualitatively drawn as an additional line in Figure 6.

Paradox. At early times, the page curve reflects what we expect from Hawking's result: the emitted radiation quanta are completely entangled with the interior and the entropy $H(\hat{\sigma}_\mathrm{r})$ rises monotonously with a maximal rate. But come the page time, the page curve diverges from this rise: at this time, the emitted radiation is no longer totally entangled with the interior any more, and entanglement starts to decrease, until the end of the evaporation.

On the one hand, we have the qualitative shape of the page curve, in particular its eventual decline forced upon us by the decreasing thermodynamic entropy $H_{TD,\sim M}$ of the black

[27] Ideally, because the argument of shrinking $\dim \mathcal{H}_{\sim M}$ as M decreases reaches its limits when $M \sim \Delta M$, where ΔM is the uncertainty in mass. But even now, only a very small amount of information is potentially lost; this is not so bad, since after all we do not have a theory of quantum gravity available to describe this regime.

[28] The goal of Page's paper [55] was not to uncover a new version of the paradox (although this ended up happening), but to provide an argument for how information could be conserved by not coming out of the black hole until much time has passed, in a way that previous perturbative methods could not rule out.

[29] Note that another typicality argument, Hayden's and Preskill's *random unitaries* [34] gives a comparable result.

[30] This shows that indeed information conservation is possible, even if initially no information escapes, which was the goal of Page's paper [55].

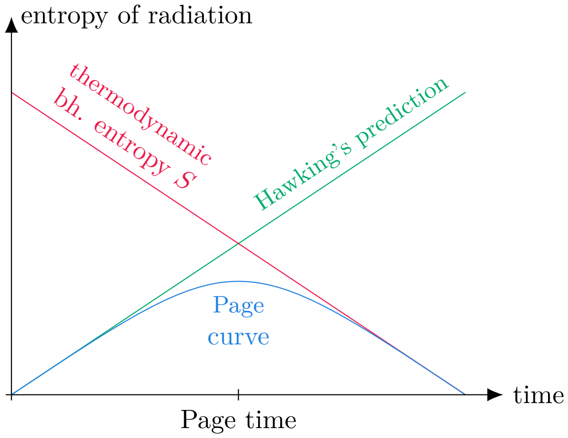


Figure 6: The Page curve version of the black hole information loss paradox.

Green curve: qualitative evolution of the von Neumann entropy $H(\hat{\sigma}_{\mathrm{r}})$ of escaping Hawking radiation during evaporation, as predicted by HAWKING. Red curve: qualitative evolution of the thermodynamic entropy S of the black hole. Blue curve (page curve): qualitative evolution of the von Neumann entropy $H(\hat{\sigma}_{\mathrm{r}})$ assuming that at early times HAWKING's prediction holds true, but information is eventually conserved. PAGE's argument reproduces the Page curve (possibly with a different initial slope, since the argument does not take into account the precise mixed state of Hawking radiation).

The page curve version of the black hole information loss paradox occurs as a conflict between the green and red curves: $H(\hat{\sigma}_{\mathrm{r}})$ should be bounded by S while simultaneously following Hawking's prediction.

The rough shape of the Page curve is inspired by the exact plot in [55]; there, the time axis is the *thermodynamic entropy of the radiation*, defined as $\log_2(\dim \operatorname{supp} \hat{\sigma}_{\mathrm{r}})$.

hole (for concreteness, take the entropy (2.56)) and the requirement of unitarity throughout the evaporation process. On the other hand, we have HAWKING's prediction of Hawking radiation, based on quantum field theory in curved spacetime, which implies that $H(\hat{\sigma}_{\mathrm{r}})$ should increase monotonically until the black hole has evaporated completely. This is the *page curve version of the black hole information loss paradox*. It is a contradiction between the predictions of quantum field theory in curved spacetime and the predictions of black hole thermodynamics in form of thermodynamic black hole entropy [84].[31] Importantly, the paradox already occurs roughly halfway through evaporation, around the Page time, where the black hole is still macroscopic such that quantum field theory on curved spacetime can be trusted; we can discuss the paradox without ever having to invoke the presumed naked singularity X in Figure 5. In this sense, the Page curve version of the paradox is much more poignant than HAWKING's version of the paradox.

Possible Solutions. Since we will be preoccupied with discussing whether one can state the paradox or even find a context for it in analogue models, we will not need to go into much detail about arguments for and against the two conflicting views involved in the paradox. Instead, we only give a brief overview of possible solutions, with one potential solution, *replica wormholes*, considered in some more depth.

[31] Note that one can even go further and argue for *black hole statistical mechanics* instead of just black hole thermodynamics [83] [84].

Two main types of solutions to the paradox come to mind [84]:

1. The emitted Hawking radiation is not completely entangled with the interior at late times; this would mean that Hawking radiation must be dependent on the history of the black hole. The methods of quantum field theory in curved spacetime used in Section 2.5 would thus have to be amended. In short, "the green curve in the figure is wrong".

2. The maximum entropy of the black hole system does not decrease in the way predicted by thermodynamics; in particular, black hole entropy (2.56) is not an upper bound for von Neumann entropy of the black hole. This would mean revising our idea of black hole thermodynamics.[32] One way in which this could happen, is if black hole thermodynamics as seen from outside the horizon does not reflect the interior, microscopic dynamics of the black hole; but then black hole thermodynamics could hardly be called "thermodynamics" (this could for instance happen, if large amounts of information can be stored inside the black hole, in things like *baby universes* or so-called *bags of gold* [84]). In short, "the red curve in the Figure is wrong".

Both of the above solutions are ultimately based on the split (3.1) of the total Hilbert $\mathcal{H}$ space into black hole interior $\mathcal{H}_\text{b}$ and escaping radiation $\mathcal{H}_\text{r}$, or more precisely the split of the currently relevant subspace of $\mathcal{H}$ into escaping radiation and black hole of mass M $\mathcal{H}_{\sim M}$. Other solutions are potentially thinkable if this split is critically reevaluated. We will mention a recent such approach in the next section.

3.3 A Recent Approach to a Solution: Replica Wormholes

Let us briefly mention a recent (2019 – 2021) and remarkable effort due to Pennington, Shenker, Stanford, Yang, Almheiri, Hartman, Maldacena, Shaghoulian, Tajdini and others [59] [1] [2], which suggests a reinterpretation of the split (3.1); it however assumes that such a split is possible. This section will only be a very brief account without derivations and detailed explanations, mostly based on [2].

Gravitational Path Integrals and Replica Wormholes. One way to compute the entropy $H(\hat{\rho})$ of a quantum state $\hat{\rho}$ (we use $\hat{\rho}$ since $\hat{\sigma}$ will be reserved for the state of black hole plus escaping radiation, as in the previous section) is the so-called *replica trick* [2]:

$$H(\hat{\rho}) = (1 - n\partial_n) \log_2 \operatorname{tr}(\hat{\rho}^n)|_{n=1}. \tag{3.6}$$

Here it is understood that $\log_2 \operatorname{tr}(\hat{\rho}^n)$ is analytically continued to any real $n \geq 1$. That this trick works can be readily checked by writing $\operatorname{tr}(\hat{\rho}^n)$ as the sum over the n-th powers of eigenvalues of $\hat{\rho}$, and then using $\operatorname{tr}(\hat{\rho}) = 1$. This trick is called the *replica trick*, since $\hat{\rho}^n$ can be seen as n "replicas" of the state $\hat{\rho}$ multiplied together.

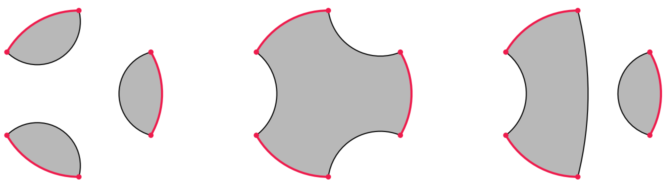
It is now possible to write $\operatorname{tr}(\hat{\rho}^n) = \operatorname{tr}(\hat{\rho}^{\otimes n}\hat{\tau}_n)$ as an expectation value of a specific observable τ_n in the n-fold tensor product state $\hat{\rho}^{\otimes n}$. We can attempt to compute this expectation value using *path integrals* with certain *boundary conditions*. We will not review path integrals here, see *e.g.* [65] for a general introduction and [68] for an introduction to path integrals in the context of black holes. Importantly, the boundary conditions of the path integral are enough for the computation, and *one does not need the actual state* $\hat{\rho}$; the missing information is essentially "filled in" appropriately by the path integral. Roughly speaking because $\hat{\rho}^{\otimes n}$ is a product of identical factors, the boundary conditions is a set of n identical copies of boundary conditions, essentially one for each factor in $\hat{\rho}^{\otimes n}$; for rigorous explanations, see [59].

[32] And revising black hole statistical mechanics [83] [84].

In our case $\hat{\rho}$ is the state of the radiation $\hat{\sigma}_{\mathrm{r}}$, and the boundary conditions are n copies of the field configuration together with the spacetime geometry at infinity.

In full quantum gravity, the path integral is expected to run over all spacetime geometries compatible with the boundary conditions (including potentially different topologies). But we do not have a full theory of quantum gravity, so we take a semi-classical approach, where instead of integrating over spacetime configurations, we only sum over those configurations whose action is stationary; these are essentially the classically allowed spacetime configurations given the boundary conditions. This makes sense, since it is known [65] that in the classical limit the dominant contributions to a path integral come from these classical contributions. But in order to treat the field quantum-mechanically, we keep the integration over all configurations of the field, despite only summing over classically allowed spacetime configurations.

Crucially, there could be more than one classically allowed spacetime configuration, as illustrated in Figure 7. Besides the most obvious geometry, where each boundary condition receives its own black hole, there is also a geometry where all boundary conditions receive a black hole, *whose interiors are connected*. This connection is called a *replica wormhole*. There are other classical geometries where only some black holes are interconnected. All contributions contribute with different weights to the path integral. See [59] [1].



(a) (b) (c)

Figure 7: Qualitative representations of some classically allowed spacetime geometries involved in the computation of the path integral for $\mathrm{tr}(\hat{\sigma}_{\mathrm{r}}^3)$. The three copies of the boundary conditions are schematically represented in red, and the spacetime geometry satisfying these conditions is indicated in grey. (a) Hawking saddle contribution: each boundary condition receives its own spacetime, with one black hole. The event horizons are not indicated, but can be thought of as lines parallel to the boundary conditions, slightly inset. (b) Replica wormhole contribution with all boundaries sharing a single spacetime, connected through wormholes in the black hole interiors. (c) Another replica wormhole contribution, where only two of the black holes are connected.

These diagrams are inspired by the ones in [59]; there, the different elements carry additional technical meaning.

The replica wormhole geometries are responsible for generating the correct Page curve (qualitatively like the blue curve in Figure 6); if we only take the simplest geometry without wormholes, the so-called *Hawking saddle*[33], we obtain Hawking's prediction (the green curve in Figure 6).

Interestingly, the path integral does not actually quite compute $\mathrm{tr}(\hat{\sigma}_{\mathrm{r}}^{\otimes n}\hat{\tau}_n)$ in the case $n > 1$, since the n boundary conditions in the presence of replica wormholes naturally give rise to states which are not necessarily n-fold tensor products. The difference between $\mathrm{tr}(\hat{\sigma}_{\mathrm{r}}^{\otimes n}\hat{\tau}_n)$ and the quantity actually computed with path integrals has recently been interpreted quantum information-theoretically as leaving out or including *reference information* [68]. The exact physical nature of this reference information is not completely understood, although

[33] "Saddle" refers to classical geometries being *stationary points* of the gravitational action.

it has been connected to the typicality arguments of [55] and [34], see [68].[34]

Island Formula for von Neumann Entropy. The result of the replica wormhole calculations is the so-called *island formula* (see [59] [1] [2], and sources therein). It computes the entropy $H(\hat{\sigma}_\mathrm{r})$:

$$H(\hat{\sigma}_\mathrm{r}) = \min_Q \left[\operatorname*{ext}_Q \left(\frac{A(Q)}{4} + H_{\text{s.-cl.}}(\Sigma_Q) \right) \right]. \tag{3.7}$$

Let us explain the different objects contained in the equation.

Q is a two-dimensional hypersurface[35], which can be completely inside the event horizon; $A(Q)$ is its area (defined in the same way as event horizon area). Σ_Q is a three-dimensional[36] hypersurface bounded by Q and by the intersection with some *cut-off surface* outside the event horizon. The "time" at which Σ_Q intersects the cut-off surface tells us when $H(\hat{\sigma}_\mathrm{r})$ will be evaluated Importantly, Σ_Q may be *disconnected*, with one part $\Sigma_Q^{\text{rad.}}$ stretching from the cut-off surface to infinity, and with one part $\Sigma_Q^{\text{i.}}$ only bounded by Q; this second part is called the *island*. These regions are illustrated in Figure 8.

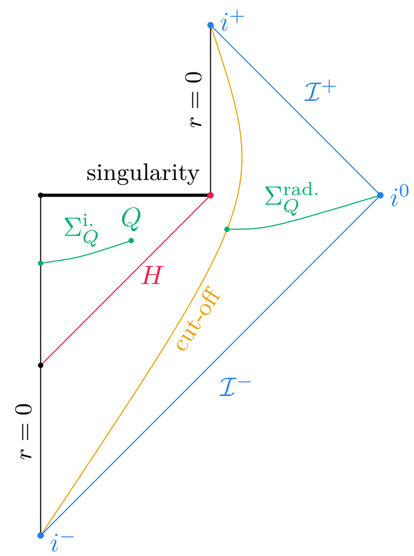


Figure 8: Typical configuration of cut-off surface, the quantum extremal surface Q, and Σ_Q (split into two connected components $\Sigma_Q^{\text{i.}}$ for the island, and $\Sigma_Q^{\text{rad.}}$ for the external radiation). We can think of Q as a topological sphere at the instant in "time" defined by Σ_Q. The island $\Sigma_Q^{\text{i.}}$ is then the interior space of this sphere. Similarly, the intersection of Σ_Q and the cut-off surface is such a sphere, and the connected component $\Sigma_Q^{\text{rad.}}$ disconnected from the island would then be the exterior space of this sphere, stretching to infinity.

$H_{\text{s.-cl.}}(\Sigma_Q)$ is the so-called *semi-classical entropy on* Σ_Q: it is the entropy of the quantum state of the matter fields and potentially of gravitational waves in the semi-classical regime (*i.e.* fixed spacetime with small amplitude, quantized gravitational waves), in the region Σ_Q.

[34] At the time of planning for this semester project, it was not very clear or at least not explicitly discussed how much analogue gravity could contribute to the issue of the black hole information loss paradox, and by extension possibly and eventually also to the issue of reference information. Thus, the idea of this work was born: to find out or report on whether black hole information loss can be discussed in the context of analogue models.

[35] More generally, its *codimension* is two.

[36] More generally, codimension one.

The island formula instructs us to find Q and Σ_Q such that the expression in round brackets is extremal; Q is thus called the *quantum extremal surface*.[37] Finally, we must take the minimum in case there are multiple extremal candidates for Q.

Increasing the island has two effects in the island formula: firstly, the term with $A(Q)$ increases; secondly, the semi-classical state on Σ_Q receives more contributions from radiation within the black hole, which is entangled with radiation outside the black hole, and thus the state becomes purer, decreasing $H_{\text{s.-cl.}}(\Sigma_Q)$. Thus, a trade-off has to be found. At early times, there is almost no entangled radiation and thus the increase in $A(H)$ cannot be offset by the decrease in $H_{\text{s.-cl.}}(\Sigma_Q)$.

As it turns out, the island vanishes for this reason roughly until the Page time, and thus the entropy is just given by $H_{\text{s.-cl.}}(\Sigma_Q^{\text{rad.}})$, following Hawking's prediction. Around the page time, the island becomes relevant, eventually reproducing the Page curve (recall Figure 6). For details, see [2] and references therein.

Finally, we note that the island formula is quite general as it applies to many gravitating systems. It can for instance be modified to give the black hole entropy $H(\hat{\sigma}_{\text{b}})$; one finds that this entropy too follows a page curve, as expected [2]. However, explicit expressions for the entropies in the semi-classical regime, let alone in the context of the full path integral, have thus far only been obtained in simplified theories of gravity, not in general relativity [59] [1] [2].

Reinterpretation of the Hilbert Space Split. Remarkably, the island formula provides the fine-grained entropy of radiation (and of the black hole), without knowledge of the actual quantum state; because of this, the formula has been called an "oracle" [47]. Nevertheless, one can attempt a qualitative interpretation of these states.

Interestingly, the entropy $H(\hat{\sigma}_{\text{r}})$ computed by the formula (3.7), has contributions from the radiation outside the horizon (the part of $H_{\text{s.-cl.}}(\Sigma_Q)$ coming from the connected component of Σ_Q outside the horizon) *but also* from the radiation inside the horizon (the part of $H_{\text{s.-cl.}}(\Sigma_Q)$ from the island) *and* from the area of the quantum extremal surface Q. This seems to suggest that while the split (3.1) of the Hilbert space $\mathcal{H} = \mathcal{H}_{\text{b}} \otimes \mathcal{H}_{\text{r}}$ into two parts holds, the two parts are not what we might naively think: concretely, the radiation part seems to also contain contributions which we would naively have attributed to $\mathcal{H}_{\text{b}}$. Particularly, the term $A(Q)/4$ on the right-hand side of the island formula (3.7), a truly geometric term, could further signify that *a part of the geometry of spacetime is to be interpreted as actually a part of the radiation state* $\hat{\sigma}_r$.

This idea can be made precise with so-called *entanglement wedges*; they describe the region of the black hole which is completely determined by the interior part of Σ_Q. At early times, the entanglement wedge is non-existent and $\hat{\sigma}_{\text{r}}$ indeed contains only radiation outside the black hole; around the page time, Q becomes non-degenerate and the wedge forms, growing from there until it encompasses almost the entire black hole region shortly before total evaporation. For details, see [2] and sources therein.

3.4 Anatomy of the Paradox

In Section 5 we will see whether the information loss paradox can be discussed in the context of analogue gravity models. In preparation for this, it will be useful to gather the crucial ingredients leading to the paradox.

[37] In [2], the quantum extremal surface is denoted by X. We chose not to use this notation because the naked singularity at the end of evaporation is already called X.

Anatomy. As we have seen in Section 3.2, the Page curve version of the black hole information loss paradox is a clash between the von Neumann entropy $H(\hat{\sigma}_\mathrm{r})$ of the escaping Hawking radiation and the thermodynamic entropy S of the black hole. Thus, we are lead to *define* the following:

Definition 3.1: Anatomy of the Black Hole Information Loss Paradox

A *context*, in which black hole information loss *can be reasonably discussed*, requires the following notions:

1. A notion of outgoing Hawking radiation and correspondingly the von Neumann entropy $H(\hat{\sigma}_\mathrm{r})$ of the radiation state at any time.
2. A notion of black hole entropy.

Let us make a few remarks on how this definition should be interpreted:

1. A "context" can be any part of any physical system. Particularly, we will be interested in analogue gravity models as contexts.
2. The notion of black hole entropy can be a thermodynamic (coarse-grained) one, such as the black hole entropy S in (2.56) in the actual information loss paradox in general relativity. Or it could be a von Neumann (fine-grained) entropy, as in the proposed solution approach we saw in Section 3.3.
3. If the black hole entropy is a coarse-grained entropy, it is merely an upper bound for the fine-grained black hole entropy (assuming that it eventually exists in some more advanced theory). To still be able to talk about whether a Page-curve type paradox occurs (recall Figure 6), the entropy of the radiation must be fine-grained: for if both the entropies of radiation and black hole were coarse-grained and thus upper bounds, no simple statement could be made. We therefore always want a fine-grained entropy for the radiation.
4. There are other, more basic requirements which we did not mention: for instance the context must have a notion of what "outgoing" means, and we need a notion of time. But since these will be automatically present in analogue models, which is our context of interest, we do not state them in the above definition.

Back-Reaction? On first sight, it might seem that we have forgotten back-reaction (recall Section 2.6) in Definition 3.1. After all, without back-reaction, the black hole mass and thus thermodynamic entropy never decreases.

But even if the mass of the black hole is constant, then the paradox still occurs: Hawking's result still predicts a monotonously increasing radiation entropy, which will eventually surpass the now constant black hole thermodynamic entropy. This situation is depicted in Figure 9. Thus, back-reaction is not necessary to state the paradox.

It is interesting to note, that even in contexts with back-reaction present, we can artificially feed the black hole with matter of our choice in order to keep the black hole mass and thermodynamic entropy constant. The information loss paradox can be discussed in that context [17].

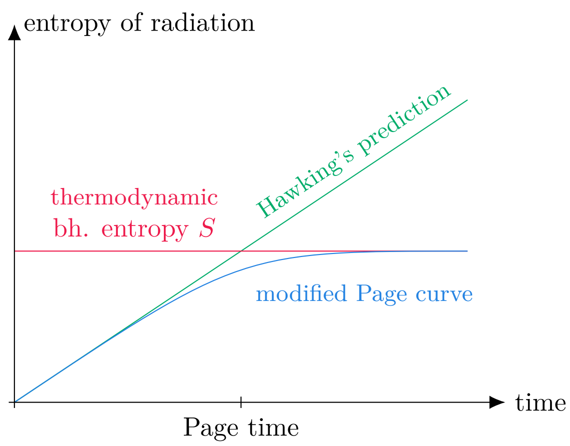


Figure 9: Page curve version of the paradox with no back-reaction. The thermodynamic entropy of the black hole entropy remains constant. Nevertheless, the paradox still occurs when the entropy of the radiation has increased enough. To prevent information loss, the radiation entropy should follow a modified page curve.

