Gravity, Entropy & Holography

Oliver Friedrich, Varun Kushwaha

Fakultät für Physik, Ludwig-Maximilians Universität München,
Geschwister-Scholl-Platz 1, 80539 München, Germany

E-mail: oliver.friedrich@physik.lmu.de, varun.kushwaha@physik.uni-muenchen.de

Abstract. Test

Contents

1 Horizons, horizon entropy & the holographic principle - an incomplete overview **2**
1.1 Units & notation 2
1.2 Null surfaces and black holes 2
1.3 Key events in the development of horizon thermodynamics 4
1.4 Exercises for lecture 1 & feedback form 9

Part I: The classical holographic principle **10**

2 Bekenstein entropy and the Bekenstein bound **10**

3 Hawking radiation **12**
3.1 Scalar field in Minkowski space and decomposition into field modes 12
3.2 Particle detectors & why positive frequency modes are "particles" 14
3.3 Exercises for lecture 2 & feedback form 15
3.4 Unruh radiation 16
3.5 Hawking radiation: easy but slightly flawed derivation 18
3.6 Exercises for lecture 3 & feedback form 20
3.7 Spirit of Hawking's original derivation 20
3.8 Exercises for lecture 4 & feedback form 22

4 The Einstein equations as a 1st law **22**
4.1 Exercises for lecture 5 & feedback form 29

5 Entropy bounds and the holographic principle **29**
5.1 The spherical entropy bound 29
5.2 The covariant entropy bound (Bousso's bound) 32
5.2.1 A simplified version of the covariant entropy bound 32
5.2.2 Bousso's original definition of light-sheets and the covariant entropy bound 33
5.2.3 Space-like projection theorem 34
5.2.4 Application to flat Friedmann Universe 34
5.3 The species problem 36
5.4 The holographic principle 36
5.5 Exercises for lecture 6 & feedback form 37

Part II: Quantum implementations of the holographic principle **41**

6 Basics of quantum information **42**
6.1 Entanglement 1st law 44
6.2 Example: modular Hamiltonian and entropy perturbation in Rindler wedge 44

7 Quantum mereology (a philosophical backdrop of what's to come) **45**

8 Entanglement entropy of quantum fields across surfaces **47**
8.1 Exercises for lecture 7 & feedback form 50

9 Space from Hilbert space & Bulk entanglement gravity **50**
9.1 Original Cao & Carroll axioms 56
9.2 Exercises for lecture 8 & feedback form 57
9.3 Three derivations of Einstein's equations in different frameworks 57

10 AdS/CFT & error correcting codes **57**
10.1 anti-de Sitter space 57
10.2 Exercises for lecture 9 & feedback form 58
10.3 Conformal field theory, conformal boundary of AdS 58
10.4 The dictionary of the AdS/CFT correspondence 60
10.5 A simple toy model for AdS/CFT 62
10.6 Exercises for lecture 11 & feedback form 65

11 Overlapping qubits & holographic Weyl field in flat space **66**
11.1 Pauli algebra and Clifford algebra 66
11.2 overlapping qubits 68
11.3 Holographic Weyl field 69

A The scalar field as a collection of oscillators **77**

B The Weyl field as a collection of qubits **78**
B.1 Weyl field basics 78
B.2 Decomposition into qubits 79

C Penrose diagrams **80**

D Horizon definitions **80**

E Flow across a surface **80**

1 Horizons, horizon entropy & the holographic principle - an incomplete overview

1.1 Units & notation

In this lecture we will for the most part use natural units, where the reduced Planck constant $\hbar$, the speed of light c and Boltzmann's constant k are all set to 1. We will keep the appropriate factors of Newton's constant G in most expressions. The reason for this is that G has mass dimension

$$[G] = \frac{1}{\mathrm{mass}^2} \ , \tag{1.1}$$

so multiplication by G can change the mass dimension of an expression.

Throughout the script, boldface letter like e.g. $\boldsymbol{x}$ will represent 3-dimensional vectors, while usual lower case math font may represent (3+1)-dimensional vectors (e.g. $x = (x^0, x^1, x^2, x^3)$). Indices i, j, k, l will label the components of 3-dimensional vectors and -tensors while indices a, b, c, d will label the components of 4-vectors and -tensors. Occasionally, also greek indices α, β, μ, ν will label components of 4-dimensional objects.

1.2 Null surfaces and black holes

Consider a 4-dimensional manifold $\mathcal{M}$ equipped with coordinates $x = (x^0, x^1, x^2, x^3)$ and a pseudo-Riemannian line element

$$\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu \ , \tag{1.2}$$

where $g_{\mu\nu}$ are the components of the (pseudo-Riemannian) metric tensor on $\mathcal{M}$. Any real-values function $S : \mathcal{M} \to \mathbb{R}$ induces a family of hypersurfaces on $\mathcal{M}$ via

$$\Sigma_{S=r} := \{x \in \mathcal{M} \ \text{ s.t. } \ S(x) = r\} \ . \tag{1.3}$$

Except for potentially existing local extrema of the function S, these surfaces will be 3-dimensional (since the condition $S(x) = r$ "removes" one of the dimensions).

All "horizons" that we will talk about in this lecture are *null surfaces*. These are surfaces whose normal vectors are null vectors (i.e. light-like). At any location $x \in \Sigma_{S=r}$ the components of the normal covector to $\Sigma_{S=r}$ is given by $\partial_\mu S(x)$ (the normal vector would be $n^\mu = g^{\mu\nu}\partial_\nu S$), and in order for the normal vector to be null these components need to satisfy

$$g^{\mu\nu}\,\partial_\mu S\,\partial_\nu S = 0 \ . \tag{1.4}$$

We could choose a coordinate system where, e.g. , $S(x) \equiv x^1$. In this case, Equation 1.4 becomes a condition for the 11-component of the inverse metric, i.e.

$$g^{11} = 0 \ . \tag{1.5}$$

This will come in handy in a minute, when we try to identify null surfaces of black holes.

In general, a reason why null surfaces are interesting because they *bifurcate* a given spacetime into two regions $\mathcal{A}$ and $\mathcal{B}$ such that light-like geodesics cannot ("light rays") cannot travel from $\mathcal{A}$ to $\mathcal{B}$ (though they may still be able to travel from $\mathcal{B}$ to $\mathcal{A}$). This in itself is not a unique feature of null surfaces, because any space-like boundary that separates $\mathcal{M}$ into a "past" region and a "future" region would also act as a on-way-surface for light. But for a null-surface the region $\mathcal{B}$ may be one with an infinite past and future, i.e. an observer may have lived in $\mathcal{B}$ since forever and continue to live there forever but still never receive signals from the region $\mathcal{A}$. In Section 3 we will be more nuanced about our definition of "horizon" and distinguish between *event horizons*, *particle horizons* and *Killing horizons*. But for now let us stick with the notion that a horizon is a null surface and identify such a surface in a black hole spacetime.

We consider a stationary, non-rotating black hole of mass M. The latter is e.g. described by the *Schwarzschild metric*, whose line element is

$$\mathrm{d}s^2 = -\left(1 - \frac{r_s}{r}\right)\mathrm{d}t^2 + \frac{1}{1-\frac{r_s}{r}}\,\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2 \ , \tag{1.6}$$

where we defined the *Schwarzschild radius* as

$$r_s := 2GM \ . \tag{1.7}$$

Note that in the above coordinates the metric is diagonal, such that $g^{rr} = 1/g_{rr}$. Remembering Equation 1.5, this allows us to identify as null the surface where $r = r_s$ which is of course the event horizon of the black hole.

A coordinate system that makes the null-nature of the hypersurface $r \equiv r_s$ more transparent are *Kruskal-Szekeres coordinates*. They are given in terms of the Schwazschild coordinates t and r as

$$T \ = 2r_s\left|1-\tfrac{r}{r_s}\right|^{\frac{1}{2}}\exp\left(\tfrac{r}{2r_s}\right)\cdot\begin{cases}\sinh\left(\frac{t}{2r_s}\right) & \text{for } r > r_s\\ \cosh\left(\frac{t}{2r_s}\right) & \text{for } r \leq r_s\end{cases} \tag{1.8}$$

$$R = 2r_s\left|1-\tfrac{r}{r_s}\right|^{\frac{1}{2}}\exp\left(\tfrac{r}{2r_s}\right)\cdot\begin{cases}\cosh\left(\frac{t}{2r_s}\right) & \text{for } r > r_s\\ \sinh\left(\frac{t}{2r_s}\right) & \text{for } r \leq r_s\end{cases} , \tag{1.9}$$

and with these coordinates the line element becomes

$$\mathrm{d}s^2 = \frac{r_s}{r(R,T)}\exp\left(-\frac{r(R,T)}{r_s}\right)\left(-\mathrm{d}T^2 + \mathrm{d}R^2\right) + r(R,T)^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right) \ . \tag{1.10}$$

(We do not replace r by its explicit expression in terms of R and T because it is tedious and because it doesn't offer any additional insight at the moment.) A convenient feature of Kruskal-Szekeres coordinates is the fact that T is a timelike coordinate and R is a spacelike coordinate throughout spacetime. In contrast the Schwarzschild coordinates r and t switch their roles and temporal and spatial coordinates when crossing the horizon. Additionally, any light-like geodesic that is radial, i.e. for which $\mathrm{d}\theta = 0 = \mathrm{d}\phi$, will have $\mathrm{d}T = \pm\mathrm{d}R$. Hence, such geodesics will appear as lines of 45° in a diagram of R and T. Together with the following exercise this enables us to see that $r \equiv r_s$ is indeed a null surface.

Exercise 1

Show that the black hole horizon $r \equiv r_s$ is located at the diagonal $T = R$ of a diagram of the Kruskal-Szekeres coordinates T and R. Hint: you cannot directly take the limit $r \to r_s$ in Equations 1.8 and 1.9. Find a useful way to combine the two equations first.

In Figure 1 we compare the locations of the horizon $r = r_s$ and the location of an observer outside the horizon in Schwarzschild and Kruskal-Szekeres coordinates. In the latter, the

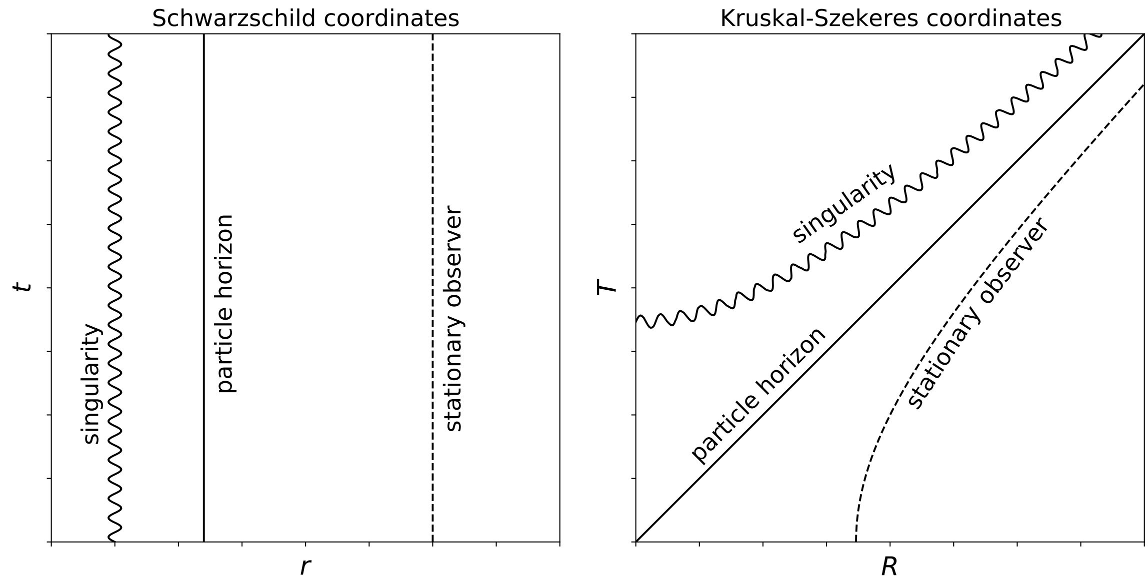

Figure 1: Sketching the location of singularity, particle horizon and a stationary observer outside the horizon in Schwarzschild coordinates and Kruskal-Szekeres coordinates of a non-rotating black hole. The Kruskal-Szekeres coordinate T is time-like everywhere in spacetime, and R is space-like everywhere in spacetime. Another benefit of Kruskal-Szekeres coordinates is the fact that light-like (and radial) geodesics would appear as lines of a 45° angle wrt. the coordinate axes.

horizon does not in fact appear as a very special surface. The fact that g_{rr} diverges in the Schwarzschild picture is thus only a so called coordinate singularity, and one can for example show that spacetime curvature is completely well behaved at the horizon surface. The only true singularity (i.e. point where spacetime curvature itself diverges) appears at $r = 0$. In Kruskal-Szekeres coordinates this is a future boundary for observers entering the black hole.

Exercise 2

Derive an equation $T = R(T)$ for the location of the black hole singularity (i.e. for the "wavy" line in Figure 1) in Kruskal-Szekeres coordinates.

1.3 Key events in the development of horizon thermodynamics

It is almost common knowledge that a black hole with horizon area A_{bh} supposedly carries an entropy

$$S_{\mathrm{bh}} = \frac{A_{\mathrm{bh}}}{4\,\ell_{\mathrm{Planck}}^2}\ , \tag{1.11}$$

where ℓ_{Planck} is the Planck-length. And sentences like the following are popular when trying to motivate this entropy assignment: "If some amount of matter, together with the entropy that it carries, crosses the horizon of a black hole, then this decreases the entropy of the observable Universe and thus breaks the 2nd law of thermodynamics. The horizon entropy then restores the 2nd law."

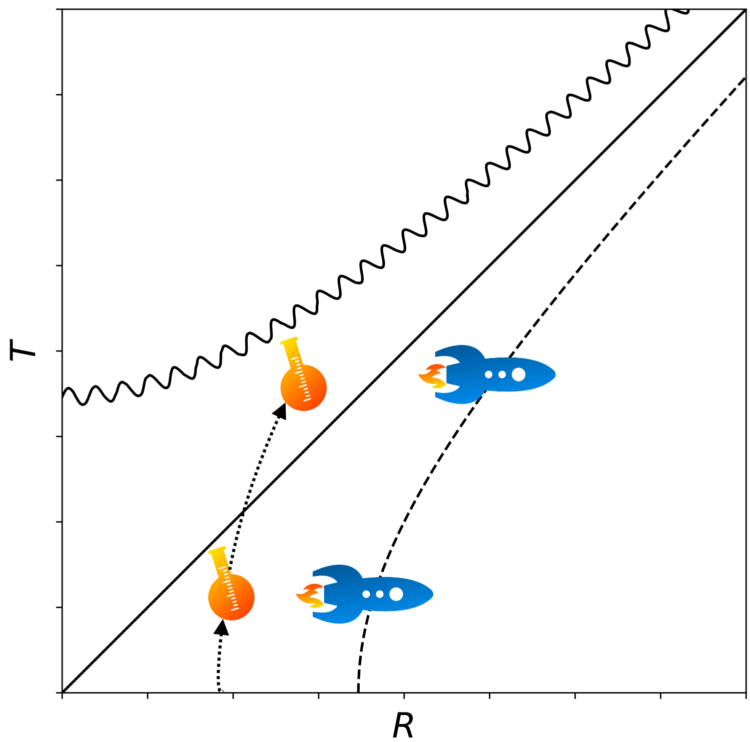

Figure 2: Information can travel at most at the speed of light, which corresponds to lines of 45° in Kruskal-Szekeres coordinates. When a container of matter - and the corresponding entropy - moves across a black hole horizon, then an outside observer loses the ability to receive any information about that container. Does that constitute a violation of the 2nd law of thermodynamics? This question is somewhat missing the point of the holographic principle.

We sketch this situation in Figure 2, and in Kruskal-Szekeres coordinates such a proclaimed breakdown of the 2nd law becomes questionable. At any T = const. surface of spacetime, the entropy of the matter falling into the black hole is still perfectly present in the Universe (except potentially for the moment, when the matter falls onto the singularity, but let us not bother with this moment of which there is no common sense understanding yet). Claiming that matter crossing the horizon breaks the 2nd law is a somewhat observer-centric point-of-view (which, to be honest, your lecturer holds). But it is also somewhat missing the point of the assignment of entropy to a black hole horizon! The fact that horizons seem to carry entropy is interesting not primarily because it restores the 2nd law, but e.g. for the following reasons:

A) The fact Equation 1.11 restores the 2nd law for outside observers means that for some reason the horizon area A_{bh} is an upper bound for how much entropy has been thrown into the horizon! This is actually how Bekenstein first motivated his entropy formula: he tried to through entropy into a black hole in a way that minimizes the growth of A_{bh} (cf. Section 2).

B) The entropy assignment in Equation 1.11 together with the horizon temperature that was later derived by Hawking actually makes it so that the combined system of black holes and ordinary matter don't just satisfy a 2nd law but also a *1st law*! And it was later shown by Jacobson, that demanding the validity of this 1st law for all local horizons is in fact *equivalent to the full Einstein equations!* We will investigate this in more detail in Section 4, but let us already get a glimpse of such a 1st law by looking at the Schwarzschild black hole. The inner energy of the latter will be given by its mass M, while its entropy is given by

$$S = \frac{A_{\mathrm{bh}}}{4G} = \frac{4\pi r_s^2}{4G} = 4\pi G M^2 \ . \tag{1.12}$$

We will further more see in Section 3, that the temperature of the Hawking radiation emitted by the black hole horizon is given by

$$T = 1/8\pi G M \ . \tag{1.13}$$

If we now drop an infinitesimal mass $\mathrm{d}M$ into the black hole, then temperature, entropy and inner energy satisfy the equation

$$T\mathrm{d}S = \mathrm{d}M \ . \tag{1.14}$$

This equation describes the reaction of spacetime geometry to a flow of matter across the horizon, but it also seems to represent a 1st law of horizon thermodynamics.

C) The fact that horizon area (and not e.g. the volume enclosed by the horizon) bounds the entropy inside the horizon seems to be in conflict with standard particle physics! The entropy that can be carried by a quantum field inside a given volume V is typically proportional to V and not to the boundary area of that volume.

D) Hawking radiation is an effect of quantum field theory on a fixed curved spacetime background. In particular, its derivation does not require that this background spacetime satisfies the Einstein equations. How then does quantum field theory know to produce horizon radiation with a temperature that exactly makes the Einstein equations look like a 1st law of thermodynamics? This apparent coincidence has sparked the hope that horizon thermodynamics is a pathway towards understanding quantum gravity itself. This hope is e.g. summarized by the following quote from professor Padmanabhan [1]:

> *"... if spacetime can be hot, it must have microstructure."*

In Table 1 you can find a (very incomplete!) timeline of important scientific work that lead to the above realisations A) - D).

Exercise 3

In Appendix A it is shown that a scalar quantum field in side a box of side length ℓ_{box} can be interpreted as a collection of quantum harmonic oscillators.

a) Show that the number of these oscillators scales as the volume of the box, i.e.

$$N_{\mathrm{osc}} \sim \ell_{\mathrm{box}}^3 \ .$$

The dimension of the Hilbert space of a quantum harmonic oscillator is infinite, so the Hilbert space of a quantum field in a box must also be infinite. Assume that quantum gravitational effects regularize this Hilbert space dimension such that each oscillator only lives in a Hilbert space of dimension d (independent of Fourier mode $\boldsymbol{k}$).

b) *What is the dimension of the full Hilbert space of the quantum field as a function of d?*

A generalized notion of a quantum state is given by a density matrix. *This is a Hermitian operator $\hat{\rho}$ that is positive (semi-)definite and has* $\mathrm{Tr}\,\hat{\rho} = 1$. *If $\hat{\rho}$ is defined on a Hilbert space $\mathcal{H}$ of dimension d and of $\{\lambda_1\ ,\ \dots\ ,\ \lambda_d\}$ are the eigenvalues of $\hat{\rho}$, then the von-Neumann entropy of $\hat{\rho}$ is defined as*

$$S(\hat{\rho}) := -\sum_j \lambda_j \ln(\lambda_j)\ . \tag{1.15}$$

c) *Show that the maximum entropy that any $\hat{\rho}$ on $\mathcal{H}$ can have is* $\ln(d)$. (Hint: the condition $\mathrm{Tr}(\hat{\rho}) = 1$ means that $\sum_j \lambda_j = 1$, which you can add to your optimization problem via the method of Lagrange multipliers.)

d) *What is the maximum entropy of a scalar quantum field in a box (assuming that the dimension of the individual oscillators is again regularized to be d)? Discuss your result in the light of Bekenstein's proposal that the maximum entropy that can be stored in a given volume is proportional to the boundary area of that volume.*

Table 1 (incomplete!) timeline of key developments in horizon thermodynamics

1971	Demonstration that no classical process can reduce the sum of black hole horizon areas: • Hawking 1971, "Gravitational Radiation from Colliding Black Holes"
1972-1974	Bekenstein proposes that black holes have an entropy, and that this entropy is proportional to horizon area: • Bekenstein 1972, "Black holes and the second law" • Bekenstein 1973, "Black holes and entropy" • Bekenstein 1974, "Generalized second law of thermodynamics in black-hole physics"
1974	Derivation that black hole horizons have a temperature and emit thermal radiation: • Hawking 1974, "Black hole explosions?"
1977	Discovery that entropy and temperature can also be assigned to the cosmological horizon: • Gibbons & Hawking 1977, "Cosmological event horizons, thermodynamics, and particle creation"
1981	It is proposed that the maximum entropy that can be accumulated in a spherical volume is bounded by the surface area of that volume: • Bekenstein 1981: "Universal upper bound on the entropy-to-energy ratio for bounded systems"
1995	Discovery that the entire Einstein equations follow from demanding the validity of a 1st law for the combined system of horizons & outside matter: • Jacobson 1995: "Thermodynamics of Spacetime: The Einstein Equation of State" • many works of Padmanabhan; see e.g. `https://arxiv.org/pdf/0706.1654` for an overview
1999	Formulation of a generally covariant version of Bekenstein's original entropy bound (the *holographic principle*): • Bousso 1999: "A Covariant Entropy Conjecture"

1.4 Exercises for lecture 1 & feedback form

Lecture 1 is accompanied by exercises 1 - 3. Also, you can find the feedback form for the lecture here:
https://cloud.physik.lmu.de/index.php/apps/forms/s/eLefQDX4LrZjweCsKtmRdD8S

Part I: The classical holographic principle

2 Bekenstein entropy and the Bekenstein bound

In [2] Bekenstein proposed to assign an entropy

$$S_{\rm bh} = \eta A_{\rm bh}/G \tag{2.1}$$

to the horizon of black holes, where $A_{\rm bh}$ and η a constant of order unity (in natural units). The proportionality to $A_{\rm bh}$ was motivated by the findings of e.g. [3] that black hole area cannot be decreased via classical processes.

Bekenstein then proposed that the is some value for η such that the above definition of black hole entropy resurrects the 2nd law of thermodynamics in the sense that the sum of black hole entropy and the entropy of matter systems residing outside black holes never decreases:

$$\mathrm{d}(S_{\rm bh} + S_{\rm matter}) > 0 \ . \tag{2.2}$$

The above condition is called the *generalised 2nd law*. In order to estimate the value of η that is required for the validity of this law, Bekenstein pursued the thought experiment of shooting a beam of radiation with a given entropy $S_{\rm rad}$ into a black hole in such a way that the increase in horizon area is minimized.

Imagine that we were indeed to shoot a light-ray of typical frequency ω and total energy $\mathrm{d}E$ into a Schwarzschild black hole of Mass M. This changes the horizon area and the entropy of the black hole as

$$\mathrm{d}A_{\rm bh} = 4\pi \mathrm{d}(4M^2G^2) = 32\pi G^2 M \mathrm{d}E \tag{2.3}$$

$$\Rightarrow \mathrm{d}S_{\rm bh} = 32\pi\eta G M \mathrm{d}E \ , \tag{2.4}$$

where we have assumed that $\mathrm{d}E \ll M$, such that we can e.g. ignore effects from recoil. Note that the above increase in area is independent of the entropy of the light-beam. We now want to try to sink as much entropy as possible into the black hole, while keeping the energy of the beam fixed to $\mathrm{d}E$ and still having ω as the typical frequency of the beam. This means that we prepare the beam such that its photons are in a thermal state with some temperature T, and in order to still have the spectral distribution of these photons peak at ω this temperature must be given by $T \sim \omega$. We can hence consider our beam as a photon gas in a thermal state, and the entropy of such a gas is given by

$$\mathrm{d}S_{\rm beam} = \frac{4}{3}\frac{\mathrm{d}E}{T} \sim \frac{4}{3}\frac{\mathrm{d}E}{\omega} \tag{2.5}$$

and consequently, the matter entropy that is lost when the beam crosses the horizon is

$$\mathrm{d}S_{\rm matter} \sim -\frac{4}{3}\frac{\mathrm{d}E}{\omega} \ . \tag{2.6}$$

In order to focus our beam into the black hole we need to be in (or at least close to) the limit of geometric optics, i.e. the typical wavelength $\lambda \sim 1/\omega$ needs to be (much) smaller than the typical scale $r_s = 2GM$ over which the gravitational field varies. If this was not the case, then diffraction around the black hole would make it so that part of our beam would scatter and not enter the black hole. So we require

$$GM \gg 1/\omega \ . \tag{2.7}$$

Putting 2.4, 2.6 and 2.7 together yields

$$\mathrm{d}S_{\mathrm{bh}} + \mathrm{d}S_{\mathrm{matter}} > \mathrm{d}S_{\mathrm{bh}} - \frac{4}{3}GM\mathrm{d}E = \left(32\pi\eta - \frac{4}{3}\right)GM\mathrm{d}E \ . \tag{2.8}$$

As long as $\eta \sim \mathcal{O}(1)$, this is indeed positive.

However, there is a way how one can minimize the increase of the horizon area even further while keeping the same beam energy $\mathrm{d}E$, and this is to consider a rotating black hole. Such a black hole is described by the Kerr metric, whose line element is given by [4]

$$\begin{aligned}\mathrm{d}s^2 = &-\left(1 - \frac{2Mr}{\rho^2}\right)\mathrm{d}t^2 - \frac{4M^2ar\sin^2\theta}{\rho^2}\mathrm{d}t\mathrm{d}\phi + \frac{\rho^2}{\Delta}\mathrm{d}r^2 + \rho^2\mathrm{d}\theta^2 \\ &+ \left(r^2 + M^2a^2 + \frac{2M^3a^2r\sin^2\theta}{\rho^2}\right)\sin^2\theta\,\mathrm{d}\phi^2 \ .\end{aligned} \tag{2.9}$$

Here M is the total mass energy of the black hole, while a is given in terms of the angular momentum $\boldsymbol{L}$ of the black hole as

$$a := \frac{|\boldsymbol{L}|}{M^2} \ . \tag{2.10}$$

Furthermore, we have defined

$$\rho^2 := r^2 + M^2a^2\cos^2\theta \tag{2.11}$$

$$\Delta := r^2 - 2Mr + M^2a^2 \ . \tag{2.12}$$

It is straight forward to show that also in this metric $g^{rr} = 1/g_{rr}$. This means that we can find a null surface at

$$\begin{aligned} 0 &= \frac{\Delta}{\rho^2} \\ \Rightarrow 0 &= r^2 - 2Mr + M^2a^2 \\ \Rightarrow r &= M\cdot\left(1 \pm \sqrt{1-a^2}\right) \ .\end{aligned} \tag{2.13}$$

If we choose the plus sign in the above expression then this gives the correct Schwarzschild limit, so we take $r \equiv M\cdot\left(1+\sqrt{1-a^2}\right)$ to be the null surface that represents the horizon of the Kerr black hole.