4 Analogue Gravity

Analogue gravity models are the second main topic of this work. All we needed so far of them were some of their basic properties introduced in Section 1, especially the form of fluid-flow metrics encountered in the context of fluid-flow analogue models. Recall that we are mostly interested in fluid-flow analogue models, since they readily allow for apparent horizons.

We will now properly introduce analogue gravity models, starting with general classical models in Section 4.1. In Section 4.2 we then derive one of the most important classical models: the fluid-flow analogue model of *linear sound*. In Section 4.3 we will encounter another important example, this time of a quantum fluid-flow analogue model: linear sound in a *Bose-Einstein condensate*. We then discuss a central limitation of fluid-flow analogue models in Section 4.4: the fact that they must obey some sort of continuity equation makes it impossible to perfectly model Schwarzschild spacetime. Finally, we briefly encounter without derivation also in Section 4.3 the fully quantum model of sound in a Bose-Einstein condensate introduced by [45] [46], as an example of fully quantum-mechanical analogue models which do not come from quantizing a classical model.

Note that the present treatment of analogue gravity focuses mainly on fluid-flow analogue models, with classical or quantum sound propagation, since they are the simplest relevant models appropriate for describing horizons. There are however many more complex models, such as the fully quantum one in Section 4.3, which also have this capability. While we simply cannot treat all models here due to their sheer variety, we will already be able to arrive at some conclusions about more general models from considering the models treated here. For more complete treatments of analogue gravity, see [7], [53] and [78].

4.1 General Classical Models

As mentioned above, there is an enormous wealth of analogue gravity models; this is true even when only considering classical models. We show here that classical analogue gravity models occur quite generally when considering scalar fields obeying a local principle of least action leading to wave propagation.

Hyperbolic Partial Differential Equations. We wish to study scalar fields that allow for wave propagation. A useful class of partial differential equations (PDEs) with this property are *linear hyperbolic PDEs*:

Definition 4.1: Linear Hyperbolic PDE

Let D be a linear partial differential operator acting on the scalar field ϕ, with at most second derivatives, *i.e.*

$$D\phi = h^{\mu\nu}\partial_\mu\partial_\nu\phi + \mathcal{O}(\partial\phi), \tag{4.1}$$

where we may always choose $h^{\mu\nu}$ to be symmetric. We say that D is *hyperbolic*, if the signature of $h^{\mu\nu}$ is $-+++$. We then also call the PDE $D\phi = 0$ *hyperbolic*.

Note that the case of signature $+---$ can trivially be turned into $-+++$ without changing the physics by multiplying the PDE with a factor of -1.

Of particular physical interest are PDEs set in flat spacetime; *i.e.* defined on a subset $\Omega \subset \mathbb{R} \times \mathbb{R}^d$, where the first factor is time and the second is d-dimensional space. With this we choose a preferred inertial reference frame.

We have not yet restricted the direction in spacetime in which $h^{\mu\nu}$ is negative definite (*i.e.* the reason for the $-$ in the signature). In light of the following lemma it makes sense to only consider hyperbolic PDEs such that

$$h^{00} < 0, \quad h^{ii} > 0 \text{ for } i = 1, 2, 3. \tag{4.2}$$

These PDEs are interesting, because they yield wave-like solutions: information about perturbations in the field propagates at a *finite speed*. Equivalently, a field vanishing ($\phi = 0$, $\partial_\mu \phi = 0$) in some subset of space remains zero at least for a finite amount of time until potential disturbances from outside had the time to reach the region of interest. Formally:

Lemma 4.2: Finite Propagation Speed for Linear Hyperbolic PDEs

Let $D\phi = 0$ be a linear hyperbolic PDE defined on the domain Ω. For the purpose of this lemma we call a vector ξ^μ at some point of our domain $h^{\mu\nu}$*-timelike*, if $h^{\mu\nu}\xi_\mu\xi_\nu < 0$, and $h^{\mu\nu}$*-spacelike*, if $h^{\mu\nu}\xi_\mu\xi_\nu > 0$.

Let $p \in U \subset \Sigma \subset \Omega$, where U is open in some $h^{\mu\nu}$-spacelike hypersurface Σ. Assume $\phi = 0$ and $\partial_\mu \phi = 0$ on U.

Then, if c is a regular, $h^{\mu\nu}$-timelike curve through p and parametrized by t with $c(0) = p$, there exists $\varepsilon > 0$ such that $\phi(c(t)) = 0$ for $|t| < \varepsilon$.

So if we require (4.2), a particular set of regular $h^{\mu\nu}$-timelike curves are the integral curves of the time coordinate vector field (the "flow of time" in our chosen reference frame); furthermore, surfaces of constant time ("space at an instant" in our frame) make valid candidates for Σ. The lemma then implies that in the chosen reference frame information carried by the field ϕ has a finite propagation speed.

Note that while the coordinate basis vectors are $h^{\mu\nu}$-timelike (-spacelike) if and only if they are timelike (spacelike) due to (4.2), this is not necessarily true for all vectors. Thus, there might exist observers, for which solutions do not appear as propagating waves at all. This is no problem, since our PDE must in no way be Lorentz-invariant.

Lemma 4.2 follows from Proposition 8.1 of [72]; let us give a short account of its proof and consequence here. There, roughly speaking, the PDE is first converted into a PDE in curved spacetime with inverse metric density $h^{\mu\nu} = \sqrt{|g|}\, g^{\mu\nu}$, thus replacing notions of timelike and spacelike with $h^{\mu\nu}$-timelike and $h^{\mu\nu}$-spacelike. This is also the first hint at a possible analogue model. One can then derive an energy estimate (for an appropriately defined notion of energy, quadratic and positive definite in $\partial_\mu \phi$) in a spacetime volume enclosed by two spacelike hypersurfaces (in the curved spacetime, corresponding to $h^{\mu\nu}$-spacelike in our terminology), based only on initial data given on the past hypersurface; this is the content of Proposition 8.1 in [72]. From there it follows that the energy remains zero for a finite amount of time after the field vanished, in the way described in the lemma above. Since energy is quadratic and positive definite in the field derivatives, and the spacetime region as well as the hypersurfaces are arbitrary, it follows that the field must remain constant for a finite time (along timelike curves in curved spacetime, *i.e.* along $h^{\mu\nu}$-timelike curves in our terminology). Given the initial conditions, the field must vanish for a finite amount of time.

Lagrangian Formulation. Let us further restrict ourselves to PDEs which follow from a local principle of least action, since those are even easier to work with and still very general. We assume that the PDE follows from $\delta S = 0$, where $S = \int \mathrm{d}^4x\, \mathcal{L}$ is the action and $\mathcal{L}$ the Lagrangian density. Since we want our PDEs to be linear, $\mathcal{L}$ can at most contain terms quadratic in the field ϕ and/or its derivatives. Furthermore, since the PDE is to be hyperbolic and thus second order, we can have at most first derivatives of ϕ in $\mathcal{L}$. The most general Lagrangian density is thus

$$\mathcal{L} = \frac{1}{2} h^{\mu\nu} (\partial_\mu \phi)(\partial_\nu \phi) + A^\mu (\partial_\mu \phi)\phi - \frac{1}{2} B\phi^2 + F\phi - G^\mu \partial_\mu \phi, \tag{4.3}$$

where $h^{\mu\nu}$, A^μ, F and G^μ are arbitrary functions, and $h^{\mu\nu}$ is without loss of generality symmetric. The field equations are the Euler-Lagrange equations:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0, \tag{4.4}$$

yielding

$$\partial_\mu (h^{\mu\nu} \partial_\nu \phi) + [(\partial_\mu A^\mu) + B]\phi = F + \partial_\mu G^\mu. \tag{4.5}$$

Note that it is *impossible* to obtain a term of the form $C^\mu \partial_\mu \phi$ in the PDE, with C^μ arbitrary functions; the best we can hope for is the term $(\partial_\mu h^{\mu\nu})\partial_\nu \phi$ resulting from rewriting the first term in the PDE (4.5), but it is not independent of the term quadratic in $\partial_\mu \phi$. This simple consequence of the Euler-Lagrange equations will have important implications for analogue models further down. Note also that the last term in the Lagrangian density is largely useless, since its effects can be replicated by the second-to-last term.

Finally, we note that our requirements of hyperbolicity as well as (4.2) above directly translate to requirements on the functions $h^{\mu\nu}$ in the Lagrangian density, hence the choice of symbol for the first term in the Lagrangian density.

Analogue Models. The first term in (4.5) resembles the d'Alambertian $\Box$ (*i.e.* the *wave operator*) in curved spacetime: given a metric g_{ab}, it is defined as

$$\Box := \frac{1}{\sqrt{|g|}} \partial_a \left(\sqrt{|g|}\, g^{ab}\, \partial_b \right). \tag{4.6}$$

This leads us to the following result:

Theorem 4.3: Classical Analogue Model from Linear Hyperbolic Lagrangian PDE

Consider the Lagrangian density

$$\mathcal{L} = \frac{1}{2} h^{\mu\nu} (\partial_\mu \phi)(\partial_\nu \phi) + A^\mu (\partial_\mu \phi)\phi - \frac{1}{2} B\phi^2 + F\phi \tag{4.7}$$

leading to a linear, hyperbolic field equation for the scalar field ϕ, as described above. Assume that the dimension of space satisfies $d > 1$.

This field equation is then mathematically identical to the field equation

$$\Box\phi + \frac{1}{\sqrt{|g|}} [(\partial_\mu A^\mu) + B]\phi = \frac{F}{\sqrt{|g|}}, \tag{4.8}$$

where $\Box$ is the d'Alambertian (4.6) in the curved spacetime with inverse metric

$$g^{\mu\nu} = h^{\mu\nu} / \sqrt{|g|}, \qquad |g| = |\det g_{\mu\nu}| = |\det h^{\mu\nu}|^{2/(d-1)}. \tag{4.9}$$

In other words: the original linear hyperbolic PDE in flat spacetime can be seen as a Klein-Gordon equation in curved spacetime, with general mass and source terms, both potentially depending on spacetime position.

Proof. We begin by identifying $h^{\mu\nu} =: \sqrt{|g|}\, g^{\mu\nu}$. Taking the absolute value of the determinant on both sides yields $|\det h^{\mu\nu}| = |g|^{-1+d/2+1/2}$, that is, $|g| = |\det h^{\mu\nu}|^{2/(d-1)}$. To recover the full d'Alambert operator (4.6) in the first term of the field equation, we divide (4.5) by $\sqrt{|g|}$. Note that $|g| \neq 0$, because $\det h^{\mu\nu} \neq 0$, which in turn follows from our assumption of the signature of $h^{\mu\nu}$. □

Theorem 4.3 can be roughly used in two ways:

1. One may choose a Klein-Gordon equation of interest, with some general mass and source terms in some curved spacetime with chosen coordinate system, and rewrite it as a hyperbolic PDE in flat spacetime. Furthermore, one may be interested to see whether there exist known physical systems in flat spacetime which implement this PDE. Such systems are commonly called *analogue gravity models*. They are of prime experimental interest, since they may be implemented in a lab, while arbitrarily curved spacetimes cannot. It should be noted that, strictly speaking, these analogue models are *not* (despite the name) analogies for the dynamics of gravity, the Einstein equations, but merely of the Klein-Gordon equation in the already fixed curvature background.

2. Of course one may also use the theorem in the direction we have stated it. That is, we begin with a hyperbolic Lagrangian PDE and arrive at some Klein-Gordon equation in some curved spacetime, in the coordinates inherited from the theorem. Of course, one may then *impose* diffeomorphism invariance and switch to another coordinate system, where the equation is easier to solve. Turning PDEs into Klein-Gordon equations in curved spacetime is even useful for the purely mathematical study of PDEs (as is done in [72]).

Although we have imposed various constraints on our PDE (linearity, hyperbolicity, condition (4.2), obeying an action principle), we argue that we have not greatly reduced the number of physically interesting PDEs that fulfil the requirements for Theorem 4.3:

1. Assuming hyperbolicity and (4.2) of course limits us to PDEs with wave-like solutions. But such PDEs are still very common in physics: *e.g.* sound waves in fluids and solids, electromagnetic waves, *etc.*

2. Linearity restricts us to even simpler wave equations. Nevertheless, many fundamental systems exhibit linear behaviour. And small excitations can be approximated as linear phenomena.

3. Any PDE lending itself to quantization should be derivable from an action principle, because quantization generally involves a Lagrangian or Hamiltonian formulation of the classical system [43] [65]. Many interesting PDEs are of course quantizable, thus we should have no shortage of PDEs from which we can construct analogue models. And when studying quantized Klein-Gordon equations on curved spacetimes through analogue models, the analogue model must be quantizable and the assumption of an action principle is anyway required. Quantizable systems are exactly what we are interested in when studying Hawking radiation.

   One can also ask more generally when a given PDE follows from an action principle. This question is the *inverse problem of Lagrangian field theory*.[38] We will not discuss it further.

With these arguments in mind, Theorem 4.3 shows that we should not be too surprised by the existence of analogue models. And if we also allow more complicated models not described by the theorem (such as coupled scalar fields, complex scalar fields, vectors fields, *etc.*), we are lead to expect a large wealth of analogue models; this is indeed the case [7], [53], [78].

[38] The corresponding problem for Lagrangian mechanics (only time as independent variable instead of position and time) has been solved [22].

But let us remark that while Theorem 4.3 is very general and captures a large class of scalar field analogue models, it is not the most useful way to explicitly construct analogue models: one usually works with the field equations of the analogue system directly, instead of the Lagrangian. See [7] and [53] for extensive treatments of models constructed in this way. We introduced analogue models via Theorem 4.3, since we wish to keep the discussion more general. Note that a discussion similar to ours can be found in [6]. There the focus lies on hyperbolic PDEs obtained from a general Lagrangian through linearization around a solution; once linearized, methods similar to our theorem are used. In the Sections 4.2 and 4.3 we will see examples of such linearizations followed by an application of the theorem.

Mass- and Sourceless Models, Gauge Invariance. We often do not want a general Klein-Gordon equation, but one with specific mass and source terms; particularly interesting are analogue models describing the mass- and sourceless Klein-Gordon equation

$$\Box\phi = 0 \tag{4.10}$$

in some given curved spacetime. For this we need $[(\partial_\mu A^\mu) + B]/\sqrt{|g|} = 0$ and $F/\sqrt{|g|} = 0$, *i.e.* the mass and source terms must vanish.

Let us investigate the direct meaning of the mass and source terms: The term $\partial_\mu A^\mu$ stems from the term $A^\mu(\partial_\mu\phi)\phi$ in the Lagrangian density. Such a term requires a distinguished direction in the analogue system, providing A^μ; the analogue system must be anisotropic. It must also be inhomogeneous, since otherwise A^μ is constant, *i.e.* $\partial_\mu A^\mu = 0$. To summarize: $\partial_\mu A^\mu$ may be non-zero in inhomogeneous, anisotropic systems, which is often the case for analogue model systems (for instance the flowing fluids of the next section). B arises from $-B\phi^2/2$, the mass term of the analogue system. Overall, there seems to be no simple condition for masslessness. The source term $F/\sqrt{|g|}$ in the Klein-Gordon equation follows from the source term $F\phi$ in the analogue system Lagrangian and the induced metric. The source term in the Klein-Gordon equation vanishes if and only if it vanishes in the analogue system.

But there is a simpler condition for the mass term to vanish: if ϕ possesses the *Gauge freedom*

$$\phi \rightsquigarrow \phi + \tilde{\phi}, \qquad \tilde{\phi} = \text{const}, \tag{4.11}$$

then a non-zero mass term would be unphysical and thus cannot occur. So a very useful way of constructing massless analogue models using Theorem 4.3 is to only consider systems with such a gauge freedom; to our best knowledge, this method has not been used before. It has the potential of providing a very simple and intuitive reason for why certain analogue models are massless. We will apply it to the models of the Sections 4.2 and 4.3.

Dimension. An interesting observation is the condition $d > 1$ in the theorem. In the case $d = 1$ the decomposition $h^{\mu\nu} = \sqrt{|g|}\, g^{\mu\nu}$ fails since the equality can only hold up to a factor, and $g^{\mu\nu}$ is also fixed only up to a factor. Another way of seeing this is to note that with $d = 1$, the expression $\sqrt{|g|}g^{\mu\nu}$ becomes *conformally invariant* (if $g^{\mu\nu}$ scales by a conformal factor of χ, then $\sqrt{|g|}$ scales by a factor of χ^{-1}).

Einsteinian Dynamics? The name "analogue gravity model" is slightly misleading, because these models really are models of Klein-Gordon fields in curved spacetime, and not of gravity itself.

Importantly, analogue models only provide a metric as a background for the Klein-Gordon field to evolve on, but do not model the dynamics of the metric, *i.e.* the Einstein field equations (0.1) [7, Section 7.4]. The dynamics inherent in the model (such as the *continuity* and *Euler equations* for a fluid) rather have the effect of placing constraints (with usually no gravitational meaning) on the metrics achievable by the Model. We will see an example of this in Section 4.4.

We should note that to our knowledge two instances of analogue models (albeit very complex ones) have been able to reproduce the Einsteinian dynamics at least partially. For one, it is possible to model the dynamics of certain near-extremal Reissner-Nordström black holes (Kerr-Newman black holes with $J = 0$) [4] [5]; see also [7, Section 7.1] and sources therein. For the other, an Einstein-Hilbert action term was shown to occur [78] in certain phases of Fermi-liquids such as ^{3}He; however problems occur. See also [7, Section 7.8] and sources therein.

4.2 Linear Sound in Irrotational, Barotropic, Perfect Fluids

As stated in Section 1.2, we here derive the classical fluid-flow analogue model discovered by Unruh [74] and Visser [75]. For this, we consider the *propagation of sound in an irrotational, barotropic, perfect fluid.*

This model, like most, is most straightforwardly derived directly from rearranging the equations of motion of the analogue system, the fluid equations in this case. But this relatively standard approach is not very intuitive: it does not reveal why at the end one would expect an analogue model for the mass- and sourceless Klein-Gordon equation to come out; instead, some terms almost "magically" cancel to give the final result. We will therefore employ the method explained in the previous section and use Theorem 4.3 together with gauge freedom present in the system to derive the analogue model.

Fluid Description. Consider a fluid described by the local flow velocity $\mathbf{v}(t, \mathbf{x})$, density $\rho(t, \mathbf{x})$ and pressure $p(t, \mathbf{x})$, flowing in d spatial dimensions. The pressure is given by the *barotropic* equation of state

$$p = p(\rho). \tag{4.12}$$

We will also allow an external potential force $\mathbf{f}(t, \mathbf{x}) = -\rho(t, \mathbf{x})\,\nabla U(t, \mathbf{x})$ (such as gravity) to act on the fluid. That the fluid is *irrotational* means that we can introduce the *velocity potential* Φ and write

$$\mathbf{v} =: -\nabla\Phi. \tag{4.13}$$

In the cases $d = 2$ and $d = 3$ this is equivalent to $\nabla \times \mathbf{v} = 0$, hence the name "irrotational". Note that the velocity potential has the gauge freedom

$$\Phi \rightsquigarrow \Phi + f(t), \tag{4.14}$$

where f is a function of time only, because such changes to Φ do not change the velocity field $\mathbf{v}$ and thus the physics of the fluid.

For later we note that this fluid obeys the continuity equation (enforcing the conservation of mass)

$$\frac{\partial\rho}{\partial t} + \nabla \cdot (\rho\mathbf{v}) = 0 \tag{4.15}$$

and the Euler equation (enforcing the conservation of momentum)

$$\frac{\partial}{\partial t}(\rho\mathbf{v}) + \sum_{k=1}^{d} \frac{\partial}{\partial x^k}(\rho\mathbf{v}\,\mathrm{v}_k) + \nabla p = -\rho\nabla U. \tag{4.16}$$

Since we also deal with spacetime coordinates, it is important to note that $\partial_t = \partial/\partial_t$ is the time derivative without factor of c, *i.e.* $\partial_t = c\partial_0$. For irrotational flows, the Euler equation is commonly rewritten as the *Bernoulli equation*:

$$-\frac{\partial\Phi}{\partial t} + \frac{1}{2}(\nabla\Phi)^2 + P + U = 0, \qquad P(\rho) := \int^{\rho} \frac{1}{\rho'}\frac{\partial p}{\partial\rho}(\rho')\,\mathrm{d}\rho'. \tag{4.17}$$

Importantly, this rewriting forces us to partially fix the gauge $\Phi \rightsquigarrow \Phi + f(t)$, with a function f whose derivative $f'(t)$ is fixed. Thus, the remaining gauge freedom is

$$\Phi \rightsquigarrow \Phi + \tilde{\Phi}, \tag{4.18}$$

where $\tilde{\Phi}$ is a constant.

For an extensive treatment of fluid dynamics including the derivation of these equations, see for instance [39].

Linear Sound. We will be interested in *sound*, that is, solutions (Φ, ρ) of (4.15) and (4.17) (or alternatively of (4.16)) which are small perturbations around another given solution (Φ_0, ρ_0), the so-called *background flow*. For this we write

$$\Phi = \Phi_0 + \varepsilon\Phi_1, \qquad \rho = \rho_0 + \varepsilon\rho_1, \tag{4.19}$$

with a small perturbation parameter $\varepsilon \ll 1$. The perturbations Φ_1, ρ_1 describe the *propagation of sound on top of the background flow*.

Expanding the fluid equations in powers of ε gives the equations of *sound propagation*. Taking only terms of first order in ε gives the equations of *linear sound*; these are a good approximation, if ε is small (faint sound) as we have assumed. In the case $\varepsilon \sim 1$ (loud sound), the approximation breaks down, and the distinction between background flow and sound blurs out.

Fetter-Walecka Lagrangian. The method of obtaining linear sound equations described just above is the one most often employed (see *e.g.* [74], [75], [7], [53] in the context of analogue models and [39] in the context of general fluid dynamics). But especially when constructing analogue models, this method does not provide much insight and does not explain why the resulting Klein-Gordon equation is mass- and sourceless. We instead wish to use the more intuitive approach of applying Theorem 4.3 and hopefully using the gauge invariance of Φ. We thus need a (much less common) Lagrangian approach to fluid dynamics.

While it is possible to describe the fluid by considering infinitely many *infinitesimal fluid elements* as a limiting case of a multi-particle system and construct a Lagrangian density in this way [50], a more exotic approach fits our use-case better: we consider a version of the *Fetter-Walecka Lagrangian density* [25]

$$\mathcal{L} = \rho\frac{\partial\Phi}{\partial t} - \frac{1}{2}\rho(\nabla\Phi)^2 - \int^{\rho} P(\rho')\,\mathrm{d}\rho' - \rho U, \tag{4.20}$$

where the action $S = \int \mathrm{d}^4x\mathcal{L}$ is to be varied in Φ and ρ.

The Euler-Lagrange equation (4.4) for Φ yields the continuity equation (4.15):

$$\frac{\partial\rho}{\partial t} - \nabla\cdot(\rho\nabla\Phi) = 0. \tag{4.21}$$

From the Euler-Lagrange equation for ρ we obtain the Bernoulli equation (4.17):

$$\frac{\partial\Phi}{\partial t} - \frac{1}{2}(\nabla\Phi)^2 - P(\rho) - U = 0, \tag{4.22}$$

So the Lagrangian density (4.20) really gives the correct fluid equations.

Lagrangian Description of Linear Sound. Assume that the background flow Φ_0 and ρ_0 is given and is a solution of the fluid equations. To derive equations for Φ_1 and ρ_1, we require that the full fields Φ, ρ are also a solution of the fluid equations. For this, we insert the decomposition (4.19) into the Lagrangian density (4.20) and vary the action with respect to Φ_1 and ρ_1.

So that we can easily read off the linear sound equations, we also expand $\mathcal{L}$ in powers of ε before varying:

$$\begin{aligned}\mathcal{L} = \mathcal{L}_0 + \varepsilon\left[\rho_0\frac{\partial\Phi_1}{\partial t} + \rho_1\frac{\partial\Phi_0}{\partial t} - \rho_0(\nabla\Phi_0)\cdot(\nabla\Phi_1) - \frac{1}{2}\rho_1(\nabla\Phi_0)^2 - P(\rho_0)\rho_1 - \rho_1 U\right] \\ + \varepsilon^2\left[\rho_1\frac{\partial\Phi_1}{\partial t} - \frac{1}{2}\rho_0(\nabla\Phi_1)^2 - \rho_1(\nabla\Phi_0)\cdot(\nabla\Phi_1) - \frac{1}{2}\frac{1}{\rho_0}\frac{\partial p}{\partial\rho}(\rho_0)\rho_1^2\right] \\ + \mathcal{O}(\varepsilon^3),\end{aligned} \tag{4.23}$$

with $\mathcal{L}_0$ a constant, and we have used

$$\frac{\partial P}{\partial\rho}(\rho_0) = \frac{1}{\rho_0}\frac{\partial p}{\partial\rho}(\rho_0) \tag{4.24}$$

in the second line. At first order in ε the variation of the action with respect to Φ_1 and ρ_1 simply asserts that Φ_0, ρ_0 is a solution. The linear sound equations must thus emerge at the next higher order. Varying the term in the action coming from the Lagrangian term quadratic in ε yields the equations

$$\frac{\partial\rho_1}{\partial t} - \nabla\cdot\big(\rho_0\nabla\Phi_1 + \rho_1\nabla\Phi_0\big) = 0 \tag{4.25}$$

(from varying Φ_1) and

$$\frac{\partial\Phi_1}{\partial t} - (\nabla\Phi_0)\cdot(\nabla\Phi_1) - \frac{1}{\rho_0}\sigma^2\rho_1 = 0. \tag{4.26}$$

(from varying ρ_1). Here we have introduced the *speed of sound* σ:

$$\sigma^2 := \frac{\partial p}{\partial\rho}(\rho_0). \tag{4.27}$$

Its meaning will become clear further down.

Equations (4.25) and (4.26) are the equations of linear sound. We note that one can solve for ρ_1 in (4.26), insert into (4.25), and thus obtain a linear, second order PDE for Φ_1 alone, the so-called *linear sound equation*. This fact can already be deduced without much calculation from the fact that ρ_1 enters the $\mathcal{O}(\varepsilon^2)$-part of $\mathcal{L}$ only undifferentiated, which in turn follows from ρ appearing undifferentiated in $\mathcal{L}$. Thus, linear sound is completely described by the perturbation Φ_1 of the velocity potential alone. And as we will see, it is the linear sound equation for Φ_1 that leads to the analogue model. We will not need to derive the linear sound equation at the present time, however.

To apply Theorem 4.3, we must find a Lagrangian density from which the linear sound equation follows. But this is now very simple: all we need to do, is insert the expression for ρ_1 obtained from (4.26) into the Lagrangian density (4.23), making it only dependent on Φ_1; varying the action with respect to Φ_1 then gives the linear sound equation when considering the terms of order $\mathcal{O}(\varepsilon^2)$. Thus, the Lagrangian responsible for the linear fluid equation is

$$\begin{aligned}\mathcal{L}_{\text{sound}} = \frac{\rho_0}{\sigma^2}\frac{\partial\Phi_1}{\partial t}\left[\frac{\partial\Phi_1}{\partial t} - (\nabla\Phi_0)\cdot(\nabla\Phi_1)\right] - \frac{1}{2}\rho_0(\nabla\Phi_1)^2 \\ - \frac{\rho_0}{\sigma^2}\left[\frac{\partial\Phi_1}{\partial t} - (\nabla\Phi_0)\cdot(\nabla\Phi_1)\right](\nabla\Phi_0)\cdot(\nabla\Phi_1) \\ - \frac{1}{2}\frac{\rho_0}{\sigma^2}\left[\frac{\partial\Phi_1}{\partial t} - (\nabla\Phi_0)\cdot(\nabla\Phi_1)\right]^2.\end{aligned} \tag{4.28}$$

As expected, it is of the form required by Theorem 4.3. Furthermore, it does not have a source term. We can read off

$$h^{\mu\nu} = \frac{\rho_0}{2}\frac{c^2}{\sigma^2}\left(\begin{array}{c|c} 1 & -\partial_i\Phi_0/c \\ \hline -\partial_i\Phi_0/c & (\partial_i\Phi_0)(\partial_j\Phi_0)/c^2 - \delta_{ij}\,\sigma^2/c^2 \end{array}\right), \qquad i,j = 1,\dots,d. \tag{4.29}$$

Note that the time derivatives in (4.7) are not ∂_t but ∂_0, we thus had to add the appropriate factors of c, the speed of light.

Finally, we must check the signature of $h^{\mu\nu}$ to check hyperbolicity, before applying Theorem 4.3. We note that $h^{\mu\nu}$ has signature $+---$ in the simple case $\mathbf{v} = -\nabla\Phi_0 = 0$ of vanishing background flow; in this case the linear sound equation becomes a familiar wave equation, and we can see that σ is indeed the speed of sound. Since the signature is the (ordered) list of signs of the eigenvalues of the matrix, which themselves depend continuously on $\nabla\Phi_0$, ρ_0 and σ^2, the signature can only become different from $+---$ if at least one of the eigenvalues becomes zero for some combination of $\nabla\Phi_0$, ρ_0 and σ^2.

To show that this is not the case, we compute $\det h^{\mu\nu}$, which will anyway be useful to determine $g_{\mu\nu}$ later. For this, we use the block matrix determinant formula

$$\det\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) = \det(A)\cdot\det(D - CA^{-1}B) \tag{4.30}$$

to arrive at

$$\begin{aligned}\det h^{\mu\nu} = \left(\frac{\rho_0}{2}\frac{c^2}{\sigma^2}\right)^{d+1}\cdot 1\cdot\det\left(-\frac{\sigma^2}{c^2}\delta_{ij} + \frac{1}{c^2}(\partial_i\Phi_0)(\partial_j\Phi_0) - \frac{1}{c^2}(\partial_i\Phi_0)(\partial_j\Phi_0)\right)& \\ = \left(\frac{\rho_0}{2}\right)^{d+1}\cdot(-1)^d\cdot\frac{c^2}{\sigma^2} \neq 0.& \end{aligned} \tag{4.31}$$

So there is never a zero eigenvalue, and the signature of $h^{\mu\nu}$ is always $+---$. As noted above, this is equivalent to $-+++$, making the linear sound equation hyperbolic, and we can apply Theorem 4.3.

Alternative Intuitive Reasoning. Note that we could have already intuitively guessed that the coupled equations (4.25) and (4.26) for Φ_1 and ρ_1 would be combinable into one equation for one of the two scalar fields, without at all having to derive them, since we expect sound to only have one polarization. So we could have guessed that sound behaves according to a linear, second order PDE of one scalar field. Furthermore, we expect wave propagation and could thus have reasonably guessed that this PDE was hyperbolic.

It is perhaps surprising that this scalar field is Φ_1 and not ρ_1, since we commonly associate sound with oscillations in density and might not immediately think of the necessary oscillations in the velocity field entailed by it. But we could have just as well solved one equation for Φ_1 and inserted into to other, in order to obtain an equation for ρ_1 (although this would have been a bit harder due to the explicit form of the equations); and the resulting equation would be a linear, second order PDE all the same. That the linear sound equation follows from a Lagrangian action principle, would then of course also follow. The only difficulty would then be that we can no longer use the gauge invariance of Φ_1 following from the gauge invariance of Φ, and the vanishing mass term in the final result would have to come about in some other way.

Finally, we can argue that the resulting PDE should be sourceless: assuming that the Lagrangian description of sound comes from the perturbation of the Lagrangian description of the entire fluid around a background flow solution for the fluid, the relevant Lagrangian for sound comes only at second order in the perturbation and thus can only contain terms quadratic in Φ_1 and its derivatives (or ρ_1 and its derivatives).

So with a bit of intuition, almost all computations in the previous section could have been skipped. But in order to get the explicit equations, they are of course necessary.

Analogue Model. Applying Theorem 4.3 to the linear sound Lagrangian density $\mathcal{L}_{\text{sound}}$ (4.28) gives a Klein-Gordon equation for Φ_1, in the spacetime given by the metric components obtained from $h^{\mu\nu}$ of (4.29). We will come back to the explicit form of the metric below.

Having successfully described linear sound by a Lagrangian for Φ_1, we can now use the gauge freedom

$$\Phi_1 \rightsquigarrow \Phi_1 + \tilde{\Phi}, \tag{4.32}$$

with $\tilde{\Phi}$ a constant, following from the gauge freedom (4.18) of Φ. It follows that the resulting Klein-Gordon equation *must be masseless.*[39] And we have also seen that it is sourceless.

In conclusion: Linear sound in an irrotational, barotropic, perfect fluid is an analogue model for the source- and massless Klein-Gordon equation (4.10), the mass term vanishing due to the gauge freedom inherent in Φ_1, and the source term disappearing because of sound being a perturbation around a background flow. If we had worked directly with the fluid equations and never asked about the Lagrangian density, we would of course have obtained the same result; but it would have been entirely unclear, why terms "miraculously" cancel to give the source- and massless Klein-Gordon equation (4.10). The Lagrangian approach also permits us to trade most computations for some intuitive assumptions, if we like, while still allowing us to arrive at this conclusion. It allows us to understand the analogue model of linear sound almost completely intuitively.

Induced Metric. Let us now turn to the part where intuitive reasoning does not help us much further: the explicit form of the metric $g_{\mu\nu}$. For this we first note that $h^{\mu\nu}$ has dimension of mass $\cdot$ length^{-3}, as opposed to no units, which one would expect of an inverse metric density. We can remedy this by multiplying with an arbitrary constant $2C$ of opposite dimension: $h^{\mu\nu} \rightsquigarrow -2C \cdot h^{\mu\nu}$.[40] This does not change the field equations or the geometry of the metric, it being an overall factor. Note that we then have

$$\det h^{\mu\nu} = -(C\rho_0)^{d+1} \cdot \frac{c^2}{\sigma^2}. \tag{4.33}$$

We now identify $h^{\mu\nu} = \sqrt{|g|} g^{\mu\nu}$. Then, $\det h^{\mu\nu} = \det(\sqrt{|g|} g^{\mu\nu}) = \sqrt{|g|}^{\,d+1} \det(g^{\mu\nu})$. So with $|g| = |\det(g_{\mu\nu})| = |\det(g^{\mu\nu})^{-1}|$, and using (4.33), we have $\sqrt{|g|}^{\,d-1} = (C\rho_0)^{d+1} \cdot c^2/\sigma^2$, and hence

$$\sqrt{|g|} = \left((C\rho_0)^{d+1} \cdot \frac{c^2}{\sigma^2} \right)^{1/(d-1)}. \tag{4.34}$$

Therefore, with $v_{0,j} = -\partial_j \Phi_0$ reinstated, we get

$$g^{\mu\nu} = -\frac{(C\rho_0)^{1-(d+1)/(d-1)}}{(\sigma/c)^{2-2/(d-1)}} \cdot \left(\begin{array}{c|c} 1 & v_{0,j}/c \\ \hline v_{0,i}/c & -\delta_{ij}\, \sigma^2/c^2 + v_{0,i} v_{0,j}/c^2 \end{array} \right). \tag{4.35}$$

[39] Since we already have the explicit expression for $\mathcal{L}_{\text{sound}}$ in (4.28), we can explicitly see that this is really the case. Had we taken the intuitive route we would however not have that expression; the argument would still work of course.

[40] The factor 2 is to get rid of the factor 1/2 in (4.29), and the -1 changes the signature from $+---$ to $-+++$.

Using $1-(d+1)/(d-1)=-2/(d-1)$ and inverting the metric, we get

$$g_{\mu\nu} = -\left(C\rho_0 \cdot \frac{c}{\sigma}\right)^{2/(d-1)} \cdot \frac{\sigma^2}{c^2} \cdot \left(\begin{array}{c|c} 1-\mathbf{v}_0^2/\sigma^2 & v_{0,j}\, c/\sigma^2 \\ \hline v_{0,i}\, c/\sigma^2 & -\delta_{ij}\, c^2/\sigma^2 \end{array}\right) \tag{4.36}$$

$$= \left(C\rho_0 \cdot \frac{c}{\sigma}\right)^{2/(d-1)} \cdot \left(\begin{array}{c|c} \mathbf{v}_0^2/c^2 - \sigma^2/c^2 & -v_{0,j}/c \\ \hline -v_{0,i}/c & \delta_{ij} \end{array}\right). \tag{4.37}$$

Upon closer inspection, this has the form (1.28) of a fluid-flow metric, with

$$\mathbf{V} = \frac{\mathbf{v}_0}{c}, \qquad c_s = \frac{\sigma}{c}, \qquad \Theta = \left(C\rho_0 \cdot \frac{c}{\sigma}\right)^{2/(d-1)}. \tag{4.38}$$

Quantization. To obtain a quantum fluid flow analogue gravity model, one can now proceed by quantizing the sound waves. This is done by replacing Φ_1 by a quantum field $\hat{\Phi}_1$, and instating the correct commutators, as we did in Section 2.5. The emerging particles are so-called *phonons*, and, if an apparent horizon is present, we obtain Hawking radiation.

4.3 Linear Sound in Bose-Einstein Condensates

In this section we consider another important example of an analogue gravity model: linear sound in a *Bose-Einstein condensates.* We first apply Theorem 4.3 and the reasoning of Section 4.1 combined with second quantization to obtain a *quantum fluid-flow analogue model* in the so-called *Bogoliubov approximation*; we will see that here too, gauge invariance prohibits a mass term. Finally, we also mention some of the more cutting-edge analogue gravity models involving Bose-Einstein condensates recently constructed in [45] [46]; these are examples of fully quantum analogue models.

Bose-Einstein Condensates. An interacting, dilute, non-relativistic *Bose-Einstein condensate* can be described using the formalism of second quantization as a quantum field $\hat{\phi}(x)=\hat{\phi}(t,\mathbf{x})$ in the Heisenberg picture, satisfying the bosonic commutation relations

$$[\hat{\phi}(t,\mathbf{x}), \hat{\phi}^\dagger(t,\mathbf{y})] = \delta^3(\mathbf{x}-\mathbf{y}), \qquad [\hat{\phi}(t,\mathbf{x}), \hat{\phi}(t,\mathbf{y})] = 0, \tag{4.39}$$

evolving in time according to the Hamiltonian

$$\hat{H} = \int \mathrm{d}^3\mathbf{x} \left[\hat{\phi}^\dagger(x)\left(-\frac{\nabla^2}{2m}\hat{\phi}(x)\right) + \frac{\lambda}{2}\hat{\phi}^\dagger(x)\hat{\phi}^\dagger(x)\hat{\phi}(x)\hat{\phi}(x)\right], \tag{4.40}$$

that is as

$$i\frac{\partial}{\partial t}\hat{\phi}(x) = \left[\hat{\phi}(x), \hat{H}\right] = -\frac{\nabla^2}{2m}\hat{\phi}(x) + \lambda\hat{\phi}^\dagger(x)\hat{\phi}(x)\hat{\phi}(x). \tag{4.41}$$

Specifically, we have taken a delta-function interaction potential with strength $\lambda > 0$, reflecting elastic collisions of point particles with mass m. See *e.g.* [66] for a derivation and discussion of this description.

Description from Classical Field Theory. We note that this description of a Bose-Einstein condensate can also be derived from a classical action principle for the complex field $\phi(x)=\phi(t,\mathbf{x})$, followed by quantization. More specifically, ϕ is described by the Lagrangian density

$$\mathcal{L} = i\left[(\partial_t\phi)\phi^* - (\partial_t\phi)^*\phi\right] - \frac{1}{m}(\nabla\phi)^*\cdot(\nabla\phi) - \lambda|\phi|^4. \tag{4.42}$$

Note that the first term is $2\mathrm{Im}\,(\phi^* \partial_t \phi)$ and so $\mathcal{L}$ is indeed real. Considering ϕ and ϕ^* as independent fields (as is usually done for complex fields), we find that

$$\frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial(\partial_t \phi^*)} = -i\frac{\partial}{\partial t}\phi, \qquad \nabla \cdot \frac{\partial \mathcal{L}}{\partial(\nabla \phi^*)} = -\frac{1}{m}\nabla^2 \phi, \qquad \frac{\partial \mathcal{L}}{\partial \phi^*} = i\frac{\partial}{\partial t}\phi - 2\lambda\phi|\phi|^2, \tag{4.43}$$

and, as claimed,

$$i\frac{\partial}{\partial t}\phi = -\frac{\nabla^2}{2m}\phi + \lambda\phi|\phi|^2. \tag{4.44}$$

The corresponding Hamiltonian description is hard, because $\partial\mathcal{L}/\partial(\partial_t\phi) = -i\phi^*$ does not contain any time derivatives, making the Legendre transform impossible without further modification. Luckily, we do not need the Hamiltonian description, not even for quantization: we can simply replace the classical field ϕ by a quantum field $\hat{\phi}$, and *impose* the commutation relations (4.39) *manually* to get the right quantum model of our Bose-Einstein condensate. This model will inherit the properties of the underlying classical field, and gain specifically quantum properties due to (4.39). We will take up quantization of the model further down.

Notice that the Lagrangian density is invariant under the global gauge transformation $\phi \rightsquigarrow \mathrm{e}^{\mathrm{i}\alpha}\phi$, with $\alpha \in \mathbb{R}$. This gauge invariance will be responsible for preventing a mass term, although at the moment it looks very different from the gauge invariance employed in the previous section.

Linear Sound: Classical Fluid-Flow Analogue Model. Motivated by the fact that, in the condensate phase, probability densities derived from wave functions become actual densities (since macroscopically, many particles are described by the ground state wave function), we decompose ϕ into a real density $\rho(x)$ and a real phase $\theta(x)$:

$$\phi(x) = \sqrt{\rho(x)}\,\mathrm{e}^{\mathrm{i}\theta(x)}. \tag{4.45}$$

We now find that our complex field equation (4.44) can be written as two real equations:

$$\frac{\partial}{\partial t}\rho + \frac{1}{m}\nabla \cdot (\rho\nabla\theta) = 0, \tag{4.46}$$

$$\frac{1}{m}\frac{\partial}{\partial t}\theta + \frac{1}{m^2}(\nabla\theta)^2 - \frac{1}{2m^2}\frac{1}{\sqrt{\rho}}\nabla^2\sqrt{\rho} + \frac{\lambda}{m}\rho = 0. \tag{4.47}$$

Comparing these equations to the equations (4.15) and (4.17) of fluid dynamics, we can interpret (4.46) as a continuity equation, with density ρ and flow velocity $\mathbf{v} = \nabla\theta/m$, and (4.47) as the corresponding Bernoulli equation (as is done in [45][41]). Indeed, the last two terms together on the left-hand side can be seen as a positive pressure term, monotonously increasing with density (*i.e.* eventually yielding a positive speed of sound), if

$$\lambda\sqrt{\rho}^3 \gg \nabla^2\sqrt{\rho}/2m. \tag{4.48}$$

In order to understand the physical meaning of this regime, we refer back to the classical field equation (4.44), which, in light of eventual quantization, is the Schrödinger equation of the wave functions involved in second quantization. Ignoring the phase contribution $\mathrm{e}^{\mathrm{i}\theta}$ to the wave function, we see that (4.48) holds, and (4.47) really is a Bernoulli equation with physical pressure term, if the interaction term $\lambda\phi|\phi|^3$ dominates the kinetic term $-\nabla^2\phi/2m$ in (4.44); in other words, it holds in the low-energy regime. At low energies we thus expect our Bose-Einstein condensate to emerge through quantization from a classical fluid governed by the continuity equation (4.46) and a Bernoulli equation (4.47) with a physically reasonable pressure term.