Exercise 4

Show that the horizon area of the Kerr black hole is given by

$$\begin{aligned} A_{\mathrm{bh}}(M,a) &= 8\pi M^2\left(1+\sqrt{1-a^2}\right) \\ &= 8\pi M^2\left(1+\sqrt{1-\boldsymbol{L}^2/M^4}\right) \ .\end{aligned} \tag{2.14}$$

Hint: *you can obtain the 2D induced metric on the horizon by setting* $\mathrm{d}t = 0 = \mathrm{d}r$ *in the line element. Then integrate* $\int\sqrt{g_{2\mathrm{D}}}\,\mathrm{d}\theta\mathrm{d}\phi$*.*

Exercise 5
With the result from Exercise 4, show that the infinitesimal increase in $A_{\rm bh}$ *by shooting our light beam of energy* $\mathrm{d}E$ *into the black hole is given by*

$$\mathrm{d}A_{\rm bh} = 16\pi M\left(1+\frac{1}{\sqrt{1-a^2}}\right)\mathrm{d}E - \frac{8\pi}{M^2}\frac{\boldsymbol{L}\mathrm{d}\boldsymbol{L}}{\sqrt{1-a^2}}\ . \tag{2.15}$$

According to Equation 2.15, we can minimize the area increase (and thus the entropy increase of the black hole) if we maximize the angular momentum that our beam adds to the black hole. In order to achieve that, we need to shoot the beam inside the equatorial plane of the black hole in such a way, that is is only barely captured by the black hole. Bekenstein shows that this strategy is most successful for extremal Kerr black holes, i.e. when $|\boldsymbol{L}| \to M^2$. In this case he shows that

$$\mathrm{d}S_{\rm bh} = 16\pi(1-\sqrt{3}/2)\eta G M \mathrm{d}E\ . \tag{2.16}$$

For $\eta \sim \mathcal{O}(1)$ this still doesn't break the generalised 2nd law.

Exercise 6
In table I of
https://dcc.ligo.org/public/0122/P1500218/014/PhysRevLett.116.241102.pdf
you find information about GW150914 *which was the first gravitational wave event detected by LIGO. This event is thought to have been caused by the merger of two black holes. According to the analysis of the LIGO team, the masses of these black holes were*

$$m_1 \approx 36\, M_\odot\ ,\ \ m_2 \approx 29\, M_\odot$$

while their dimensionless spin magnitudes were determined to be (with quite large uncertainty)

$$a_1 \approx 0.32\ ,\ \ a_2 \approx 0.57\ .$$

The final, merged black hole then had mass and spin magnitude

$$M \approx 62\, M_\odot\ ,\ \ s \approx 0.67\ .$$

Calculate the total horizon area before and after the merger. Is the generalized 2nd law consistent with the LIGO findings?

3 Hawking radiation

In this section we follow the textbooks [4, 5] as well as the original literature [6–9] to understand Hawking radiation as well as related phenomena.

3.1 Scalar field in Minkowski space and decomposition into field modes

Consider a real, massless scalar field in Minkowski space. Such a field will be subject to the Klein-Gordon equation

$$\Box\phi = 0\ . \tag{3.1}$$

There is a complete set of basis functions that satisfy this equation, i.e. the plane waves

$$f_{\boldsymbol{k}}(t,\boldsymbol{x}) = \frac{1}{\sqrt{2\omega_{\boldsymbol{k}}}}e^{-i\omega_{\boldsymbol{k}}t+i\boldsymbol{k}\cdot\boldsymbol{x}}\ ,\ \ g_{\boldsymbol{k}}(t,\boldsymbol{x}) = \frac{1}{\sqrt{2\omega_{\boldsymbol{k}}}}e^{i\omega_{\boldsymbol{k}}t+i\boldsymbol{k}\cdot\boldsymbol{x}}\ , \tag{3.2}$$

where $\omega_{\boldsymbol{k}} = |\boldsymbol{k}|$, and where we have chosen a standard normalization $1/\sqrt{2\omega_{\boldsymbol{k}}}$. The functions $f_{\boldsymbol{k}}$ are typically referred to as the *positive frequency* solutions, because they are eigenfunctions of $i\partial/\partial t$ with positive eigenvalue $\omega_{\boldsymbol{k}}$. Similarly, the $g_{\boldsymbol{k}}$ are called *negative frequency* solutions. We can decompose a general solution ϕ in terms of this basis as

$$\begin{aligned}\phi(t,\boldsymbol{x}) &= \int \frac{\mathrm{d}^3k}{(2\pi)^3}\ \{a_{\boldsymbol{k}} f_{\boldsymbol{k}}(t,\boldsymbol{x}) + b_{\boldsymbol{k}} g_{\boldsymbol{k}}(t,\boldsymbol{x})\} \\ &= \int \frac{\mathrm{d}^3k}{(2\pi)^3}\ \left\{a_{\boldsymbol{k}} e^{-i\omega_{\boldsymbol{k}} t} + b_{\boldsymbol{k}} e^{i\omega_{\boldsymbol{k}} t}\right\}\ e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\ . \end{aligned} \tag{3.3}$$

From this it is clear that $\left(a_{\boldsymbol{k}} e^{-i\omega_{\boldsymbol{k}} t} + b_{\boldsymbol{k}} e^{i\omega_{\boldsymbol{k}} t}\right)/\sqrt{2\omega_{\boldsymbol{k}}}$ is the Fourier transform $\tilde{\phi}(t,\boldsymbol{k})$ of the field ϕ. At the same time, since ϕ is real, we know that $\tilde{\phi}(t,\boldsymbol{k}) = \tilde{\phi}(t,-\boldsymbol{k})^*$ and thus

$$\begin{aligned} a_{\boldsymbol{k}} e^{-i\omega_{\boldsymbol{k}} t} + b_{\boldsymbol{k}} e^{i\omega_{\boldsymbol{k}} t} &= a^*_{-\boldsymbol{k}} e^{i\omega_{\boldsymbol{k}} t} + b^*_{-\boldsymbol{k}} e^{-i\omega_{\boldsymbol{k}} t} \\ \Rightarrow a_{\boldsymbol{k}} &= b^*_{-\boldsymbol{k}}\ . \end{aligned} \tag{3.4}$$

We can use this to re-express ϕ as

$$\begin{aligned}\phi(t,\boldsymbol{x}) &= \int \frac{\mathrm{d}^3k}{(2\pi)^3}\ \frac{1}{\sqrt{2\omega_{\boldsymbol{k}}}} \left\{a_{\boldsymbol{k}} e^{-i\omega_{\boldsymbol{k}} t} + a^*_{-\boldsymbol{k}} e^{i\omega_{\boldsymbol{k}} t}\right\}\ e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \\ &= \int \frac{\mathrm{d}^3k}{(2\pi)^3}\ \frac{1}{\sqrt{2\omega_{\boldsymbol{k}}}} \left\{a_{\boldsymbol{k}} e^{-i\omega_{\boldsymbol{k}} t + i\boldsymbol{k}\cdot\boldsymbol{x}} + a^*_{\boldsymbol{k}} e^{i\omega_{\boldsymbol{k}} t - i\boldsymbol{k}\cdot\boldsymbol{x}}\right\}\ . \end{aligned} \tag{3.5}$$

In the following, symmetries will often allow us to study a given situation in a 1+1 dimensional metric as opposed to a 1+3 dimensional one. In that case, the above decomposition becomes

$$\begin{aligned}\phi(t,\boldsymbol{x}) &= \int_{-\infty}^{\infty} \frac{\mathrm{d}k}{(2\pi)}\ \frac{1}{\sqrt{2\omega_k}} \left\{a_k e^{-i\omega_k t + ikx} + a^*_k e^{i\omega_k t - ikx}\right\} \\ &= \int_0^{\infty} \frac{\mathrm{d}k}{(2\pi)}\ \frac{1}{\sqrt{2\omega_k}} \left\{a_k e^{-i\omega_k (t-x)} + a^*_k e^{i\omega_k (t-x)}\right\} \\ &\quad + \int_{-\infty}^{0} \frac{\mathrm{d}k}{(2\pi)}\ \frac{1}{\sqrt{2\omega_k}} \left\{a_k e^{-i\omega_k (t+x)} + a^*_k e^{i\omega_k (t+x)}\right\}\ . \end{aligned} \tag{3.6}$$

The last line of Equation 3.6 warrants the definition of the so called light-cone coordinates

$$u := t - x\ ,\ \ v := t + x \tag{3.7}$$

in terms of which the decomposition of ϕ becomes

$$\begin{aligned}\phi(t,\boldsymbol{x}) &= \int_0^{\infty} \frac{\mathrm{d}k}{(2\pi)}\ \frac{1}{\sqrt{2\omega_k}} \left\{a_k e^{-i\omega_k u} + a^*_k e^{i\omega_k u}\right\} \\ &\quad + \int_{-\infty}^{0} \frac{\mathrm{d}k}{(2\pi)}\ \frac{1}{\sqrt{2\omega_k}} \left\{a_k e^{-i\omega_k v} + a^*_k e^{i\omega_k v}\right\}\ . \end{aligned} \tag{3.8}$$

This classical field motivates the definition of a quantum scalar field, where a_k and a^*_k are replaced by operators $\hat{a}_k$ and $\hat{a}^\dagger_k$ that satisfy the commutation relations [5]

$$[\hat{a}_k, \hat{a}^\dagger_{k'}] = \delta_D(k - k')\ . \tag{3.9}$$

For positive k the action of the operator $\hat{a}_k$ is to *annihilate* a right-moving particle of energy ω_k while for negative k $\hat{a}_k$ annihilates a left-moving particle of the same energy. In the following section we explore, why the positive frequency excitation are indeed considered the *particles* of a quantum field.

3.2 Particle detectors & why positive frequency modes are "particles"

Consider a laboratory on board a spaceship that is traveling along a world line $x(\tau)$, where τ is the proper time in the spaceship and x are coordinates in some spacetime coordinate system. The laboratory contains a detector with the purpose of detecting a scalar field ϕ (which we take to be a classical, i.e. non-quantum field for now). We follow [8] (see also [9]) in modeling such a detector as a quantum system with Hamiltonian in the lab frame given by

$$\hat{H} = \hat{H}_0 + g\phi \ , \tag{3.10}$$

where $\hat{H}_0$ is the self-Hamiltonian of the detector and the term $g\phi$ represents the coupling to ϕ with coupling strength g. For definiteness we assume that the detector consists of a single, non-relativistic quantum particle moving inside some potential, i.e.

$$\hat{H}_0 = \frac{\hat{\boldsymbol{P}}^2}{2m} + V(\hat{\boldsymbol{Q}}) \ , \tag{3.11}$$

where $\hat{\boldsymbol{P}} = (\hat{P}_1, \hat{P}_2, \hat{P}_3)$ and $\hat{\boldsymbol{Q}} = (\hat{Q}_1, \hat{Q}_2, \hat{Q}_3)$ are the momentum and position operators of the particle. This self-Hamiltonian $\hat{H}_0$ will have some ground state with energy E_0 and wave function $\langle \boldsymbol{q}|\psi_0\rangle = \psi_0(\boldsymbol{q})$ as well as higher energy eigenstates with wave functions $\langle \boldsymbol{q}|\psi_n\rangle = \psi_n(\boldsymbol{q})$.

When we switch on the detector at proper time τ_0, we assume it to start in the state $|\psi_0\rangle$. At some later time τ we switch the detector off again, and if we find it in a state $|\psi_n\rangle$ different from $|\psi_0\rangle$ then we consider this a detection.

Exercise 7

Show that for small coupling (i.e. $|g\phi| \ll |\hat{H}_0|$) the transition amplitude from the initial state $|\psi_0\rangle$ to a different state $|\psi_n\rangle$ is approximated by

$$\langle\psi_n|\, \hat{U}(\tau,\tau_0)\, |\psi_0\rangle \propto -ig \int_{\tau_0}^{\tau} \mathrm{d}\tau' \int \mathrm{d}^3q \ \psi_n(\boldsymbol{q})^* \psi_0(\boldsymbol{q}) \phi(\boldsymbol{q},\tau') \ e^{i\tau'(E_n - E_0)} \ , \tag{3.12}$$

where the time evolution operator is

$$U(\tau,\tau_0) = \mathcal{T} \exp\left(-i(\tau-\tau_0)\hat{H}_0 - ig\int_{\tau_0}^{\tau} \mathrm{d}\tau' \ \phi(\boldsymbol{Q},\tau')\right) \ . \tag{3.13}$$

You can e.g. proceed as follows:

A) *Change from the Schrödinger picture to the* interaction picture *by defining the states*

$$|\psi(\tau)\rangle_{\mathrm{I}} := \exp(i\tau\hat{H}_0)\, |\psi(\tau)\rangle \ , \tag{3.14}$$

where $|\psi(\tau)\rangle = \hat{U}(\tau,\tau_0)\, |\psi(\tau_0)\rangle$ is the usual time evolution of quantum states in the Schrödinger picture. Show that $|\psi(\tau)\rangle_{\mathrm{I}}$ satisfies the equation

$$i\partial_\tau\, |\psi(\tau)\rangle_{\mathrm{I}} = g\phi_{\mathrm{I}}\, |\psi(\tau)\rangle_{\mathrm{I}} \ , \tag{3.15}$$

where $\phi_{\mathrm{I}} = e^{i\tau\hat{H}_0}\phi e^{-i\tau\hat{H}_0}$.

B) *At each time τ, the state $|\psi_0(\tau)\rangle_{\mathrm{I}}$ can be decomposed in terms of the (time independent) basis $\{|\psi_m\rangle\}$ as*

$$|\psi_0(\tau)\rangle_{\mathrm{I}} = \sum_{m=0}^{\infty} c_m(\tau)\,|\psi_m\rangle \ . \tag{3.16}$$

Show that the coefficient $c_n(\tau)$ satisfies the equation

$$i\partial_\tau c_n = g\sum_m \langle\psi_n|\,\phi\,|\psi_m\rangle\, e^{i\tau(E_n-E_m)} c_m(\tau) \ . \tag{3.17}$$

C) *Why can you assume that $c_0(\tau) = 1 + \mathcal{O}(g)$ and $c_m(\tau) = \mathcal{O}(g)$ for $m > 0$? With that assumption, neglect all terms in Equation 3.17 that are $\mathcal{O}(g^2)$. Integrate the remaining equation to obtain 3.12.*

Inside the lab frame we can decompose ϕ into positive and negative frequency modes, as we had done in Section 3.1. Together with your result from Exercise 7 this means that

$$\begin{aligned}&\langle\psi_n|\,\hat{U}(\tau+\Delta\tau,\tau)\,|\psi_0\rangle \approx\\ &-ig\int\frac{\mathrm{d}^3k}{(2\pi)^3}\,c_{\boldsymbol{k}}\left[\int_\tau^{\tau+\Delta\tau}\mathrm{d}\tau'\ e^{i\tau'(E_n-E_0-\omega_{\boldsymbol{k}})}\right]\left[\int\mathrm{d}^3q\ \psi_n(\boldsymbol{q})^*\psi_0(\boldsymbol{q})\,e^{+i\boldsymbol{k}\boldsymbol{q}}\right]\\ &-ig\int\frac{\mathrm{d}^3k}{(2\pi)^3}\,c^*_{\boldsymbol{k}}\left[\int_\tau^{\tau+\Delta\tau}\mathrm{d}\tau'\ e^{i\tau'(E_n-E_0+\omega_{\boldsymbol{k}})}\right]\left[\int\mathrm{d}^3q\ \psi_n(\boldsymbol{q})^*\psi_0(\boldsymbol{q})\,e^{-i\boldsymbol{k}\boldsymbol{q}}\right] \ .\end{aligned} \tag{3.18}$$

Note that the integrands in the time integrals of the above expression are potentially highly oscillatory. Because $E_n - E_0$ is positive, these oscillations can only be canceled by positive frequency modes with $\omega_{\boldsymbol{k}} \approx E_n - E_0$. Hence, if the detector runtime is long enough, then only positive frequency modes can contribute to a detection. This is the reason why - once we transition to a quantum theory of ϕ - the positive frequency excitations of the field are considered “particles”. The detection event we discussed above then corresponds to the absorption of such a particle by the detector (or more specifically in our case: by the particle in the detector).

An important detail here is the following: The modes that contribute to a detection are the positive frequency modes *with respect to the lab frame.* These do not need to coincide with positive frequency modes that were defined wrt. some global coordinate system of spacetime. In particular, it may be that a vacuum state that does not contain any particles (i.e. any positive frequency excitations) wrt. such global coordinates still contains positive frequency modes wrt. to some suitably defined detector frame. In the next section, we will investigate an example of this.

Exercise 8

If the spaceship with our detector was moving on an accelerating trajectory, then the lab frame we discussed above cannot be Minkowskian. Assume that at each time τ the acceleration vector in the lab frame is $\boldsymbol{a}(\tau)$. How would this change the Hamiltonian of the detector (in a Newtonian limit, where the acceleration is small)?

3.3 Exercises for lecture 2 & feedback form

Lecture 2 is accompanied by exercises 4-8. Also, you can find the feedback form for the lecture here:
https://cloud.physik.lmu.de/index.php/apps/forms/s/eLefQDX4LrZjweCsKtmRdD8S

3.4 Unruh radiation

Let's apply the methodology we developed in Section 3.2 to a uniformly accelerating observer in Minkowski space. For simplicity, we will again work in (1+1)-dimensions with inertial coordinates $(x^0, x^1) = (t, x)$. We take uniform acceleration to mean that at each time the observer experiences in their instantaneous rest frame an acceleration a, which in that frame is described by the acceleration vector $(a^0, a^1) = (0, a)$. This in particular means that

$$a^\mu a_\mu = a^2 \ . \tag{3.19}$$

This expression is covariant, i.e. it will hold in any inertial coordinate system and we can re-write it as

$$\ddot{t}^2 - \ddot{x}^2 = -a^2 \ , \tag{3.20}$$

where $\dot{(\ )}$ means derivative wrt. the eigen-time τ of the observer. At the same time, the velocity vector of the observer satisfies

$$\begin{aligned} -1 &= u^a u_a \\ \Rightarrow \ \ \dot{t} &= \sqrt{1+\dot{x}^2} \\ \Rightarrow \ddot{t} &= \frac{\ddot{x}\,\dot{x}}{\sqrt{1+\dot{x}^2}} \ . \end{aligned} \tag{3.21}$$

Exercise 9

Using equations 3.20 and 3.21 and assuming the initial conditions $\dot{x}(\tau = 0) = 0$ and $x(\tau = 0) = 1/a$ as well as $a > 0$, show that the trajectory of the uniformly accelerated observer is given by

$$t(\tau) = \frac{1}{a}\sinh(a\tau) \tag{3.22}$$

$$x(\tau) = \frac{1}{a}\cosh(a\tau) \ . \tag{3.23}$$

A natural coordinate system for an observer that moves along any given world line $(t(\tau), x(\tau))$ is given by what we will call *radar coordinates* (though one might also call it *Padmanabhan coordinates*, cf. [4]). If an even is located at some point in spacetime, then we determine its radar coordinates as follows:

- we send out a radar pulse at eigen-time τ_1;
- this pulse gets reflected the the event in question, and we receive the reflected signal at eigen-time τ_2;
- we then assign the coordinates

$$T := \frac{\tau_2 + \tau_1}{2} \tag{3.24}$$

$$X := \pm\frac{\tau_2 - \tau_1}{2} \tag{3.25}$$

  to the detected event, where the sign in the 2nd line is "$-$" if the radar signal was sent to the left and "+" if it was sent to the right. (In 3+1 dimensions one would simply use $R = +(\tau_2 - \tau_1)/2$ as a radial coordinate and the corresponding angular coordinates would e.g. be the direction from which the radar echo was received.)

Exercise 10

Show that the radar coordinates (T, X) of the constantly accelerated observer are related to the inertial coordinate (t, x) via

$$t(T,X) = \frac{1}{a}e^{aX}\sinh(aT) \tag{3.26}$$

$$x(T,X) = \frac{1}{a}e^{aX}\cosh(aT) \ . \tag{3.27}$$

Also show that the light-cone coordinates $(u, v) = (t - x, t + x)$ are related to the "radar light-cone coordinates" $(U, V) = (T - X, T + X)$ via

$$u(U,V) = -\frac{1}{a}e^{-aU} \tag{3.28}$$

$$v(U,V) = \frac{1}{a}e^{aV} \ . \tag{3.29}$$

Equations 3.28 and 3.29 show that right-moving and left-moving modes to not mix when switching from the coordinates (t, x) to the coordinates (T, X). We can thus treat both of the types of modes separately, and we will mostly ignore the left-moving modes in the following. We can now decompose a scalar field $\hat{\phi}$ into positive and negative frequency modes both wrt. to the inertial coordinates and the radar coordinates,

$$\begin{aligned}\hat{\phi} &= \int_0^\infty \frac{\mathrm{d}\omega}{2\pi}\,\frac{1}{\sqrt{2\omega}}\left\{\hat{a}_\omega e^{-i\omega u} + \hat{a}_\omega^\dagger e^{i\omega u}\right\} \ + \ (\mathrm{left-moving}) \\ &= \int_0^\infty \frac{\mathrm{d}\Omega}{2\pi}\,\frac{1}{\sqrt{2\Omega}}\left\{\hat{A}_\Omega e^{-i\Omega U} + \hat{A}_\Omega^\dagger e^{i\Omega U}\right\} \ + \ (\mathrm{left-moving}) \ .\end{aligned} \tag{3.30}$$

The latter decomposition is the relevant one to describe particles in the lab frame of the accelerated observer, while the former decomposition is the relevant one for defining the Minkowski vacuum. In particular, $\hat{a}_\omega |0\rangle = 0$ such that there are no particle in the vacuum. But in general, one can expect that $\hat{A}_\Omega |0\rangle \neq 0$, such that there are indeed particles from the accelerated observer's perspective. **NOTE: for the above decomposition to be possible, one needs to show that $e^{-i\Omega U}/\sqrt{2\Omega}$ is indeed a solution to the Klein-Gordon equations. I will add a derivation of this to the script soon - both for the accelerated observer discussed here and for the black hole situation discussed in the next section.**

To quantify this apparent particle content, let us decompose the positive frequency mode functions $e^{-i\omega u}$ in the inertial frame in terms of the positive frequency mode functions $e^{-i\Omega U}$ in the accelerated frame. This decomposition can be written as

$$\frac{1}{\sqrt{2\omega}}e^{-i\omega u} = \int_0^\infty \mathrm{d}\Omega\,\frac{1}{\sqrt{2\Omega}}\left(\alpha_{\Omega,\omega}\,e^{-i\Omega U} + \beta_{\Omega,\omega}\,e^{+i\Omega U}\right) \tag{3.31}$$

where $\alpha_{\Omega,\omega}$ and $\beta_{\Omega,\omega}$ are coefficients which we have yet to determine. In terms of these coefficients, the mode operators in the two frames are related by

$$\hat{A}_\Omega = \int_0^\infty \mathrm{d}\omega\left\{\alpha_{\Omega,\omega}\hat{a}_\omega + \beta^*_{\Omega,\omega}\hat{a}^\dagger_\omega\right\} \ . \tag{3.32}$$

Exercise 11
Show that the particle occupation number experienced by the accelerated observer is given by

$$\langle 0| \hat{N}_\Omega |0\rangle \equiv \langle 0| \hat{A}^\dagger_\Omega \hat{A}_\Omega |0\rangle = \int_0^\infty \mathrm{d}\omega \; |\beta_{\Omega,\omega}|^2 \; . \tag{3.33}$$

$$\Rightarrow \beta_{\Omega,\omega} = \sqrt{\frac{\Omega}{\omega}} \int_{-\infty}^{\infty} \mathrm{d}U \; e^{-i\omega u - i\Omega U} \tag{3.34}$$

$$= \sqrt{\frac{\Omega}{\omega}} \int_{-\infty}^{0} \mathrm{d}u \; e^{-i\omega u} e^{(-i\Omega + a)U} \tag{3.35}$$

$$= \sqrt{\frac{\Omega}{\omega}} \int_{-\infty}^{0} \mathrm{d}u \; e^{-i\omega u} \, (-au)^{i\frac{\Omega}{a} - 1} \tag{3.36}$$

$$= \sqrt{\frac{\Omega}{\omega}} (\omega/a)^{-i\Omega/a} e^{-\frac{\pi\Omega}{2a}} \, \Gamma\left(\frac{i\Omega}{a}\right) \; . \tag{3.37}$$

Using the fact that

$$\left|\Gamma\left(i\frac{\Omega}{a}\right)\right|^2 = \frac{2\pi}{\frac{\Omega}{a}\left(\exp\left[\frac{\pi\Omega}{a}\right] - \exp\left[-\frac{\pi\Omega}{a}\right]\right)} \tag{3.38}$$

we then see that

$$|\beta_{\Omega,\omega}|^2 = \frac{2\pi}{a\omega} \, \frac{1}{\exp\left[\frac{2\pi\Omega}{a}\right] - 1} \; . \tag{3.39}$$

So the Ω dependence of $\langle 0| \hat{N}_\Omega |0\rangle$ is given by $1/(\exp\left[\frac{2\pi\Omega}{a}\right] - 1)$, which is the Bose-Einstein factor for a thermal distribution with the temperature

$$T_{\text{Unruh}} = \frac{a}{2\pi} \; . \tag{3.40}$$

So a uniformly accelerated observer will be engulfed in thermal radiation - the so called *Unruh radiation* which was first derived by [8]. Of course, to truly calculate $\langle 0| \hat{N}_\Omega |0\rangle$ we would need to integrate the right hand-side of Equation 3.39 over ω. This integral would diverge, but this divergence is only an artifact of $\langle 0| \hat{N}_\Omega |0\rangle$ counting the particles present in the entire range $-\infty < X < \infty$. If we were to properly restrict the counting of particles to e.g. within the accelerated lab, we would indeed arrive at a finite occupation number.

Note that the above derivation would be the same, had we considered left-moving instead of right-moving modes. So the Unruh radiation is perceived by the accelerated observer to be isotropic.

3.5 Hawking radiation: easy but slightly flawed derivation

We will investigate the origin of Hawking radiation from two different perspectives. In this section, we will follow a derivation which is common in text books [4, 5] and very efficient, but which inaccurate in some details. In the following section, we will then study at least the spirit of Hawkings original derivation [6, 7].

For now, consider a Schwarzschild black hole with the line element

$$\mathrm{d}s^2 = -\left(1 - \frac{r_s}{r}\right) \mathrm{d}t^2 + \frac{1}{1 - \frac{r_s}{r}} \, \mathrm{d}r^2 \; , \tag{3.41}$$

where we are again considering (1+1)-dimensions for simplicity. Introducing the new coordinates

$$T(t,r) := t \quad , \quad R(t,r) := r + r_s \ln\left(\frac{r}{r_s} - 1\right) \tag{3.42}$$

the line element becomes

$$\mathrm{d}s^2 = \left(1 - \frac{r_s}{r(R)}\right)\left(-\mathrm{d}T^2 + \mathrm{d}R^2\right) \ . \tag{3.43}$$

In Exercise 12 you are asked to show that the coordinates (T, R) are the radar coordinates of an observer who is at rest (i.e. who's world line satisfies $\mathrm{d}t = 0$) at a distance $r_{\mathrm{obs}} \gg r_s$. They are thus a suitable coordinate system to describe the positive and negative frequency modes that will be felt by an observer far away from the black hole. But these coordinates do not cover the entire black hole spacetime (they are only defined for $r_s < r < \infty$) and they are thus not suitable for describing the global vacuum state.

To define a decomposition of a scalar field $\hat{\phi}$ into positive and negative frequency modes on the entire spacetime, let us instead consider Kruskal-Szekeres coordinates, which in contrast to Section 1.2 we will here denote with $(\mathcal{T}, \mathcal{X})$.

Exercise 12

A) Show that the coordinates $(T, R - R(r_{\mathrm{obs}}))$ *are the radar coordinates of an observer who is at rest (i.e. who's world line satisfies* $\mathrm{d}r = 0$*) at a distance* $r_{\mathrm{obs}} \gg r_s$*.*

B) Show that Kruskal-Szekeres coordinates $(\mathcal{T}, \mathcal{X})$ *and radar coordinates* (T, X) *are related via*

$$\mathcal{T}(T,X) = 2r_s \exp\left(\frac{1}{2r_s}X\right) \sinh\left(\frac{1}{2r_s}T\right) \tag{3.44}$$

$$\mathcal{X}(T,X) = 2r_s \exp\left(\frac{1}{2r_s}X\right) \cosh\left(\frac{1}{2r_s}T\right) \ . \tag{3.45}$$

Note that the relation between $(\mathcal{T}, \mathcal{X})$ and (T, X) is completely analogous to that between inertial and radar coordinates for the uniformly accelerated observer! This it is clear that the vacuum state defined with respect to the positive and negative frequency modes in the coordinate system $(\mathcal{T}, \mathcal{X})$ will correspond to thermal radiation wrt. positive and negative frequency modes in the coordinate system (T, X). And the temperature of that thermal bath will be the *Hawking temperature*

$$T_{bh} = \frac{1}{2\pi \cdot 2r_s} = \frac{1}{8\pi GM} \ . \tag{3.46}$$

While this derivation somewhat straight forward, it has a number of flaws:

- It is not clear why the state that is annihilated by the Kruskal-Szekeres annihilation operators should be the true global vacuum of the black hole space time.
- In fact, the Kruskal-Szekeres vacuum is *not* the global vacuum. Just as for the uniformly accelerated observer, right- and left-moving modes will yield exactly the same thermal spectrum while Hawking radiation only consists of outgoing modes (i.e. radiation that is *leaving* the black hole)!

- Relying on the analogy with the case of an accelerated observer causes a risk of confusing Unruh radiation and Hawking radiation. E.g. , in order to stay at a constant distance from the black hole one needs to constantly accelerate. The Unruh radiation caused by that acceleration **IS NOT** the Hawking radiation!

Exercise 13

A) According to the Stefan-Boltzmann law, a thermal body of temperature T and surface area A has a Luminosity

$$L = \sigma A T^4 \ , \tag{3.47}$$

where σ is the Stefan-Boltzmann constant. For a black hole, we can interpret the luminosity as $L = -c^2 \mathrm{d}M/\mathrm{d}t$. Use this to derive $M = M(t)$, i.e. the time evolution of a black hole's mass due to Hawking evaporation.

B) A typical candle releases about 80 Watts of heat energy per time interval. What mass would a black hole need to emit at the same luminosity?

C) The temperature of today's cosmic microwave background (CMB) is about $T = 2.7K$. Which mass must a black hole have to produce radiation of the same temperature? (Black holes with a higher mass will actually grow by absorbing CMB photons, rather than evaporating due to their Hawking radiation).

3.6 Exercises for lecture 3 & feedback form

Lecture 3 is accompanied by exercises 9-13. Also, you can find the feedback form for the lecture here:
https://cloud.physik.lmu.de/index.php/apps/forms/s/eLefQDX4LrZjweCsKtmRdD8S

3.7 Spirit of Hawking's original derivation

To understand Hawkings original derivation of Hawking radiation, consider again the Schwarzschild black hole in the coordinates coordinates

$$T(t,r) := t \quad , \quad R(t,r) := r + r_s \ln\left(\frac{r}{r_s} - 1\right) \tag{3.48}$$

where the line element becomes

$$\mathrm{d}s^2 = \left(1 - \frac{r_s}{r(R)}\right)\left(-\mathrm{d}T^2 + \mathrm{d}R^2\right) \ , \tag{3.49}$$

where for now we will again consider (1+1) dimensions.

Exercise 14

In (1+1) dimensions the action of a real, massless scalar field is given by

$$S = \frac{1}{2}\int \mathrm{d}^2x \ \sqrt{-g}\, g^{ab}\partial_a\phi\partial_b\phi \ . \tag{3.50}$$

A conformally flat spacetime is a spacetime with line element

$$\mathrm{d}s^2 = F(T,X)\left(-\mathrm{d}T^2 + \mathrm{d}X^2\right) \ , \tag{3.51}$$

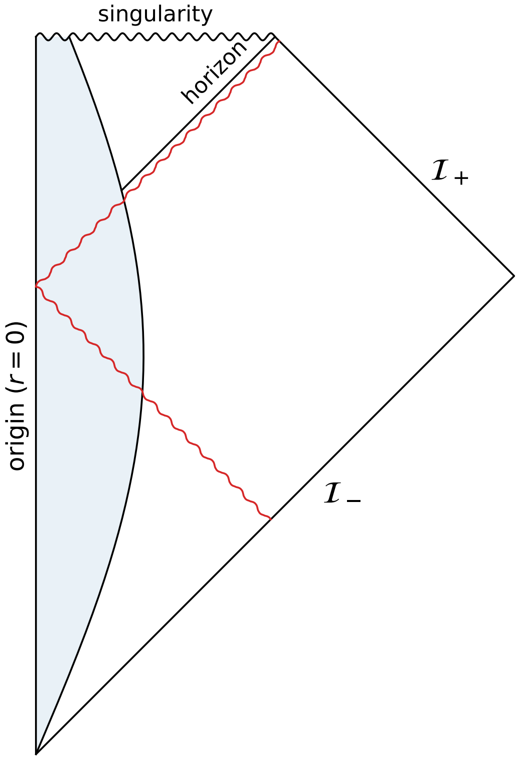

Figure 3: Penrose diagram for a real black hole, forming as a result of the collapse of a massive object (blue shaded region). In Hawkings original derivation of black hole radiation [7], he considered a particle propagating backward in time and scattering in the time-dependent metric of the collapsing object that is forming the black hole (see red, curvy line).

where $F(T, X)$ is some function of the coordinates. Show that for such a space time the Klein-Gordon equation reduces to

$$-\frac{\partial^2 \phi}{\partial T^2} + \frac{\partial^2 \phi}{\partial X^2} = 0 \ , \tag{3.52}$$

i.e. to the equations of motion that are satisfied in Minkowski space. Hint: *Show that $\sqrt{-g}\, g^{ab} = \eta^{ab}$.*

Using the result of Exercise 14, it is clear that outside of a Schwarzschild black hole (at least in our simplified, 2D scenario) the solutions to the Klein-Gordon equation that represent outgoing, positive frequencies modes for an observer far away from the black hole are given by

$$f_\omega(R, T) = \frac{1}{\sqrt{2\omega}} e^{-i\omega(T-R)} \ . \tag{3.53}$$

To be completed.

Exercise 15

De-Sitter space is an empty universe with a cosmological constant Λ*. In the so called* flat slicing *it can e.g. be described by the line element*

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + e^{2H_\Lambda t}\left[\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2\right] \tag{3.54}$$

where $H_\Lambda = \sqrt{\Lambda/3}$ *.*

A) *Find new coordinates* (T, R) *such that the line element becomes*

$$\mathrm{d}s^2 = -\left(1 - R^2H_\Lambda^2\right)\mathrm{d}T^2 + \frac{1}{1 - R^2H_\Lambda^2}\mathrm{d}R^2 + R^2\mathrm{d}\Omega^2 \ . \tag{3.55}$$

B) *Find a time-like Killing vector field for this line element! Express it both in terms of* (T, R) *and in terms of* (t, r)*. Where does de-Sitter space have a Killing horizon?*

C) *Ignoring again the angular coordinates, find a new radius coordinate* $\tilde{R}$*, such that the line element becomes*

$$\mathrm{d}s^2 = \left(1 - R^2H_\Lambda^2\right)\left[-\mathrm{d}T^2 + \mathrm{d}\tilde{R}^2\right] \ . \tag{3.56}$$

What are the solutions to the Klein-Gordon equations in this simplified 2D scenario?

D) *Guess (e.g. by analogy with Section 3.5) at what temperature the de-Sitter horizon radiates. Then consult [9] to check your guess.*

3.8 Exercises for lecture 4 & feedback form

Lecture 4 is accompanied by exercises 14 & 15A-D. Also, you can find the feedback form for the lecture here:
https://cloud.physik.lmu.de/index.php/apps/forms/s/eLefQDX4LrZjweCsKtmRdD8S

4 The Einstein equations as a 1st law

In this section, we will follow [10] as well as chapter 16 of [4] to derive Einstein's field equations from a 1st law of thermodynamics for horizon entropies. To motivate how such a 1st law might play out, let us repeat a simple example that we already looked at in Section 1.3: if a small mass $\mathrm{d}E$ is dropped (slowly) into a Schwarzschild black hole, then its horizon area and its entropy increase as

$$\mathrm{d}A_{\mathrm{bh}} = 32\pi G^2 E_{\mathrm{bh}}\mathrm{d}E \tag{4.1}$$

$$\Rightarrow \mathrm{d}S_{\mathrm{bh}} = 8\pi G E_{\mathrm{bh}}\mathrm{d}E \ , \tag{4.2}$$

where $E_{\mathrm{bh}} = M_{\mathrm{bh}}$ is the black hole mass. Equation 4.1, which determines how the horizon area reacts to a mass increase $\mathrm{d}E$, captures some aspect of Einstein's equations, so it can be seen as a primitive stand-in for these equations. Using the Hawking temperature $T_{\mathrm{bh}} = 1/8\pi G E_{\mathrm{bh}}$ associated with the black hole is, we can write this stand-in Einstein equation as

$$E_{\mathrm{bh}} = T_{\mathrm{bh}}\mathrm{d}S_{\mathrm{bh}} \ . \tag{4.3}$$

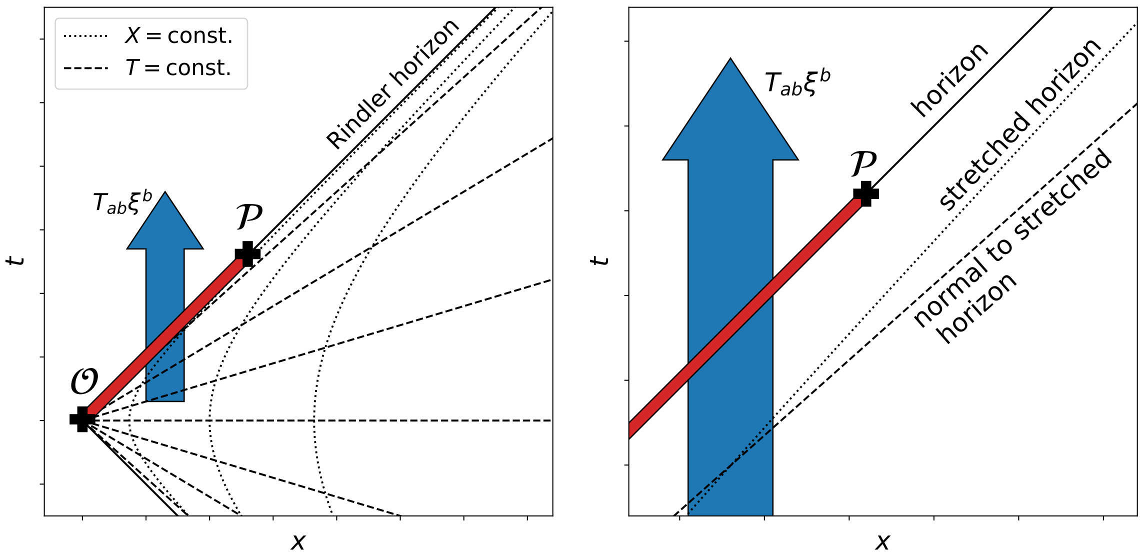

Figure 4: Matter crossing the local Rindler horizon between two spacetime points $\mathcal{O}$ and $\mathcal{P}$. The sketch also indicates the stretched horizon which is an orbit of the approximate, time-like Killing vector field corresponding to Unruh-time and which is located infinitesimally close to the actual horizon.

We want to extend this primitive example to general spacetimes that are described by some metric tensor with components g_{ab} and filled by some energy-momentum tensor with components T_{ab}. To do that, let us consider some spacetime point $\mathcal{O}$ (see also the two sketches in Figure 4). Around this point we can choose a local inertial coordinate system (t, x, y, z), i.e. an "Einstein elevator" in which the metric locally looks Minkowskian. In that local inertial frame we can further switch from (t, x) to the radar coordinates (T, X) of an observer who is uniformly accelerated in the x-direction (the coordinates y, z will not play an important role in the following).

To the accelerated observer, the null-surface $t = x$ will become a horizon (to be more precise: a particle horizon, see different horizon definitions in Appendix D), and the right-moving part of the Unruh-radiation will seem to emanate from that horizon, or rather: from an infinitesimal region just outside this horizon. Furthermore, the observer will assign an entropy density $s = 1/4\ell_{\text{Planck}}^2$ to the horizon area. We now want to investigate how some stream of matter (cf. the blue arrow in Figure 4), which is described by the energy-momentum tensor T_{ab} and which crosses the horizon between the point $\mathcal{O}$ and another point $\mathcal{P}$, changes the horizon area. Of course, the observer will never actually see how the matter crosses the horizon (just as they attribute Unruh radiation not to the horizon itself but to an infinitesimally small neighborhood of the horizon). This is why in the following we will consider a hypersurface just outside the horizon - the so called **stretched horizon** [4] - and eventually consider a limit that sends this stretched horizon to the actual horizon.

Excourse: Stretched Horizons Are More Useful Than You Think!

The concept of a stretched horizon was introduced as a theoretical construct to better understand the physical behavior of black holes, especially from the viewpoint of an observer who remains outside the event horizon. It is defined as a notional, time-like surface located a very small proper distance (typically on the order of the Planck length) outside the true event horizon (see section 3 of [11] for a rigorous definition). Unlike the event horizon, which is a null surface and marks the boundary beyond which no information can escape, the stretched horizon is conceived as a physical membrane with locally measurable properties.

The terminology was first introduced in the membrane paradigm [12], an effective theory developed to describe the external behavior of black holes using classical physics. In this framework, the stretched horizon is imagined as a viscous, conducting surface that can absorb, radiate, and process information. This allows one to model black holes as if they were ordinary thermodynamic systems, with laws like Ohm's law and the Navier-Stokes equations applying at the stretched horizon. The reason why it is considered necessary comes from black hole thermodynamics. While black holes exhibit thermodynamic properties such as entropy and temperature, these quantities are difficult to define on the event horizon itself due to its null nature and the breakdown of static time coordinates there. By placing the stretched horizon just outside the true horizon, where time still flows (albeit extremely slowly), one can meaningfully assign local thermodynamic properties to it. This provides a practical way to discuss black hole entropy and Hawking radiation in terms of standard thermodynamic reasoning. This viewpoint has the goal of providing astrophysicists with mental pictures, physical intuition, computational techniques, and other research tools which facilitates analyses of the interactions of black holes with complex astrophysical environments. This approach becomes especially valuable because, for an external observer, infalling matter appears to asymptotically approach the horizon without ever crossing it. The membrane thus serves as a stand-in for the event horizon, capturing and reprocessing infalling matter and energy [13].

Stretched horizon isn't interesting to consider just from the classical point of view but even in the quantum context, the stretched horizon becomes a key ingredient in reconciling black hole dynamics with the principles of quantum mechanics. [11] and [14] have emphasized that the formation and evaporation of a black hole—though it leads to an effective loss of information and an increase in thermal entropy—is, in principle, governed by unitary quantum evolution. The apparent loss of information stems from the practical impossibility of tracking all microscopic degrees of freedom in such highly complex systems, not from a fundamental failure of quantum theory. However, when gravitational collapse leads to the formation of an event horizon, conventional quantum field theory suggests that information is irretrievably lost—since Hawking radiation is thermal and seemingly uncorrelated with the black hole's internal state.

To counter this, [11] proposes that black hole evolution can and should be described in a way that preserves unitarity, even in the presence of gravity. This requires a phenomenological model that accounts for the way information is processed near the horizon. The stretched horizon serves this role. It acts as a "surrogate" for the global event horizon, providing a surface where incoming matter and radiation are absorbed, scrambled, and eventually re-emitted in Hawking radiation. By enforcing a description in which a distant observer never refers to events inside the event horizon, the stretched horizon preserves the causality and locality of external physics while remaining consistent with quantum theory.

Moreover, in string-theoretic models and the fuzzball proposal [15], the stretched horizon may correspond to a region where quantum gravity effects become significant, and where the classical notion of an empty event horizon is replaced by a structure rich in "microscopic degrees of freedom."

In this view, the stretched horizon becomes the locus where quantum gravity effects are effectively encoded, allowing for the black hole to behave like an ordinary thermodynamic system. It absorbs information, thermalizes it, and radiates it back in a unitary fashion, albeit in an extremely scrambled form. Each point on the stretched horizon corresponds to a point on the true horizon, maintaining a one-to-one mapping that allows the stretched surface to mirror the behavior of the global event horizon. This framework reflects the postulates of black hole complementarity, which assert that physics as seen by an external observer should be self-consistent and unitary, without requiring access to the black hole interior.

The term **"stretched"** reflects the fact that this surface is not coincident with the true horizon, but instead lies just outside it, like a membrane "stretched" over the event horizon. This allows it to be treated as a physical entity with time-like properties, capable of supporting local dynamics, in contrast to the event horizon which is fundamentally non-local and inaccessible to external measurement.

Now back to our discussion of the first law for Rindler horizons. Remember that the radar coordinates (T, X) are related to inertial coordinates (t, x) via

$$t(T, X) = \frac{1}{a} e^{aX} \sinh(aT) \tag{4.4}$$

$$x(T, X) = \frac{1}{a} e^{aX} \cosh(aT) \ , \tag{4.5}$$

where a is the constant acceleration. The line element in this coordinate system is

$$\mathrm{d}s^2 = e^{2aX}(-\mathrm{d}T^2 + \mathrm{d}X^2) + \mathrm{d}y^2 + \mathrm{d}z^2 \ , \tag{4.6}$$

where y and z are the unchanged Minkoswki coordinates that are transverse to the x-t plane. The above line element, and the coordinates (T, X), only applies to a quarter of our local Minkowski space. And in that quarter, the vector field ∂/∂_T acts as a time-like Killing vector field. The components of this field in the coordinate system (t, x) can be read off from the fact that

$$\begin{aligned}\partial_T &= \frac{\partial t}{\partial T}\partial_t + \frac{\partial x}{\partial T}\partial_x \\ &= ax\,\partial_t + at\,\partial_x & (4.7)\\ &\equiv \xi^t\,\partial_t + \xi^x\partial_x \ . & (4.8)\end{aligned}$$

Here, in the last line we have denoted the components of ∂_T as ξ^a, i.e. $\partial_T = \xi^a\partial_a$ where the index a labels inertial Minkowski coordinates. Similarly, the components of ∂_X can be read off from the fact that

$$\begin{aligned}\partial_X &= \frac{\partial t}{\partial X}\partial_t + \frac{\partial x}{\partial X}\partial_x \\ &= at\,\partial_t + ax\,\partial_x & (4.9)\\ &\equiv n^t\,\partial_t + n^x\,\partial_x \ . & (4.10)\end{aligned}$$

Here, n^a now denotes the components of ∂_X in Minkowski coordinates, i.e. $\partial_X = n^a\partial_a$.

Note that ∂_T and ∂_X are normal to each other, i.e.

$$\xi^a n_a = a^2 xt - a^2 xt = 0 \ . \tag{4.11}$$

At the same time, as one approaches the Unruh horizon at $t = x$, the two vector fields become parallel. Being simultaneously normal and parallel to each other is of course only possible, because the two fields become light-like in the limit $t \to x$.

For the uniformly accelerated observer (who monitors the universe via their radar / radar-coordinates) the matter with energy-momentum tensor T_{ab} will flow according to an energy flux $J_a = -T_{ab}\xi^b$, where the minus-sign stems from our metric signature condition $(-,+,+,+)$. As mentioned above, the observer never sees that the matter actually crosses the horizon. And also the Unruh radiation of temperature $a/2\pi$ will seem to originate to the observer not directly at the horizon, but from a close neighborhood of the horizon. We thus pick an orbit of ∂_T that passes close by the horizon, i.e. a trajectory with constant and highly negative $X = X_0$, and call it the *stretched horizon* (we will later take the limit $X_0 \to -\infty$). In the coordinate system of the accelerated observer, the normal to the stretched horizon is given by

$$\hat{n}^a \partial_a \equiv \frac{n^a}{\sqrt{n^b n_b}} \partial_a = \frac{t}{\sqrt{x^2 - t^2}} \partial_t + \frac{x}{\sqrt{x^2 - t^2}} \partial_x \ . \tag{4.12}$$

So according to the accelerated observer, the rate at which energy flows across the stretched horizon per transverse area element $\mathrm{d}A = \mathrm{d}y\mathrm{d}z$ will be given by

$$\frac{\mathrm{d}^2 E}{\mathrm{d}T\mathrm{d}A} = -J_a n^a = \hat{n}^a \xi^b T_{ab} \ , \tag{4.13}$$

where we took into account that $-\hat{n}$ is the vector pointing into the horizon. In the limit where $X_0 \to -\infty$, i.e. where the stretched horizon approaches the actual horizon, both the vectors ξ and $\hat{n}$ become parallel to the horizon generator $k = (1,1,0,0)$. Thus, in that limit we can write

$$\frac{\mathrm{d}^2 E}{\mathrm{d}T\mathrm{d}A} \approx \frac{at(X_0,T)^2}{\sqrt{x(X_0,T)^2 - t(X_0,T)^2}} \, k^a k^b T_{ab} \ . \tag{4.14}$$

To obtain a finite amount of energy δE that passes through the horizon, let us integrate the above expression over some finite cross section ΔA of the horizon and some finite interval in T. This yields

$$\delta E = \int \mathrm{d}T\mathrm{d}A \ \sqrt{-|h|} \, \frac{\mathrm{d}^2 E}{\mathrm{d}T\mathrm{d}A} \ , \tag{4.15}$$

where h_{ab} is the induced metric on the stretched horizon, and $|h|$ is its determinant.

Exercise 16

With the help of Equation 4.6 show that

$$\sqrt{-|h|} = e^{aX_0} \ . \tag{4.16}$$

Use this to show that

$$\delta E = a^2 \int \mathrm{d}T\mathrm{d}A \ t(X_0,T)^2 \, k^a k^b T_{ab} \ . \tag{4.17}$$

———

We are now in a position to take the limit $X_0 \to -\infty$. In this limit, we can change the time parameter of our integral from T to the affine parameter of our approximately light-like stretched horizon, which is $\lambda \approx t(X_0,T) \approx x(X_0,T)$.

Exercise 17

Show that

$$\delta E = a \int \mathrm{d}\lambda \mathrm{d}A \; \lambda \, k^a k^b T_{ab} \; . \tag{4.18}$$

———

This is now indeed independent of X_0 and corresponds to the flow of energy crossing a small portion of the actual horizon. We can think of this portion as being traced out by a narrow beam of horizon generators ([10] calls this a "pencil") with an area cross-section ΔA that starts at $\mathcal{O}$ and ends at $\mathcal{P}$. You can e.g. literally imagine this beam as a set of light rays coming out of a flashlight with cross-section ΔA that is turned on at $\mathcal{O}$ and that is pointing towards $\mathcal{P}$. On their way from $\mathcal{O}$ to $\mathcal{P}$ these light rays will cover a 3-dimensional region which is a sub-region of the Rindler horizon.

The cross section of the beam of generators will change as the beam propagates, i.e. $\Delta A = \Delta A(\lambda)$. To calculate this change of area we can look at the *expansion* θ of the beam which can be interpreted as

$$\theta = \frac{\mathrm{d} \ln \Delta A}{\mathrm{d}\lambda} = \frac{1}{\Delta A} \frac{\mathrm{d}\Delta A}{\mathrm{d}\lambda} \approx \frac{\mathrm{d}^2 A_{\text{total}}}{\mathrm{d}\lambda \mathrm{d}A} \; . \tag{4.19}$$

A more accurate definition of θ would be to say that it is the trace of the so called *expansion tensor* θ_{ab} which describes parallel transport of transverse vectors along a light-like geodesic. But for our purposes, Equation 4.19 suffices, and we can express the change in horizon area from $\mathcal{O}$ to $\mathcal{P}$ as

$$\begin{aligned} \delta A &= \int_{\text{beam}} \mathrm{d}\lambda \mathrm{d}A \; \frac{\mathrm{d}^2 A_{\text{total}}}{\mathrm{d}\lambda \mathrm{d}A} \\ &= \int_{\text{beam}} \mathrm{d}\lambda \mathrm{d}A \; \theta \; . \end{aligned} \tag{4.20}$$

The evolution of θ along the beam is described by the *Raychaudhuri equation*

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{1}{2}\theta^2 - \sigma^{ab}\sigma_{ab} - R_{ab} k^a k^b \; , \tag{4.21}$$

where σ_{ab} is the traceless part of the expansion tensor, and R_{ab} is the Ricci tensor. Let us choose our beam such that it's generators are parallel at $\mathcal{P}$, i.e. $\theta(0) = 0$, such that the horizon area doesn't change after all matter has passed. And let us also choose $\mathcal{P}$ close enough to $\mathcal{O}$ such that the quadratic terms θ^2 and $\sigma^{ab}\sigma_{ab}$ can be ignored along the entire path from $\mathcal{O}$ to $\mathcal{P}$ (we can do this because we are anyways trying to recover a first law only in an infinitesimal neighborhood of $\mathcal{O}$). Then the Raychaudhuri equation becomes

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} \approx -R_{ab} k^a k^b \; . \tag{4.22}$$

With our boundary conditions this has the solution

$$\theta(\lambda) = \int_{\lambda}^{\infty} \mathrm{d}\lambda' \; R_{ab} k^a k^b \; , \tag{4.23}$$

and the change in Beam cross section, and thus horizon area, is given by

$$\begin{aligned} \delta A &\approx \int_{\text{beam}} \mathrm{d}\lambda \mathrm{d}A \int_{\lambda}^{\infty} \mathrm{d}\lambda' \, k^a k^b R_{ab}(\lambda') \\ &= \int_{\text{beam}} \mathrm{d}\lambda \mathrm{d}A \, \lambda \, k^a k^b R_{ab}(\lambda) \; , \end{aligned} \tag{4.24}$$

where the 2nd line can be obtained by changing the order of integration of λ and λ'.

Jacobson's Argument now goes as follows:

A) Assign an entropy $S \sim \eta A$ to the horizon.

B) Demand that

$$\delta E = T\delta S = \eta T \delta A \ , \tag{4.25}$$

where $T = a/2\pi$ is the temperature of the horizon as inferred by the uniformly accelerating observer via Unruh radiation. This leads to

$$a \int \mathrm{d}\lambda \mathrm{d}A \ \lambda \, k^a k^b T_{ab} = \frac{a\eta}{2\pi} \int \mathrm{d}\lambda \mathrm{d}A \ \lambda \, k^a k^b R_{ab} \ . \tag{4.26}$$

C) Since we can create local Rindler horizons anywhere in spacetime and with any orientation, we can conclude that

$$k^a k^b T_{ab} = \frac{\eta}{2\pi} k^a k^b R_{ab} \tag{4.27}$$

for any light-like vector k at any location in spacetime. This means that

$$\frac{2\pi}{\eta} T_{ab} = R_{ab} + f \cdot g_{ab} \tag{4.28}$$

for some function f on spacetime.

Exercise 18

The contracted Bianci identities are given by

$$\nabla^a R = 2\nabla^b R^a{}_b \ . \tag{4.29}$$

A) Use Equation 4.29 as well as the fact that the energy momentum tensor satisfies the conservation equation $\nabla^a T_{ab} = 0$ to show that the function f is given by

$$f = -\frac{1}{2} R + \Lambda \ , \tag{4.30}$$

where Λ is an undetermined constant (the cosmological constant*). Conclude from this that*

$$\frac{2\pi}{\eta} T_{ab} = G_{ab} + \Lambda \cdot g_{ab} \ . \tag{4.31}$$

B) How do we have to "define" Newton's constant G (as a function of η) in order for this to take the form of the Einstein equations? Does the $\eta(G)$ required for this yield the Bekenstein-Hawking entropy formula we encountered in the beginning of the course?