[41] The context there is slightly different: they consider directly the equations of the already quantized Bose-Einstein condensate; mathematically, they are however completely analogous to our equations.

We can now essentially repeat the steps of Section 4.2: we split

$$\rho = \rho_0 + \varepsilon\rho_1, \qquad \theta = \theta_0 + \varepsilon\theta_1 \tag{4.49}$$

into background quantities (ρ_0 and θ_0), which themselves satisfy (4.46) and (4.47), and foreground quantities (ρ_1 and θ_1) for small $\varepsilon > 0$. Since the situation is very much analogous to a perfect fluid discussed before, we can immediately conclude that we will observe linear sound propagation in θ_1, governed by a Lagrangian density $\mathcal{L}_{\text{sound}}$ of the form required by Theorem 4.3. The field equations for ρ_1 and θ_1 are similar to the ones obtained for the perfect fluid in Section 4.2, with some differences reflecting the different pressure terms; we will not need their explicit form. The gauge invariance under $\phi(x) \rightsquigarrow e^{i\alpha}\phi(x)$, $\alpha \in \mathbb{R}$, really is the invariance under $\theta(x) \rightsquigarrow \theta(x) + \alpha$, and thus also $\theta_1(x) \rightsquigarrow \theta_1(x) + \alpha$. This is analogous to the gauge invariance under (4.32) of perfect fluids discussed before.

Applying Theorem 4.3 gives us a classical fluid-flow analogue model. Because of the gauge invariance just mentioned, there will be *no mass term*. There is also *no source term*, for the same reasons as in Section 4.2. Gauge invariance yet again intuitively paves the way to a massless analogue gravity model.

Bogoliubov Approximation. The Bose-Einstein condensate is of course obtained only once the classical fluid just discussed is quantized. Most accurately, one would want to quantize the entire fluid (background and foreground), *i.e.* $\rho \rightsquigarrow \hat{\rho}$ and $\theta \rightsquigarrow \hat{\theta}$ with corresponding commutators; we will come back to this approach below.

As an approximation, we can also only quantize the linear sound waves (foreground), *i.e.* $\rho_1 \rightsquigarrow \hat{\rho}_1$ and $\theta_1 \rightsquigarrow \hat{\theta}_1$ with corresponding commutators, while treating the background flow classically. This yields a *quantum fluid-flow analogue model*. This approximation can be understood as a combination of the so-called *Bogoliubov approximation* (otherwise known as *mean-field approximation*) and the assumption of no back-reaction from linear sound onto the background, and of course we always stick to a linear regime.

In the Bogoliubov approximation we first split the field

$$\hat{\phi}(x) = \langle\hat{\phi}(x)\rangle + \varepsilon\delta\hat{\phi}(x) \tag{4.50}$$

into its expectation value, the *mean field* describing the condensate, and quantum fluctuations $\delta\hat{\phi}$ on top. This depends on the quantum state of the Bose-Einstein condensate, of which the field expectation value is then taken. The approximation is then to only consider terms of order at most $\mathcal{O}(\varepsilon)$. Physical observables are (functions of) n-point field expectation values $\langle\hat{\phi}\cdots\hat{\phi}^\dagger\cdots\rangle$; the Bogoliubov approximation is thus exact for *coherent states*, *i.e.* $\hat{\phi}(x)\,|\psi\rangle = \langle\phi(x)\rangle\,|\psi\rangle$, and approximately correct for nearly coherent states.

One then defines $\langle\hat{\phi}(x)\rangle =: \sqrt{\rho_0(x)}e^{i\theta_0(x)}$ and $\delta\hat{\phi} =: \langle\hat{\phi}(x)\rangle\,(\hat{\rho}_1/2\rho_0 + i\hat{\theta}_1)$, see [45]. Upon closer inspection, this is just quantum version of the $\mathcal{O}(\varepsilon)$ expansion of the split (4.49).

The Bogoliubov approximation is however not enough: the equation for $\langle\hat{\phi}\rangle$ is not quite (4.44), since taking the expectation value of (4.44) yields the potential term $\lambda\,\langle\hat{\phi}\hat{\phi}\hat{\phi}^\dagger\rangle$ (up to ordering), which is not equal to $\lambda\,\langle\hat{\phi}\rangle\,|\langle\hat{\phi}\rangle|^2$. If we assume it to be equal, we recover our approximation above; it has the effect of removing the back-reaction of sound onto the background flow [45].

Fully Quantum Model. Instead of only taking sound waves as quantized, we can also quantize the *entire fluid* (including the background flow), essentially arriving at the description (4.40) (4.41). Note that this is not the same quantization procedure as described in Section 2.5, as we did not quantize the background (essentially what gives rise to the metric) there. This approach can be considerably more involved (for instance, the physics in need of quantization may be non-linear) and yields a *fully quantum analogue gravity model*; it is also no longer obvious whether a mass term eventually occurs (it does not) [45] [46]. Models like this have been used in attempts to model aspects of back-reaction of Hawking radiation (in the foreground sound waves) onto spacetime geometry (background flow) [45] [46]. We will briefly return to this in Section 5.

4.4 Continuity and the Difficulty of Schwarzschild Geometry

After introducing classical and quantum fluid-flow analogue models as well as important examples thereof in the previous three Sections, we will now attempt to apply these models to the problem of modelling Schwarzschild spacetime in Gullstrand-Painlevé coordinates (recall Section 1.1), the very metric which inspired the more general fluid-flow metrics (Section 1.3).

We will however quickly reach a roadblock: the continuity equation, inherent in the realistic models discussed in Sections 4.2 and 4.3 *prevents us from modelling Schwarzschild spacetime exactly.* As it turns out, this is a quite general feature of fluid-flow analogue models deriving from an actually flowing medium. This fact will be useful when attempting to answer the initially posed questions in Section 5.

We finish the Section by mentioning some possible approaches to nevertheless come close to a model for Schwarzschild spacetime. Particularly, we introduce our own, novel model, capable of modelling Schwarzschild spacetime with one spatial dimension, at the cost of needing a second, unphysical spatial dimension because of continuity.

Schwarzschild Spacetime in Rectilinear Gullstrand-Painlevé Coordinates. Consider Schwarzschild spacetime in Gullstrand-Painlevé coordinates, as derived in Proposition 1.3. For our purposes it is useful to introduce rectilinear coordinates:

$$X^1 := r\sin\theta\cos\phi, \qquad X^2 := r\sin\theta\sin\phi, \qquad X^3 := r\cos\theta, \tag{4.51}$$

in which the metric (1.20) becomes

$$\mathrm{d}s^2 = -\mathrm{d}T^2 + \sum_{j=1}^{3}\left(\mathrm{d}X^j + \sqrt{\frac{r_s}{r}}\frac{X^j}{r}\mathrm{d}T\right)^2. \tag{4.52}$$

In the notation of Section 1.3, this is a fluid-flow metric with flow velocity $V^j = \sqrt{r_s/r}\cdot X^j/r$, speed of sound $c_s = 1$ and conformal factor $\Theta = 1$.

Attempt at a Model based on a Perfect Fluid. We wish to obtain the metric (4.52), possibly up to a constant, from an analogue model.

Consider for this a physical system exhibiting a perfect, barotropic and irrotational fluid flow, described by a density ρ, a flow velocity $\mathbf{v} = -\nabla\Phi$, speed of sound σ, following the continuity equation (4.15) and Bernoulli equation (4.17). The systems we saw in Sections 4.2 and 4.3 (before quantization) are both of this type.

The system gives rise to an analogue model, with metric (4.37), which we copy here for convenience:

$$g_{\mu\nu} = \left(C\rho_0\cdot\frac{c}{\sigma}\right)^{2/(d-1)}\cdot\left(\begin{array}{c|c}\mathbf{v}_0^2/c^2-\sigma^2/c^2 & -v_{0,j}/c\\ \hline -v_{0,i}/c & \delta_{ij}\end{array}\right). \tag{4.53}$$

Recall that ρ_0 and $\mathbf{v}_0$ is the density and velocity of the background flow, C is a constant needed to get the correct units, and d is the dimension of space.

For (4.53) to be equivalent to (4.52), the space-space part of (4.53) can have at most a constant prefactor, and thus

$$\left(C\rho_0 \cdot \frac{c}{\sigma}\right)^{2/(d-1)} = \text{const.} \tag{4.54}$$

For $d > 1$, in particular for the case $d = 3$ relevant here, this implies $\rho_0/\sigma = \text{const.}$ To get the correct prefactor also in the time-time part would then require $c = \sigma$. This is hardly reasonable, but we can rescale the spatial coordinates X^j (or equivalently the Gullstrand-Painlevé time coordinate T) to achieve $c/\sigma = \text{const.} > 1$ (or even $c/\sigma \gg 1$). Irrespective of rescaling, we thus have $\sigma = \text{const.}$ Finally, in order for the time-space part of the metrics to match up, we must have

$$v_{0,j} = -\tilde{c}\sqrt{\frac{r_s}{r}}\frac{X^j}{r}, \tag{4.55}$$

where $\tilde{c} < c$ is a velocity arising due to the rescaling of X^j. Reverting to spherical coordinates, this is a radial flow with radial velocity $v_{0,r} = -\tilde{c}\sqrt{r_s/r}$.

In conclusion, we need a radial flow with velocity $\propto -\sqrt{r_s/r}$, constant density ρ_0 and constant speed of sound σ. This flow must satisfy the continuity equation (4.15), and since $\rho_0 = \text{const.}$ (both in space and time), we must have

$$\nabla \cdot \mathbf{v}_0 \overset{!}{=} 0. \tag{4.56}$$

However,

$$\nabla \cdot \mathbf{v}_0 = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 v_r) = -\tilde{c}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\sqrt{r_s/r}\right) \neq 0. \tag{4.57}$$

Thus, we cannot model the metric (4.52) with a perfect, barotropic and irrotational fluid, because the continuity equation cannot be satisfied.

Schwarzschild spacetime in Gullstrand-Painlevé coordinates can only be modelled using perfect, barotropic, and irrotational fluids up to a *non-trivial conformal factor* [7]. If the exact geometry of the spacetime is not needed, then this is not a concern. For instance, surface gravity and thus Hawking temperature is invariant under conformal rescaling of spacetime [7].

It is perhaps interesting to note that if we analytically extend the divergence of radial vector fields to any real dimension d, *i.e.* $\nabla \cdot \big(f(r)\mathbf{e}_r\big) = r^{-(d-1)}\partial(r^{d-1}f(r))/\partial r$, then an exact model of the radial Schwarzschild flow is possible for $d = 3/2$.

More Elaborate Approaches. Perhaps this is only an issue of the considered fluid system being too simple and that it can be remedied with more elaborate models; but as it turns out, because the continuity equation is rather fundamental, this is not at all straightforward. We quote here a few results.

One approach is to exploit the freedom of choosing coordinates before attempting to find a corresponding analogue model; with this approach Schwarzschild spacetime has been exactly modelled (although not in Gullstrand-Painlevé coordinates) [77] [21].

Another approach is to consider so-called *non-isentropic fluids*, essentially introducing the possibility of particle creation and destruction; roughly speaking, this allows for modification of the continuity equation [13]. Research in this area has been successful at modelling Schwarzschild spacetime, even in Gullstrand-Painlevé coordinates; one however needs a relativistic fluid whose flow velocity approaches the speed of light at the horizon [14]; even when introducing an external pressure gradient, the fluid has to essentially remain relativistic for the model to work [12].

Ansatz in Higher Dimensions. Another (as far as we can tell, novel) idea is to model a lower-dimensional version of Schwarzschild spacetime, within a subset of a higher-dimensional model; the additional dimensions could then be used for the sole purpose of satisfying the continuity equation.

Such an approach could be of interest in a laboratory setting where one would like to model Schwarzschild spacetime in Gullstrand-Painlevé coordinates, but one only has access to non-relativistic and isentropic fluids. The approach is detailed in Appendix C; we will however not need it for our discussion.

5 Analogue Gravity and the Information Loss Paradox

We have motivated and introduced analogue gravity models in Section 1, discussed black holes and their features in Section 2, seen the information loss paradox in Section 3, and properly treated analogue gravity models in Section 4. We are now ready to tackle the questions posed in the introduction:

Question 5.1

Can we learn about the black hole information loss paradox from analogue gravity models?

Question 5.2

Can we infer anything at all about gravity from analogue gravity models?

We begin by noting that these questions are both extremely broad. In particular, we have not precisely defined what an "analogue gravity model" is, leaving open the possibility of more general models than the ones described here. And of course, it would make little sense to restrict the meaning and potentially leave out interesting models.[42] We will therefore certainly not be able to answer the questions here.

In the following sections we will instead content ourselves with providing evidence for *possible* answers to these questions. In Section 5.1 we tackle Question 5.1. We argue that the difficulty in obtaining a useful notion of black hole entropy in the context of analogue models makes the discussion of the black hole information loss paradox in analogue models equally difficult, because such a notion is required. We will see that the main issue is the general lack of Einsteinian dynamics in analogue models. In Section 5.2 we discuss how even complex analogue models with favourable features (in particular masslessness and the occurrence of Hawking radiation) are expected to exist based on simple reasoning. We will further argue based on examples in the literature that one should expect that with enough (potentially very tedious) effort and complexity put into the model, one could perhaps obtain an analogue model with any favourable features, describing gravity arbitrarily well. All in all, we will gather evidence for the hypothesis that *analogue models do not have a very deep connection to gravity, but instead truly are analogies, that is, simply (well-understood) occurrences of the same mathematics in different places of physics.*

We will conclude that both questions are, at least for now, best answered with "no". Question 5.1 more clearly so, since a vital piece, analogue black hole entropy, is missing. Our tentative answer to Question 5.2 is less certain, not least due to the broader scope of the question.

We would like to stress that *these answers do not discredit analogue models.* It simply means that we have no guarantee of finding features in gravity, if they have been observed before in analogue models; but of course one may still search for features in gravity which were found in analogue models. If corresponding features have been found in both gravity and analogue models, then we may gain valuable insight about one by looking at the other; unsurprisingly, this is one of the main ways in which analogue models of gravity are being used [7]. For instance, the very first objective pursued in analogue gravity, understanding how Hawking radiation would react to the high-energy breakdown of the wave equation (a research program started by Unruh [74], lasting until today [7]; essentially, realistic

[42] Technically one could argue that anything between the simplest classical fluid-flow analogue model and general relativity itself, or even quantum gravity (if it exists) would qualify as an analogue model. This clearly undermines the point of both questions, so we will not be as pedantic with the meaning of "analogue model". For us, it will be sufficient to assume that there is somewhere an arbitrary and unknown line distinguishing analogue models of gravity from gravity itself.

fluids behave very differently at high energies, which manifests first in modified dispersion relations, and can end in an entirely different regime of physics, where the wave equation possibly no longer holds), falls into this category: such breakdown naturally occurs and is understood in analogue models, and is expected in theories of quantum gravity; the exact shape of the breakdown is of course not known, but this is often kept in mind. Finally, the experimental interest in being able to simulate features of curved spacetime in a laboratory through analogue models is independent of the answers to these two questions. See *e.g.* [37] for both past and future experimental efforts.

5.1 The Missing Piece: Black Hole Entropy

We saw in Section 3.4 that for the information loss paradox to *potentially* occur we need Hawking radiation and *a notion of black hole entropy*; in particular we do not need a notion of back-reaction. So if we want to talk about the black hole information paradox, or even the possibility of it occurring at all, in the context of analogue gravity, we need to understand Hawking radiation and black hole entropy in that context too.

Hawking radiation is known to occur in analogue models (see Section 2.5). We will discuss here how one might approach identifying black hole entropy in analogue models, which we will call *analogue black hole entropy* in order to distinguish it from actual black hole entropy. We will see that none of the approaches are truly satisfying or even useful, and that this is mainly due to the lack of Einsteinian dynamics in analogue models. Thus, it will be very hard if not impossible, to even sensibly talk about the black hole information loss paradox in the context of analogue models.

Features of Black Hole Entropy. Let us quickly summarize the features of black hole entropy S. We saw in Section 2 that:

1. The area $A(H)$ of the event horizon satisfies the area theorem, Theorem 2.6 (Section 2.2).
2. If we define $S = f(A(H))$ for a monotonously increasing function f, then S, together with other thermodynamic-like quantities such as total mass as energy and surface gravity as temperature, satisfies also a zeroth, first, and third law (Section 2.3).
3. S can also be motivated information-theoretically, providing a very good estimate for f (Section 2.4).
4. With the temperature of black holes fixed through Hawking radiation (Section 2.5), f is also fixed.

Ideally, analogue black hole entropy should have as many of those features as possible.

Analogue Black Hole Entropy from the Analogue System? If the physical system underlying the analogue model (*e.g.* a fluid) has a notion of entropy (or entropy density in the case of a fluid [39]), then one can attempt to define the analogue black hole entropy S_A as simply the entropy of the physical system, including sound waves inside the system, within the black hole region.

In the case of a fluid-flow analogue model, fluid and thus entropy is constantly flowing into the black hole region, increasing the analogue black hole entropy even if the black hole remains stationary, and thus has a constant horizon area $A(H)$. Thus, it cannot hold that $S_A = f_A(A(H))$ for some monotonously increasing f_A. Not just is this very far from the case of actual black holes in general relativity, it also means that the paradox might not even occur, since both the analogue black hole entropy and the Hawking radiation entropy are increasing.

Furthermore, it is not clear whether two systems with different entropy densities cannot give rise to exactly the same analogue model, thus rendering the interpretation of system entropy as black hole entropy difficult. For instance, the models discussed in Section 4.1 are completely indifferent to the microscopic properties of the fluid and hence its entropy density.

These considerations show that defining analogue black hole entropy via notions of entropy already present in the analogue model is not particularly useful. Rather, we have to find a notion of analogue black hole entropy which does not necessarily depend on the entropy notions already present and which better matches the black hole entropy of general relativity. In fact, if a successful candidate for analogue black hole entropy S_A were to be found, the differences between S_A and other entropies naturally present in the analogue system could give valuable insight into the black hole information loss paradox.

Analogue Black Hole Entropy from Black Hole Dynamics? Another approach is to see whether the four laws of black hole dynamics (Sections 2.2 and 2.3) also hold in (some) analogue models, and to define analogue black hole entropy in this way.

If the area theorem, Theorem 2.6, holds in a given analogue gravity model, we could tentatively define analogue black hole entropy as $S_A := f_A(A(H))$, with f_A an unknown, monotonously increasing function. But the area law is hardly enough to relate $A(H)$ to entropy (see our list above). We need a quantitative reason to call S_A an entropy: we need to find the explicit form of f_A.

f_A could be fixed through the derivation of Hawking radiation, as it was done for black holes in regular general relativity, and luckily, Hawking radiation occurs in analogue models. But recalling the arguments towards the end of Section 2.5, this identification of f_A only works if S_A also satisfies at least the first law (see Section 2.3).

The first law of black hole dynamics in the form of Theorem 2.7 is a statement about the relations between differential changes of properties of Kerr-Newman black holes, which are recognized as black holes in equilibrium through the no-hair-theorem (see Section 2.3 and Appendix A). The more general form of the first law [8] also allows for external matter.

If we are to identify something resembling a first law in analogue models, we also need a notion of equilibrium black holes. Since the no-hair-theorem requires the Einstein field equations (after all, the Kerr-Newman spacetimes are solutions of those), and because analogue models do generally not have Einsteinian dynamics (see the end of Section 4.1), this approach to a first law is highly unlikely to work in the context of an analogue model.[43]

Another approach is to instead embrace the dynamics of the analogue model and try to build a thermodynamic picture from them. But these are then simply thermodynamics of the analogue system, with no guaranteed connection at all to black holes, limiting the usefulness of such an approach. Furthermore, is not clear whether such an approach can even yield thermodynamics.

Finally, the zeroth law allows us to define a surface gravity for some types of black holes, which are not necessarily solutions to the Einstein field equations; in fact, the Einstein field equations are not needed anywhere in the derivation of the law (see Section 2.3). It could therefore hold for certain analogue gravity models. It is indeed easy to construct analogue gravity models with an event horizon on which the surface gravity κ is constant. But without the first law, the zeroth law does not help us identify black hole entropy.

To sum up, defining analogue black hole entropy through the dynamics of analogue black holes is not as effective as for black holes in general relativity, due to large differences in the dynamics between analogue models and general relativity: importantly, the Einstein field

[43] As we have noted in Section 4.1, there are rare exceptions which however come with their own problems. And even with these exceptions, our argument still holds quite generally.

equations are generally missing in analogue models.

Analogue Bekenstein Entropy? A completely different way towards an analogue black hole entropy could be possible through extending Bekenstein entropy (Section 2.4) to analogue models. But as with the first law, this requires the Einstein field equations, as we have seen. So this does not seem to be an option either.

Traces of a Solution to the Paradox? Before closing the discussion of Question 5.1, let us mention another approach to answering it: instead of trying to establish a context in analogue models where one can talk about the paradox (an endeavour which after these paragraphs seems almost hopeless), we can investigate whether certain aspects of potential solutions to the paradox can be found within analogue models, deferring the search for the analogy of the paradox.

As an example, let us consider the analogue models described in [45] [46]. There, quantized linear sound waves in a Bose-Einstein condensate are considered, including interaction between the background flow and the sound waves. It is found that as emission of Hawking radiation proceeds, entanglement between the background and linear sound builds; this is of course possible because in a Bose-Einstein condensate we may also treat the background flow, in particular the condensate, as quantum. This entanglement between "geometrical degrees of freedom" (those of the background flow) is argued to be expected of quantum gravity theories [45]. We would expect the same from any system obeying the Page curve. Furthermore, a form of back-reaction could be demonstrated [46]: the condensate is slowly *depleted* due to emission of Hawking radiation.

So do these models exhibit parts of a possible solution? While entanglement between geometry and radiation in an analogue model is remarkable, and models a situation that one would expect in quantum gravity, we believe that one has to be careful with such a conclusion.[44] Firstly, the depletion of the condensate can only continue as long as the condensate is still occupied. Now the analogue metric could still contain a black hole even when the condensate is completely unoccupied; this suggests that back-reaction could look very different depending on whether the condensate is occupied. This is not a feature we immediately expect from actual gravity, since there is no analogue of the condensate. Finally, more advanced features such as the entanglement wedge of replica wormholes (recall Section 3.3) remain completely out of reach even for these models.

Advanced models such as these raise the question of whether analogue models have any deep connections to gravity, if it is always possible to obtain models with more features simply by making the model system more complicated. We will come back to this question in the next section.

Conclusion. As it seems, there is no known and satisfying way of defining analogue black hole entropy (in the sense of our list of desirable features such an entropy should have), and consequently, it is not easily possible to talk about whether the information loss paradox even occurs in the context of analogue gravity. Furthermore, while some features of a potential solution to the paradox (namely entanglement between geometry and radiation, as is implied by the replica wormhole approach to the paradox) can be found in some analogue models, these features are quite generic and expected to occur if both the scalar field and the background geometry are of quantum nature. We therefore answer Question 5.1 with a "no" at this time.

[44] It is important to note that the authors of [45] [46] never make such claims or even discuss the present question; we simply use their models as an example for our discussion. We use their models because they are relatively accessible yet advanced.

5.2 Analogue Gravity Models are Simply Models

After answering Question 5.1 preliminarily with "no" in the previous section, let us now turn to Question 5.2. We will attempt to argue that analogue models should really be seen as what their name implies: that they are just *models* of gravitational physics, *i.e.* systems whose mathematics resembles that of certain aspects of gravity, but which do otherwise not have a profound connection to gravity.

The Price of Analogue Gravity Models. Analogue models are physical systems themselves and must thus follow their own physical laws, which might have nothing to do with gravity. The properties of models obtained from these laws then may not help us in modelling gravity, they are "extra baggage" our model carries. In the worst case, such extra baggage, required for the model itself, even hinders the description of gravity. We have seen a striking example of this in Section 4.4: due to the continuity equation which fluid-flow models have to fulfil, it is hard to model the Schwarzschild metric exactly (one has to leverage coordinate transformations [77] [21]), arguably the most important black hole metric in general relativity. Looking at the wealth of analogue models known today, the existence of such "extra baggage" is indeed a general trend; see for instance [7, Section 4].

One could argue that describing gravity is far from the most natural use of the analogue systems. Cynically, one could even claim that analogue models are just another way to describe aspects of gravity (Klein-Gordon field dynamics and Hawking radiation), and a worse one than classical gravity at that, since for one, not all aspects of gravity are covered, and models contain additional baggage, useless for the description of gravity. We do not take such a radical perspective here, but simply acknowledge the price paid in inconveniences when describing gravity using analogue models; we take this as evidence towards our hypothesis that analogue models likely do not possess a deep connection to gravity.

The Ubiquity of Analogue Models. Another argument for our hypothesis comes from the relative ease with which one can find analogue models. As we have seen in Section 4.1, and then again in the Sections 4.2 and 4.3, at least the simplest analogue models (which also often serve as basis for more complicated models [7]) are not rare at all, but are to be expected in all corners of physics. Concretely, we saw that a hyperbolic PDE coming from an action principle, a very common sight in physics, was enough to obtain an analogue model. Furthermore, with little gauge invariance (which is present in many systems, in particular fluid systems) such models even become massless. Especially the relatively small effort with which one obtains massless analogue models of gravity is surprising, since masslessness at first seems like a very specific and deep property a model can have.[45]

Understanding why certain types of analogue models occur very often in physics "demystifies" their connection to gravity: we see that these models are well-understood cases of the same mathematics occurring in different parts of physics. In this sense, the connection to gravity of such models is not a deep one. In particular, it is very well possible (and as the previous section has shown for the example of black hole information loss) that the analogy breaks down as one tries to extend it further, *i.e.* that the mathematics again become different in the two parts of physics.

The Generality of Analogue Models. We have seen how it is relatively easy to construct massless analogue models, albeit with additional baggage imposed by the physics of the analogue system. Accepting further such baggage, we can of course always make the model more complex, in order for it to exhibit more features of gravity. For instance, the

[45] For instance, there are very profound differences between massless quantum field theories and massive quantum field theories with a very small mass [65].

models described in [78] are very complex, but also allow for more features (famously, they even contain a trace of Einsteinian dynamics).[46]

Thus, in many cases we should also not be surprised to find advanced features of gravity beyond scalar field propagation and Hawking radiation in analogue models, if the analogue model is complex enough.

Conclusion. As it seems, analogue models, even with complex features are to be expected. They seem to have little deep connections with gravity itself, and we should rather see them really as *models* of gravity. This suggests that for the time being, analogue models cannot make deep predictions about gravity, and we should answer Question 5.2 with "no" for now. A similar conclusion has been drawn on more philosophical grounds [20].

Finally, we stress again that these analogies are still very powerful: once established *on both sides*, they can be used to transfer knowledge from one part of physics to the other. Answering Question 5.2 with "no" only means that (without further assumptions) analogue models should not be expected to predict yet unknown features of gravity.

[46]We must mention that many parts of these models are incredibly elegant and interesting. They are not to be dismissed simply because they are complicated.

6 Conclusion and Outlook

We have provided an introduction to the black hole information loss paradox and to the field of analogue gravity models, before studying the possible applications of analogue models to the paradox.

We have found that a crucial ingredient for the paradox, black hole entropy, has yet not been identified in analogue models and that there is no clear way towards such an identification.

Furthermore, we investigated the connection between analogue gravity models and gravity itself. We argued that analogue models can be perfectly understood as mathematical similarities between gravity and analogue systems, and that we should not expect a much deeper connection between analogue systems and gravity.

We have thus answered the two originally posed questions preliminarily with "no". This (especially the second "no") however does not mean that analogue models are useless; but rather, that we should simply treat them as models, as which they can be (and have been [7]) immensely useful. A quick web search shows that the field of analogue gravity is currently very active. It will thus be interesting to watch future developments; perhaps the limitations leading to "no" in this work can be eventually overcome with yet unknown ways and analogue systems, allowing us to model black hole information loss (or preservation) in analogue systems. Until then, the main barrier seems to be the difficulty of obtaining dynamics for the analogue model which are comparable to Einstein field equations.

Besides this main result we have also obtained other results along the way. Firstly, the approach of constructing analogue models with vanishing mass terms in 4.2 and 4.3 using gauge invariance is novel, as far as we can tell.[47] This allowed for an intuitive introduction to two of the most important analogue models. It would be interesting to see how far this approach can be extended in order to intuitively derive other analogue models, or whether variations of the approach can be used to generate other useful features of models. Secondly, the derivation of Hawking radiation in Section 2.5 and Appendix B is a new mixture of more traditional canonical quantization approaches, *e.g.* [31] and [69], and the modern and minimalist approach [76]. Since it focuses on the essential ingredients of Hawking radiation (following [76]) but still contains the extensive quantization explanations of traditional derivations, we believe that it could serve as a pedagogically valuable first derivation of Hawking radiation. Thirdly, the new analogue model discussed in Appendix C, while not being very important for our main argument, is interesting in its own right.

Let us close with a more adventurous piece of outlook. We have mentioned in Section 1.3 that fluid-flow metrics are a special case of the forms of metrics encountered in the ADM initial value formalism of general relativity. The ADM metrics are more general, concretely allowing for a non-Euclidean spatial part [3] [49, Chapter 21], but one may nevertheless wonder whether some interpretation of spacetime as a flowing fluid is also possible in the general case. Unfortunately, this avenue of research was quite outside the scope of this project, and we thus did not pursue it further.

[47] Although the contents of Theorem 4.3 has been known before [6].

A Details on the Zeroth and First Law of Black Hole Dynamics

We provide here some technical details to complement the surface-level discussion in Section 2.3.

Stationary Electrovac Black Holes. An asymptotically flat spacetime is said to be *stationary* if it admits an asymptotically timelike Killing vector field ξ^a, *i.e.* a Killing vector field that becomes timelike in the asymptotic limit.[48] Intuitively, stationary spacetimes are "time-independent" with respect to the "time" standard provided by the affine parameter of integral curves of ξ^a. In the asymptotic region (at least in the asymptotic limit) ξ^a becomes timelike and its integral curves are worldlines of observers, the so-called *stationary observers*.

An asymptotically flat spacetime is called *axisymmetric* if it possesses an asymptotically spacelike Killing vector field ψ^a, whose integral curves are closed (the "circles of revolution").

As we have seen in Section 2.3, stationary, electrovac spacetimes are also axisymmetric and fully described by the Kerr-Newman metric, parametrized by M, J and Q. We will explain the physical meaning of these parameters below.

The Killing vector fields ξ^a and ψ^a are not unique. It is however possible to choose ξ^a to be orthogonal to some spacelike hypersurface in the asymptotic limit (we say such a ξ^a is *asymptotically hypersurface orthogonal*), and ψ^a to lie in that hypersurface in the same limit. Intuitively, this choice prevents the observers following the integral curves of ξ^a from rotating around the black hole's symmetry axis, making them *static observers*. It also fixes the Killing vector fields at least in the asymptotic limit up to normalization (which we may take to be ± 1). We will assume this choice from now on. In the Schwarzschild case the choice makes ξ^a hypersurface orthogonal everywhere outside the horizon; consequently, in Schwarzschild geometry, static observers exist everywhere outside the horizon, not just at infinity.

Finally, one can show [33] that any asymptotically flat, stationary, electrovac spacetime also possesses a Killing vector field χ^a parallel to the horizon generators and that can be obtained as a linear combination of ξ^a and ψ^a:

$$\chi^a = \xi^a + \Omega_H \psi^a. \tag{A.1}$$

Ω_H is called the *angular velocity of the horizon*.

Kerr-Newman Metric. It is particularly useful to work in *Boyer-Lindquist coordinates* (t, r, θ, ϕ), first employed by Boyer and Lindquist [16] for the $Q = 0$ case (the so-called *Kerr metric*). In these coordinates, the Kerr-Newman metric reads [81, Section 12.3]

$$\begin{aligned} \mathrm{d}s^2 = -\left(\frac{\Delta - a^2 \sin^2\theta}{\rho^2}\right)\mathrm{d}t^2 - \frac{2a\sin^2\theta\,(r^2+a^2-\Delta)}{\rho^2}\,\mathrm{d}t\,\mathrm{d}\phi \\ + \left[\frac{(r^2+a^2)^2 - \Delta a^2 \sin^2\theta}{\rho^2}\right]\sin^2\theta\,\mathrm{d}\phi^2 + \frac{\Sigma}{\Delta}\mathrm{d}r^2 + \rho^2\,\mathrm{d}\theta^2, \end{aligned} \tag{A.2}$$

where $a := J/M$ is the *spin parameter* and

$$\rho^2 := r^2 + a^2\cos^2\theta, \qquad \Delta := r^2 + a^2 + Q^2 - 2Mr. \tag{A.3}$$

[48] Schwarzschild spacetime is for instance stationary, with $\xi^a = (\partial_t)^a$ the Schwarzschild time coordinate vector field. This vector field is timelike outside the horizon, but becomes null on the horizon and spacelike inside the black hole; hence it is in particular timelike in the asymptotic limit, as required.

The potential one-form of the corresponding electromagnetic field is

$$A = -\frac{Qr}{\rho^2}\left(\mathrm{d}t - a\,\sin^2\theta\,\mathrm{d}\phi\right). \tag{A.4}$$

The event horizon is located at

$$r_+ = M + \sqrt{M^2 - a^2 - Q^2}. \tag{A.5}$$

It only exists if

$$M^2 \geq J^2/M^2 + Q^2. \tag{A.6}$$

To prevent naked singularities, we always assume (A.6). The angular velocity of the event horizon is [81, Section 12.3]

$$\Omega_H = \frac{J}{r_+^2 + a^2}. \tag{A.7}$$

One can show that for the horizon null congruence expansion, we have $\theta = 0$ everywhere; tracing out the steps in the proof of the area theorem 2.6 shows that the event horizon area is *independent* of the choice of spacelike hypersurface Σ for Kerr-Newman black holes. In particular, the area theorem holds with an equality. A particularly simple choice of Σ is a surface with $t = \text{const}$ (although this surface is otherwise unnatural, since it becomes null at the horizon); it yields the event horizon area

$$A(H) = \int_{r=r_+} \sqrt{g_{\theta\theta}g_{\phi\phi}}\,\mathrm{d}\theta\,\mathrm{d}\phi = 4\pi\cdot(r_+^2 + a^2). \tag{A.8}$$

Taking a closer look at (A.2) we see that the coordinates are adapted to the asymptotic flatness and approach spherical coordinates of flat spacetime at large values of r. Furthermore, the stationary Killing vector field is provided by $\xi^a = (\partial_t)^a$ and the axisymmetric Killing vector field by $\psi^a = (\partial_\phi)^a$. ξ^a is hypersurface orthogonal and perpendicular to ψ^a at infinity as described above. ξ^a becomes timelike for $r > r_E$, where

$$r_E(\theta) = M + \sqrt{M^2 - a^2\,\cos^2\theta} \tag{A.9}$$

is the *static limit*. Stationary observers can only exist at $r > r_E$. For $J = 0$, $Q = 0$ the metric becomes the Schwarzschild metric, $r_+ = 2M$ becomes the Schwarzschild radius, and $A = 4\pi\cdot(2M)^2$, as expected.

Mass, Angular Momentum and Charge. As explained in Section 2.3, we use static observers to define M, J and Q. These observers are completely determined by the above choice for the Killing vector fields ξ^a and ψ^a. So the quantities M, J and Q must be expressible through these vector fields, the metric g_{ab} (for M and J, since these can be read off the metric in the linear regime) and the electromagnetic field strength F^{ab} (for Q), in a *coordinate-independent fashion*; for the necessary choice of coordinates is already contained in the choice of the Killing vector fields. We thus define [81, Sections 11.2, 12.3]:

Definition A.1: Energy, Angular Momentum and Charge

Consider an asymptotically flat, stationary and axisymmetric spacetime, with Killing vector fields ξ^a and ψ^a. Let $S \subset \Sigma$ be a two-dimensional spacelike hypersurface homeomorphic to the two-sphere, contained in an asymptotically flat spacelike Cauchy surface Σ.

The *total energy* E, *total angular momentum* J and *total electric charge* Q enclosed by S are defined by

$$E := -\frac{1}{8\pi}\int_S \epsilon_{abcd}\nabla^c\xi^d, \tag{A.10}$$

$$J := \frac{1}{16\pi}\int_S \epsilon_{abcd}\nabla^c\psi^d, \tag{A.11}$$

$$Q := \frac{1}{8\pi}\int_S \epsilon_{abcd}F^{cd}. \tag{A.12}$$

If the stationary Killing vector field ξ^a is hypersurface orthogonal in the asymptotic limit, we call E the *total mass enclosed by* S and denote it by M.

It can be shown (see [81, Section 11.2]) that these reduce to the values one would expect from the linear regime, in cases where this regime applies to all of spacetime; they are thus physically reasonable. For example, the energy E can be obtained in the linear regime by integrating the acceleration $a^b = (\xi^c/V)\nabla_c(\xi^b/V)$ felt by static observers over S, where $V = \sqrt{-\xi_a\xi^a}$ is the redshift factor between the observer in question and observers at infinity; this is essentially the generalization of Gauss' law of Newtonian gravity to the linear regime of general relativity. More precisely, we consider *static* observers suspended from ropes attached at infinity, such that we can measure the required acceleration at infinity. This has the effect of introducing an additional redshift factor V into the integrand, and it leads to a definition of energy independent of S in vacuum black hole spacetimes (as long as the event horizon is enclosed by S). Since ξ^a might not be hypersurface orthogonal at S, we generalize the definition of energy and use *stationary* observers instead (which become static once ξ^a is hypersurface orthogonal).

Note that the definition of E really only requires a stationary spacetime, J requires only an axisymmetric spacetime, and Q requires neither. Also, these definitions do not require vacuum or electrovac spacetimes. Therefore, they apply to more spacetimes than just the Kerr-Newman family of spacetimes.

A few explicit computations in the Kerr-Newman metric with ξ^a asymptotically hypersurface orthogonal to some spacelike hypersurface Σ (*i.e.* the standard of "space" set by Σ and the standard of "time" set there by ξ^a at infinity are compatible) show that the three quantities $M = E$, J and Q of Definition A.1 are precisely the three parameters of the Kerr-Newman metric [81, Section 12.3], hence the symbols and names used for the parameters. In particular, the quantities are *independent of* S, as long as S encloses the event horizon (*i.e.* encloses $H \cap \Sigma$). Thus, we do not have to worry about there being no static observers at finite distances, since we can always use the case where S approaches infinity to recover the physical meaning of M, J and Q.

Using *Stokes' Theorem*, the integral in (A.10) may be converted into an integral over the three-dimensional volume $\mathrm{int}(S) - B \subset \Sigma$ laying between horizon and S, and an integral over the intersection $H \cap \Sigma$ of the horizon with Σ. Using the general result $\nabla_a\nabla^a v_b = -R_b{}^c v_c$ valid for Killing vector fields v^a, as well as the Einstein field equations (0.1), allows us to write

$$E = \frac{1}{4\pi}\int_{\mathrm{int}(S)-B}\left(T^a{}_b - \frac{1}{2}T^f{}_f\,\delta^a{}_b\right)\xi^b\,\epsilon_{acde} - \frac{1}{8\pi}\int_{H\cap\Sigma}\epsilon_{abcd}\nabla^c\xi^d, \tag{A.13}$$

The first term corresponds to the contribution of the energy-momentum tensor to E, and it looks like the expression one would expect from linearized gravity [81, Exercise 11.5]. It vanishes in vacuum spacetimes. The second term can be seen as a contribution coming purely from the black hole; we denote it by E_H. Contrary to the first term, we did not use the Einstein field equations to write it down. See [81, Sections 11.2 & 12.5].

Similarly, we can split J into a contribution from T_{ab} and a contribution

$$J_H = \frac{1}{16\pi} \int_{H\cap\Sigma} \epsilon_{abcd} \nabla^c \psi^d \tag{A.14}$$

from the horizon; see [8].

Zeroth Law of Black Hole Dynamics. Rewriting (A.13) with $\chi^a = \xi^a + \Omega_H \psi^a$, we get

$$E = \frac{1}{4\pi} \int_{\mathrm{int}(S)-B} \left(T^a{}_b - \frac{1}{2} T^f{}_f \, \delta^a{}_b \right) \xi^b \epsilon_{acde} + 2\,\Omega_H J_H - \frac{1}{8\pi} \int_{H\cap\Sigma} \epsilon_{abcd} \nabla^c \chi^d. \tag{A.15}$$

Note that $\epsilon_{abcd} = \epsilon_{ab} N_{[c}\chi_{d]}$ and thus $\epsilon_{abcd}\nabla^c\psi^d = 2\epsilon_{ab}N_c\chi_d\nabla^c\chi^d = \epsilon_{ab}N_c\nabla^c(\chi_d\chi^d)$, where we have used Killing's equation $\nabla_b\chi_a = \nabla_a\chi_b$.

$\chi_d\chi^d$ is zero everywhere on the horizon and hence $\nabla_b(\chi_a\chi^a)$ is normal to the horizon; it must therefore be proportional to χ^b. See [81, Section 12.5]. We can thus define:

Definition A.2: Surface Gravity

Let H be the event horizon in an (asymptotically flat) spacetime with the property that there exists a Killing vector field χ^a which is tangent to the horizon generators on H. In particular, $\nabla_b(\chi_b\chi^a)$ is proportional to χ_b.

The *surface gravity* κ is the function defined on H such that

$$\nabla_b(\chi_a\chi^a) = -2\kappa\chi_b. \tag{A.16}$$

Note that stationarity and axisymmetry is not necessary for this definition, but they together imply the existence of χ^a.

With N^a given (*i.e.* with a choice of Σ), the surface gravity is

$$\kappa = -N^a\chi^b\nabla_a\chi_b, \tag{A.17}$$

see [8]. And for a Kerr-Newman black hole, the surface gravity is given by

$$\kappa = \frac{\sqrt{M^2 - a^2 - Q^2}}{2M(M + \sqrt{M^2 - a^2 - Q^2}) - Q^2}, \tag{A.18}$$

in particular it is *constant on the horizon.* More generally, the surface gravity satisfies the following theorem:

Theorem A.3: (Bardeen, Carter, Hawking) Zeroth Law of Black Hole Dynamics

For a stationary black hole spacetime satisfying the condition

$$\chi_{[\sigma}R_{\rho]}{}^{\mu}\chi_\mu = 0 \tag{A.19}$$

on the horizon, the surface gravity κ is constant on the horizon.