—

Let us close this section with a couple of remarks.

- Our derivation here was in fact somewhat different from Jacobson's original approach in [10]. We considered matter falling *into* the observer's Rindler horizon, while he considered matter coming *out of* past null infinity of the Rindler coordinate system. A line of argument that is closer to this section can e.g. be found in [16]. In that paper you can also find an analysis of the formation of caustics (and avoidance thereof), which we ignored here.

- Instead of using Equation 4.23 as a solution to Equation 4.22 we could have e.g. used

$$\theta(\lambda) = -\int_{-\infty}^{\lambda} \mathrm{d}\lambda' \; R_{ab} k^a k^b \; . \tag{4.32}$$

  This would be the solution where the horizon generators are parallel *before* the matter streams across the horizon. The horizon area would then seemingly decrease as matter crosses. To justify our choice we may e.g. also invoke the 2nd law of horizon thermodynamics. So Einstein's equations would be equivalent to a combination of 1st and 2nd law. But ultimately, our main conceptual problem is the task of defining the area and area-increase of an infinite surface. This problem can be addressed in a mathematically more elegant manner by considering virtual horizon displacements at equal time as opposed to considering the change of horizon area over time. This is e.g. described in chapter 16 of [4]. At least for closed horizons, one can however unambiguously define area change and show that such horizons satisfy the 1st law within Einstein gravity [17].

- Jacobson himself has attempted to retrieve Einstein's equations from thermodynamic arguments involving closed surfaces. You can explore this in the following exercise.

Exercise 19

(extra) Jacobson has now considered an alternative version for deriving Einstein's equations from thermodynamic arguments, which he detailed in
https://arxiv.org/abs/1505.04753 .
Have a look at this paper and try to explain it's main idea to your fellow students. Speculate why in the lecture we have still focused on his older argument.

4.1 Exercises for lecture 5 & feedback form

Lecture 5 is accompanied by exercises 16-19. Also, you can find the feedback form for the lecture here:
https://cloud.physik.lmu.de/index.php/apps/forms/s/eLefQDX4LrZjweCsKtmRdD8S

5 Entropy bounds and the holographic principle

In this section we follow [18] as well as the works of [19] and [20] and consider entropy bounds that follow from demanding the validity of the generalized 2nd law.

5.1 The spherical entropy bound

Imagine a matter system of total energy E and characteristic radius R, which as [20] we take to be the radius of the smallest sphere that fits around the system. In order for this system to not yet have collapsed into a black hole, we must have

$$E < \frac{R}{2G} \; . \tag{5.1}$$

If we drop additional matter with energy $\Delta E = R/2G - E$, then the system will form a black hole of entropy

$$S_{\mathrm{bh}} = \pi R^2/G \; . \tag{5.2}$$

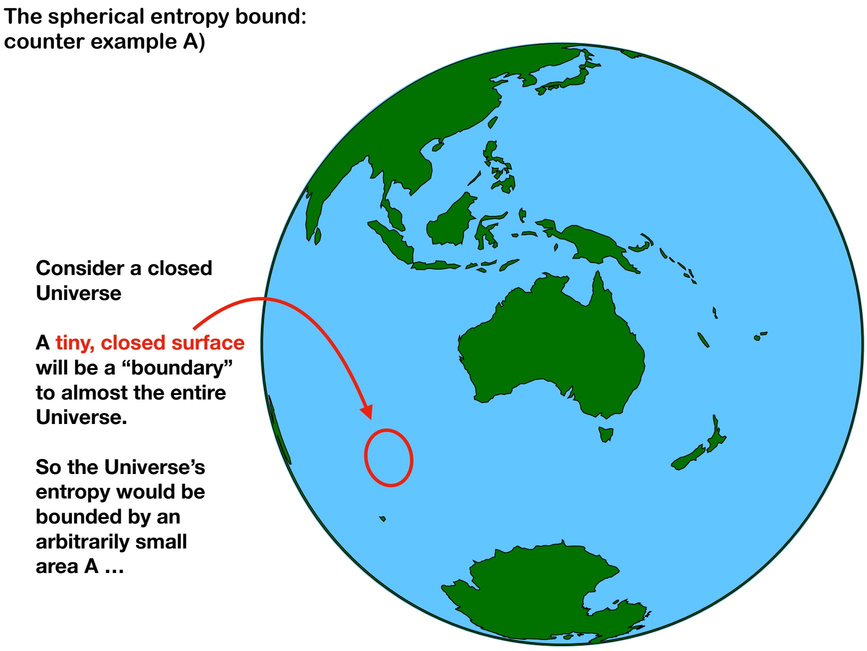

Figure 5: A closed universe will break the spherical antropy bound.

At the same time, the process of dropping additional matter into the system can have only increased the entropy present inside the radius R. Thus, if the generalized 2nd law is indeed valid, then the entropy S_{matter} of the original matter system must have been smaller than the final black hole entropy, i.e.

$$S_{\text{matter}} < \pi R^2 / G \ . \tag{5.3}$$

This is the so called *spherical entropy bound* [20]. It can be interpreted as saying that the maximum entropy that can be accumulated inside a spherical region of radius R is given by on quarter of the boundary area of that region (in units of Planck area).

Note that this bound - just as the generalized 2nd law itself - is not strictly derived from GR. In the absence of such a derivation, its only source of its credibility could then be a lack of convincing counter examples. This is however bad news for the spherical entropy bound, because there are in fact a number of counter examples. We will list here only a selection of the examples given by [18].

- **Counter example A: closed spaces**

  In a closed Friedmann universe $\mathcal{U}$, the boundary of any tiny spherical region $\mathcal{R}$ is also the boundary of another spherical region: the rest of the universe, $\mathcal{U} - \mathcal{R}$ (cf. Figure 5). So the entropy everywhere outside $\mathcal{R}$ would need to be bounded by

  $$S_{\mathcal{U}-\mathcal{R}} < \frac{\partial \mathcal{R}}{4} \ . \tag{5.4}$$

  We can shrink the right-handside of this inequality to 0 by considering a smaller and smaller region $\mathcal{R}$. This would seem to indicate the the entropy of the entire Universe vanishes.

- **Counter example B: flat Friedmann universe**

Now consider an expanding, flat Friedmann universe. The equal-time slices of that spacetime are infinite spaces that are filled with some finite and spatially constant entropy density σ. The total entropy insider a spherical volume with radius R is then

$$S_R = \frac{4\pi}{3} R^3 \sigma \ . \tag{5.5}$$

For $R \geq 4/3G\sigma$ this will break the spherical entropy bound.

- **Counter example C: a cloud with (EXTREMELY) many particle species**

  Consider a gas of thermal photons, forming a cloud of radius R. In the weak gravity regime, where spacetime is close to Minkowskian, the total energy of this cloud will be

$$E = \frac{4\pi^3}{45} R^3 T^4 \ , \tag{5.6}$$

  where T is the temperature of the cloud. At the same time, the entropy of the cloud will be given by

$$S = \frac{4}{3}\frac{E}{T} = \frac{16\pi^3}{135} R^3 T^3 \ . \tag{5.7}$$

  We can eliminate T from these two equations to arrive at

$$S = \left(\frac{4^5 \pi^3}{3^7\, 5}\right)^{1/4} R^{3/4} E^{3/4} \ . \tag{5.8}$$

  If the cloud is close to collapsing into a black hole, i.e. $E \sim R/2G$ then

$$S \sim \left(\frac{2}{3^7\, 5}\right)^{1/4} (A/G)^{3/4} \ . \tag{5.9}$$

  At first glance, this seems to (over) satisfy the spherical entropy bound. But the above derivation was for a gas of only one species of photons. If our Universe contained a number Z of different, massless vector fields, and if all of the different vector species were equally excited in our cloud, then the cloud's total entropy would become

$$S \sim \left(\frac{2}{3^7\, 5}\right)^{1/4} Z\,(A/G)^{3/4} \ , \tag{5.10}$$

  which for large enough Z can be made to break the spherical entropy bound.

- **Further counter argument: general covariance**

  The way we defined the spherical entropy bound is not generally covariant. In particular, the size of the boundary area of some space-like region will generally depend on one's choice of coordinates. In [18] you will e.g. find an example where the coordinates around a collapsing star are chosen such that the star is enclosed by an arbitrarily large boundary surface.

Both counter examples A & B as well as the criticism of a lack of general covariance are resolved by replacing the spherical entropy bound with the so called covariant entropy bound (cf. Section 5.2). Counter example C persists with this bound, but we argue in Section 5.3 (following [18]) that it is not physically relevant.

5.2 The covariant entropy bound (Bousso's bound)

A straight forward way how an observer can account for the entropy that surrounds them, is by collecting data about this surrounding along the observer's backward light-cone. In other words: a snap-shot of the Universe that an observer can have access to will never represent a space-like hypersurface of spacetime, but will be at best as subset of a null hypersurface (of the observer's backward light-cone). This motivates a version of the entropy bound that counts entropy on light-cones.

Counter example A) from above also indicates that such a bound will in certain situations only apply to part of the light-cone. To understand this, let us e.g. denote by $L^-_{\mathcal{O}}$ the past light-cone of a point $\mathcal{O}$ in spacetime. At that point, we can choose an inertial frame such that light-like vectors $k_{\hat{\boldsymbol{n}}} = (-1, \hat{\boldsymbol{n}})$ label the directions of the different light-like geodesics that generate $L^-_{\mathcal{O}}$, with $\hat{\boldsymbol{n}}$ being 3-dimensional unit vectors. Any point $\mathcal{P}$ on the light-cone will then be parametrized by the $\hat{\boldsymbol{n}}$ corresponding to the geodesic that connects $\mathcal{O}$ and $\mathcal{P}$, and by a parameter that tells us how far to move along that geodesic to reach $\mathcal{P}$. A natural choice of such a parameter would be the affine parameter λ corresponding to $k_{\hat{\boldsymbol{n}}}$. But let us be slightly more "observer-centric" by assuming that the observer whose light-cone we are looking at has build a radar-coordinate system according to the world-line that they have been following. Every point on the light-cone has then a unique radar-time coordinate τ. Let us denote by $\mathcal{B}_\tau$ all points on the light-cone with the same value for τ. The $\mathcal{B}_\tau$ are then space-like surfaces that cut through the light-cone, and we will denote the boundary area of these cuts by $A_\tau \equiv |\mathcal{B}_\tau|$. Now let $L_\tau \subset L^-_{\mathcal{O}}$ be the part of the light-cone that ends at $\mathcal{B}_\tau$. Then there can be situations - such as e.g. the closed universe of counter example A) - where

$$L_{\tau_1} \subset L_{\tau_2}$$

but

$$A_{\tau_1} > A_{\tau_2} \ .$$

If in such a situation the entropy on L_{τ_i} where to be bounded by A_{τ_i}, then somehow L_{τ_2} could carry less information then L_{τ_1} despite the fact that it completely contains L_{τ_1}.

The above thoughts lead to the definition of so-called *light-sheets*, which are the parts of light-like hypersurfaces to which the so-called *covariant entropy bound* applies [18, 21]. In Section 5.2.1 we give a simplified - but potentially intuitive - version of covariant entropy bound and in Section 5.2.2 we present the actual definition of light-sheets used by [21] to formulate the bound.

5.2.1 A simplified version of the covariant entropy bound

Let us again consider the past light-cone $L^-_{\mathcal{O}}$ of a spacetime point $\mathcal{O}$. We can again think of $\mathcal{O}$ as a location along the world-line of an observer, and $L^-_{\mathcal{O}}$ is the piece of spacetime about which the observer can instantaneously collect information. Let τ parametrize cross-sections of $L^-_{\mathcal{O}}$ as described above. (For definiteness, we assume that the range $0 \leq \tau < \infty$ covers the entire light-cone.) Then the covariant entropy bound can be stated as follows:

Covariant entropy bound (simplified):
The entropy S_τ that can be contained on a piece $L_\tau \subset L^-_{\mathcal{O}}$ is bounded by

$$S_\tau < \frac{A_\tau}{4G} \tag{5.11}$$

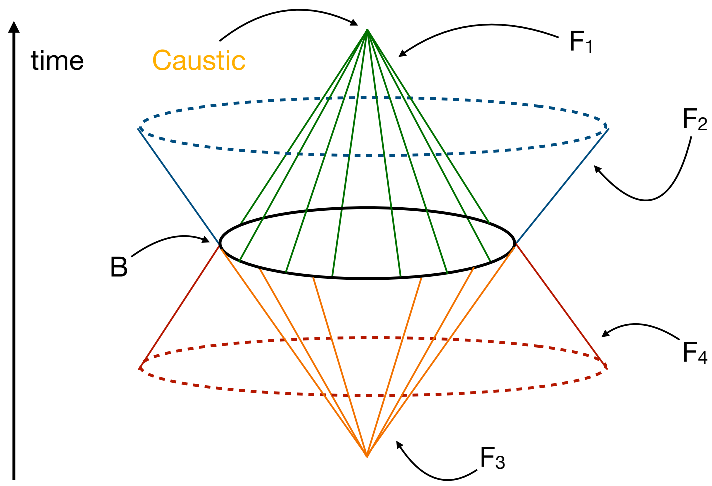

Figure 6: Sketch of the different light-like hypersurfaces emanating from a space-like surface $\mathcal{B}$ (cf. Section 5.2.2 for details).

as long as $A_{\tau'}$ *monotonically increases between* $\tau' = 0$ *and* $\tau' = \tau$ *.*

The entropy S_τ here should be understood as the integral of the entropy density s (of a thermal fluid that fills spacetime) over the piece of the light-cone L_τ .

5.2.2 Bousso's original definition of light-sheets and the covariant entropy bound

For the purpose of this course, the simplified version of the covariant entropy bound given above will be sufficient. But for completeness we nevertheless want to summarize the more far-reaching statement given by [21] regarding entropy on light-sheets.

Light-sheets:
Let $\mathcal{B}$ *be a closed, spacelike 2D surface in a given spacetime. Then there will be four different light-like hypersurfaces that emanate orthogonally from* $\mathcal{B}$*, which we will denote by* $\{F_1\,, F_2\,, F_3\,, F_4\}$*. Intuitively, you can think of these as the hypersurfaces generated by light-like geodesics that travel to the "future-inside", "future-outside", "past-inside" and "past-outside" of* $\mathcal{B}$ *(cf. Figure 6) assuming that we can properly define what "inside" means. From these four surfaces we pick the so-called "light-sheets" as follows:*

A) Select the "inside-going" F_i*'s. Formally these can be defined as the* F_i *whose geodesic generators have expansion* $\theta \geq 0$ *as they leave from* $\mathcal{B}$*, or in other words: the cross-sections of the surface shrink as we move away from* $\mathcal{B}$*.*

B) But even if these cross-sections shrink initially, in curved spaces it is not guarantied that they continue to shrink. (In fact, even in flat spacetime the geodesics might meet in a focal point after which the expansion will change sign - cf. again Figure 6.) We choose to cut the inside-F_i *before that happens, i.e. we consider the light-sheets to consist only of those parts of the inside-*F_i *where the expansion is* ≥ 0*.*

With the above definition of light-sheets we are not in the position to formulate the covariant entropy bound as it fas formulated by Raphael Bousso e.g. in [21]:

Covariant entropy bound (Bousso's bound:
Let $\mathcal{B}$ be a closed, spacelike 2D surface and let F be a light-sheet emanating from $\mathcal{B}$ with boundary area A. Then the total entropy on the light-sheet is always bounded by

$$S < \frac{A}{4} \ . \tag{5.12}$$

(Note that the boundary area A above may be bigger than $|\mathcal{B}|$. This will be the case if the criterion $\theta \leq 0$ cuts the light-sheet not in a single point but along an entire 2nd boundary surface.)

5.2.3 Space-like projection theorem

(will be completed soon)

5.2.4 Application to flat Friedmann Universe

A minimum requirement in order for the covariant entropy bound to avoid the problem of counter example B in Section 5.1 is that light-sheets in Friedmann Universes should be finite. Let us see whether this is indeed the case! The line-element of the Friedmann metric can e.g. be given in the form

$$\mathrm{d}s^2 = a(\eta)^2 \left(-\mathrm{d}\eta^2 + \mathrm{d}\chi^2 + f(\chi)^2 \mathrm{d}\Omega^2\right) \ , \tag{5.13}$$

where the 2nd line uses the so-called conformal time coordinate η and the co-moving radial distance coordinate χ, and the the function $f(\chi)$ differs depending on the sign of the spatial curvature of the universe. In particular: $f(\chi) = \chi$ for spatially flat universes, $f(\chi) = \sin\chi$ for universes with positive spatial curvature and$f(\chi) = \sinh\chi$ for universes with negative spatial curvature.

In this universe we consider an observer located at $\chi = 0$ and their backward light-cone at some time η_0. Because of radial symmetry, we can choose the parameter τ that labels different orthogonal cuts to the light-cone (cf. the beginning of Section 5.2) to be affine. Remember that the expansion scalar along the past light-like geodesics is defined as

$$\theta = \frac{\mathrm{d}\Delta A}{\Delta A \mathrm{d}\lambda} \ , \tag{5.14}$$

where λ is an affine parameter (in particular: $\tau = \lambda$) and ΔA is the cross-section of an infinitesimal light-beam traveling along the geodesics. To test the simplified version of the covariant entropy bound given in Section 5.2.1, we want to follow the above light-cone until its cross-sectional areas A_λ no-longer increase. And according to Equation 5.14, this means that we are looking for the shell at which $\theta = 0$.

Along backward light-cones the coordinates χ and η change such that

$$\frac{\partial\eta}{\partial\lambda} = -\frac{\partial\chi}{\partial\lambda} \ . \tag{5.15}$$

And the infinitesimal area elements perpendicular to radial geodesics are

$$\Delta A \sim a(\eta)^2 f(\chi)^2 \sin(\vartheta)\mathrm{d}\vartheta\mathrm{d}\phi \ . \tag{5.16}$$

Exercise 20

A) Use Equations 5.15 and 5.16 to show that the condition $\theta = 0$ *is equivalent to*

$$\frac{f'(\chi)}{f(\chi)} - \frac{\dot{a}(\eta)}{a(\eta)} = 0 \; . \tag{5.17}$$

B) If our light-sheet terminates at η^**, then the solution* $\chi^* = \chi^*(\eta^*)$ *of Equation 5.17 is the radial coordinate of that termination-surface. Show that this satisfies*

$$f(\chi^*) = \frac{1}{\sqrt{\mathcal{H}(\eta^*)^2 + k}} \; , \tag{5.18}$$

where $\mathcal{H} = \dot{a}/a$ *, and where* $k = 0$ *for a spatially flat universe,* $k = 1$ *for a universe with positive curvature and* $k = -1$ *for a universe with negative (spatial) curvature.*

C) Show that the boundary area of the light-sheet is given by

$$A(\eta^*) = \frac{3}{2\rho(\eta^*)} \; , \tag{5.19}$$

where ρ *is the mass-energy density that is filling the universe.* Hint: *use the 1st Friedmann equation in the form* $\mathcal{H}^2 = \frac{8\pi}{3}\rho a^2 - k$ *, which means that*

$$f(\chi^*) = 1/\sqrt{\frac{8\pi}{3}\rho(\eta^*)a(\eta^*)^2} \; . \tag{5.20}$$

Let us now investigate whether the entropy on light-sheets in the early phase of our own Universe is indeed bounded by $A(\eta^*)/4$ (we are dropping any factors of G for the remainder of this sub-section). For this test, we approximate the primordial quark-gluon plasma as a mixture of ultra-relativistic particle species in thermal equilibrium. The entropy density of such a plasma can be approximated as [22]

$$s(T) \approx \frac{2\pi^2}{45} g_* T^3 \; , \tag{5.21}$$

where g_* is the effective number of relativistic degrees-of-freedom in the plasma. If all particle species have the same temperature, then this can be approximated as

$$g_* = \sum_{\text{Boson species } i} g_{\text{Boson},i} + \frac{7}{8} \sum_{\text{Fermion species } i} g_{\text{Fermion},i} \; , \tag{5.22}$$

where the sums are over the different Boson and Fermion species in the plasma, and the $g_{\dots,i}$ count the internal degrees of freedom of these species (e.g. $g = 2$ for photons). In the standard model at high temperatures one has $g_* \approx 106.75$ [22]. With that same g_*, the mass-energy density of the plasma can be approximated as

$$\rho(T) \approx \frac{\pi^2}{30} g_* T^4 \; . \tag{5.23}$$

If the Universe evolves adiabatically, then the entropy on the light-sheet will be equal to the entropy in the constant-η piece of the Universe that is bounded by $A(\eta^*)$. We thus want to test whether

$$\frac{A(\eta^*)}{4} \stackrel{?}{>} \frac{4\pi}{3}[a(\eta^*)\chi^*]^3\, s \ , \tag{5.24}$$

where we already used the fact that our Universe is spatially flat, i.e. $f(\chi) = \chi$.

Exercise 21

Use Equation 5.20 to show that the condition 5.24 is equivalent to

$$\sqrt{\frac{405}{16\pi}} > \sqrt{g_*}\, T \ . \tag{5.25}$$

———

Note that we have done all calculations in this section in Planck units (even $G = 1$ here). Thus, the above equation measures T in units of the Planck temperature and in epochs where the standard model is valid we have $T \ll 1$. Thus, for realistic values of g_* the entropy bound is indeed satisfied.

5.3 The species problem

Counter example C from Section 5.1 is still not circumvented by the covariant entropy bound, i.e. a collapsing relativistic gas cloud that is made up of many different particle species can still break the bound. There are however two reasons why we may regard this as an unphysical counter example. The first reason regards the formation time of black holes compared to their evaporation time due to Hawking radiation. The intensity of Hawking radiation is approximately proportional to Z, since each particle species will radiate for itself (at least in the limit where the Hawking temperature is higher than the masses of the different species). It can be shown that for Z that are high-enough to break the covariant entropy bound the evaporation time becomes shorter than the typical black-hole formation time [23]. This, in such a scenario, the concept of black-hole horizons itself becomes questionable.

The second reason why counter example C can be questioned is the sheer number of species required to actually break the covariant entropy bound. You will calculate this number in the following exercise.

Exercise 22

Assume that the gas cloud in counter example C has a radius similar to that of earth. How big would Z need to be in order to break the entropy bound? How big would it need to be if the radius of the cloud was that of a single proton?

5.4 The holographic principle

The covariant entropy bound seems to be satisfied in any physically meaningful situation within classical general relativity. But there is no derivation of the bound from within GR. And the principles that gave rise to GR - general covariance and the equivalence principle - do not seems to imply or at least motivate any kind of entropy bounds. The *holographic principle* expresses the hope, that the mystery of why classical GR satisfies the covariance entropy bound be explained by an eventual theory of quantum gravity. Concretely, it may be formulated as follows [18, 20, 24].

The holographic principle: *In its low-energy limit, the theory of quantum gravity will give rise to semi-classical geometry and to effective matter fields on that geometry in such a way that the dimension d of the Hilbert space describing matter degrees-of-freedom on a light-sheet in the emergent geometry is bounded by*

$$\ln d < \frac{A}{4G} \,, \tag{5.26}$$

where A is the boundary area of the light-sheet.

5.5 Exercises for lecture 6 & feedback form

Lecture 6 is accompanied by exercises 20-22. Also, you can find the feedback form for the lecture here:
https://cloud.physik.lmu.de/index.php/apps/forms/s/eLefQDX4LrZjweCsKtmRdD8S

Excourse: More On Covariant Entropy Bound

In the lecture, we examined the covariant entropy bound primarily in the context of normal surfaces, where its application is most straightforward. At first glance, one might suspect that the bound could fail in more complex scenarios—such as those involving trapped or anti-trapped surfaces—where the behavior of light-sheets is less intuitive. However, as we will demonstrate through a concrete example, the bound continues to hold even in such situations. This example will not only reinforce your understanding of how the covariant entropy bound operates across different spacetime geometries but also build your confidence in applying it more broadly. You will begin to appreciate that the mechanisms enforcing the bound can be surprisingly subtle. Rather than following a single elegant principle, they often depend on a delicate interplay of factors specific to the physical context—sometimes appearing almost conspiratorial in how they conspire to uphold the bound. Before diving into the example, let us first clarify a few key definitions.

Trapped Surface: It is a compact, two-dimensional, spacelike surface (usually taken to be topologically a 2-sphere) such that **both** the outgoing and ingoing future-directed null geodesic congruences orthogonal to the surface are **converging**, i.e., their expansions are negative.

This means light rays emitted in **both** directions (inward and outward) from the surface are converging — as happens inside the **event horizon** of a black hole.

Anti-Trapped Surface: It is the **time-reversed** analog of a trapped surface. It is a spacelike 2-surface for which **both** future-directed null congruences **diverge**, i.e., their expansions are positive.

Such surfaces typically occur in **white holes** or **early cosmological epochs** (like during inflation), where light emitted in both directions is diverging due to expansion.

Now consider a spherical surface enclosing a spherically symmetric distribution of matter. In this idealized setup, future-directed ingoing light-rays from the surface converge toward a caustic at the center, much like in the case of an empty sphere. If the matter distribution deviates from perfect spherical symmetry, some light-rays are deflected into angular directions,

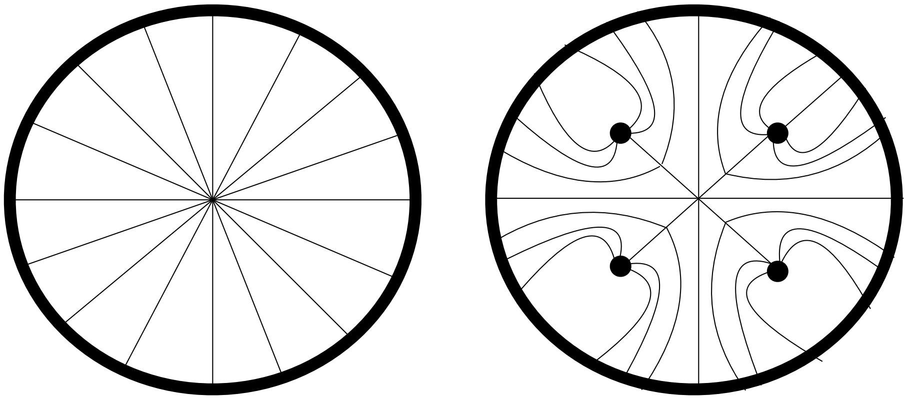

Figure 7: Light-rays incident on a massive object from a spherical surface. For a symmetric interior (left), rays focus at a central caustic. In the asymmetric case (right), mass inhomogeneities deflect rays, producing angular caustics (thick lines) near dense regions.

forming angular caustics (see Fig. 7). This does not prevent the light-sheet from probing the interior entirely—between any two deflected rays, infinitely many continue inward. However, the deflections increase the affine time required for many rays to reach the center.

In static systems, this delay is of no consequence—the light-sheet eventually traverses the whole region. The situation is markedly different during gravitational collapse, where light-rays can encounter a singularity after a finite affine parameter, limiting how far the light-sheet can extend. Consider a spherically symmetric collapsing ball, and construct a future-directed ingoing light-sheet from its outer surface once it lies within its Schwarzschild radius. It is possible to arrange things such that the light-sheet reaches both the central caustic and the singularity at $r = 0$ simultaneously (see Fig. 8). Now suppose we have a collapsing ball of the same mass and radius at the time the light-sheet is launched, but with a highly irregular interior. Although the system remains roughly spherical on large scales, internal inhomogeneities deflect light-rays, forcing them to take intricate, angular paths—effectively undergoing a kind of percolation.

In such highly entropic systems, internal disorder enhances this effect. The spherical cross-sections of the light-sheet become distorted, eventually forming angular caustics that signal the end of the light-sheet. Rays not terminated by caustics may still fail to probe the interior effectively, since they waste affine time navigating angular deviations and may reach the singularity before covering significant depth. Even rays heading straight inward may take longer in underdense regions, experiencing weaker focusing due to the Raychaudhuri equation. The net result is that the light-sheet samples less entropy in highly disordered systems than in more uniform ones.