For a proof starting with the Definition (A.16), see [81, Section 12.5]; the original proof [8] instead starts with (A.17). Both proofs make use of the fact that χ^a is a Killing vector and thus satisfies $\nabla_a\nabla_b\chi_c = -R_{bcad}\chi^d$. Both proofs conclude by requiring a similar condition on the Ricci tensor on the horizon; we will discuss this condition further down. Importantly, the Einstein field equations have not been used in the proof.

In the Schwarzschild case, ξ^a is hypersurface orthogonal everywhere outside the event horizon; this allows for the intuitive definition of surface gravity as a force felt on a string left dangling towards the black hole, which we mentioned in Section 2.3. See [81, Section 12.5].

Finally, let us make some observations:

1. Like the second law, the zeroth law in the form of Theorem A.3 does not require the Einstein field equations.
2. By using the Einstein field equations, we can however transform the rather obscure condition (A.19) into a more meaningful energy condition. One can show [8] that the *dominant energy condition*

$$\begin{pmatrix} T_{ab}v^a v^b \geq 0 \quad \text{and} \quad -T^a{}_a v^a \\ \text{future-directed, timelike or null} \end{pmatrix} \qquad \text{for all timelike vectors } v, \tag{A.20}$$

   implies (A.19) via continuity. Note that the dominant energy condition implies the weak energy condition (2.11). $-T^b{}_a v^a$ is the four-momentum locally seen by an observer with four-velocity v^a. The strong energy condition thus extends the weak energy condition by assuming that observers always measure future-directed, timelike or lightlike four-momenta. This is reasonable in almost all situations. Since the process of Hawking radiation violates the weak energy condition as mentioned above, it also violates the dominant energy condition.

First Law of Black Hole Dynamics. Assume that our black hole is stationary, axisymmetric and such that the zeroth law holds. The last term in Equation (A.13) can be rewritten in terms of the surface gravity κ. By varying the equation one may (after some involved computations) obtain a relation of differentials similar to the *first law of thermodynamics* (see *e.g.* [40]). The result thus obtained by Bardeen, Carter and Hawking [8] works for all stationary, axisymmetric black holes, with the zeroth law holding, and with energy-momentum tensor in the form of a perfect fluid.

We will not need this generality and instead focus only on Kerr-Newman black holes, where $J = J_H$. For this it is easiest to vary instead equation (A.8) and to solve for $\mathrm{d}M$; after a short calculation we thus obtain [9] the first law of black hole dynamics, Theorem 2.7.

The zeroth and first law of thermodynamics are compatible in that they both apply to the same set of systems: those in equilibrium. Similar parallels exist between the zeroth, first and second law of black hole dynamics: if the first law applies to a black hole spacetime, then so does the zeroth law; and the entropy identified in the first law would be proportional to the horizon area $A(H)$, which is precisely the object described by the second law of black hole dynamics.

B Details on the Derivation of Hawking Radiation

We gather here details of some computations left out in Section 2.5.

Spherical Symmetry. Since the metric (1.31) is spherically symmetric, with angular part behaving like in flat spacetime, it makes sense to separate solutions of (2.22) according to

$$\Phi(t,r,\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{U_{l,m}(t,r)}{r\sqrt{\Theta(t,r)}} Y_{l,m}(\theta,\phi), \tag{B.1}$$

where $Y_{l,m}$ are the spherical harmonics. For $l = 0$ we anticipate modes to fall-off roughly as $\propto 1/r\sqrt{\Theta}$, since a continuity equation should hold, and the sphere at coordinate radius r has an area of $4\pi\Theta r^2$, as we can see from (1.31). In preparation for this, we have already extracted a factor of $1/r\sqrt{\Theta}$. Because Φ is real, we must have

$$U_{l,m}(t,r) = \bar{U}_{l,-m}(t,r). \tag{B.2}$$

In particular, $U_{0,0}(t,r) \in \mathbb{R}$.

Out- and Ingoing Null Coordinates. Before plugging this decomposition into

$$\partial_\mu \left(\sqrt{|g|}\, g^{\mu\nu} \partial_\nu \Phi \right) = 0, \tag{B.3}$$

((2.22) in the main text) to obtain a differential equation for $U_{l,m}$, it is worthwhile to consider qualitatively the solutions we expect for $U_{l,m}$. In flat spacetime (B.3) reduces to the well-known three-dimensional wave equation in polar coordinates; the spherically symmetric ($l = 0$) solutions for $U_{0,0}$ are then superpositions of ingoing waves $\propto \exp(i\omega(t-r))$ and outgoing waves $\propto \exp(i\omega(t+r))$, with $\omega \in \mathbb{R}$. That this is the case can be most directly seen by changing to the *outgoing* and *ingoing null coordinates* $u = t - r$ and $v = t + r$; this transforms the equation for $U_{0,0}$ into $\partial_u \partial_v U_{0,0} = 0$. Thus, $U_{0,0}(u,v) = f(u) + g(v)$ for arbitrary real functions f and g; Fourier-transforming these functions yields the solutions mentioned above.[49] Since we will be interested mostly in the $l = 0$ case, it makes sense to try finding corresponding null coordinates u and v for our metric (1.31).

Recall from the main text that we assumed V, c_s and Θ to be time-independent at this stage. We focus on the t-r sector of (1.31):

$$\begin{aligned} \mathrm{d}s_{t,r}^2 &= \Theta \cdot \left[(V^2 - c_s^2)\, \mathrm{d}t^2 - 2V\, \mathrm{d}t\mathrm{d}r + \mathrm{d}r^2 \right] \\ &= \Theta \cdot (V^2 - c_s^2) \cdot \left[\left(\mathrm{d}t - \frac{\mathrm{d}r}{c_s + V} \right) \cdot \left(\mathrm{d}t + \frac{\mathrm{d}r}{c_s - V} \right) \right] =: \Theta \cdot (V^2 - c_s^2) \cdot \mathrm{d}u\, \mathrm{d}v. \end{aligned} \tag{B.4}$$

We thus define the *out- and ingoing null coordinates* according to

$$u := t - \int^r \frac{\mathrm{d}r'}{c_s(r') + V(r')}, \qquad v := t + \int^r \frac{\mathrm{d}r'}{c_s(r') - V(r')}. \tag{B.5}$$

To prevent problems at the horizon, we introduce $u_<$ and $u_>$, as described in the main text:

$$u_< := t - \int_{r_<}^r \frac{\mathrm{d}r'}{c_s(r') + V(r')}, \qquad u_> := t - \int_{r_>}^r \frac{\mathrm{d}r'}{c_s(r') + V(r')}. \tag{B.6}$$

These equations are (2.23) and (2.24) in the main text. Note that ∂_u is independent of the choices of $r_<$ and $r_>$, since those only induce shifts in $u_<$ and $u_>$.

[49] Additionally, we obtain some reality condition linking modes of ω and $-\omega$ because Φ is real.

The full metric is now

$$ds^2 = \Theta \cdot \left[(V^2 - c_S^2)\,\mathrm{d}u\,\mathrm{d}v + r^2 \mathrm{d}\Omega^2\right], \tag{B.7}$$

hence

$$|g| = \Theta^4 \cdot r^4 \sin^2\theta \cdot \frac{1}{4}(V^2 - c_s^2)^2, \tag{B.8}$$

$$\sqrt{|g|}g^{\mu\nu} = \Theta \cdot r^2 \sin\theta \cdot \mathrm{diag}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \frac{1}{2}(V^2 - c_s^2) \cdot \mathrm{diag}\left(r^{-2}, r^{-2}\sin^{-2}\theta\right)\right]. \tag{B.9}$$

Field Modes. Because the metric components do not depend on time, computations are simplified: any function $f(r)$ only depending on r, such as r itself and now by assumption also c_s, V and Θ, must depend on u and v through $f = f(u - v)$. Thus, $\partial_u f = -\partial_v f$, $\partial_u^2 f = \partial_v^2 f = -\partial_u \partial_v f$, *etc.*

Inserting the decomposition (B.1) into (B.3) with (B.9) and assuming time-independence yields, after some algebra:

$$\partial_u \partial_v U_{l,m} - \frac{\partial_u \partial_v (r\sqrt{\Theta})}{r\sqrt{\Theta}} U_{l,m} = -\frac{1}{4}(V^2 - c_s^2) \cdot \frac{l(l+1)}{r^2} U_{l,m}, \tag{B.10}$$

where $U_{l,m}$ is now viewed as a function of the new coordinates (u, v). In particular, we have used that $Y_{l,m}$ are eigenfunctions of the angular Laplacian $\triangle_{S^2}$:

$$\triangle_{S^2} Y_{l,m} = -l(l+1)\,Y_{l,m}, \qquad \triangle_{S^2} := \frac{1}{\sin\theta}\partial_\theta\left(\sin\theta\,\partial_\theta\right) + \frac{1}{\sin^2\theta}\partial_\phi^2. \tag{B.11}$$

The factor $1/r\sqrt{\Theta}$ has also helped in simplifying (B.10). However, the actual fall-off of solutions is not quite $\propto 1/r\sqrt{\Theta}$. Firstly, l is known to produce different fall-off rates in flat spacetime; this also happens here due to the term on the right-hand side of (B.10), which in the present case is further modified by the properties of the flow, which among other things are responsible for curvature. For $l = 0$ this term vanishes. Secondly, the flow properties also influence the fall-off directly via the second term on the left-hand side; this happens because u and v contain V and c_s implicitly. These two terms make (B.10) very difficult to solve in general; even for the Schwarzschild case, it is a formidable task (see *e.g.* [24]).

Note that without the two terms just discussed, we simply have $\partial_u \partial_v U_{l,m} = 0$, whose solution is of the form $f_>(u_>) + f_<(u_<) + g(v)$, where $f_>$, $f_>$ and g are arbitrary functions (defined on the respective domain of the variables). Thus, the solution space is spanned by functions $\propto \mathrm{e}^{-\mathrm{i}\omega\alpha}$, $\omega \in \mathbb{R}$, $\alpha = u_<, u_>, v$, and it is understood that the exponentials with $u_<$ and $u_>$ are set to zero outside the domain of $u_<$ and $u_>$ respectively. Now since both non-trivial terms in (B.10) are multiplications with functions of r, the solution space of (B.10) will be spanned by the functions $F^{l,m}_{\omega,\alpha}(r)\,\mathrm{e}^{-\mathrm{i}\omega\alpha}$, where $F^{l,m}_{\omega,\alpha}(r)$ are unknown functions of r. Investigating (B.10), we see that $F^{l,m}_{\omega,\alpha}$ does not depend on m, and we simply write $F^{l}_{\omega,\alpha}$. The dependence on α occurs in part because we need $F^{l}_{\omega,u_>}(r) = 0$ for $r < r_H$, and similarly $F^{l}_{\omega,u_<}(r) = 0$ for $r > r_H$. Furthermore, through choosing the phase of $A^{l,m}_{\omega,\alpha}$, we can always make the $F^{l}_{\omega,\alpha}$ real-valued, which we will do from now on. Thus, the solution to (B.10) is

$$U_{l,m}(t, r) = \int_{-\infty}^{\infty} \mathrm{d}\omega \sum_{\alpha = u_<, u_>, v} A^{l,m}_{\omega,\alpha}\, F^{l}_{\omega,\alpha}(r)\, \mathrm{e}^{-\mathrm{i}\omega\alpha}, \qquad A^{l,m}_{\omega,\alpha} \in \mathbb{C}. \tag{B.12}$$

Due to asymptotic flatness, we have $V \to 0$, and $c_s \to \mathrm{const}$ for $r \to \infty$, and (B.10) becomes $\partial_u \partial_v U_{l,m} = \mathcal{O}(r^{-1})$ for $r \to \infty$. We must thus have that $F^{l}_{\omega,u_>}(r), F^{l}_{\omega,v}(r) = \mathrm{const.} + \mathcal{O}(r^{-1})$ for $r \to \infty$. These functions contain all the information about the shape of the r-fall-off of solutions; luckily we will not need their exact expression.

It is useful to choose the constants in $F^l_{\omega,u_>}(r)$ and $F^l_{\omega,v}(r)$ for $r \to \infty$ such that $F^l_{\omega,u_>}(r)$, $F^l_{\omega,v}(r) \to 1 + \mathcal{O}(r^{-1})$ for $r \to \infty$. This is possible, since in flat spacetime, spherical waves fall off as $1/r$ due to the continuity equation [65], and we have already extracted a factor $1/r$ from our modes. We also extract a factor $1/\sqrt{\omega}$ from all $F^l_{\omega,\alpha}$, as is usual [65].

Due to the reality condition (B.2), the $A^{\omega,\alpha}_{l,m}$ are not fully independent, but must satisfy

$$A^{l,m}_{\omega,\alpha} = \bar{A}^{l,-m}_{-\omega,\alpha}. \tag{B.13}$$

In particular, $A^{0,0}_{\omega,\alpha} = \bar{A}^{0,0}_{-\omega,\alpha}$.

In conclusion, the general solution to the Klein-Gordon equation (B.3) under the assumptions above can be written as a mode expansion

$$\Phi(t,r,\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{Y_{l,m}(\theta,\phi)}{r\sqrt{\Theta(r)}} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{\sqrt{\omega}} \sum_{\alpha=u_<,u_>,v} A^{l,m}_{\omega,\alpha}\, F^l_{\omega,\alpha}(r)\, \mathrm{e}^{-\mathrm{i}\omega\alpha}, \tag{B.14}$$

with the reality condition (B.13) imposed; this shows (2.25) and (2.26) in the main text. A similar expansion is obtained for the special case of a collapsing spherical star geometry in Hawking's original paper [31].

The spherically symmetric case ($l = 0$) is

$$\Phi(t,r,\theta,\phi) = \frac{1}{r\sqrt{\Theta(r)}} \int_0^{\infty} \frac{\mathrm{d}\omega}{\sqrt{\omega}} \sum_{\alpha=u_<,u_>,v} F_{\omega,\alpha}(r) \cdot \left[A_{\omega,\alpha}\mathrm{e}^{-\mathrm{i}\omega\alpha} + \bar{A}_{\omega,\alpha}\mathrm{e}^{\mathrm{i}\omega\alpha}\right], \tag{B.15}$$

which is (2.28) in the main text.

Finally, we can allow for slow time dependence (2.27) without changing our results, since this allows us to effectively ignore the time derivatives of V, c_s and Θ.

Extension of the u-Coordinate Through the Horizon. The Feynman-$i\varepsilon$ prescription roughly speaking instructs us to add an imaginary mass term $m^2 = -\mathrm{i}\varepsilon$ in the Klein-Gordon equation, *i.e.* $|g|^{-1/2}\partial_\mu(|g|^{1/2}g^{\mu\nu}\partial_\nu\Phi) - \mathrm{i}\varepsilon\Phi = 0$, to obtain the correct propagator. This has the effect of adding $-\mathrm{i}\varepsilon\sqrt{|g|}/r\sqrt{\Theta} \cdot U_{l,m}$ to the right-hand side of (B.10).

The small imaginary mass term of course changes the mode decomposition (B.15). One possibility is to absorb the change into the functions $F_{\omega,\alpha}$. But this hides the potentially very large influence of the $i\epsilon$-prescription: after all, it plays a crucial role in the expression of the propagator, by making a divergent quantity finite. We expect a similarly drastic impact on the modes of (B.15). Another approach is thus to try to include the change in the exponential factors $\mathrm{e}^{\pm\mathrm{i}\omega\alpha}$, by slightly changing the definitions $u_<$, $u_>$ and v. As we will see, this allows us to connect $u_<$ and $u_>$ into a single coordinate, which is what we originally wanted to achieve in the main text.

Consider the *eikonal limit* $\omega \to \infty$, where the differential equation (B.3) in the new coordinates naturally decomposes into an equation for α and one for $F_{\omega,\alpha}$; by moving the $\mathrm{i}\varepsilon$-term into the equation for α, we move all the influence of $\mathrm{i}\varepsilon$ from $F_{\omega,\alpha}$ into α. Note that this *a priori* only works in the eikonal limit. In that limit we have the *eikonal equation* for α [76]:

$$\omega^2 g^{\mu\nu}(\partial_\mu\alpha)(\partial_\nu\alpha) - \mathrm{i}\varepsilon = 0. \tag{B.16}$$

Note that two factors of i due to the derivatives have introduced a minus sign. It is then possible to absorb $\Theta(r)^{-1}$ from $g^{\mu\nu}$ into ε. Although this makes $\varepsilon \rightsquigarrow \varepsilon(r)$ r-dependent, this does not matter, since the $\mathrm{i}\varepsilon$-prescription is ultimately to be understood with the eventual limit $\varepsilon \to 0$ in mind; so if Θ is sufficiently well-behaved, $\lim_{\varepsilon\to 0} = \lim_{\varepsilon(r)\to 0}$ pointwise. Let us thus simply write $\varepsilon = \varepsilon(r)$ and forget about the conformal factor in the $g^{\mu\nu}$.[50] With $\alpha = \alpha(t, r)$, the eikonal equation gives (see also [76])

$$(\omega - V(r)k_\alpha(r))^2 = c_s(r)^2 k_\alpha(r)^2 + \mathrm{i}\varepsilon, \qquad k_\alpha(r) := -\omega \frac{\partial \alpha}{\partial r}. \tag{B.17}$$

Thus,

$$\omega - V(r)k(r) = \pm \left(c_s(r)k(r) + \frac{1}{2}\mathrm{i}\varepsilon \right) + \mathcal{O}(\varepsilon^2) = \pm(1 + \mathrm{i}\varepsilon)c_s(r)k(r), \tag{B.18}$$

where we have redefined ε in the last step. Thus,

$$k_\alpha^\pm(r) = \frac{\omega}{V(r) \pm (1 + \mathrm{i}\varepsilon)c_s(r)}. \tag{B.19}$$

We recognize that in the limit $\varepsilon \to 0$, $k_\alpha^+ = -\omega\partial u/\partial r$ and $k_\alpha^- = -\omega\partial v/\partial r$ (think of $u_<$ or $u_>$ wherever they are defined). So solving the eikonal equation (B.16) leads us to *redefine* u and v according to

$$u := t - \int_{r_>}^{r} \frac{\mathrm{d}r'}{(1 + \mathrm{i}\varepsilon)c_s(r') + V(r')}, \tag{B.20}$$

$$v := t + \int_{r_>}^{r} \frac{\mathrm{d}r'}{(1 + \mathrm{i}\varepsilon)c_s(r') - V(r')} = t + \int_{r_>}^{r} \frac{\mathrm{d}r'}{c_s(r') - V(r')} + \mathcal{O}(\varepsilon). \tag{B.21}$$

Note that for v, the iε-prescription does not change the definition in the eventual limit $\varepsilon \to 0$. u on the other hand is regularized by it and can now be extended through the horizon; and for $r \to \infty$, $u \to u_>$.

So far, this only works for $\omega \to \infty$. We instate these definitions of u and v for all ω; this requires us to modify the $F_{\omega,\alpha}$ for low values of ω, and away from $r \to \infty$. But this does not bother us, since we have found a regularization of u which is what we originally wanted. Let us now explore how this connects the coordinates $u_>$ and $u_<$.

Since $c_s > 0$, the iε-prescription leads us to regularize the integral of u by moving the pole from $r = r_H$ out of the way into the lower half plane. This is equal to (still for $r < r_H$)

$$u = -\int_{\Gamma_\varepsilon} \frac{\mathrm{d}r'}{c_s(r') + V(r')} + \mathcal{O}(\varepsilon) + u_<, \tag{B.22}$$

where Γ_ε is the contour shown in Figure 10: trace the real line starting from $r_>$ until $r_H + \varepsilon$, followed by a semicircle γ_ε of radius ε into the upper half plane in order to circumvent the pole at $r = r_H$, and finish by tracing the real line from $r_H - \varepsilon$ to $r_<$. Also, it is understood that c_s and V have been sufficiently analytically extended into the complex plane; note that this is always possible if c_s and V are for instance analytical as real functions (see *e.g.* [51]).

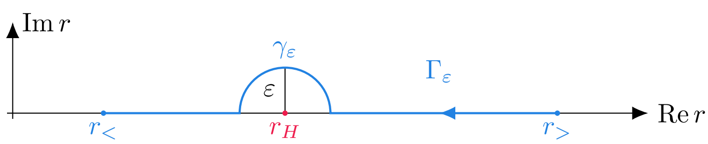


Figure 10: The integration contour Γ_ε.

This contour integral contains a real contribution C due to two legs of the integration along the real line. In circumventing the pole, the contour integral also picks up an imaginary

[50] Another reason for why the conformal factor should not matter here, is that it does not matter for the eikonal equation of the massless field. And since our field is nearly massless, even more so because ω is large, it should only matter very little here.

contribution: close to the horizon, the denominator of the integrand is (2.36) and thus for small ε the imaginary contribution is

$$\int_{\gamma_\varepsilon} \frac{\mathrm{d}r'}{\kappa \cdot (r' - r_H)} = \frac{\mathrm{i}\pi}{\kappa}. \tag{B.23}$$

Thus,

$$u = u_< - C - \frac{\mathrm{i}\pi}{\kappa} + \mathcal{O}(\varepsilon). \tag{B.24}$$

The constant C can be set to zero by a suitable choice of $r_>$ and $r_<$, which we will assume. This shows Lemma 2.11 in the main text.