Crucially, this example involves a trapped surface: both ingoing and outgoing future-directed null expansions are negative, a hallmark of black hole formation. The analysis shows that even in this extreme case—where the surface area is decreasing and the entropy content is high—the covariant entropy bound remains valid. The subtle way the bound is preserved

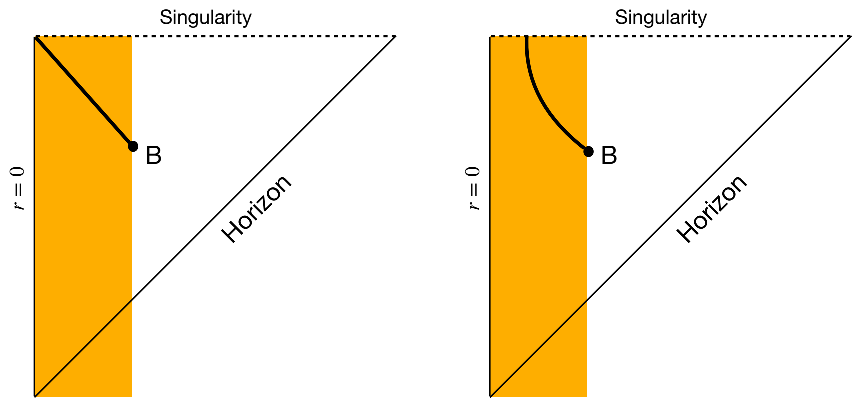

Figure 8: Penrose diagrams for the collapse of two spherical objects with identical mass. Each point represents a two-sphere; thick lines depict future-directed ingoing light-rays. In the symmetric case (left), rays from the surface B reach the center just before hitting the singularity. In the disordered case (right), angular deflections prevent the rays from fully probing the interior before encountering the singularity. The light-ray paths appear timelike due to suppressed angular dimensions in the diagram.

highlights the robustness of the bound.

This reasoning also offers insight into the case of anti-trapped surfaces, where both null expansions are positive, as in the early universe during inflation. There, one considers past-directed light-sheets. While the spacetime geometry is very different, the key idea remains: geometric effects, such as the rapid expansion of space or inhomogeneities in the matter distribution, limit how much entropy a light-sheet can access before it reaches its terminal caustic or hits the initial singularity. Thus, the same intuitive mechanisms that protect the bound in the trapped case extend naturally to anti-trapped regions. Despite their different causal structure, both regimes respect the covariant entropy bound—underscoring its universality across very different physical scenarios. For a simple quantitative estimate supporting this conclusion, see [21].

Why the Holographic Principle Concerns Degrees of Freedom Rather Than Entropy?

In thermodynamics, entropy is a quantity that carries a built-in arrow of time: according to the second law, the entropy of a closed system never decreases and typically increases toward the future. This feature stands in contrast with the time-reversal invariance of most fundamental physical laws, such as those of classical mechanics, quantum field theory, and general relativity. The *covariant entropy bound*, also known as the *Bousso bound*, imposes a universal limit on the entropy that can pass through certain null surfaces—called *light-sheets*—constructed from a given spacelike two-surface B. Specifically, it states that the entropy S on any such light-sheet cannot exceed one quarter of the area A of the surface B, in Planck units:

$$S \leq \frac{A}{4}. \tag{5.27}$$

The surprising aspect of this bound is that it is formulated in a time-reversal invariant way—it can be applied equally to light-sheets propagating into the past or into the future. This appears puzzling because entropy, as a thermodynamic quantity, is not time-symmetric.

The resolution lies in interpreting the covariant entropy bound not as a direct statement about thermodynamic entropy in the usual sense, but as a deeper statement about the number of fundamental degrees of freedom present in a given region of spacetime. In statistical mechanics, entropy is given by

$$S = \ln \Omega, \tag{5.28}$$

where Ω is the number of microstates consistent with the macroscopic configuration. Hence, entropy is bounded above by the logarithm of the number of degrees of freedom accessible in a region. If the entropy on a light-sheet is bounded by $A/4$, then so too is the number of degrees of freedom that can reside on that light-sheet. Importantly, the number of degrees of freedom is not a quantity that depends on the direction of time—it is a timeless, structural feature of the system. Thus, the time-reversal invariance of the covariant entropy bound becomes natural when one interprets the bound as a constraint on information content or state space size, rather than a dynamical statement about entropy flow.

Moreover, the derivation and application of the covariant entropy bound do not rely on any specific microscopic assumptions about matter. No particular field content, symmetries, or ultraviolet cutoffs are invoked. Nevertheless, the consistency of the bound across all physically reasonable scenarios implies that nature cannot accommodate an arbitrarily large number of degrees of freedom per spacetime region. The only viable conclusion is that the number of degrees of freedom is fundamentally finite—a central tenet of the holographic principle. This idea, originally proposed in works by 't Hooft, Susskind, and later Bousso, underlies many modern approaches to quantum gravity and suggests that the true description of spacetime must involve a radical reduction of degrees of freedom compared to standard local quantum field theory.

Part II:
Quantum implementations of the holographic principle

6 Basics of quantum information

Consider the tensor product

$$\mathcal{H} = \mathcal{H}_A + \mathcal{H}_B \tag{6.1}$$

of two Hilbert space $\mathcal{H}_A$ and $\mathcal{H}_B$. If $\{|a_i\rangle\}$ is a basis of $\mathcal{H}_A$ and $\{|b_i\rangle\}$ is a basis of $\mathcal{H}_A$, then $\{|e_{ij}\rangle \equiv |a_i\rangle \otimes |b_j\rangle\}$ is a basis for $\mathcal{H}$.

Given a state (i.e. a density matrix) ρ, the reduced density matrix of system A is given by

$$\rho_A = \mathrm{Tr}_B(\rho) \equiv \sum_j \langle b_j| \rho |b_j\rangle \ , \tag{6.2}$$

and an analogous definition holds for ρ_B.

The von-Neumann entropy of any density matrix is defined as

$$\begin{aligned} S(\rho) &\equiv -\langle \ln \rho \rangle \\ &= -\mathrm{Tr}(\rho \cdot \ln \rho) \\ &= -\sum_i \lambda_i \ln \lambda_i \ , \end{aligned} \tag{6.3}$$

where λ_i are the eigenvalues of ρ. If ρ on the Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ is a pure state (i.e. $\rho = |\psi\rangle\langle\psi|$ for some $|\psi\rangle$), then the von-Neumann entropies $S(\rho_A)$ and $S(\rho_B)$ are equal and we refer to them as entanglement entropy.

Now consider a Hilbert space that splits into many factors, i.e.

$$\mathcal{H} = \bigotimes_i \mathcal{H}_i \ . \tag{6.4}$$

Between two of the factors $\mathcal{H}_i$ and $\mathcal{H}_j$ we define the so-called quantum mutual information as

$$I(i:j) \equiv S(\rho_i) + S(\rho_j) - S(\rho_{ij}) \ , \tag{6.5}$$

where $S(\rho_i)$, $S(\rho_j)$ and $S(\rho_{ij})$ are the von-Neuman entropies of the reduced density matrices in $\mathcal{H}_i$, $\mathcal{H}_j$ and $\mathcal{H}_i \otimes \mathcal{H}_j$ respectively. $I(i:j)$ is symmetric and strictly positive, except if there is no correlation between i and j, i.e. if the reduced density matrix ρ_{ij} is just the product $\rho_i \otimes \rho_j$, in which case $I(i:j) = 0$. So $I(i:j)$ quantifies the degrees of correlations between the subsystems i and j. This can also be seen from the fact that it constitutes an upper bound for the correlation of any two operators O_i and O_j, i.e.

$$\frac{(\langle O_i O_j \rangle - \langle O_i \rangle \langle O_j \rangle)^2}{2\,\|O_i\|^2\,\|O_j\|^2} \leq I(i:j) \ . \tag{6.6}$$

Here, the expectation values are wrt. to the same quantum state wrt. which also the von-Neumann entropies above were computed.

Now consider the Hilbert space

$$\mathcal{H} = \mathcal{H}_q \otimes \mathcal{H}_{\mathrm{rest}} \ , \tag{6.7}$$

where $\mathcal{H}_q$ is the Hilbert space of a qubit. Let σ_x, σ_y, $\sigma_z = -i\sigma_x\sigma_y$ be a Pauli algebra defined on $\mathcal{H}_q$. We can extend these to a representation of the Pauli algebra on $\mathcal{H}$ by defining

$$\Sigma_x \equiv \sigma_x \otimes \mathbb{I}_{\mathrm{rest}} \ , \tag{6.8}$$

$$\Sigma_y \equiv \sigma_y \otimes \mathbb{I}_{\mathrm{rest}} \ , \tag{6.9}$$

$$\Sigma_z \equiv \sigma_z \otimes \mathbb{I}_{\mathrm{rest}} \ , \tag{6.10}$$

where $\mathbb{I}_{\text{rest}}$ is the unit matrix on $\mathcal{H}_{\text{rest}}$.

Now suppose we are only given the operators Σ_x, Σ_y and ρ, but not the corresponding tensor structure of $\mathcal{H}$ (e.g. because Σ_x, Σ_y and ρ are given to us in an arbitrary basis that obscures the product structure). Can we still calculate the reduced density matrix on the factor $\mathcal{H}_q$? We indeed can, as you as asked to show in the following exercise.

Exercise 23

A) *Let $\{|\uparrow\rangle, |\downarrow\rangle\}$ be eigenstates of $\hat{\sigma}_x$ with eigenvalues $\{+1, -1\}$, and let $\rho_q = \text{Tr}_{\text{rest}}$ be the reduced density matrix on the qubit factor. Show that the matrix elements of ρ_q, when expressed in the basis $\{|\uparrow\rangle, |\downarrow\rangle\}$, are given by*

$$\begin{aligned}
\rho_{q,\uparrow\uparrow} &= \text{Tr}\left(|\uparrow\rangle\langle\uparrow| \otimes \mathbb{I}_{\text{rest}} \cdot \rho\right) \\
\rho_{q,\uparrow\downarrow} &= \text{Tr}\left(|\downarrow\rangle\langle\uparrow| \otimes \mathbb{I}_{\text{rest}} \cdot \rho\right) \\
\rho_{q,\downarrow\uparrow} &= \text{Tr}\left(|\uparrow\rangle\langle\downarrow| \otimes \mathbb{I}_{\text{rest}} \cdot \rho\right) \\
\rho_{q,\downarrow\downarrow} &= \text{Tr}\left(|\downarrow\rangle\langle\downarrow| \otimes \mathbb{I}_{\text{rest}} \cdot \rho\right) \ ,
\end{aligned} \tag{6.11}$$

Hint: *Choose an orthonormal basis $\{|a_n\rangle\}$ for $\mathcal{H}_{\text{rest}}$ to spell out what Tr_{rest} means.*

B) *Show that*

$$|\uparrow\rangle\langle\uparrow| \otimes \mathbb{I}_{\text{rest}} = \tfrac{1}{2}\left(\mathbb{I} + \Sigma_x\right) \tag{6.12}$$

$$|\downarrow\rangle\langle\downarrow| \otimes \mathbb{I}_{\text{rest}} = \tfrac{1}{2}\left(\mathbb{I} - \Sigma_x\right) \tag{6.13}$$

$$|\downarrow\rangle\langle\uparrow| \otimes \mathbb{I}_{\text{rest}} = \tfrac{1}{2}\left(\Sigma_z + i\Sigma_y\right) \tag{6.14}$$

$$= \tfrac{i}{2}\left(\Sigma_y - \Sigma_x\Sigma_y\right) \ . \tag{6.15}$$

Consequently, the elements of the reduced density matrix are given by

$$\rho_{q,\uparrow\uparrow} = \tfrac{1}{2}\text{Tr}\left[\left(\mathbb{I} + \Sigma_x\right)\rho\right] \tag{6.16}$$

$$\rho_{q,\downarrow\downarrow} = \tfrac{1}{2}\text{Tr}\left[\left(\mathbb{I} - \Sigma_x\right)\rho\right] \tag{6.17}$$

$$\rho_{q,\uparrow\downarrow} = \tfrac{i}{2}\text{Tr}\left[\left(\Sigma_y - \Sigma_x\Sigma_y\right)\rho\right] \tag{6.18}$$

$$\rho_{q,\downarrow\uparrow} = \tfrac{-i}{2}\text{Tr}\left[\left(\Sigma_y + \Sigma_x\Sigma_y\right)\rho\right] \ . \tag{6.19}$$

Hint: *This is most straightforwardly in the standard representation of the Pauli matrices and their eigenvectors: $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $|\uparrow\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $|\downarrow\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$.*

C) *Now show that the eigenvalues of ρ_q are given by*

$$\lambda = \tfrac{1}{2}\left(1 \pm \sqrt{\left(\text{Tr}\left[\Sigma_x\rho\right]\right)^2 + \left(\text{Tr}\left[\Sigma_y\rho\right]\right)^2 + \left|\text{Tr}\left[\Sigma_x\Sigma_y\rho\right]\right|^2}\right) \tag{6.20}$$

$$= \tfrac{1}{2}\left(1 \pm \sqrt{\left(\text{Tr}\left[\Sigma_x\rho\right]\right)^2 + \left(\text{Tr}\left[\Sigma_y\rho\right]\right)^2 + \left(\text{Tr}\left[\Sigma_z\rho\right]\right)^2}\right) \ . \tag{6.21}$$

6.1 Entanglement 1st law

Any density matrix ρ can be written in the form

$$\rho \equiv \exp(-K) \ , \tag{6.22}$$

where the operator $K = -\ln \rho$ is the so called *modular Hamiltonian.* Now consider a small perturbation to an initial density matrix ρ.

$$\rho \to \rho + \delta\rho \ . \tag{6.23}$$

In order for the new density matrix to still have a trace of 1 we must have

$$\rho \to \rho + \delta\rho \ . \tag{6.24}$$

This can be used to show that

$$\begin{aligned} \delta S &= -\delta \mathrm{Tr}(\rho \ln \rho) \\ &= -\mathrm{Tr}(\delta\rho \ln \rho) - \mathrm{Tr}(\rho\rho^{-1}\delta\rho) \\ &= \mathrm{Tr}(\delta\rho K) \\ &= \delta\langle K\rangle \ . \end{aligned} \tag{6.25}$$

The equation $\delta S = \delta\langle K\rangle$ is called the *entanglement 1st law*, because of its resemblance of the thermodynamic 1st law $\mathrm{d}S = T\mathrm{d}E$.

6.2 Example: modular Hamiltonian and entropy perturbation in Rindler wedge

We had already seen that the Killing vector that preserves the Rindler metric is given by

$$\boldsymbol{\xi} = x_3 \boldsymbol{\partial}_t + t\boldsymbol{\partial}_3 \ , \tag{6.26}$$

where we have set the acceleration of the Rindler observer to be $a = 1$ for simplicity. If $\hat{T}_{\mu\nu}$ is the stress-energy operator (e.g. of some quantum field), then the energy flowing across Rindler time is given by the Rindler Hamiltonian, which at $t = 0$ is given by

$$\begin{aligned} \hat{H}_\xi(t=0) &= \int_{x_3>0} \mathrm{d}^3x \ \xi^0 \ \hat{T}_{00}(0,\mathbf{x}) \\ &= \int_{x_3>0} \mathrm{d}^3x \ x_3 \ \hat{T}_{00}(0,\mathbf{x}) \ . \end{aligned} \tag{6.27}$$

In the global vacuum state, we know that the Rindler observer perceives a thermal state with temperature $T = (2\pi)^{-1}$. Thus, the density matrix of that observer is given by

$$\rho \propto \exp\left(-2\pi \hat{H}_\xi\right) \ . \tag{6.28}$$

Thus we see that the modular Hamiltonian of the Rindler observer is given by $\hat{H}_{\mathrm{mod}} = 2\pi\hat{H}_\xi$. If we now slightly perturb the vacuum state, then the additional entropy that will be registered by the Rindler observer will be given by

$$\begin{aligned} \delta S &= 2\pi \, \delta\langle \hat{H}_\xi\rangle \\ &= 2\pi \int_{x_3>0} \mathrm{d}^3x \ x_3 \ \delta\langle \hat{T}_{00}(0,\mathbf{x})\rangle \qquad (6.29) \\ &\equiv 2\pi \int_{x_3>0} \mathrm{d}^3x \ x_3 \ \delta T_{00}(0,\mathbf{x}) \ . \qquad (6.30) \end{aligned}$$

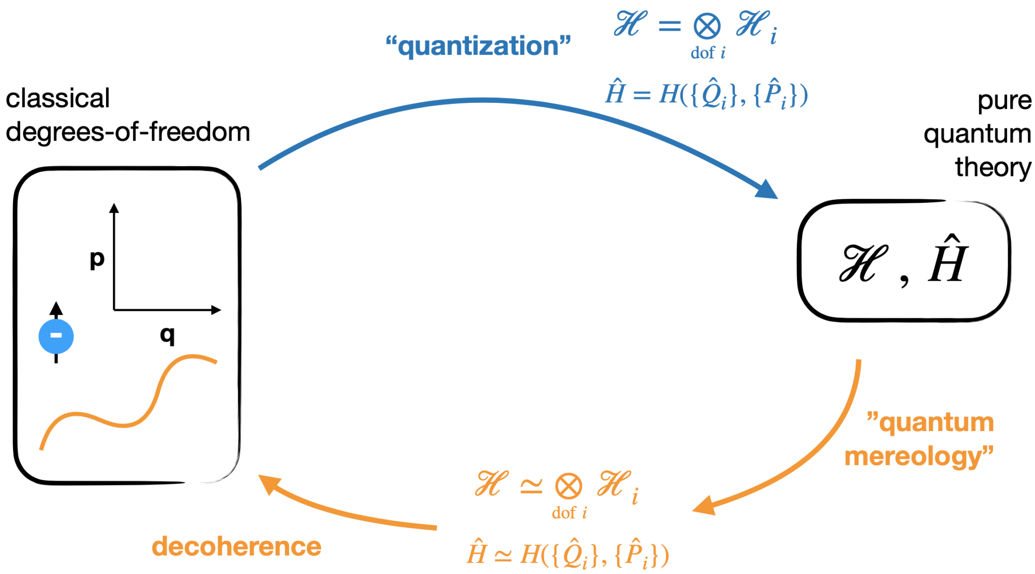

Figure 9: The process of quantisation turns a classical theory into a quantum theory. Quantum mereology (+ decoherence) on the other hand, tries to identify classical dynamics in a Hamiltonian that was defined without any reference to quantised classical variables (e.g. just a random matrix).

7 Quantum mereology (a philosophical backdrop of what's to come)

We now enter the 2nd part of the course, in which we will look into attempts to identify or implement the holographic entropy bound in quantum theories. In this context, the holographic principle is often treated as (part of) a road map to find the theory of quantum gravity. The hope is that if we build quantum theories of matter that obey the holographic principle then we will automatically arrive at something that at least resembles quantum gravity.

At several points in this half of the course we will invoke certain conditions or demand that models have certain properties that may seem ad-hoc and poorly motivated. In this section, we will look at a philosophical backdrop which will serve as an excuse whenever such arbitrary seeming modeling choices appear - quantum mereology [25]. The latter tries to address a somewhat paradoxical situation in modern physics: On the one hand, we claim that the framework of quantum theory (i.e. unitary evolution of states in a Hilbert space, and Hermitian operators as observables) is the most fundamental description of physics. But on the other hand, all actual quantum theories that we ever consider are just "quantized" versions of classical theories (i.e. we "put hats" on classical phase space variables, cf.). Why should the Universe care about what kind of classical phase space dynamics we Humans can imagine? Why is the Hamiltonian $\hat{H}$ of the entire Universe not just some arbitrary Hermitian operator that is defined without any reference to a set of classical degrees of freedom?

The idea of quantum mereology is to do the opposite of quantization. Given any Hamiltonian $\hat{H}$ defined on some Hilbert space $\mathcal{H}$, mereology tries to identify a Hilbert space factorisation

$$\mathcal{H} \simeq \bigotimes_i \mathcal{H}_i \tag{7.1}$$

and conjugate operator pairs $\hat{Q}_i$ and $\hat{P}_i$ on the factors $\mathcal{H}_i$ such that the Hamiltonian $\hat{H}$ induces dynamics for the $\hat{Q}_i$ and $\hat{P}_i$ that look close to classical dynamics. If e.g. the dimension of $\mathcal{H}$ is of the form $\dim \mathcal{H} = m \cdot n$, and if $|e_1\rangle$, ... , $|e_{m\cdot n}\rangle$ is an orthonormal basis of $\mathcal{H}$ then we can e.g. *factorize* $\mathcal{H}$ into two factors $\mathcal{H}_A$ and $\mathcal{H}_B$ via the identification

$$\begin{aligned}
|e_1\rangle &\leftrightarrow |a_1\rangle\,|b_1\rangle \\
|e_2\rangle &\leftrightarrow |a_1\rangle\,|b_2\rangle \\
&\vdots \\
|e_n\rangle &\leftrightarrow |a_1\rangle\,|b_n\rangle \\
|e_{n+1}\rangle &\leftrightarrow |a_2\rangle\,|b_1\rangle \\
&\vdots \\
|e_{m\cdot n}\rangle &\leftrightarrow |a_m\rangle\,|b_n\rangle \ ,
\end{aligned} \tag{7.2}$$

where $\{|a_i\rangle\}$ and $\{|b_i\rangle\}$ are ONBs for $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively. The above identification of course highly depends on our initial choice of the basis $\{|e_i\rangle\}$, and different bases will lead to a factorization of $\mathcal{H}$ into different factors. The goal of mereology is now to find the basis, such that the factorization splits $\mathcal{H}$ into two parts $\mathcal{H}_A$ and $\mathcal{H}_B$ such that the dynamics of these parts resembles a yet to be defined notion of classicality.

By now, there has been a number of studies trying to implement quantum mereology, based on different criteria of what it means for the parts of a Hilbert space to behave classically. We list here an incomplete selection of publications about the topic:

- [26] Zanardi et al. 2004: "Quantum Tensor Product Structures are Observable Induced"
- [27] Piazza 2010: "Glimmers of a Pre-geometric Perspective"
- [28] Cotler et al. 2019: "Locality from the Spectrum"
- [25] Carroll & Singh 2021: "Quantum Mereology: Factorizing Hilbert Space into Subsystems with Quasi-Classical Dynamics"
- [29] Loizeau et al. 2023: "Unveiling Order from Chaos by approximate 2-localization of random matrices"
- [30] Loizeau & Dries 2024: "Quantum mereology and subsystems from the spectrum"
- [31] Zanardi et al. 2024: "Operational Quantum Mereology and Minimal Scrambling"
- [32] Adil et al. 2024: "A Search for Classical Subsystems in Quantum Worlds" .

One quantity that some - but not all - of the above papers invoke in order to judge whether a factorization $\mathcal{H} \simeq \mathcal{H}_A \otimes \mathcal{H}_B$ represents a split of $\mathcal{H}$ into two independent parts is the interaction part of a given Hamiltonian wrt. this split. In the absence of any factorization, the Hamiltonian $\hat{H}$ is of course just some Hermitian matrix (or operator), without any reference to a tensor structure. But once we identify a basis of $\mathcal{H}$ with the basis of a product space as above, then a Hamiltonian can always be de-composed as

$$\hat{H} \;=\; \hat{H}_A \otimes \mathbb{I}_B + \mathbb{I}_A \otimes \hat{H}_B + \tfrac{\mathrm{Tr}(\hat{H})}{d} \cdot \mathbb{I} + \hat{H}_{\mathrm{int}} \ , \tag{7.3}$$

where the self-Hamiltonians $\hat{H}_A$ and $\hat{H}_B$ are traceless and where already the partial traces of the interaction Hamiltonian $\hat{H}_{\text{int}}$ vanish, i.e. $\text{Tr}_B(\hat{H}_{\text{int}}) = 0$ and $\text{Tr}_A(\hat{H}_{\text{int}}) = 0$. Defining an operator norm as $\left\|\hat{O}\right\|^2 \equiv \text{Tr}(\hat{O}^\dagger\hat{O})$, we can e.g. compare the goodness of different factorizations via the loss function[1]

$$\mathcal{L}(\text{factorization}) = \left\|\hat{H}_{\text{int}}\right\|^2 . \tag{7.4}$$

Exercise 24

A) In Equation 7.3 a Hamiltonian operator $\hat{H}$ is decomposed into two different terms. Reflect: what is the meaning of each of these terms?

B) In Equation **??** *of Exercise* **??** *we had already encountered a Hamiltonian in a product Hilbert space. What are the 4 terms in this case? What is the value of the loss function $\left\|\hat{H}_{\text{int}}\right\|^2$ as a function of k (setting $\mu = 1$)? How does that compare to the norms of the self-Hamiltonians of the two qubits?*

C) If you are given a Hamiltonian $\hat{H}$, show how to calculate each of the four terms in Equation 7.3 with the help of the operations Tr, Tr_A *and* Tr_B *and* $\otimes$.

bonus) Apply a random unitary transformation $\hat{H} \to \hat{U}^\dagger\hat{H}\hat{U}$ to the Hamiltonian from Equation 7.3 - e.g. using the python module scipy.stats.unitary_group *- and calculate $\hat{H}_{\text{int}}$ and the loss function again. Did the "classicality" of the factorization get better or worse?*

8 Entanglement entropy of quantum fields across surfaces

In this section we summarize and discuss the results of Srednicki [33], who computed the entanglement entropy of a massless scalar field across the boundary A of a spherical region in flat space (see also [34] for ealier, related work by different authors).

Consider a sub-region $\mathcal{R}$ of space, and its complement $\bar{\mathcal{R}}$. We can think of the Hilbert space of a quantum field that is defined on the full space $\mathcal{R} \times \bar{\mathcal{R}}$ as the product

$$\mathcal{H} = \mathcal{H}_{\mathcal{R}} \otimes \mathcal{H}_{\bar{\mathcal{R}}} , \tag{8.1}$$

where $\mathcal{H}_{\mathcal{R}}$ holds the degrees-of-freedom of the field in side $\mathcal{R}$ and vice versa. A pure quantum state $|\psi\rangle \in \mathcal{H}$ may carry entanglement between $\mathcal{H}_{\mathcal{R}}$ and $\mathcal{H}_{\bar{\mathcal{R}}}$, such that the reduced density matrices

$$\rho_{\mathcal{R}} = \text{Tr}_{\bar{\mathcal{R}}} |\psi\rangle\langle\psi| \quad , \quad \rho_{\bar{\mathcal{R}}} = \text{Tr}_{\mathcal{R}} |\psi\rangle\langle\psi| \tag{8.2}$$

are mixed states. The entanglement entropy induced by $|\psi\rangle$ is the entropy of either the density matrix $\rho_{\mathcal{R}}$ or the density matrix $\rho_{\bar{\mathcal{R}}}$ which is well defined because for pure states in the global Hilbert space we have $S_{\mathcal{R}} = S_{\bar{\mathcal{R}}}$.

As observed by [33], the fact that $S_{\mathcal{R}} = S_{\bar{\mathcal{R}}}$ is already an indication that entanglement entropy might scale like the boundary area $A_{\mathcal{R}} = |\partial\mathcal{R}|$ of the region $\mathcal{R}$, since this is a quantity

[1] In machine learning the term "loss function" commonly denotes a function that is to be minimized by the machine.

that is shared by both $\mathcal{R}$ and $\bar{\mathcal{R}}$. Since entropy is a unitless quantity, the relation between $S_{\mathcal{R}}$ and $A_{\mathcal{R}}$ should be of the form

$$S_{\mathcal{R}} \approx \eta\, A_{\mathcal{R}}/\ell^2 \ , \tag{8.3}$$

where ℓ has the unit of length and η is a constant. A natural choice for ℓ would be $1/\Lambda$ where Λ is the UV-cut of the quantum field we are considering. Srednicki goes on to prove this initial intuition by explicitly calculating the entanglement entropy across a spherical surface (though the argument extends to any sufficiently regular closed surface). We do not repeat his full derivation here (which even he himself could only find numerically). But the following exercise let's you at least explore (part of) the first step of this derivation.

Exercise 25

Entanglement entropy between two coupled harmonic oscillators

Consider the Hamiltonian of two coupled harmonic oscillators,

$$\hat{H} = \frac{\hat{P}_1^2}{2} + \frac{\hat{P}_2^2}{2} + \frac{\omega^2}{2}(\hat{Q}_1^2 + \hat{Q}_2^2) + \frac{k}{2}(\hat{Q}_1 - \hat{Q}_2)^2 \ . \tag{8.4}$$

A) *Introducing the new operators*

$$\hat{Q}_+ = \frac{\hat{Q}_1 + \hat{Q}_2}{\sqrt{2}} \quad , \quad \hat{Q}_- = \frac{\hat{Q}_1 - \hat{Q}_2}{\sqrt{2}} \tag{8.5}$$

as well as their canonically conjugate momenta

$$\hat{P}_+ = \frac{\hat{P}_1 + \hat{P}_2}{\sqrt{2}} \quad , \quad \hat{P}_- = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{2}} \ , \tag{8.6}$$

show that the Hamiltonian becomes

$$\hat{H} = \frac{\hat{P}_+^2}{2} + \frac{\hat{P}_-^2}{2} + \frac{\omega_+^2}{2}\hat{Q}_+^2 + \frac{\omega_-^2}{2}\hat{Q}_-^2 \ , \tag{8.7}$$

which is the Hamiltonian of two decoupled harmonic oscillators with frequencies $\omega_+ = \omega$ *and* $\omega_- = \sqrt{\omega^2 + 2k}$.

B) *Conclude from A) that the ground state wave function* $\psi_0(q_1, q_2) = \langle q_1, q_2|0\rangle$ *is given by*

$$\psi_0(q_1, q_2) = \frac{1}{(\pi\omega_+)^{1/4}(\pi\omega_-)^{1/4}}\ \exp\left(-\frac{\omega_+ q_+^2}{2} - \frac{\omega_- q_-^2}{2}\right) \ . \tag{8.8}$$

C) *We now want to calculate the reduced density matrix of the 2nd oscillator after tracing out the first. Explain why the elements* $\rho_2(q_2, q_2') \equiv \langle q_2|\, \rho_2\, |q_2'\rangle$ *of this density matrix are given by*

$$\rho_2(q_2, q_2') = \int \mathrm{d}q_1 \psi_0(q_1, q_2)^* \psi_0(q_1, q_2') \ . \tag{8.9}$$

Evaluate the right handside of this equation!

D) *Now look at equation (5) of Srednicki's paper* https://arxiv.org/pdf/hep-th/9303048 . *There he defines what the eigenfunctions* $f_n(q_2)$ *and corresponding eigenvalues* p_n *of the density matrix* $\rho_2(q_2, q_2')$ *are, i.e. the* f_n *satisfy*

$$p_n f_n(q_2) = \int \mathrm{d}q_2'\ \rho_2(q_2, q_2') f_n(q_2') \ . \tag{8.10}$$

Verify his ansatz for the case $n = 0$.

Still to be completed

Exercise 26

Questions after lecture 7

Remember the following questions that arose after lecture 7:

A) *Srednicki found that* $S_{\mathcal{R}} \approx 0.3 A/\ell^2$. *Why is the prefactor* 0.3 *and now* 0.25*?*

B) *If we consider* Z *different fields, the entanglement entropy should increase by a factor of* Z. *Why could Srednicki's result nevertheless - i.e. even for many fields - conform to Bekenstein's bound?*

C) *At the same time - do we even need Srednicki's result to agree with Bekenstein's? Are they talking about the same entropy?*

Feel free pause before reading the following comments regarding the above posed questions, if you still want to take some time to think about these questions yourself.

The following are possible answersto questions A) - C) from the above exercise:

A) Note that Srednicki's derivation was based on a particular regularization scheme of quantum field theory - similar to the "pixelization" of the quantum field that we talked about in the lecture. A different regularization - or even the same regularisation but with a discretisation length ℓ that is only slightly different from the Planck length - could easily change Srednicki's result to exactly meet the Bekenstein bound.

B) A number of studies - see e.g. [35–39] for a selection - seem to reveal that quantum field theory must already break down at an energy scale given by the so called *species scale*

$$\Lambda_{\text{sp}} \sim \Lambda_{\text{Planck}}/\sqrt{Z} \ , \tag{8.11}$$

where Z corresponds to the number of distinct particle species. Thus, the correct discretization scale to be used in Srednicki's derivation should be of the form $\ell \sim \sqrt{Z}\ell_{\text{Planck}}$, such that the total entanglement entropy of all quantum fields would again be close to the Bekenstein bound.

C) The Bekenstein bound we discussed earlier in this course referred to ordinary entropy of matter systems, without any mention of entanglement - which is an additional and separate source of entropy e.g. for an observer sitting outside a horizon. In fact, quantum mechanically, such an observer will perceive a total entropy given by

$$S_{\text{total}} = S_{\text{matter}} + S_{\text{ent.}} \ , \tag{8.12}$$

where $S_{\text{ent.}}$ is the entanglement entropy that the observer perceived because of their ignorance of the horizon interior. Assuming that indeed $S_{\text{ent.}} = A/4G$ - as indicated by our discussion about regularization scale as well as the findings regarding the species scale - then the generalised 2nd law that Bekenstein discovered is simply the statement that the total entropy of an observer outside a horizon never decreased. This is very surprising, because GR should in principle not know about entanglement and the entropy associated with it! Why does gravity increase horizon sizes just enough such that the increased entanglement entropy across these horizons compensates the matter entropy flowing across the same horizons? To the best of our knowledge, and answer to this question is still outstanding.

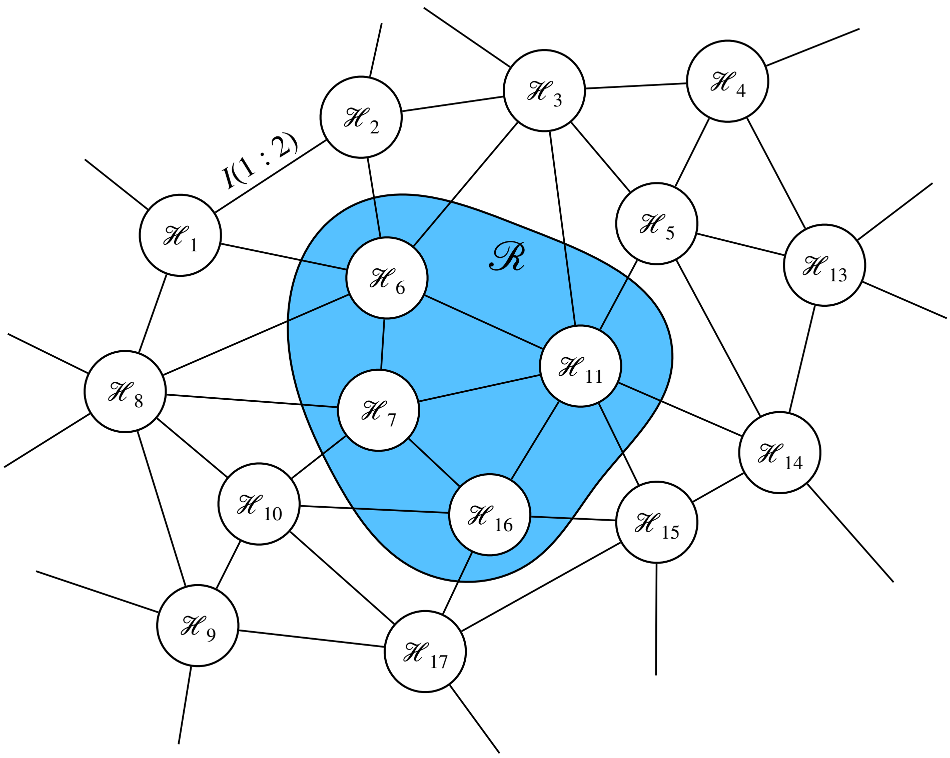

Figure 10: Graph of mutual information between different factors $\mathcal{H}_r$ in the factorization $\mathcal{H} = \bigotimes_r \mathcal{H}_r$. We can assign a boundary area to a region $\mathcal{R}$ consisting of several factors via Equation 9.4.

8.1 Exercises for lecture 7 & feedback form

Lecture 7 is accompanied by exercises 25-28. Also, you can find the feedback form for the lecture here:
https://cloud.physik.lmu.de/index.php/apps/forms/s/eLefQDX4LrZjweCsKtmRdD8S

9 Space from Hilbert space & Bulk entanglement gravity

In this section we follow [40] and [41] who attempt to recover concepts of geometry - and eventually even the Einstein equations - from a pure quantum setting with the help of the entanglement structure between different Hilbert space factors.

Consider a Hilbert space that factorizes as

$$\mathcal{H} = \bigotimes_r \mathcal{H}_r \ . \tag{9.1}$$

In principle, there is no notion of geometry inherent to such a factorization. But in the following we want to interpret the different Hilbert spaces $\mathcal{H}_r$ as holding the degrees of freedom of quantum fields in small regions of space r (you can think of these regions as being chosen just small enough so that classical geometry can still be meaningfully applied to describe each region, i.e. the regions should be bigger than the species scale we discussed in the previous section). And the mutual information $I(r : r')$ between any two factors $\mathcal{H}_r$ and

$\mathcal{H}_{r'}$ at least induced a natural graph structure, as depicted in Figure 10 (see also Section 6 for a summary of the definition and properties of $I(r:r')$).

How can we define a notion of (pre-)geometry from this graph of mutual information? The two papers [40] and [41] give two different but related answers:

Answer 1: distance = mutual information

An answer given by [40] is that a measure for distance between any two factors factor $\mathcal{H}_r$ and $\mathcal{H}_{r'}$ might be given by the quantum mutual information between these factors, which is defined as (cf. Section 6)

$$I(r:r') \equiv S(r) + S(r') - S(rr') \ , \tag{9.2}$$

where $S(r)$, $S(r')$ and $S(rr')$ are the von-Neuman entropies of the reduced density matrices in $\mathcal{H}_r$, $\mathcal{H}_{r'}$ and $\mathcal{H}_r \otimes \mathcal{H}_{r'}$ respectively. In ground states of quantum field theories the distance $d(r, r')$ between any two regions r and r' is a function of the correlation function between the field operators in these regions. We had seen in Equation 6.6 that the mutual information between two regions is an upper bound to such correlation functions. That motivates [40] to postulate a distance measure of the form

$$d(r, r') = \ell\, \Phi\left(\frac{I(r:r')}{I_{\max}}\right) \ , \tag{9.3}$$

where the function Φ is a monotonically decreasing function with $\Phi(1) = 0$, $I_{\max}$ is the mutual information that two maximally entangled factors would have, and ℓ is the size scale we attribute to the regions r (e.g. the species scale). For the purpose of building intuition, [40] propose to think of $\Phi(x)$ as $\approx -\ln(x)$, since this is approximately the behavior of ground state correlation functions in massive QFTs.

Answer 2: area = mutual information

We had seen in Section 8 that in the vacuum state of a QFT the von-Neumann entropy within a closed surface is proportional to the area of that surface. This motivates [41] to define a geometry from an entanglement graph as follows. Consider any larger region $\mathcal{R}$ that includes a number of factors $\mathcal{H}_r$ (cf. Figure 10). Then they postulate that the boundary area between $\mathcal{R}$ and its complement $\bar{\mathcal{R}}$ be given by

$$A(\mathcal{R}, \bar{\mathcal{R}}) = \alpha\, S_{\mathrm{RC}}(\mathcal{R}) \ , \tag{9.4}$$

where α is a yet to be determined constant and $S_{\mathrm{RC}}(\mathcal{R})$ is the so called *redundancy constrained* part of the relative entropy of the region $\mathcal{R}$, which is given by

$$S_{\mathrm{RC}}(\mathcal{R}) \equiv \frac{1}{2} \sum_{r\in\mathcal{R}, r'\in\bar{\mathcal{R}}} I(r:r') \ . \tag{9.5}$$

More generally, given any two regions $\mathcal{R}_1$ and $\mathcal{R}_2$, the area of the (potential) boundary between the two regions would be given by

$$A(\mathcal{R}_1, \mathcal{R}_2) = \frac{\alpha}{2} I_{\mathrm{RC}}(\mathcal{R}_1 : \mathcal{R}_2) \ , \tag{9.6}$$

where

$$I_{\text{RC}}(\mathcal{R}_1 : \mathcal{R}_2) \;\equiv \sum_{r\in\mathcal{R}_1, r'\in\mathcal{R}_2} I(r:r') \; . \tag{9.7}$$

Exercise 27

Consider a region $\mathcal{R}$ that consists of two non-overlapping sub-regions $\mathcal{R}_1$ and $\mathcal{R}_2$. Show that Equations 9.4 and 9.6 lead to consistent definitions of the 3 areas $A(\mathcal{R},\bar{\mathcal{R}})$, $A(\mathcal{R}_1,\bar{\mathcal{R}}_1)$ and $A(\mathcal{R}_2,\bar{\mathcal{R}}_2)$.

States for which $S_{\text{RC}}(\mathcal{R}) = S(\mathcal{R})$, i.e. for which the relative entropy of any region $\mathcal{R}$ is given completely in terms of the mutual information between all the pairs of parts of $\mathcal{R}$ and $\bar{\mathcal{R}}$, are called *redundancy constraint* states. We can think of these states as "vacuum" states, since their entanglement structure satisfies exactly the area scaling that we observed for QFT ground state in Section 8. But of course most states will not be redundancy constrained and generally we expect $S(\mathcal{R})$ to be of the form

$$S(\mathcal{R}) = S_{\text{RC}}(\mathcal{R}) + S_{\text{sub}}(\mathcal{R}) \; , \tag{9.8}$$

where $S_{\text{RC}}(\mathcal{R})$ is given by Equation 9.5 and $S_{\text{sub}}(\mathcal{R})$ is a correction to that. Following [41], we will assume $S_{\text{sub}} \ll S_{\text{RC}}$ and we will comment at the end of this section on how realistic this is, resp. how this would be ensured in practice.

Exercise 28

Redundancy constraint states

A) Consider a Hilbert space that consists of 3 factors, i.e. $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$. Show that any pure state is redundancy constraint wrt. to this factorization.

B) Now consider the Hilbert space $\mathcal{H} = \mathcal{H}_{\text{qubit}} \otimes \mathcal{H}_{\text{rest}}$, where the factor $\mathcal{H}_{\text{qubit}}$ is the Hilbert space of a qubit and $\mathcal{H}_{\text{rest}}$ is some d-dimensional Hilbert space (such that $\mathcal{H}$ has dimension $2d$). Define the operators $\hat{\Sigma}_x$, $\hat{\Sigma}_y$ and $\hat{\Sigma}_z$ as in Equations 6.8 to 6.10, and let $|\downarrow\rangle, |\uparrow\rangle \in \mathcal{H}_{\text{qubit}}$ be eigenvectors of σ_x as in Exercise 23.
Show that

$$\hat{\Sigma}_z \, |\downarrow\rangle \otimes |a\rangle = |\uparrow\rangle \otimes |a\rangle \quad \text{and} \quad \hat{\Sigma}_z \, |\uparrow\rangle \otimes |a\rangle = |\downarrow\rangle \otimes |a\rangle \tag{9.9}$$

for any $|a\rangle \in \mathcal{H}_{\text{rest}}$.

C) Let $\{|a_i\rangle\}$ be an ONB for $\mathcal{H}_{\text{rest}}$. We can view the states $|\downarrow\rangle \otimes |a_i\rangle$ as $2d$-dimensional vectors in $\mathcal{H}$. Let $\boldsymbol{A}$ be the $2d \times d$ matrix that has these vectors as columns, i.e. the ith column of $\boldsymbol{A}$ is $|\downarrow\rangle \otimes |a_i\rangle$.
Now let $\boldsymbol{\rho}$ be the density matrix of the total Hilbert space and $\boldsymbol{\rho}_{\text{rest}} = \text{Tr}_{\text{qubit}}\, \boldsymbol{\rho}$ be the reduced density matrix in $\mathcal{H}_{\text{rest}}$. Show that - in the basis $\{|a_i\rangle\}$ - $\boldsymbol{\rho}_{\text{rest}}$ is given by

$$\boldsymbol{\rho}_{\text{rest}} = \boldsymbol{A}^\dagger \boldsymbol{\rho} \boldsymbol{A} + (\hat{\Sigma}_z \boldsymbol{A})^\dagger \boldsymbol{\rho} (\hat{\Sigma}_z \boldsymbol{A}) \; . \tag{9.10}$$

Also show that if $\hat{O} = \mathbb{I} \otimes \hat{o}$ is some operators on $\mathcal{H}$, then the corresponding operator $\hat{o}$ on $\mathcal{H}_{\text{rest}}$ is - in the basis $\{|a_i\rangle\}$ - given by

$$\hat{o} = \frac{1}{2}\left(\boldsymbol{A}^\dagger \hat{O} \boldsymbol{A} + (\hat{\Sigma}_z \boldsymbol{A})^\dagger \hat{O} (\hat{\Sigma}_z \boldsymbol{A})\right) \; . \tag{9.11}$$

D) *Use your results from B and C to simulate a 4-qubit system* $\mathcal{H} = \mathcal{H}_{q1} \otimes \mathcal{H}_{q2} \otimes \mathcal{H}_{q3} \otimes \mathcal{H}_{q4}$ *with* python *and check whether random pure states are redundancy constraint.* Hint: *from the output of the function* numpy.linalg.eigh *you can extract the matrix* $\boldsymbol{A}$ *. Note that any state* $|\psi\rangle$ *is a complex, normalised vector. You can e.g. compute the density matrix* $\rho = |\psi\rangle\langle\psi|$ *from such a pure state with* numpy.outer .

bonus) *Instead of Equation 9.10 we could also have used*

$$\boldsymbol{\rho}_{\text{rest}} = \boldsymbol{A}^\dagger \boldsymbol{\rho} \boldsymbol{A} + \boldsymbol{B}^\dagger \boldsymbol{\rho} \boldsymbol{B} \ , \tag{9.12}$$

where $\boldsymbol{B}$ *is the matrix with the vectors* $|\uparrow\rangle \otimes |a_i\rangle$ *as columns. What problem would this produce, once we try to implement the calculation of* $\boldsymbol{\rho}_{\text{rest}}$ *with* python*?*

Exercise 29

Preparation for next week

Let $g_{ij} = \delta_{ij} + h_{ij}$ *be the metric of a 2-dimensional Riemannian space, where the perturbation* h_{ij} *ontop of the Euclidean metric* δ_{ij} *is assumed to be small. Show that*

$$\sqrt{|g|} = 1 + \frac{\delta^{ij} h_{ij}}{2} + \mathcal{O}(h^2) \ . \tag{9.13}$$

Now let x^i *be the coordinates wrt. which the metric has the components* g_{ij}*. Compute the change in area* δA *of the space compared to the Euclidean metric at 1st order in the perturbation* h*.*

What we have achieved so far is to define boundary areas for all possible regions of our mutual information graph. According to [41] this is sufficient to fix the Riemannian geometry of a wide class of 2-dimensional spaces. And they refer to results from the mathematical literature that indicate that this may also be true for 3-dimensional spaces, i.e. : it is likely that knowledge of the area of all surfaces in a 3D Riemannian manifold is sufficient to infer the metric tensor of that manifold. We will not attempt to follow the details of these statements, and instead rely on the following assumptions for the rest of this section.

- We assume that there is an exactly redundancy constraint state $|\psi_0\rangle$ such that the mutual information graph can be embedded isometrically into 3D flat space. This means that the points of the graph can be given coordinates in $\mathbb{R}^3$ such that the relative entropy $S(\mathcal{R})$ of any region $\mathcal{R}$ of the graph is equal to α times the area of a closed surface that actually encloses the points of $\mathcal{R}$ in $\mathbb{R}^3$.
- We assume that the actual state of the total Hilbert space is given by $|\psi\rangle = |\psi_0\rangle + |\delta\psi\rangle$ where $|\delta\psi\rangle$ is a small correction to the redundancy constraint state $|\psi_0\rangle$. In particular, the full state $|\psi\rangle$ will contain a small component $S_{\text{sub}} \ll S_{\text{RC}}$ in the total entropy of all the regions of the mutual information graph.
- We assume that the S_{RC} part of the entropy can be attributed to "spacetime degrees-of-freedom" while the S_{sub} part of the entropy can be attributed to "matter degrees-of-freedom" as described by an effective field theory. This means that on top of the factorization $\mathcal{H} = \bigotimes_r \mathcal{H}_r$ each factor can be further decomposed as

$$\mathcal{H}_r \simeq \mathcal{H}_{r,\text{space}} \otimes \mathcal{H}_{r,\text{matter}} \ . \tag{9.14}$$

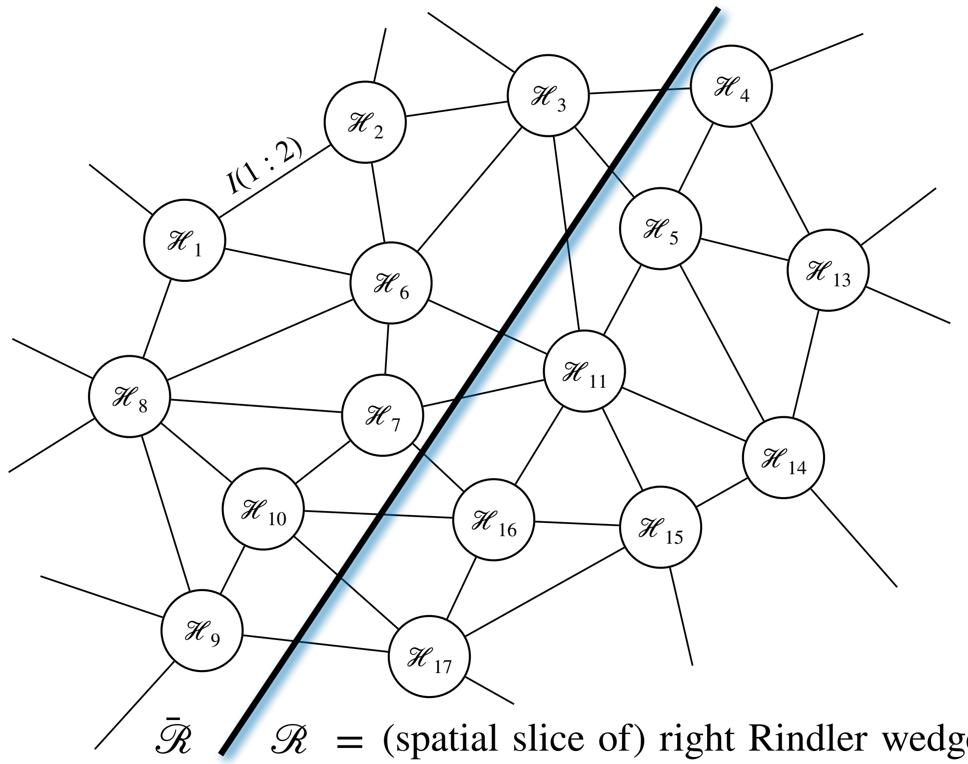

Figure 11: Same as Figure 10, but with the region $\mathcal{R}$ now representing a constant-time slice of a Rindler wedge of the emergent spacetime.

The combined Hilbert space of all "space factors"

$$\mathcal{H}_{\text{space}} = \bigotimes_r \mathcal{H}_{r,\text{space}} \tag{9.15}$$

carries an entanglement structure that is exactly redundancy constrained and thus - per our previous assumption regarding the RC state $|\psi\rangle$ - defines a geometry. And the combined Hilbert space of all matter factors

$$\mathcal{H}_{\text{EFT}} = \bigotimes_r \mathcal{H}_{r,\text{matter}} \tag{9.16}$$

carries an effective field theory that is defined on this geometry.

Starting from the above assumptions, [41] show that small perturbations away from Minkowski space - i.e. away from the exactly redundancy constrained state $|\psi_0\rangle$ - satisfy Einstein's equations is the following entanglement equilibrium condition is met:

$$\delta S_{\text{total}} \equiv \delta S_{\text{RC}} + \delta S_{\text{sub}} \stackrel{!}{=} 0\ . \tag{9.17}$$

Here δ means perturbation away from Minkowski space.

To arrive at their conclusion, let us consider a small perturbation of Minkowski space, described by the metric $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ with $h_{\mu\nu} \ll 1$. It can be shown that in this case the

00-component of the Einstein tensor becomes

$$G_{00} = \frac{1}{2}\left(\partial^i\partial^j h_{ij} - \delta^{ij}\Delta h_{ij}\right) \; + \; \mathcal{O}(h^2) \; . \tag{9.18}$$

We take the RC-part of the entanglement graph to describe a constant time slice in this space. And we now want to split this slide into a region $\mathcal{R}$ and its complement $\bar{\mathcal{R}}$. For simplicity (and following [41]) we choose $\mathcal{R}$ to be one half of the entire space, i.e. using coordinates $\mathbf{x} = (x_1, x_2, x_3)$ we cut the 3D space with a plane at constant $x_3 = p$ and consider $\mathcal{R}$ to be the region with $x_3 > p$ (cf. Figure 11). Of course, the area $A(\mathcal{R}, \bar{\mathcal{R}})$ of the cutting plane is infinite. But area increase due to the metric perturbation h may still be finite, and in Exercise 29 you had shown that this increase is given by

$$\delta A_p = \frac{1}{2}\int_{x_3=p} \mathrm{d}x_1 \mathrm{d}x_2 \; (h_{11} + h_{22}) \; + \; \mathcal{O}(h^2) \; . \tag{9.19}$$

Exercise 30

Using Equation 9.18, show that a linear order in the metric perturbation, δA can be connected to G_{00} via

$$-\frac{\mathrm{d}^2}{\mathrm{d}p^2}\delta A_p = \int_{x_3=p} \mathrm{d}x_1 \mathrm{d}x_2 \; G_{00} \; . \tag{9.20}$$

Green's function for the differential operator $\mathrm{d}^2/\mathrm{d}p^2$ is given by

$$g(p) = \frac{|p|}{2} \; , \tag{9.21}$$

and thus we can invert Equation 9.20 to see that

$$\delta A_p = -\frac{1}{2}\int \mathrm{d}^3x \; |x_3 - p| \; G_{00} \; . \tag{9.22}$$

So the perturbation in the RC-part of the entanglement entropy between $\mathcal{R}$ and $\bar{\mathcal{R}}$ is given by

$$\delta S_{\mathrm{RC}}(\mathcal{R}) = -\frac{1}{2\alpha}\int \mathrm{d}^3x \; |x_3 - p| \; G_{00} \; . \tag{9.23}$$

To also calculate δS_{sub} we make use of the assumption that S_{sub} is carried by a part of the Hilbert space that describes an EFT (i.e. a quantum field theory). To this theory, the region $\mathcal{R}$ acts as a constant-time slice of the right Rindler wedge. Following our discussion of the entanglement 1st law in Section 6.2 we can then write the entanglement entropy perturbation as

$$\begin{aligned} \delta S_{\mathrm{sub}} &= \delta S_{\mathrm{Rindler,right}} \\ &= 2\pi\, \delta\langle \hat{H}_\xi \rangle \\ &= 2\pi \int_{x_3>p} \mathrm{d}^3x \; (x_3 - p) \; \delta T_{00} \; , \end{aligned} \tag{9.24}$$

where $\delta T_{00}(0,\mathbf{x}) \equiv \delta\langle \hat{T}_{00}(0,\mathbf{x})\rangle$ is the perturbation of the (quantum-)expectation value of the field's stress-energy tensor. If the field is in a pure state, then we must have $\delta S_{\text{Rindler,right}} = \delta S_{\text{Rindler,left}}$ and thus

$$\begin{aligned}\delta S_{\text{sub}} &= \frac{1}{2}\left(\delta S_{\text{Rindler,right}} + \delta S_{\text{Rindler,left}}\right) \\ &= \pi \int \mathrm{d}^3x \; |x_3 - p| \; \delta T_{00} \; ,\end{aligned} \tag{9.25}$$

where the part of the integral with $x_3 < p$ represents $\delta S_{\text{Rindler,left}}$.