Analytical Continuation of u-Modes. This also implies that

$$\theta(r - r_H)\mathrm{e}^{-\mathrm{i}\omega u_>} + \theta(r_H - r)\mathrm{e}^{-\mathrm{i}\omega u_<}\mathrm{e}^{-\pi\omega/\kappa} \tag{B.25}$$

is analytical in the upper half-plane of r (since this is the half-plane through which we regularized u); here, we extend the Heaviside-functions θ into to upper half plane simply by $\theta(z) := \theta(\mathrm{Re}\,(z))$. Therefore, we can also consider the modes (up to a factor $1/r\sqrt{\Theta}$)

$$A_{\omega,+}F_\omega(r) \cdot \big[\theta(r - r_H)\mathrm{e}^{-\mathrm{i}\omega u_>} + \theta(r_H - r)\mathrm{e}^{-\mathrm{i}\omega u_<}\mathrm{e}^{-\pi\omega/\kappa}\big] \tag{B.26}$$

and their complex conjugates; the $A_{\omega,+}$ are the amplitude of these new modes. Here, $F_{\omega,>}(r)$ and $F_{\omega,<}(r)$ satisfy the same differential equation and thus agree at r_H up to a constant shift, which we chose to be zero; therefore we simply wrote $F_\omega(r)$. We note that modes of the form

$$A_{\omega,-}F_\omega(r) \cdot \big[\theta(r - r_H)\mathrm{e}^{\mathrm{i}\omega u_>} + \theta(r_H - r)\mathrm{e}^{\mathrm{i}\omega u_<}\mathrm{e}^{\pi\omega/\kappa}\big], \tag{B.27}$$

and their complex conjugates are also analytical. These are (2.40) and (2.41) in the main text. These modes are *independent* of each other: the $(+)$-mode is essentially an analytical continuation of the $u_>$-mode, and the $(-)$-mode a continuation of the complex conjugate of the $u_<$-mode; but these modes were originally independent. Instead of $u_<$- and $u_>$-modes, we can expand our field also in terms of $(+)$- and $(-)$-modes.

Comparison of coefficients gives

$$A_{\omega,u_>} = \frac{1}{2}\big[A_{\omega,+} + \bar{A}_{\omega,-}\big], \tag{B.28}$$

$$A_{\omega,u_<} = \frac{1}{2}\big[A_{\omega,+}\mathrm{e}^{-\pi\omega/\kappa} + \bar{A}_{\omega,-}\mathrm{e}^{\pi\omega/\kappa}\big], \tag{B.29}$$

with inverses

$$A_{\omega,+} = \frac{1}{\sinh(\pi\omega/\kappa)}\big[\mathrm{e}^{\pi\omega/\kappa}A_{\omega,u_>} - A_{\omega,u_<}\big], \tag{B.30}$$

$$\bar{A}_{\omega,-} = \frac{1}{\sinh(\pi\omega/\kappa)}\big[A_{\omega,u_<} - \mathrm{e}^{-\pi\omega/\kappa}A_{\omega,u_>}\big]. \tag{B.31}$$

Once quantized, these relations will become relations between operators, so-called *Bogoliubov transformations*.

Because the new modes are independent, their commutator must vanish:

$$0 = [\hat{A}_{\omega,+}, \hat{A}_{\omega,-}] = \frac{1}{\sinh^2(\pi\omega/\kappa)}\big[-C_{u_>}(\omega) - C_{u_<}(\omega)\big], \tag{B.32}$$

and thus

$$C_{u_<}(\omega) = -C_{u_>}(\omega). \tag{B.33}$$

We can now compute

$$[\hat{A}_{\omega,\pm}, \hat{A}^\dagger_{\omega,\pm}] = \frac{C_{u_>}(\omega)}{\sinh^2(\pi\omega/\kappa)}\big[\pm\mathrm{e}^{\pm 2\pi\omega/\kappa} \mp 1\big] \cdot \delta(\omega - \omega') \cdot \hat{\mathrm{id}}. \tag{B.34}$$

Since the prefactor is positive, $\hat{A}_{\omega,\pm}$ can be seen as annihilation operators.

Comments on the Analytically Extended Modes. Two comments on the new modes are in order:

Firstly, despite u now being analytically continued in the complex plane above $r = r_H$, its behaviour is still singular at $r = r_H$ and $\mathrm{e}^{-\mathrm{i}\omega u}$ oscillates uncontrollably close to the horizon. This is related to the so-called *trans-Planckian problem*, the fact that derivations of Hawking radiation seem to require infinitely blue-shifted modes close to the horizon during calculations, even if these modes have no physical impact on the result. See *e.g.* [81].

Secondly, in the main text we have identified somewhat non-rigorously both the (+)- and (−)-modes as positive frequency modes as seen by the infalling observer. To show that these modes really are of positive frequency, one must first switch to a coordinate system which is non-singular around the horizon (such as local Minkowski coordinates of an inertial reference frame) and find a set of positive frequency modes there; one then shows that the (+)- and (−)-modes are linear combinations of only those modes. We will not do the computations here; see [76] and especially sources therein for details.

Annihilation Operators for Infalling Observer. As argued in the main text, $\hat{A}_{\omega,\pm}$ can be seen as annihilation operators for the infalling observer. It will however be useful to define the operators for the observer as normalized versions of $\hat{A}_{\omega,\pm}$ such that the commutation relations are the same as for particles at infinity; we thus set

$$\hat{b}_{\omega,\pm} := \frac{\sinh(\pi\omega/\kappa)}{\sqrt{\mp 1 \pm \mathrm{e}^{\pm 2\pi\omega/\kappa}}} \hat{A}_{\omega,\pm}, \tag{B.35}$$

giving the Bogoliubov transformation (these are equations (2.45) and (2.46) in the main text)

$$\hat{b}_{\omega,+} = \frac{1}{\sqrt{2\sinh(\pi\omega/\kappa)}} \left[\mathrm{e}^{\pi\omega/2\kappa} \hat{A}_{\omega,u_>} - \mathrm{e}^{-\pi\omega/2\kappa} \hat{A}_{\omega,u_<}\right], \tag{B.36}$$

$$\hat{b}_{\omega,-} = \frac{1}{\sqrt{2\sinh(\pi\omega/\kappa)}} \left[\mathrm{e}^{\pi\omega/2\kappa} \hat{A}^\dagger_{\omega,u_<} - \mathrm{e}^{-\pi\omega/2\kappa} \hat{A}^\dagger_{\omega,u_>}\right]. \tag{B.37}$$

The inverse transformation is

$$\hat{A}_{\omega,u_>} = \frac{1}{\sqrt{2\sinh(\pi\omega/\kappa)}} \left[\mathrm{e}^{\pi\omega/2\kappa} \hat{b}_{\omega,+} + \mathrm{e}^{-\pi\omega/2\kappa} \hat{b}^\dagger_{\omega,-}\right], \tag{B.38}$$

$$\hat{A}_{\omega,u_<} = \frac{1}{\sqrt{2\sinh(\pi\omega/\kappa)}} \left[\mathrm{e}^{-\pi\omega/2\kappa} \hat{b}_{\omega,+} + \mathrm{e}^{\pi\omega/2\kappa} \hat{b}^\dagger_{\omega,-}\right]. \tag{B.39}$$

The first of these equations is (2.50) in the main text.

C Fluid Analogue Model for 1+1-Dimensional Schwarzschild Spacetime

We have seen in Section 4.4 that modelling Schwarzschild spacetime in Gullstrand-Painlevé coordinates using a flowing fluid is difficult, because the continuity equation has to be satisfied. Unfortunately, the continuity equation is a fundamental part of most analogue systems based on fluid flow and cannot easily be removed. We have also seen that there are approaches to circumvent this issue, but they are complicated, often involving complicated fluids.

We present here a new approach: Instead of modelling Schwarzschild spacetime in a fluid of the same spatial dimension, we propose to model 1+1-dimensional Schwarzschild spacetime in Gullstrand-Painlevé coordinates, using a 2-dimensional fluid flow; Schwarzschild spacetime will then only be represented by a *subset* of the model, while the rest of the model is required to satisfy the continuity equation.

Solving the Laplace Equation. We saw in Section 4.4 that to model Schwarzschild spacetime in Gullstrand-Painlevé coordinates using a perfect, irrotational fluid forces us to choose constant density ρ_0 for the background flow and constant speed of sound σ. The continuity equation of the background flow then becomes $\nabla \cdot \mathbf{v}_0 = 0$ (recall (4.56)), thus, expressed in terms of the velocity potential $\mathbf{v}_0 = -\nabla\Phi_0$, we have

$$\triangle\Phi_0 = 0. \tag{C.1}$$

In other words, the velocity potential is a *harmonic function.*

Since we are working in two dimensions, we can make use of parallels between harmonic functions and holomorphic functions. Concretely, we will use the following Lemma:

Lemma C.1

Let $I \subset \mathbb{R}$ be an open interval and $f : I \to \mathbb{R}$ an analytical function. Then there exists an open neighbourhood $O \subset \mathbb{R}^2$ of $I \times \{0\} \subset \mathbb{R}^2$ and a smooth function $\Phi : O \to \mathbb{R}$ such that

(a) $\triangle\Phi(x, y) = 0 \ \forall(x, y) \in O$.
(b) $\Phi(x, 0) = f(x) \ \forall x \in I$.
(c) $\partial_y\Phi(x, 0) = 0 \ \forall x \in I$.

Proof. We begin by interpreting I as a subset of the real axis in the complex plane: $I \subset \mathbb{R} \subset \mathbb{C}$. We then holomorphically continue f to $F : U \to \mathbb{C}$, with U some open neighbourhood of I. More precisely:

We use the fact that the Taylor expansion of f around any $x \in I$ converges for an open disk around x (the disk of convergence) since f is analytical, to extend f to holomorphic functions f_x on these disks around every point $x \in I$. Since two such holomorphic functions $f_x, f_{x'}$ with overlapping domains agree on an interval in I (where they both agree with f), the identity theorem ensures that f_x and $f_{x'}$ agree everywhere their domains overlap, and they can be "glued together" to a single holomorphic function, which still agrees with f on I. We can therefore glue all f_x together and thus obtain $F : U \to \mathbb{C}$, and U is the union of all convergence disks. Of course, we might still be able to extend F and U further, but this is not required here.

Without loss of generality, we can take U to be symmetric around the real axis, *i.e.* $z \in U \Leftrightarrow \bar{z} \in U$. Consider F^+, the restriction of F to $\{z \in \mathbb{C} \mid \mathrm{Im}\,(z) \geq 0\} \cap U$. F^+ fulfils the requirements of the Schwarz reflection principle, allowing us to holomorphically extend F^+ to $\tilde{F}$ defined in all of U; furthermore, this extension has the property that

$$\tilde{F}(\bar{z}) = \overline{\tilde{F}(z)}.$$

But $\tilde{F}$ agrees with F on the upper half of U and invoking the identity theorem again, it must hold that $\tilde{F} = F$. Therefore, $F(\bar{z}) = \overline{F(z)}$.

The function $G : U \to \mathbb{R}$,

$$G(z) := \frac{1}{2}\big(F(z) + F(\bar{z})\big), \tag{C.2}$$

thus is well-defined as it takes on real values. Writing $F(x + \mathrm{i}y) = u(x, y) + \mathrm{i}v(x, y)$, with real-valued u and v, we then have

$$G(x + \mathrm{i}y) = u(x, y). \tag{C.3}$$

Since F is holomorphic, u and v satisfy the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}, \tag{C.4}$$

and therefore,

$$\frac{\partial^2}{\partial x^2} u(x, y) = \frac{\partial^2}{\partial x \partial y} v(x, y) = -\frac{\partial^2}{\partial y^2} u(x, y), \tag{C.5}$$

it follows that G is harmonic, *i.e.*

$$\triangle G(x + \mathrm{i}y) = 0. \tag{C.6}$$

Let us now define

$$O := \{(x, y) \in \mathbb{R}^2 \mid x + \mathrm{i}y \in U\} \subset \mathbb{R}^2, \qquad \Phi : O \to \mathbb{R}, \quad \Phi(x, y) := G(x + \mathrm{i}y). \tag{C.7}$$

O is an open neighbourhood of $I \times \{0\} \subset \mathbb{R}^2$, since U is an open neighbourhood of $I \subset \mathbb{C}$. Furthermore, $\triangle\Phi(x, y) = 0 \ \forall (x, y) \in O$, since $\triangle G(x + \mathrm{i}y) = 0 \ \forall x + \mathrm{i}y \in U$. Property (a) thus holds.

Property (b) also holds, because Φ agrees with f on $I \times \{0\}$, since G agrees with f on I. Finally, we can compute

$$\partial_y \Phi(x, 0) = \frac{1}{2}\big(\mathrm{i}F'(x) - \mathrm{i}F'(x)\big) = 0, \tag{C.8}$$

with a prime denoting complex differentiation, demonstrating property (c). □

Building the Model. We can use Lemma C.1 to build analogue models:

Theorem C.2

Let $I \subset \mathbb{R}$ be an open interval and $g : I \to \mathbb{R}$ such that $-\int^x \mathrm{d}x'\, g(x')$ is analytical in x. There exists a 2D potential flow $\mathbf{v} = -\nabla\Phi$ with:
(a) the flow satisfies the continuity equation for constant density, *i.e.* $\triangle\Phi = 0$.
(b) $v_x(x, 0) = g(x) \ \forall x \in I$.
(c) $v_y(x, 0) = 0 \ \forall x \in I$.

Proof. Define $f : I \to \mathbb{R}$, $f(x) := -\int^x \mathrm{d}x'\, g(x')$. f is analytical by assumption, and we can apply Lemma C.1 to obtain the velocity potential. The properties (a)-(c) are then simply the properties (a)-(c) provided by the lemma. □

Note that it suffices if g is analytical, because one can proof using Morera's theorem that $-\int^x \mathrm{d}x'\, g(x')$ must then also be analytical.

Schwarzschild Spacetime in Gullstrand-Painlevé Coordinates. Of primary interest is the case $g : (0, \infty) \to \mathbb{R}$, $g(x) = -\sqrt{r_s/x}$, corresponding to the flow of Schwarzschild spacetime in Gullstrand-Painlevé coordinates with one spatial dimension. According to Theorem C.2, there exists a fluid flow with constant density, and velocity potential Φ_0 providing the velocity profile $\mathbf{v}_0(x, 0) = g(x)$ on $(0, \infty)$.

Tracing the proofs of the lemma and theorem, we see that

$$f(x) = \int^x \mathrm{d}x' \sqrt{\frac{r_s}{x'}} = 2\sqrt{r_s x}. \tag{C.9}$$

Furthermore, if $\sqrt{\cdot}$ denotes the complex-valued square root defined on $\mathbb{C}$ with branch cut along the negative real axis (this is the holomorphic extension of the real square root function), we find that

$$\Phi_0(x, y) = \sqrt{r_s} \cdot \left(\sqrt{x + \mathrm{i}y} + \sqrt{x - \mathrm{i}y}\right). \tag{C.10}$$

Indeed, on the subset $x > 0$ and $y = 0$, this potential induces the flow

$$v_{0,x}(x, 0) = -\sqrt{\frac{r_s}{x}}, \qquad v_{0,y}(x, 0) = 0, \tag{C.11}$$

as advertised.

The full flow is best expressed in polar coordinates:

$$v_{0,x}(r, \theta) = -\sqrt{\frac{r_s}{r}} \cos\frac{\theta}{2}, \qquad v_{0,y}(r, \theta) = -\sqrt{\frac{r_s}{r}} \sin\frac{\theta}{2}. \tag{C.12}$$

For this we have used that $\sqrt{r\mathrm{e}^{\mathrm{i}\theta}} = \sqrt{r}\mathrm{e}^{\mathrm{i}\theta/2}$ for $r \geq 0$ and $\theta \in (-\pi, \pi)$. Figure 11 shows the flow.

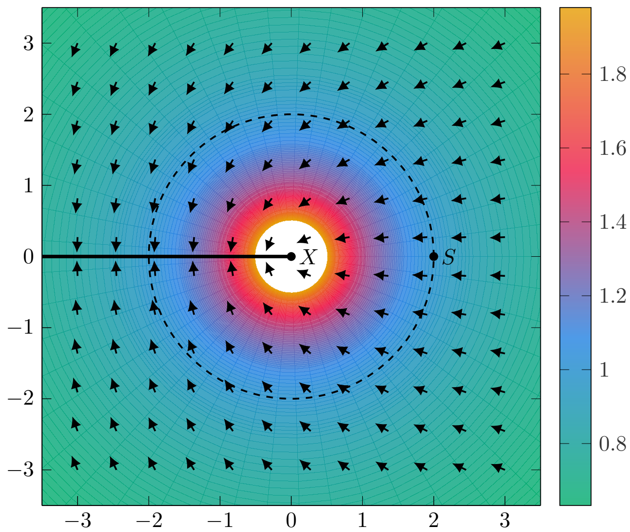


Figure 11: The potential fluid flow described in the text for $M = 1$. Colour indicates the flow velocity. A circle marks the ergo-region; note that the boundary of the ergo-region is not an apparent horizon. Also marked are the sonic point $S = (r_s, 0)$, corresponding to the horizon of Schwarzschild spacetime with one spatial dimension, and the singularity $X = (0, 0)$. The black line on the negative x-axis is the branch cut of the complex square root.

The analogue metric induced by this flow is (up to a constant) (see (4.37), with $\sigma = \text{const}$, $\rho_0 = \text{const}$)

$$g_{\mu\nu} = \begin{pmatrix} r_s/r - r & \sqrt{r_s/r}\cos(\theta/2) & \sqrt{r_s/r}\sin(\theta/2) \\ & 1 & 0 \\ * & & 1 \end{pmatrix}, \tag{C.13}$$

where $*$ us a placeholder for the symmetric part. As we can now clearly see, this becomes the Schwarzschild metric in Gullstrand-Painlevé coordinates with one spatial dimension when restricted to $x > 0$ and $y = 0$, that is $\theta = 0$.

Apparent Horizon? Notice that the flow is supersonic for $r < r_s$ and sonic for $r = 2Mr_s$. Therefore, any apparent horizon must lie inside the ergo-region $r \leq r_s$. The flow velocity is not orthogonal to the ergo-surface $r = r_s$, except at the point $(r_s, 0)$. This point is of interest, because it corresponds to the event horizon of Schwarzschild spacetime with one spatial dimension. It is thus natural to ask whether $(r_s, 0)$ is part of an apparent horizon also in our model with one additional spatial dimension. As it turns out, this is not the case.

A necessary condition for a curve $C \subset \mathbb{R}^2$ to be an apparent horizon is that the magnitude $|\mathbf{v}_\perp|$ of the component of the velocity field normal to C be equal to the speed of sound: $|\mathbf{v}_\perp| = c$. We will show that this condition cannot be satisfied for any curve passing through $(r_s, 0)$.

Assume that there is an apparent horizon through $(r_s, 0)$. At least locally around $(r_s, 0)$ we can parametrize it as a polar curve $r(\theta)$. Let $\mathbf{n}(\theta)$ be the outwards-pointing normal. The condition $|\mathbf{v}_\perp| = c$ is then

$$\mathbf{n} \cdot \mathbf{v} = -|\mathbf{n}|, \tag{C.14}$$

since the flow is inflowing while $\mathbf{n}$ is outwards-pointing.

For a general polar curve $r(\theta)$ the outwards-pointing normal is found via

$$\mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}\theta}(r(\theta)\cos\theta)\,\mathrm{d}\theta = (r'(\theta)\cos\theta - r(\theta)\sin(\theta))\,\mathrm{d}\theta, \tag{C.15}$$

$$\mathrm{d}y = \frac{\mathrm{d}}{\mathrm{d}\theta}(r(\theta)\sin\theta)\,\mathrm{d}\theta = (r'(\theta)\sin\theta + r(\theta)\cos\theta)\,\mathrm{d}\theta. \tag{C.16}$$

And thus (up to normalization)

$$\mathbf{n}(\theta) = \begin{pmatrix} r'(\theta)\sin\theta + r(\theta)\cos\theta \\ -r'(\theta)\cos\theta + r(\theta)\sin\theta \end{pmatrix}. \tag{C.17}$$

Equation (C.14) now becomes, after some trigonometric simplifications:

$$\sqrt{r_s}\left(\sqrt{r}\cos(\theta/2) + \frac{r'}{\sqrt{r}}\sin(\theta/2)\right) = \sqrt{r'^2 + r^2}. \tag{C.18}$$

We now solve for r' to bring the equation into explicit form:

$$r' = \left[\cos(\theta/2)\sin(\theta/2) \pm \sqrt{\frac{r}{r_s}\left(1 - \frac{r}{r_s}\right)}\right] \cdot \left(\frac{1}{r_s} - \frac{1}{r}\sin^2(\theta/2)\right)^{-1}. \tag{C.19}$$

Since this step involved squaring both sides, it is *a priori* not clear whether both solutions lead to the correct initial equation. Note also that (C.19) violates the Lipschitz condition at $r = 2M$, for both choices of sign; thus, there is no guarantee for a unique solution. Numerical simulations now show that neither $(+)$ nor $(-)$ lead to a real-valued solution with $r(0) = r_s$.