Inserting Equation 9.23 and Equation 9.25 into the entanglement equilibrium condition of Equation 9.17 then yields

$$\int \mathrm{d}^3x \; |x_3 - p| \; G_{00} \;\; = \;\; 2\pi\alpha \int \mathrm{d}^3x \; |x_3 - p| \; \delta T_{00} \; . \tag{9.26}$$

Since the above equation should hold for any plane that is cutting space into two halves, the authors of [41] arguee that it implies the 00-component of the Einstein equations when setting $\alpha = 4G$. Coincidentally, this is also the value for α that is needed to make Equation 9.4 agree with the Bekenstein-Hawking entropy.

9.1 Original Cao & Carroll axioms

We list here the original axioms of [41].

A1) **Factorization:** The Hilbert space comes with a preferred factorisation $\mathcal{H} = \bigotimes_i \mathcal{H}_i$.

A2) **Redundancy constraint:** We are in a state $|\psi\rangle$ such that any subspace $\mathcal{B}$ has entropy

$$\begin{aligned}S_{\mathcal{B}} &= \frac{1}{2}\sum_{i\in\mathcal{B}, j\notin\mathcal{B}} \mathcal{I}(i:j) \; + \; S_{\text{sub}} \\ &\equiv S_{\text{RC}} + S_{\text{sub}} \; ,\end{aligned} \tag{9.27}$$

where $\mathcal{I}(i:j)$ is the mutual information between $\mathcal{H}_i$ and $\mathcal{H}_j$ in the state $|\psi\rangle$ and $S_{\text{sub}} \ll S_{\mathcal{B}}$. They call such a state *approximately redundancy constrained* and S_{RC} is the redundancy constrained part of the entropy.

A3) **Area from mutual information:** The redundancy constrained part of the entropy of $\mathcal{B}$ can be identified with 4 times the boundary area of $\mathcal{B}$, i.e.

$$S_{\text{RC}} = \frac{\mathcal{A}_{\mathcal{B}}}{4} \; . \tag{9.28}$$

A4) **Entanglement equilibrium:** Under small perturbations, the configurations and subregions we consider have vanishing variation of the entanglement entropy, i.e.

$$0 = \delta S = \delta S_{\text{RC}} + \delta S_{\text{sub}} \; . \tag{9.29}$$

A5) **Emergent field theory:** The sub-leading correction to the RC entropy can be mapped to the variation in entanglement entropy in some EFT, i.e.

$$\delta S_{\text{sub}} = \delta S_{\text{EFT}} \; . \tag{9.30}$$

A6) **Dynamics:** "There exists a consistent dynamical theory, e.g., a Hamiltonian or a quantum circuit, that generates a sequence of such configurations, each admitting an emergent spatial geometry. Furthermore, there is a way to organize these emergent geometries to create a consistent Lorentzian spacetime geometry via time evolution."

A7) **Lorentz invariance:** "The above assumptions hold for any constant-time slice of the emergent Minkowski space, and the overall theory is Lorentz-invariant in an appropriate limit."

9.2 Exercises for lecture 8 & feedback form

Lecture 8 is accompanied by exercises 29-31. Also, you can find the feedback form for the lecture here:
https://cloud.physik.lmu.de/index.php/apps/forms/s/eLefQDX4LrZjweCsKtmRdD8S

9.3 Three derivations of Einstein's equations in different frameworks

The following three papers present very similar arguments for how to derive Einstein's equations, but from complementary starting points:

Faulkner et al.: "Gravitation from Entanglement in Holographic CFTs"
https://arxiv.org/abs/1312.7856

Jacobson: "Entanglement Equilibrium and the Einstein Equation"
https://arxiv.org/abs/1505.04753

Cao & Carroll: "Bulk Entanglement Gravity without a Boundary: Towards Finding Einstein's Equation in Hilbert Space"
https://arxiv.org/abs/1712.02803

10 AdS/CFT & error correcting codes

10.1 anti-de Sitter space

Anti-de Sitter space is a solution to the Friedmann equations that describes a universe with no matter content but with a negative cosmological constant Λ. In this case, the 1st Friedmann equation becomes

$$H^2 = \frac{\Lambda}{3} - \frac{k}{a^2} \, , \tag{10.1}$$

where $H = \dot{a}/a$ is the (Hubble) expansion rate and k is the sign of the spatial curvature of the universe. For $\Lambda < 0$ there can only be a solution to Equation 10.1 if $k = -1$, i.e. anti-de Sitter space is a Universe whose spatial slices are negatively curved.

Exercise 31

Show that Equation 10.1 is solved by

$$a(t) = \ell_\Lambda \sin(t/\ell_\Lambda) \, , \tag{10.2}$$

where $\ell_\Lambda = \sqrt{3/\Lambda}$ *is the curvature scale of anti-de Sitter space.*

Using the result from the above exercise, we can see that the line element of anti-de Sitter space is

$$\mathrm{d}s^2 \;=\; -\mathrm{d}t^2 + \ell_\Lambda^2 \sin\left(\frac{t}{\ell_\Lambda}\right)^2 \left(\mathrm{d}\chi^2 + \sinh(\chi)^2 \left[\mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2\right]\right) \;. \tag{10.3}$$

It seems like this universe has two singularities - the big bang at $t = 0$ and a collapse at $t = \pi/\ell_\Lambda$. But it turns out that these are just coordinate singularities, and there exist coordinates which extend AdS spacetime beyond these singularities (similar to how the Scharzschild radius in black hole spacetimes is just a coordinate singularity, that is e.g. aleviated by Kruskal-Szekeres coordinates). Without proof, we state here that the entirety of Anti-de Sittere space can be covered with coordinated that give the line element

$$\mathrm{d}s^2 = \ell_\Lambda^2 \left(-\cosh(\rho)^2 \mathrm{d}\tau^2 + \mathrm{d}\rho^2 + \sinh(\rho)^2 \left[\mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2\right]\right) \;. \tag{10.4}$$

Exercise 32

Consider a spatial slice of anti-de Sitter space with constant $t = \frac{\pi}{2}\ell_\Lambda$. The spatial line element on this slice is

$$\mathrm{d}\ell^2 \;=\; \mathrm{d}r^2 + \ell_\Lambda^2 \sinh(r/\ell_\Lambda)^2 \left[\mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2\right] \;, \tag{10.5}$$

where we have introduced the physical radius coordinate $r = \ell_\Lambda \chi$. Now consider a sphere in this slice with radius $R \gg \ell_\lambda$. Show that the volume and boundary area of that sphere are approximately given by

$$V(R) \approx \frac{\pi}{2} \ell_K^3 \exp(2R/\ell_K) \tag{10.6}$$

$$A(R) \approx \pi \ell_K^2 \exp(2R/\ell_K) \;. \tag{10.7}$$

The fact that both $V(R)$ and $A(R)$ scale as $\exp(2R/\ell_K)$ eases the problem that QFT seems to over-count degrees-of-freedom compared the the entropy bounds of GR. But it doesn't completely solve this problem either: naively, we expect the number of QFT degrees-of-freedom to be given by $\sim V(R)/\ell_{\text{Planck}}^3$, while the Bekenstein bound would suggest that the actual number of degrees-of-freedom should be $\sim A(R)/\ell_{\text{Planck}}^2$. So QFT is still over-counting degrees-of-freedom by a factor of $\ell_\Lambda/\ell_{\text{Planck}}$. And we expect $\ell_\Lambda \gg \ell_{\text{Planck}}$ in order for the AdS geometry to meaningfully exist. But at least the over-counting problem is more under control here, since it doesn't get worse and worse as the radius R of the region under consideration increases. Heuristically, this a major reason why a quantum implementation of the holographic principle was first achieved via the AdS/CFT correspondence, while an implementation for more realistic spacetimes was (even) more difficult.

10.2 Exercises for lecture 9 & feedback form

Lecture 9 is accompanied by exercises 32-34. Also, you can find the feedback form for the lecture here:
https://cloud.physik.lmu.de/index.php/apps/forms/s/eLefQDX4LrZjweCsKtmRdD8S

10.3 Conformal field theory, conformal boundary of AdS

Consider a manifold $\mathcal{M}$ with Pseudo-Riemannian metric g and line element

$$\mathrm{d}s^2 = g_{\mu\nu}(x)\,\mathrm{d}x^\mu \mathrm{d}x^\nu \;. \tag{10.8}$$

A *conformal transformation* is a change of the metric

$$g_{\mu\nu}(x) \to \tilde{g}_{\mu\nu}(x) = \Omega(x) g_{\mu\nu}(x) \tag{10.9}$$

such that the line element changes to

$$\mathrm{d}\tilde{s}^2 = \Omega(\tilde{x}) g_{\mu\nu}(\tilde{x})\, \mathrm{d}\tilde{x}^\mu \mathrm{d}\tilde{x}^\nu \ . \tag{10.10}$$

Here Ω is some scalar function on $\mathcal{M}$.

The language of conformal transformations allows us to define the *conformal boundary* of AdS. For that purpose, let us consider the global AdS line element in Equation 10.4 for $\rho \gg 1$, i.e.

$$\begin{aligned} \mathrm{d}s^2 &\approx \ell_\Lambda^2 \left(-\frac{1}{4} e^{2\rho} \mathrm{d}\tau^2 + \mathrm{d}\rho^2 + \frac{1}{4} e^{2\rho} \left[\mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2 \right] \right) \\ &= \frac{\ell_\Lambda^2}{4} e^{2\rho} \left(-\mathrm{d}\tau^2 + \left[\frac{1}{2} e^{-\rho} \mathrm{d}\rho \right]^2 + \left[\mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2 \right] \right) \ . \end{aligned} \tag{10.11}$$

Defining the new variable

$$z \equiv -\frac{1}{2} e^{-\rho} \tag{10.12}$$

this becomes

$$\mathrm{d}s^2 \approx \frac{\ell_\Lambda^2}{z^2} \left(-\mathrm{d}\tau^2 + \mathrm{d}z^2 + \left[\mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2 \right] \right) \ . \tag{10.13}$$

The above line element is conformally equivalent to

$$\mathrm{d}\tilde{s}^2 \approx \ell_\Lambda^2 \left(-\mathrm{d}\tau^2 + \mathrm{d}z^2 + \left[\mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2 \right] \right) \ , \tag{10.14}$$

which now has a well defined limit at $z = 0$ i.e. at $\rho = \infty$. The hypersurface at $z = 0$ is called the conformal boundary of AdS and has the induced line element

$$\mathrm{d}s_{\text{bound}}^2 = \ell_\Lambda^2 \left(-\mathrm{d}\tau^2 + \left[\mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2 \right] \right) \ . \tag{10.15}$$

Thus, the topology of the boundary is that of $\mathbb{R} \times S^2$, and the metric is that of a compactified Minkowski space - which is also called a *Minkowski sphere.*

Now a *conformal field theory,* is a field theory on $\mathcal{M}$ that is invariant under conformal transformations $\mathcal{M}$. In practice, this e.g. means that the equations of motion of the fields in the theory do not change under conformal transformations of the metric. Examples of conformal field theories are:

- the massless scalar field in $1 + 1$ dimensions (as you had seen in previous homework!);
- the Ising model at its critical point;
- most renormalizable quantum field theories at the fixed point of their renormalisation group.

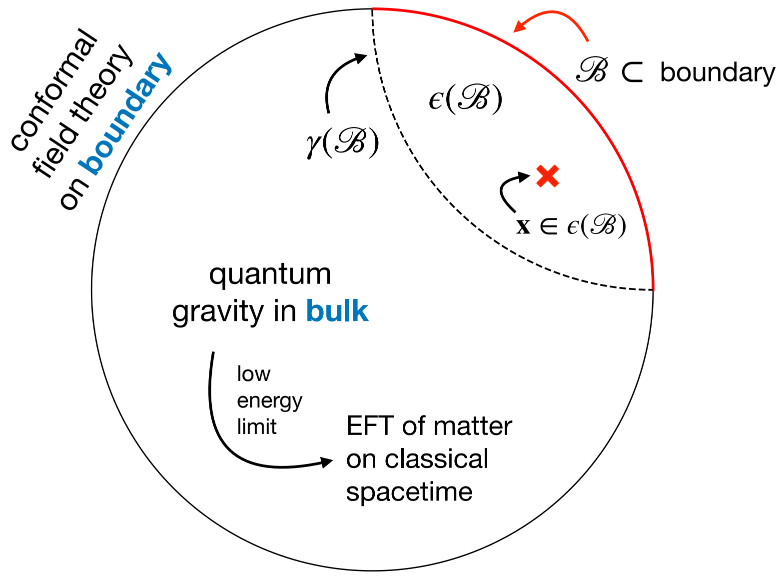

Figure 12: Sketch of the AdS/CFT correspondence

10.4 The dictionary of the AdS/CFT correspondence

The *Anti-de Sitter – Conformal Field Theory* correspondence (AdS/CFT correspondence for short) is a conjectured duality between the theory of quantum gravity on anti-de Sitter space and a conformal field theory defined on the conformal boundary of AdS. Such a duality was first proposed by [42] who took quantum gravity to be represented by a certain string theory model (see also [43–45] for related work that closely followed this initial suggestion). We sketch the setup of this duality in Figure 12. Quantum gravity on AdS lives in the bulk region of this sketch, which you can think of as a hyperbolic plane. And the conformal field theory lives in the boundary of that plane. We had seen in the previous section, that up to conformal transformations, this boundary has the topology and metric of a Minkowski sphere.

A major task in studies of this supposed AdS/CFT correspondence is to build a *dictionary* between quantities in the conformal field theory and their corresponding counter parts in the bulk theory. In Table 2 we list three entries in this dictionary, which can be summarized as follows.

- **correspondence of states:**

  There will be a unitary mapping

$$U_{\text{AdS/CFT}} : \mathcal{H}_{\text{bulk}} \to \mathcal{H}_{\text{CFT}} \tag{10.16}$$

  which maps states in the Hilbert space of quantum gravity on AdS, $\mathcal{H}_{\text{bulk}}$ to the Hilbert space of the boundary CFT, $\mathcal{H}_{\text{CFT}}$. You can think of the bulk Hilbert space as the product

$$\mathcal{H}_{\text{bulk}} = \mathcal{H}_{\text{matter}} \otimes \mathcal{H}_{\text{space}} \ , \tag{10.17}$$

/	bulk quantum gravity	boundary CFT
States	$\vert\psi\rangle_{\text{bulk}}$	$\vert\psi\rangle_{\text{boundary}} = U^{\dagger}_{\text{AdS/CFT}}\, \vert\psi\rangle_{\text{bulk}}$
(local) operators	field $\hat{\phi}(\mathbf{x})$ with $\mathbf{x} \in \epsilon(\mathcal{B})$ or any operator $\hat{O}_{\text{bulk}}$ with support only in $\epsilon(\mathcal{B})$	an operator $\hat{O}_{\text{CFT}}$ that has support only in $\mathcal{B}$
geometry from entanglement	$S(\rho_{\text{EFT},\epsilon(\mathcal{B})}) + \frac{A_{\gamma(\mathcal{B})}}{4}$ RT-Formula (in low energy limit)	$= S(\rho_{\text{CFT},\mathcal{B}})$

Table 2: Entries in the dictionary of the AdS/CFT duality

where $\mathcal{H}_{\text{space}}$ hosts geometrical degrees-of-freedom while $\mathcal{H}_{\text{matter}}$ hosts matter degrees-of-freedom that live on this geometry. In a low energy limit, this theory will be approximated by a classical geometry (governed by the Einstein equations, cf. [46]) and an effective field theory of the matter that populates this geometry.

- **Subregion duality:**

  To define *subregion duality* duality, we first have to acknowledge, that in studies of AdS/CFT usually only a sub-region of anti-de Sitter space with a finite radius R is considered. On the bulk theory, cutting out a finite region from AdS acts as an IR cut-off to quantum gravity, and it has been argued that this actually corresponds to a UV-cutoff for the boundary theory [45].

  Now consider a sub-region $\mathcal{B}$ of the boundary. The so called *Ryu-Takayanagi surface*, or RT-surface, $\gamma(\mathcal{B})$ is the smallest co-dimension one surface in the bulk whose edge coincided with the edge of $\mathcal{B}$ [47]. The red segment of the boundary in Figure 12 e.g. represents such a sub-region $\mathcal{B}$, while the dashed line in the figure represents the corresponding $\gamma(\mathcal{B})$. (Note that in the figure this dashed line is not simply a straight line between the end points of $\mathcal{B}$, because of the hyperbolic geometry of the bulk.)

  Now subregion duality states that to any operator $\hat{O}_{\text{bulk}}$ in the bulk that has support only in the region $\epsilon(\mathcal{B})$ that is enclosed by $\mathcal{B}$ and $\gamma(\mathcal{B})$ there exists a corresponding operator $\hat{O}_{\text{CFT}}$ in the boundary theory that has support only in $\mathcal{B}$. In particular,

$$\hat{O}_{\text{bulk}} = U_{\text{AdS/CFT}}\, \hat{O}_{\text{CFT}}\, U^{\dagger}_{\text{AdS/CFT}}\, . \tag{10.18}$$

- **correspondence of geometry and entanglement:**

  Let $\hat{\rho}_{\text{CFT}}$ be a state (i.e. a density matrix) of the CFT in the low energy limit where the bulk theory behaves like an EFT on a classical spacetime. The bulk EFT will then also be described by some density matrix $\hat{\rho}_{\text{EFT}}$. Now let $\mathcal{B}$, $\gamma(\mathcal{B})$ and $\epsilon(\mathcal{B})$ be defined as before. Then the entropies of the reduced density matrices in $\mathcal{B}$ and $\epsilon(\mathcal{B})$ are related by the so-called Ryu-Takayanagi formula

$$S(\rho_{\text{CFT},\mathcal{B}}) \;=\; S(\rho_{\text{EFT},\epsilon(\mathcal{B})}) + \frac{A_{\gamma(\mathcal{B})}}{4}\, . \tag{10.19}$$

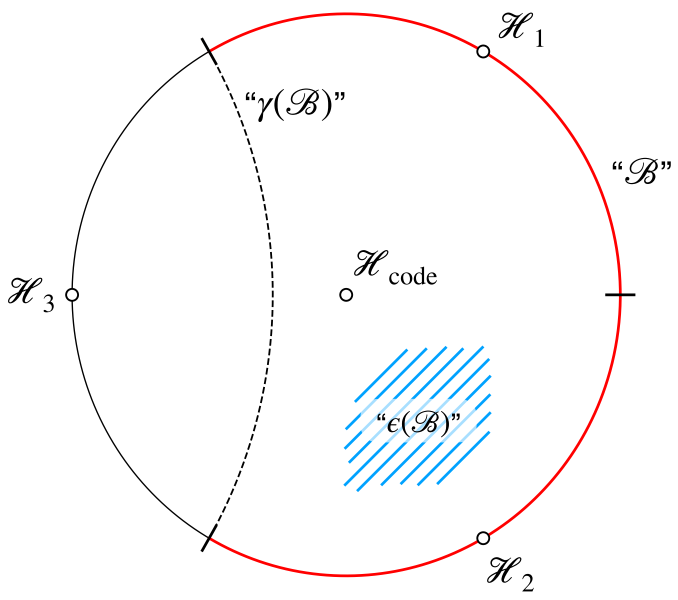

Figure 13: test

10.5 A simple toy model for AdS/CFT

We want to understand the concepts introduced in the previous sub-section by looking at a toy example of the AdS/CFT correspondence that is e.g. described in [48].

Consider a Hilbert space

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \; , \tag{10.20}$$

where each of the three factors represents a three-level system, i.e.

$$\mathcal{H}_n \simeq \mathbb{C}^3 \; . \tag{10.21}$$

For each of these factors we can pick an orthonormal basis which we will denote with $\{|0\rangle \;,\; |1\rangle \;,\; |2\rangle\}$. A basis for the full space is then given by states of the form

$$|i\rangle\,|j\rangle\,|k\rangle \equiv |ijk\rangle \quad . \tag{10.22}$$

As depicted in Figure 13, we want to think of the three Hilbert space factors $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ as three distinct subregions of the boundary of AdS. This, the total Hilbert space $\mathcal{H}$ is supposed to carry the boundary CFT in this toy-model of the AdS/CFT correspondence. At the same time $\mathcal{H}$ is going to carry the bulk theory, i.e. in this toy example $U_{\text{AdS/CFT}}$ is simply the identity on $\mathcal{H}$.

We now define a subspace of $\mathcal{H}_{\text{code}} \subset \mathcal{H}$ as the space spanned by the states

$$|\tilde{0}\rangle \equiv \frac{1}{\sqrt{3}} \left(|000\rangle + |111\rangle + |222\rangle \right) \tag{10.23}$$

$$|\tilde{1}\rangle \equiv \frac{1}{\sqrt{3}} \left(|012\rangle + |120\rangle + |201\rangle \right) \tag{10.24}$$

$$|\tilde{2}\rangle \equiv \frac{1}{\sqrt{3}} \left(|210\rangle + |102\rangle + |021\rangle \right) \ . \tag{10.25}$$

We want to think of $\mathcal{H}_{\text{code}}$ as carrying the matter degrees-of-freedom of the bulk theory, but we use the subscript "code" because our toy model was motivated from analogies between AdS/CFT and quantum error correcting codes - see e.g. [48].

So there is in fact two ways to interpret the above construction:

Viewpoint A: The Hilbert space factors $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ represent subregions of the boundary theory, while $\mathcal{H}_{\text{code}}$ represents matter degrees-of-freedom of the bulk theory. Combining any pair of the boundary factors $\mathcal{H}_n$ yields a boundary region $\mathcal{B}$ whose RT-surface encloses the bulk factor $\mathcal{H}_{\text{code}}$ - cf. the red line in Figure 13. Subsystem duality then tells us, then any operator in $\mathcal{H}_{\text{code}}$ should be representable in terms of a boundary operator that has support only in two of the three boundary factors.

Viewpoint B: $\mathcal{H}_{\text{code}}$ represents the *code subspace* that carries a message, and it is shared between the three systems $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ in such a way that anyone who looses access to any one of the three systems can still decode the message.

To see that subsystem duality indeed holds in our example, note that there is a unitary operator $\hat{U}_{12}$ with support only in $\mathcal{H}_1 \otimes \mathcal{H}_2$ which acts on the states $|\tilde{i}\rangle$ as

$$\hat{U}_{12}\, |\tilde{i}\rangle = |i\rangle\, |\chi\rangle \ , \tag{10.26}$$

where $|i\rangle$ is one of the basis vectors in $\mathcal{H}_1$ and the state $|\chi\rangle \in \mathcal{H}_2 \otimes \mathcal{H}_3$ is given by

$$|\chi\rangle = \frac{1}{\sqrt{3}} \left(|00\rangle + |11\rangle + |22\rangle \right) \ . \tag{10.27}$$

The statement that $\hat{U}_{12}$ has support only in $\mathcal{H}_1 \otimes \mathcal{H}_2$ means that it can be written as

$$\hat{U}_{12} = \hat{V}_{12} \otimes \mathbb{I}_3 \ , \tag{10.28}$$

where $\hat{V}_{12}$ is a unitary operator in $\mathcal{H}_1 \otimes \mathcal{H}_2$ and $\mathbb{I}_3$ is the unit operator in $\mathcal{H}_3$. From viewpoint B we can then interpret Equation 10.26 as follows: The message "i" has been encoded in $\mathcal{H}_{\text{code}}$ and we would like to read this message. This could e.g. be done by measuring the operator

$$\tilde{M} = 0 \cdot |\tilde{0}\rangle \langle\tilde{0}| \, + \, 1 \cdot |\tilde{1}\rangle \langle\tilde{1}| \, + \, 2 \cdot |\tilde{2}\rangle \langle\tilde{2}| \ . \tag{10.29}$$

This operator has support in all three boundary factors, i.e. to actually perform this measurement we would need access to all three systems $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$. But we could in fact still access the message, even if we lost access to e.g. $\mathcal{H}_3$, by applying the unitary $\hat{U}_{12}$ to $\mathcal{H}_1 \otimes \mathcal{H}_2$ and then measuring in $\mathcal{H}_1$ the operator

$$\hat{M}_1 = 0 \cdot |0\rangle \langle 0| \, + \, 1 \cdot |1\rangle \langle 1| \, + \, 2 \cdot |2\rangle \langle 2| \ . \tag{10.30}$$

So far, we have only postulated the existence of the operator $\hat{U}_{12}$, but in the following exercise you will convince yourself that it indeed exists.

Exercise 33

Let us again write $\hat{U}_{12} = \hat{V}_{12} \otimes \mathbb{I}_3$, where $\hat{V}_{12}$ is a unitary operator in $\mathcal{H}_1 \otimes \mathcal{H}_2$ and $\mathbb{I}_3$ is the unit operator in $\mathcal{H}_3$. According to [49], the action of $\hat{V}_{12}$ is given by

$$\hat{V}_{12}\,|00\rangle = |00\rangle \tag{10.31}$$
$$\hat{V}_{12}\,|01\rangle = |12\rangle \tag{10.32}$$
$$\hat{V}_{12}\,|02\rangle = |21\rangle \tag{10.33}$$
$$\hat{V}_{12}\,|10\rangle = |22\rangle \tag{10.34}$$
$$\hat{V}_{12}\,|11\rangle = |01\rangle \tag{10.35}$$
$$\hat{V}_{12}\,|12\rangle = |10\rangle \tag{10.36}$$
$$\hat{V}_{12}\,|20\rangle = |11\rangle \tag{10.37}$$
$$\hat{V}_{12}\,|21\rangle = |20\rangle \tag{10.38}$$
$$\hat{V}_{12}\,|22\rangle = |02\rangle \quad . \tag{10.39}$$

Show that with this definition Equation 10.26 indeed holds.

The above recovery property extends not just to individual basis states $|\tilde{i}\rangle$ but to all operators acting on $\mathcal{H}_{\text{code}}$. Let e.g. $\tilde{O}$ be such an operator with matrix elements

$$\langle\tilde{j}|\,\tilde{O}\,|\tilde{i}\rangle \equiv o_{ij} \quad . \tag{10.40}$$

The we can choose an operator $\hat{O}_1$ on $\mathcal{H}_1$ with the same matrix elements wrt. the basis states $|i\rangle$ of $\mathcal{H}_1$.

Exercise 34

Show that on $\mathcal{H}_{\text{code}}$ the operator $\hat{U}_{12}^{\dagger} \cdot (\hat{O}_1 \otimes \mathbb{I}_{23}) \cdot \hat{U}_{12}$ has the matrix elements

$$\langle\tilde{j}|\;\hat{U}_{12}^{\dagger} \cdot (\hat{O}_1 \otimes \mathbb{I}_{23}) \cdot \hat{U}_{12}\;|\tilde{i}\rangle = \langle\tilde{j}|\,\tilde{O}\,|\tilde{i}\rangle \quad . \tag{10.41}$$

The result of the above exercise tells us, that we can measure any operator on $\mathcal{H}_{\text{code}}$, even if we loose access to the boundary factor $\mathcal{H}_3$. And by symmetry of the above construction, that is also true for any other boundary factor. This is the spirit of the subregion duality in AdS/CFT: we can recover any localised operator in the bulk theory from just a part of the boundary, as long as the entire support of the bulk operator lies between that boundary piece and its RT-surface.