To analytically see why this is the case, we expand (C.18) in orders of θ around $\theta = 0$. Given the initial condition $r(0) = r_s$, we find $r'(0) = 0$ from (C.19), both signs resulting in the same r' due to the vanishing square root. For the expansion we thus have

$$r = r_s + \frac{r''(0)}{2}\,\theta^2 + \mathcal{O}(\theta^4), \qquad r' = r''(0)\,\theta + r'''(0)\,\theta^2 + \mathcal{O}(\theta^3) \tag{C.20}$$

and therefore, using $(1+x)^\alpha = 1 + \alpha x + \mathcal{O}(x^2)$,

$$\sqrt{r}^{\,\pm 1} = \sqrt{r_s}^{\,\pm 1}\left(1 \pm \frac{r''(0)}{4r_s}\,\theta^2\right) + \mathcal{O}(\theta^3). \tag{C.21}$$

With the expansions for sine and cosine we further find

$$\frac{r'}{\sqrt{r}}\sin(\theta/2) = \frac{1}{2\sqrt{r_s}}r''(0)\,\theta^2 + \mathcal{O}(\theta^3), \qquad \sqrt{r}\cos(\theta/2) = \sqrt{r_s} + \left(\frac{r''(0)}{4\sqrt{r_s}} - \frac{\sqrt{r_s}}{2}\right)\theta^2 + \mathcal{O}(\theta^3). \tag{C.22}$$

Finally,

$$r^2 = r_s^2 + r_s r''(0)\,\theta^2 + \mathcal{O}(\theta^3), \qquad r'^2 = r''(0)^2\,\theta^2 + \mathcal{O}(\theta^3), \tag{C.23}$$

and

$$\sqrt{r^2 + r'^2} = r_s \cdot \left(1 + \frac{1}{2}\left(\frac{r''(0)^2}{r_s^2} + \frac{r''(0)}{r_s}\right)\theta^2\right) + \mathcal{O}(\theta^3). \tag{C.24}$$

Therefore, equation (C.18) becomes, to second order in θ:

$$\frac{r''(0)^2}{2r_s} - \frac{r''(0)}{4} + \frac{r_s}{2} = 0. \tag{C.25}$$

Solving for $r''(0)$ gives

$$r''(0) = \frac{r_s}{4} \pm \sqrt{\frac{r_s^2}{16} - r_s^2} = \frac{r_s}{4} \pm r_s\frac{\mathrm{i}\sqrt{15}}{4}. \tag{C.26}$$

Since this is not a real number, any solution to (C.18) with $r(0) = r_s$ cannot be a real-valued solution. There is thus no apparent horizon passing through $(r_s, 0)$.

Scalar Field Propagation. A scalar field $\Phi_1(t, x, y)$ propagates in the metric (C.13) according to the massless Klein-Gordon equation

$$\partial_\mu\left(h^{\mu\nu}\partial_\nu\Phi_1\right) = 0, \tag{C.27}$$

with

$$h^{\mu\nu} = \sqrt{|g|}g^{\mu\nu} \propto \begin{pmatrix} 1 & -\sqrt{r_s/r}\cos(\theta/2) & -\sqrt{r_s/r}\sin(\theta/2) \\ & -1 + (r_s/r)\cos^2(\theta/2) & (r_s/r)\sin(\theta/2)\cos(\theta/2) \\ * & & -1 + (r_s/r)\sin^2(\theta/2) \end{pmatrix}. \tag{C.28}$$

For small values of y, we have $\cos(\theta/2) = 1 + \mathcal{O}(y^2/x^2)$, $\sin(\theta/2) = y/2 + \mathcal{O}(y^3/x^3)$, and $r = x + \mathcal{O}(y^2/x^2)$. Thus:

$$h^{\mu\nu} \propto \begin{pmatrix} 1 & -\sqrt{r_s/x} & -\sqrt{r_s/x}\cdot y/2x \\ & -1 + r_s/x & (r_s/x)\cdot y/2x \\ * & & -1 \end{pmatrix} + \mathcal{O}(y^2/x^2). \tag{C.29}$$

The massless Klein Gordon equation becomes

$$\begin{aligned}
&\frac{\partial^2}{\partial t^2}\Phi_1 + \frac{\partial}{\partial x}\left(-2\sqrt{\frac{r_s}{x}}\frac{\partial}{\partial t}\Phi_1 + \Big(-1 + \frac{r_s}{x}\Big)\frac{\partial}{\partial x}\Phi_1\right) \\
&\quad = \frac{\partial^2}{\partial y^2}\Phi_1 + \frac{\partial}{\partial y}\left(\sqrt{\frac{r_s}{x}}\cdot\frac{y}{2x}\frac{\partial}{\partial t}\Phi_1\right) - \frac{\partial}{\partial x}\left(r_s\frac{y}{2x^2}\frac{\partial}{\partial y}\Phi_1\right) - \frac{\partial}{\partial y}\left(r_s\frac{y}{2x^2}\frac{\partial}{\partial x}\Phi_1\right). \qquad \text{(C.30)}
\end{aligned}$$

For $y = 0$, we have

$$\begin{aligned}
&\frac{\partial^2}{\partial t^2}\Phi_1 + \frac{\partial}{\partial x}\left(-2\sqrt{\frac{r_s}{x}}\frac{\partial}{\partial t}\Phi_1 + \Big(-1 + \frac{r_s}{x}\Big)\frac{\partial}{\partial x}\Phi_1\right) \\
&\qquad\qquad = \frac{\partial^2}{\partial y^2}\Phi_1 + \sqrt{\frac{r_s}{x}}\cdot\frac{1}{2x}\frac{\partial}{\partial t}\Phi_1 - \frac{r_s}{2x^2}\frac{\partial}{\partial x}\Phi_1. \qquad \text{(C.31)}
\end{aligned}$$

Here, the left-hand side are the terms we should expect in Schwarzschild spacetime with one spatial dimension; the right-hand side is purely due to the presence of the second dimension.

It seems that, even for $y = 0$, the second dimension, originally introduced in order to satisfy the continuity equation, has a non-trivial influence on the propagation of sound waves. However, the right-hand side can in principle be forced to become zero if $\partial^2\Phi_1/\partial y^2$ is adjusted accordingly; then scalar field propagation in the region $x > 0$, $y = 0$ should indeed simulate scalar field propagation in Schwarzschild spacetime with Gullstrand-Painlevé coordinates and one spatial dimension. We have not investigated this possibility further; particularly interesting would be to understand how the two-dimensional flow has to be restricted, and whether this can be done experimentally, such that the right-hand side vanishes.

Finally, we note that in the eikonal limit of high frequencies, sound propagation reduces to the propagation of null geodesics in the metric (C.13); and the geodesics restricted to $y = 0$ are precisely those of Schwarzschild spacetime with one spatial dimension. In the eikonal limit, the influence of the second dimension vanishes.

Acknowledgements

I would like to thank Giulia Mazzola for her amazing support, supervision, and the many detailed discussions we had. I also want to thank Prof. Dr. Renato Renner for his time in the several meetings for this project and for providing me with the opportunity of doing a semester project in the quantum information theory group at ETH Zürich. 2022 has been especially hard for me due to personal reasons, and I am grateful that Giulia and Renato have allowed me to extend the project well past the usual due date.

List of Figures

1 Apparent Horizon in Fluid-Flow Metric 15
2 Penrose Diagram of Flat Spacetime 19
3 Penrose Diagrams of Two Black Hole Spacetimes 20
4 Outgoing Modes Seen by Different Observers 37
5 Penrose Diagram of Black Hole Evaporation 43
6 Page Curve Version of the Black Hole Information Loss Paradox 48
7 Replica Wormholes 50
8 Spacetime Regions Occurring in the Island Formula 51
9 Page Curve Paradox Without Back-Reaction 54
10 Integration Contour Used for u-Mode Regularization 86
11 Fluid Flow of the Schwarzschild Analogue Model 91

References

[1] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini. Replica Wormholes and the Entropy of Hawking Radiation. *Journal of High Energy Physics*, 13, May 2020. doi:10.1007/JHEP05(2020)013.

[2] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini. The Entropy of Hawking Radiation. *Reviews of Modern Physics*, 93:035002, Jul 2021. doi:10.1103/RevModPhys.93.035002.

[3] R. Arnowitt, S. Deser, and C. W. Misner. Dynamical Structure and Definition of Energy in General Relativity. *Physical Review*, 116:1322–1330, Dec 1959. doi:10.1103/PhysRev.116.1322.

[4] R. Balbinot, S. Fagnocchi, and A. Fabbri. Quantum Effects in Acoustic Black Holes: The Backreaction. *Physical Review D*, 71:064019, Mar 2005. doi:10.1103/PhysRevD.71.064019.

[5] R. Balbinot, S. Fagnocchi, A. Fabbri, and G. P. Procopio. Backreaction in Acoustic Black Holes. *Physical Review Letters*, 94:161302, Apr 2005. URL: https://link.aps.org/doi/10.1103/PhysRevLett.94.161302, doi:10.1103/PhysRevLett.94.161302.

[6] C. Barceló, S. Liberati, and M. Visser. Analogue Gravity from Field Theory Normal Modes? *Classical and Quantum Gravity*, 18(17):3595–3610, 2001. doi:10.1088/0264-9381/18/17/313.

[7] C. Barceló, S. Liberati, and M. Visser. Analoge Gravity. *Living Reviews in Relativity*, 14:3, May 2011. doi:10.12942/lrr-2011-3.

[8] J. M. Bardeen, B. Carter, and S. W. Hawking. The Four Laws of Black Hole Mechanics. *Communications in Mathematical Physics*, 31:161, 1973. doi:10.1007/BF01645742.

[9] J. D. Bekenstein. Black Holes and Entropy. *Physical Review D*, 7(8):2333, Apr 1973. doi:10.1103/PhysRevD.7.2333.

[10] J. D. Bekenstein. Generalised Second Law of Thermodynamics in Black-Hole Physics. *Physical Review D*, 9(12):3292, Jun 1974. doi:10.1103/PhysRevD.9.3292.

[11] J. D. Bekenstein. Universal Upper Bound on the Entropy-to-Energy Ratio for Bounded Systems. *Physical Review D*, 23(2):287–298, Jan 1981. doi:10.1103/PhysRevD.23.287.

[12] N. Bilić. Analog Schwarzschild-Like Geometry in Fluids with External Pressure. *Phys. Rev. D*, 105:064052, Mar 2022. doi:10.1103/PhysRevD.105.064052.

[13] N. Bilić and H. Nikolić. Analog Gravity in Nonisentropic Fluids. *Classical and Quantum Gravity*, 35(13):135008, jun 2018. doi:10.1088/1361-6382/aac5c6.

[14] N. Bilić and H. Nikolić. Analog Schwarzschild Black Hole from a Nonisentropic Fluid. *Universe*, 7(11), 2021. doi:10.3390/universe7110413.

[15] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1982. doi:10.1017/CBO9780511622632.

[16] R. H. Boyer and R. W. Lindquist. Maximal Analytical Extension of the Kerr Metric. *Journal of Mathematical Physics*, 8(2):265–281, Feb 1967. doi:10.1063/1.1705193.

[17] L. Buoninfante, F. Di Filippo, and S. Mukohyama. On the Assumptions Leading to the Information Loss Paradox. *Journal of High Energy Physics*, 2021, 2021. doi:10.1007/JHEP10(2021)081.

[18] D. Christodoulou. Reversible and Irreversible Transformations in Black-Hole Physics. *Physical Review Letters*, 25:1596–1597, Nov 1970. doi:10.1103/PhysRevLett.25.1596.

[19] D. Christodoulou and R. Ruffini. Reversible Transformations of a Charged Black Hole. *Physical Review D*, 4:3552–3555, Dec 1971. doi:10.1103/PhysRevD.4.3552.

[20] K. Crowther, N. S. Linnemann, and C. Wüthérich. What We Cannot Learn from Analogue Gravity Experiments. *Synthese*, 198:3701 – 3726, 2021. doi:10.1007/s11229-019-02190-0.

[21] C. C. de Oliveira, R. A. Mosna, J. P. M. Pitelli, and M. Richartz. Analogue Models for Schwarzschild and Reissner-Nordström Spacetimes. *Phys. Rev. D*, 104:024036, Jul 2021. doi:10.1103/PhysRevD.104.024036.

[22] J. Douglas. Solution of the Inverse Problem of the Calculus of Variations. *Transactions of the American Mathematical Society*, 50:71–128, 1941. doi:10.1090/S0002-9947-1941-0004740-5.

[23] A. Einstein. Die Grundlage der Allgemeinen Relativitätstheorie. *Annalen der Physik*, 49:769–822, 1916.

[24] E. Elizalde. Series Solutions for the Klein-Gordon Equation in Schwarzschild Space-Time. *Physical Review D*, 36:1269–1272, Aug 1987. doi:10.1103/PhysRevD.36.1269.

[25] A. L. Fetter and J. D. Walecka. *Theoretical Mechanics of Particles and Continua*. Dover Publications, New York, 2003.

[26] S. A. Fulling. Nonuniqueness of Canonical Field Quantization in Riemannian Space-Time. *Physical Review D*, 7:2850–2862, May 1973. doi:10.1103/PhysRevD.7.2850.

[27] E. Giorgi, S. Klainerman, and J. Szeftel. Wave Equations Estimates and the Nonlinear Stability of Slowly Rotating Kerr Black Holes, 2022. doi:10.48550/ARXIV.2205.14808.

[28] A. Gullstrand. Allgemeine Lösung des statischen Einkörperproblems in der Einsteinschen Gravitationstheorie. *Arkiv för Matematik, Astronomi och Fysik*, 16:1, 1921.

[29] J. B. Hartle and S. W. Hawking. Path-Integral Derivation of Black-Hole Radiance. *Physical Review D*, 13(8):2188, Apr 1976.

[30] S. W. Hawking. Gravitational Radiation From Colliding Black Holes. *Physical Review Letters*, 43:1344, 1971. doi:10.1103/PhysRevLett.26.1344.

[31] S. W. Hawking. Particle Creation by Black Holes. *Communications in Mathematical Physics*, 43:199, Apr 1975. doi:10.1007/BF02345020.

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

[33] S. W. Hawking, G. F. R. Ellis, and P. V. Landshoff. *The Large Scale Structure of Space-Time*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1973.

[34] P. Hayden and J. Preskill. Black Holes as Mirrors: Quantum Information in Random Subsystems. *Journal of High Energy Physics*, 2007(09):120, Sep 2007. doi:10.1088/1126-6708/2007/09/120.

[35] W. Israel. Third Law of Black-Hole Dynamics: A Formulation and Proof. *Physical Review Letters*, 57:397–399, Jul 1986. doi:10.1103/PhysRevLett.57.397.

[36] J. M. Renes. *Quantum Information Theory: Concepts and Methods*. De Gruyter Oldenbourg, Berlin, Boston, 2022. doi:doi:10.1515/9783110570250.

[37] M. J. Jacquet, S. Weinfurtner, and F. König. The Next Generation of Analogue Gravity Experiments. *Philosophical Transactions A*, 378, Jul 2020. doi:10.1098/rsta.2019.0239.

[38] M. D. Kruskal. Maximal Extension of Schwarzschild Metric. *Physical Review*, 119:1743–1745, Sep 1960. doi:10.1103/PhysRev.119.1743.

[39] L. D. Landau and E. M. Lifschitz. *Lehrbuch der Theoretischen Physik: Hydrodynamik*. Akademie-Verlag, Berlin, 1966.

[40] L. D. Landau and E. M. Lifschitz. *Lehrbuch der Theoretischen Physik: Statistische Physik Teil 1, 8. Auflage*. Europa Lehrmittel, Haan-Gruiten, 1990.

[41] L. D. Landau and E. M. Lifschitz. *Lehrbuch der Theoretischen Physik: Mechanik, 14. Auflage*. Europa Lehrmittel, Haan-Gruiten, 2016.

[42] L. D. Landau and E. M. Lifschitz. *Lehrbuch der Theoretischen Physik: Klassische Feldtheorie, 12. Auflage*. Europa Lehrmittel, Haan-Gruiten, 2018.

[43] L. D. Landau and E. M. Lifschitz. *Lehrbuch der Theoretischen Physik: Quantenmechanik, 9. Auflage*. Europa Lehrmittel, Haan-Gruiten, 2019.

[44] G. Lemaître. L'Univers en Expansion. *Annales de la Société Scientifique de Bruxelles*, A53:51–85, 1933.

[45] S. Liberati, G. Tricella, and A. Trombettoni. The Information Loss Problem: An Analogue Gravity Perspective. *MDPI: Entropy*, 21:940, Sep 2019. doi:1099-4300/21/10/940.

[46] S. Liberati, G. Tricella, and A. Trombettoni. Back-Reaction in Canonical Analogue Black Holes. *MDPI: Applied Sciences*, 10(10):8868, Dec 2020. doi:10.3390/app10248868.

[47] J. Maldacena. Black Hole Entropy and Quantum Information. Presented as the 2nd 2022 Wolfgang Pauli Lecture, ETH Zurich, 2022.

[48] K. Martel and E. Poisson. Regular Coordinate Systems for Schwarzschild and other Spherical Spacetimes. *American Journal of Physics*, 69(4):476–480, 2001. doi:10.1119/1.1336836.

[49] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. W. H. Freeman, 1973.

[50] P. J. Morrison. Hamiltonian Description of the Ideal Fluid. *Reviews of Modern Physics*, 70:467–521, Apr 1998. doi:10.1103/RevModPhys.70.467.

[51] T. Needham. *Visual Complex Analysis*. Oxford University Press, 1997.

[52] E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence. Metric of a Rotating, Charged Mass. *Journal of Mathematical Physics*, 6(6):918–919, 1965. doi:10.1063/1.1704351.

[53] M. Novello, M. Visser, and G. Volovik. *Artificial Black Holes*. World Scientific, Singapore, 2002.

[54] T. Padmanabhan. Lessons from Classical Gravity about the Quantum Structure of Spacetime. *Journal of Physics: Conference Series*, 306, 2011. doi:10.1088/1742-6596/306/1/012001.

[55] D. N. Page. Information in Black Hole Radiation. *Physical Review Letters*, 71(23):3743, Jun 1993. doi:10.1103/PhysRevLett.71.3743.

[56] D. N. Page. Black Hole Information. In *Proceedings of the 5th Canadian Conference on General Relativity and Relativistic Astrophysics*, page p. 1. R. B. Mann and R. G. McLenaghan, 1994. arXiv:hep-th/9305040.

[57] D. N. Page. Time Dependence of Hawking Radiation Entropy. *Journal of Cosmology and Astroparticle Physics*, 2013(09):028, Sep 2013. doi:10.1088/1475-7516/2013/09/028.

[58] M. P. Painlevé. La Mécanique Classique et la Théorie de la Relativité. *C. R. Acad. Sci. (Paris)*, 173:677, Oct 1921.

[59] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang. Replica Wormholes and the Black Hole Interior. 2019. arXiv:hep-th/1911.11977, doi:10.48550/arXiv.1911.11977.

[60] R. Penrose. Structure of Space-Time. In DeWitt C. M., editor, *Battelle Rencontres*, New York, 1968. W. A. Benjamin.

[61] R. Penrose. Gravitational Collapse: the Role of General Relativity. *Rivista del Nuovo Cimento, Numero Speziale I*, 257, 1969.

[62] R. Penrose. Republication of: Conformal Treatment of Infinity. *General Relativity and Gravitation*, 43:901–922, 2011. doi:10.1007/s10714-010-1110-5.

[63] R. Penrose and H. Bondi. Zero Rest-Mass Fields Including Gravitation: Asymptotic Behaviour. *Proceedings of the Royal Society of London A*, 284(1397):159–203, 1965. doi:10.1098/rspa.1965.0058.

[64] R. Penrose and R. M. Floyd. Extraction of Rotational Energy from a Black Hole. *Nature Physical Science*, 229:177–179, February 1971.

[65] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. CRC Press, 1995.

[66] C. J. Pethick and H. Smith. *Bose-Einstein Condensation in Dilute Gases*. Cambridge University Press, 2004.

[67] M. Planck. *Vorlesungen über Thermodynamik, 11. Auflage*. De Gruyter, 1964.

[68] R. Renner and J. Wang. The Black Hole Information Puzzle and the Quantum de Finetti Theorem. 2021. arXiv:hep-th/2110.14653.

[69] J. Robertson. The Theory of Hawking Radiation in Laboratory Analogues. *Journal of Physics B*, 45:42, 2012.

[70] K. Schwarzschild. Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, 7:189–196, 1916.

[71] G. Szekeres. On the Singularities of a Riemannian Manifold. *Publicationes Mathematicae Debrecen*, 7:285–301, 1960.

[72] M. E. Taylor. *Partial Differential Equations I, 2nd edition*. Springer, 2011. doi:10.1007/978-1-4419-7055-8.

[73] W. G. Unruh. Notes on Black-Hole Evaporation. *Physical Review D*, 14:870–892, Aug 1976. doi:10.1103/PhysRevD.14.870.

[74] W. G. Unruh. Experimental Black-Hole Evaporation? *Physical Review Letters*, 46:1351, May 1981. doi:10.1103/PhysRevLett.46.1351.

[75] M. Visser. Acoustic Propagation in Fluids: an Unexpected Example of Lorentzian Geometry. 1993. arXiv:gr-qc/9311028.

[76] M. Visser. Essential and Inessential Features of Hawking Radiation. *International Journal of Modern Physics D*, 12(04):649–661, 2003. doi:10.1142/S0218271803003190.

[77] M. Visser and S. Weinfurtner. Vortex analogue for the equatorial geometry of the kerr black hole. *Classical and Quantum Gravity*, 22(12):2493, 2005. doi:10.1088/0264-9381/22/12/011.

[78] G. E. Volovik. *The Universe in a Helium Droplet*. Oxford University Press, Oxford, 2003.

[79] R. M. Wald. Gedanken Experiments to Destroy a Black Hole. *Annals of Physics*, 82(2):548–556, 1974. doi:10.1016/0003-4916(74)90125-0.

[80] R. M. Wald. On Particle Creation by Black Holes. *Communications in Mathematical Physics*, 45:9–34, Feb 1975. doi:10.1007/BF01609863.

[81] R. M. Wald. *General Relativity*. The University of Chicago Press, Chicago, 1984.

[82] D. Wallace. The Case for Black Hole Thermodynamics, Part I: Phenomenological Thermodynamics. 2017. arXiv:gr-gc/1710.02724.

[83] D. Wallace. The Case for Black Hole Thermodynamics, Part II: Statistical Mechanics. 2018. arXiv:gr-gc/1710.027225.

[84] D. Wallace. Why Black Hole Information Loss is Paradoxical. 2018. arXiv:gr-qc/1710.03783.