Finally, we are also going to discover a primitive form of the RT-formula in our toy example. Let $\tilde{\rho}$ be a density matrix on $\mathcal{H}_{\text{code}}$. In terms of the basis vectors $|\tilde{i}\rangle$ this can be written as

$$\tilde{\rho} = \sum_{ij} \rho_{ij}\;|\tilde{j}\rangle\,\langle\tilde{i}| \quad . \tag{10.42}$$

Exercise 35

Show that

$$\tilde{\rho} = \hat{U}_{12}^\dagger \left(\hat{\rho}_1 \otimes |\chi\rangle \langle\chi|\right) \hat{U}_{12} \ , \tag{10.43}$$

where $\hat{\rho}_1$ is a density matrix on $\mathcal{H}_1$ given by

$$\hat{\rho}_1 = \sum_{ij} \rho_{ij} \ |j\rangle \langle i| \ . \tag{10.44}$$

Since $\hat{U}_{12}$ does not act on $\mathcal{H}_3$ the reduced density matrix $\tilde{\rho}_{12} \equiv \mathrm{Tr}_3\, \tilde{\rho}$ is given by

$$\begin{aligned}
\tilde{\rho}_{12} &= \mathrm{Tr}_3 \left(\hat{U}_{12}^\dagger \left(\hat{\rho}_1 \otimes |\chi\rangle \langle\chi|\right) \hat{U}_{12}\right) \\
&= \hat{U}_{12}^\dagger \left[\mathrm{Tr}_3 \left(\hat{\rho}_1 \otimes |\chi\rangle \langle\chi|\right)\right] \hat{U}_{12} \\
&= \hat{U}_{12}^\dagger \left(\hat{\rho}_1 \otimes \mathrm{Tr}_3 \left(|\chi\rangle \langle\chi|\right)\right] \hat{U}_{12} \\
&= \frac{1}{3} \hat{U}_{12}^\dagger \left(\hat{\rho}_1 \otimes \mathbb{I}_2\right) \hat{U}_{12} \ .
\end{aligned} \tag{10.45}$$

Exercise 36

Show that the von-Neumann entropy of $\tilde{\rho}_{12}$ is given by

$$S(\tilde{\rho}_{12}) = S(\tilde{\rho}) + \ln 3 \ . \tag{10.46}$$

Hint: *Since $\tilde{\rho}$ and ρ_1 have the same matrix elements in the bases of $\mathcal{H}_{\text{code}}$ resp. $\mathcal{H}_1$, they will have the same von-Neumann entropy.*

Another hint: *Unitary rotations don't change the von-Neumann entropy, i.e.*

$$S(\tilde{\rho}_{12}) = S(\hat{\rho}_1 \otimes \mathbb{I}_2/3) \ .$$

According to [48] the term $\ln 3$ in Equation 10.46 plays the role of the area term in the RT-formula. Similarly, $\tilde{\rho}$ plays the role of the bulk matter density matrix and $\tilde{\rho}_{12}$ plays the role of the density matrix of the boundary theory in a piece of that boundary.

Exercise 37

Extra / further reading: *One of the original formulations of the holographic principle (cf. Section 5.4 of this script) was the paper*

- *"Dimensional Reduction in Quantum Gravity"*
  https://arxiv.org/abs/gr-qc/9310026

by Gerard 't Hooft [24]. Skim through this paper and summarize its main statements for your fellow students.

10.6 Exercises for lecture 11 & feedback form

Lecture 11 is accompanied by exercises 35-38 as well as the "extra" exercise 39. Also, you can find the feedback form for the lecture here:
https://cloud.physik.lmu.de/index.php/apps/forms/s/eLefQDX4LrZjweCsKtmRdD8S

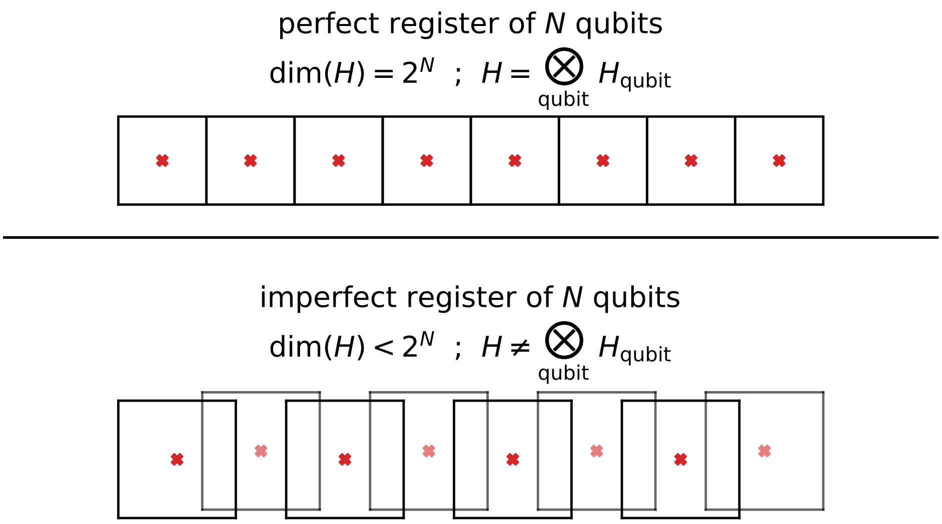

Figure 14: Figure from [50]: The upper sketch illustrates a perfect qubyte consisting of eight non-overlapping qubits. The Hilbert space of the entire qubyte is the tensor product of the individual qubit Hilbert spaces. The lower sketch illustrates the situation when the total Hilbert space is strictly smaller than such a tensor product. In this case the qubits must be overlapping, i.e. the Pauli algebras that define the different qubits are not (anti-)commuting.

11 Overlapping qubits & holographic Weyl field in flat space

In Section 10 we learned about the AdS/CFT correspondence as an attempt to implement the holographic principle (cf. Section 5.4) in a model for quantum gravity. One downside of this approach was the fact that it seems to be specific to anti-de Sitter space, which in turn seems to be quite different from our own Universe. At the same time, the area scaling of maximum entropy (or rather: the Bousso bound) applies in any spacetime. Thus, in non-AdS spacetimes, the overcounting of degrees-of-freedom by quantum field theory persists.

In this section we will look at the approach of [50], who set out to show that the volume many degrees-of-freedom of a quantum field theory in flat space can be "squeezed" into a Hilbert space of log-dimension only scales as area.

11.1 Pauli algebra and Clifford algebra

An complex algebra is a vector space V over the complex number $\mathbb{C}$ that is not just equipped with the usual operations

$$\begin{aligned} \boldsymbol{v}, \boldsymbol{w} \in V \ &\mapsto \ \boldsymbol{v} + \boldsymbol{w} \in V \\ \boldsymbol{v} \in V \,,\, \alpha \in \mathbb{C} \ &\mapsto \ \alpha\, \boldsymbol{v} \in V \end{aligned} \tag{11.1}$$

but also with a (bilinear) product operation

$$\boldsymbol{v}, \boldsymbol{w} \in V \ \mapsto \ \boldsymbol{v} \cdot \boldsymbol{w} \in V \ . \tag{11.2}$$

You can e.g. think of such an algebra as the vector space of complex matrices in some dimension. Clearly, such matrices can be added and multiplied by a complex scalar, but there is also the additional (bilinear) operation of matrix multiplication.

A set of elements $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \in V$ are called the *generators* of the algebra V if any element of V can be formed from the generators via the above three operations. To define the structure of a given algebra is then often sufficient to define how the multiplication operation acts between different generators. For example, the *Pauli algebra* has two generators - which we can e.g. call σ_x and σ_y - that satisfy

$$\{\sigma_x, \sigma_y\} \equiv \sigma_x \cdot \sigma_y + \sigma_y \cdot \sigma_x = 0 \tag{11.3}$$

$$\{\sigma_x, \sigma_x\} = 2 \tag{11.4}$$

$$\{\sigma_y, \sigma_y\} = 2 \; . \tag{11.5}$$

The Pauli algebra is a special case of the so called *Clifford algebras*. For the purpose of this lecture, we define the Clifford algebra of degree n to have generators $C_1, \ldots, C_{2n}$ that satisfy

$$\{C_i\,, C_j\} = \delta_{ij} \; . \tag{11.6}$$

A *representation* of such algebras is a set of matrices in some Hilbert space $\mathcal{H}$ that satisfy the defining multiplication relations of the algebra. In $\mathcal{H} = \mathbb{C}^2$ the Pauli algebra can e.g. be represented by the matrices

$$\boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \boldsymbol{\sigma}_x = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \; . \tag{11.7}$$

But representations of the Pauli algebra also exist in higher dimensional Hilbert spaces, as long as they have even dimension. If $\boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y$ are the generators of such a representation in a Hilbert space $\mathcal{H}$, then there is always a product basis of this Hilbert space, such that

$$\mathcal{H} \simeq \mathbb{C}^2 \otimes \mathcal{H}_{\text{rest}} \tag{11.8}$$

$$\boldsymbol{\Sigma}_x \simeq \boldsymbol{\sigma}_x \otimes \mathbf{1}_{\text{rest}} \tag{11.9}$$

$$\boldsymbol{\Sigma}_y \simeq \boldsymbol{\sigma}_y \otimes \mathbf{1}_{\text{rest}} \; . \tag{11.10}$$

More generally, the Clifford algebra of degree n can always be represented in a Hilbert space with $\dim \mathcal{H} = 2^n$. If $\boldsymbol{C}, \cdots, \boldsymbol{C}_{2n}$ are generators of such a representation (i.e. matrices that satisfy Equation 11.6), then any pair of orthonormal vectors $\boldsymbol{v}\boldsymbol{w} \in \mathbb{R}^{2n}$ can be used to form generators of a representation of the Pauli algebra by defining

$$\boldsymbol{\Sigma}_x \equiv \sum_{i=1}^{2n} v^i \boldsymbol{C}_i \quad , \quad \boldsymbol{\Sigma}_y \equiv \sum_{i=1}^{2n} w^i \boldsymbol{C}_i \; . \tag{11.11}$$

To see that these indeed satisfy the Pauli relations, note e.g. that

$$\begin{aligned} \{\boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y\} &= \sum_{i=1}^{2n} \sum_{j=1}^{2n} v^i w^j \, \{\boldsymbol{C}_i, \boldsymbol{C}_j\} \\ &= 2 \sum_{i=1}^{2n} \sum_{j=1}^{2n} v^i w^j \, \delta_{ij} \\ &= 2 \sum_{i=1}^{2n} v^i w^j \\ &= 2\, \boldsymbol{v}^T \boldsymbol{w} \\ &= 0 \; . \end{aligned} \tag{11.12}$$

11.2 overlapping qubits

Consider a *qubyte*, i.e. a compound system consisting of 8 qubits. The Hilbert space of this system is the tensor product of the individual qubit Hilbert spaces, i.e.

$$\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \ , \tag{11.13}$$

which has the dimension $\dim \mathcal{H} = 2^8$. Each qubit comes with its own representation of the Pauli algebra, e.g.

$$\begin{aligned}
\boldsymbol{\Sigma}_{1,x} &= \boldsymbol{\sigma}_x \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \\
\boldsymbol{\Sigma}_{1,y} &= \boldsymbol{\sigma}_y \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1}
\end{aligned}$$

for the 1st qubit,

$$\begin{aligned}
\boldsymbol{\Sigma}_{2,x} &= \mathbf{1} \otimes \boldsymbol{\sigma}_x \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \\
\boldsymbol{\Sigma}_{2,y} &= \mathbf{1} \otimes \boldsymbol{\sigma}_y \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1}
\end{aligned}$$

for the 2nd qubit, and so on. Because of this exact tensor product structure, operators that belong to different qubits commute, e.g.

$$\{\boldsymbol{\Sigma}_{i,x} \ , \ \boldsymbol{\Sigma}_{j,x}\} = 0 \quad \text{for} \quad i \neq j \ . \tag{11.14}$$

The above situation describes a qubyte in a perfect quantum computer (see e.g. the upper panel of Figure 14 for a graphical depiction). But the authors of [51] considered imperfect quantum computers, where the measurements performed in different qubits influence each other. In particular, they allowed slight deviations from Equation 11.14, i.e. for $i \neq j$ they assumed that

$$\begin{aligned}
\{\boldsymbol{\Sigma}_{i,x} \ , \ \boldsymbol{\Sigma}_{j,x}\} &= \epsilon_{xx,ij} \\
\{\boldsymbol{\Sigma}_{i,x} \ , \ \boldsymbol{\Sigma}_{j,y}\} &= \epsilon_{xy,ij} \\
\{\boldsymbol{\Sigma}_{i,y} \ , \ \boldsymbol{\Sigma}_{j,y}\} &= \epsilon_{yy,ij}
\end{aligned} \tag{11.15}$$

for some small numbers $\epsilon_{xx,ij}$, $\epsilon_{xy,ij}$, $\epsilon_{yy,ij}$.

Interestingly, they found that such an imperfect qubyte can be defined in a Hilbert space with $\dim \mathcal{H} = 2^n < 2^8$. To see how this can be done, simply recall that a Hilbert space with $\dim \mathcal{H} = 2^n$ permits a representation of the Clifford algebra with generators $\boldsymbol{C}, \cdots, \boldsymbol{C}_{2n}$. We had then seen in Section 11.1 that any pair of orthonormal vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^{2n}$ induced a representation of the Pauli algebra via Equations 11.11 . And because of relations 11.8-11.10 this Pauli representation can be seen as *defining a qubit*. In order to create 8 - or more generally: N - qubits, we then simply need N orthonormal vector pairs

$$(\boldsymbol{v}_1, \boldsymbol{w}_1) \, , \ \ldots \ , \ (\boldsymbol{v}_N, \boldsymbol{w}_N) \ .$$

Exercise 38

Show that the corresponding Pauli generators $\boldsymbol{\Sigma}_{i,x}$, $\boldsymbol{\Sigma}_{i,y}$ *satisfy*

$$\{\boldsymbol{\Sigma}_{i,x} \ , \ \boldsymbol{\Sigma}_{j,x}\} = \boldsymbol{v}_i^T \boldsymbol{v}_j \tag{11.16}$$

$$\{\boldsymbol{\Sigma}_{i,x} \ , \ \boldsymbol{\Sigma}_{j,y}\} = \boldsymbol{v}_i^T \boldsymbol{w}_j \tag{11.17}$$

$$\{\boldsymbol{\Sigma}_{i,y} \ , \ \boldsymbol{\Sigma}_{j,y}\} = \boldsymbol{w}_i^T \boldsymbol{w}_j \ . \tag{11.18}$$

Note that the vectors pairs where chosen from $\mathbb{R}^{2n}$. This space has at most $2n$ mutually orthogonal vectors, i.e. n mutually orthogonal pairs of orthogonal vectors. Thus, for $n < N$ is is impossible to choose the pairs $\boldsymbol{v}_i$, $\boldsymbol{w_i}$ such that the anti-commutators from Exercise 38 are vanishing for all $i \neq j$.

In summary: we can have $N > n$ qubits in a Hilbert space of dimension 2^n, but for this we pay the price of having to allow non-zero anti-commutators between the different qubits - cf. the sketch in Figure 14. Such qubits were called *overlapping qubits* by [51]. (Though note that [51] actually defined qubits with non-zero commutators, while we have here constructed qubits with non-zero anti-commutators. The reason for this is that we will now build a Weyl field out of overlapping qubits. And the Fermionic Weyl field satisfies anti-commutation relations.)

But how big is the price we have to pay? The answer is given by [51] with the help of the *Johnson-Lindenstrauss-theorem* (you can find details about this theorems as well as references to original literature in the appendix of [50]). The question that [51] ask is: for fixed n, how many qubits N can we "squeeze" into a Hilbert space of dimension 2^n while still keeping all the (anti-)commutators smaller than some small number ϵ? The exact answer (in the context of anti-commuting qubits) is that N needs satisfy

$$N \leq \frac{\sqrt{d}}{2} \exp\left(\frac{\epsilon^2 n}{8}\right) \ , \tag{11.19}$$

where [51] give a probabilistic algorithm for finding the vectors pairs $(\boldsymbol{v}_1, \boldsymbol{w}_1)$, ... , $(\boldsymbol{v}_N, \boldsymbol{w}_N)$, such that there is a probability of $1 - \delta$ that **all pairs of qubits** have anti-commutators smaller than ϵ . The crucial part about this statement is that N has an upper bound that scales exponentially with n, i.e.

$$N_{\text{max}} \sim \exp\left(\frac{\epsilon^2 n}{8}\right) \ . \tag{11.20}$$

11.3 Holographic Weyl field

Let us consider a quantum Weyl field (the simplest Fermionic quantum field, equivalent to a massless Majorana Fermion) constrained to a finite box with side length L, which we can e.g. take to be the size of the observable Universe. In Appendix B it is shown that this field can be decomposed as

$$\hat{\psi}(\boldsymbol{x}, t) = \sum_{\boldsymbol{p}} \frac{1}{(E_p L^3)^{\frac{1}{2}}} \left\{\hat{c}_{\boldsymbol{p}}(t) \ u(\boldsymbol{p}) \ e^{i\boldsymbol{p}\boldsymbol{x}} + \hat{d}_{\boldsymbol{p}}(t)^\dagger \ u(\boldsymbol{p}) \ e^{-i\boldsymbol{p}\boldsymbol{x}}\right\} \ , \tag{11.21}$$

where $\hat{c}_{\boldsymbol{p}}$ annihilates a particle with momentum $\boldsymbol{p}$ and $\hat{d}_{\boldsymbol{p}}^\dagger$ creates an anti-particle with momentum $\boldsymbol{p}$. Because the Weyl field describes fermions, there can only be one particle and one

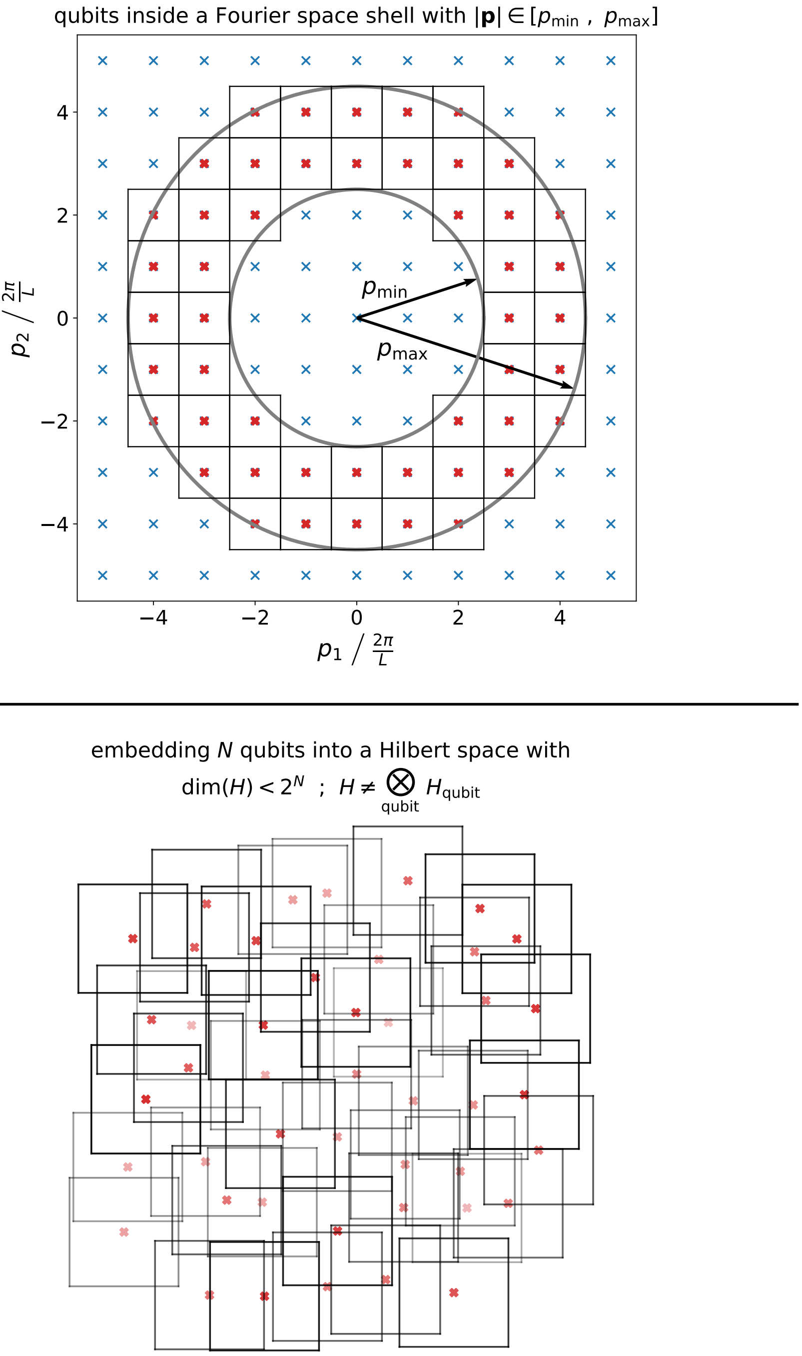

Figure 15: Figure from [50]: The Weyl field can be decomposed into a collection of qubits, which are represented by points $\boldsymbol{p}$ in Fourier space (upper sketch). In 3 dimensions, the number of qubits in a small Fourier space shell s (red crossed in upper sketch) scales as $k_s^2 \Delta_s$ where k_s is the radius and Δ_s is the width of the shell. We want to embed these qubits into a Hilbert space that is strictly smaller then the tensor product of the individual qubit Hilbert spaces in order to achieve a holographic scaling of the effective degrees of freedom in our field.

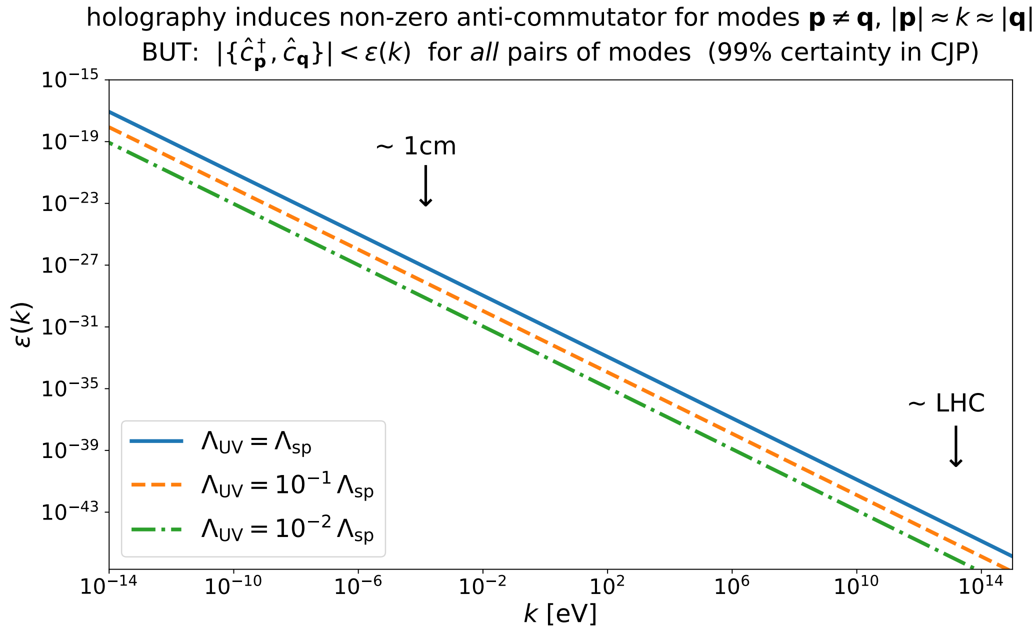

Figure 16: Figure from [50], showing their bound on the anti-commutator $\{\hat{c}^\dagger_{\mathbf{p}}, \hat{c}_{\mathbf{q}}\}$ induced for modes $\mathbf{p} \neq \mathbf{q}$, $|\mathbf{p}| \approx k \approx |\mathbf{q}|$, due to squeezing the Fourier modes of a Weyl field into a Hilbert space whose dimension satisfies an area scaling (as opposed to the volume scaling of usual quantum field theory).

anti-particle at each momentum, i.e. the occupation of particles and anti-particles is describes by a two-level system at each momentum. As is explained in more detail in Appendix B, we can interpret this such that every Fourier mode of the field consists of two qubits.

Because we consider the field in a box, the Fourier modes $\boldsymbol{p}$ are discretized with grid spacing $\delta p = \frac{2\pi}{L}$. A 2D cut through this 3-dimensional Fourier grid is shown in Figure 15, where each cross represent one Fourier mode. In this Fourier grid we can consider a shell s with width Δ_s and radius k_s. The number of Fourier modes located in this shell (represented by the red crosses in Figure 15) is approximately given by

$$N_s \approx 4\pi\, k_s^2 \Delta_s \,/\, \left(\frac{2\pi}{L}\right)^3 \,, \tag{11.22}$$

i.e. it is the volume of the shell divided by the grid spacing cubed. You can think of this as the Fourier space version of the volume scaling of degrees-of-freedom in quantum field theory.

But to satisfy the holographic entropy bounds that we encountered in the first half of the course, we could rather like an area scaling of the number of degrees-of-freedom! This is approximated in [50] by demanding that the true number of degrees-of-freedom in the particle sector of the field (and similarly for the anti-particle sector) is given by

$$n_s \;\propto\; 2\pi\, k_s \Delta_s \,/\, \left(\frac{2\pi}{L}\right)^2 \tag{11.23}$$

in each shell s. How do we achieve this compression from N_s to n_s? The authors of [50] employ the overlapping-qubits-framework of [51] to “squeeze” the N_s qubits of the field into a

Hilbert space of dimension n_s. In particular, they cover the entire Fourier space with shells, and apply the overlapping-qubits-compression to each shell separately.

Note that this is only a polynomial compression, i.e. this does not even need the exponential scaling between N and n that we encountered in the previous section. As a consequence, the upper bound ϵ for the anti-commutators of different Fourier modes can be chosen smaller and smaller as the radius of the Fourier space shells increases. In Figure 16 it is shown that this bound indeed becomes tiny over a wide range of shell radii (which at the same time correspond the the energy of excitations in the shell). In summary: thee authors of [50] used mode overlaps to construct a quantum field theory that satisfies the spherical entropy bound for the Universe, but any pair of Fourier modes still behave as almost perfectly non-overlapping degrees-of-freedom.

Acknowledgments

©2025. We are indebted to the invaluable work of the teams of the public python packages `NumPy` [52], `SciPy` [53], `mpmath` [54], `Matplotlib` [55], `ChainConsumer`[2] and `exsheets`[3]. We have also benefited from discussions with OpenAI's ChatGPT language model.

References

[1] T. Padmanabhan. Gravity and/is Thermodynamics. *arXiv e-prints*, page arXiv:1512.06546, December 2015.

[2] Jacob D. Bekenstein. Black Holes and Entropy. *Phys. Rev. D*, 7(8):2333–2346, April 1973.

[3] S. W. Hawking. Gravitational Radiation from Colliding Black Holes. *Phys. Rev. Lett.*, 26(21):1344–1346, May 1971.

[4] T. Padmanabhan. *Gravitation: Foundations and Frontiers*. Cambridge University Press, January 2010.

[5] Viatcheslav Mukhanov and Sergei Winitzki. *Introduction to Quantum Effects in Gravity*. Cambridge University Press, 2007.

[6] S. W. Hawking. Black hole explosions? *Nature*, 248(5443):30–31, March 1974.

[7] S. W. Hawking. Particle creation by black holes. *Communications in Mathematical Physics*, 43(3):199–220, August 1975.

[8] W. G. Unruh. Notes on black-hole evaporation. *Phys. Rev. D*, 14(4):870–892, August 1976.

[9] G. W. Gibbons and S. W. Hawking. Cosmological event horizons, thermodynamics, and particle creation. *Phys. Rev. D*, 15(10):2738–2751, May 1977.

[10] Ted Jacobson. Thermodynamics of Spacetime: The Einstein Equation of State. *Phys. Rev. Lett.*, 75(7):1260–1263, August 1995.

[11] Leonard Susskind, Lárus Thorlacius, and John Uglum. The stretched horizon and black hole complementarity. *Physical Review D*, 48(8):3743–3761, October 1993.

[12] José P. S. Lemos and Oleg B. Zaslavskii. Membrane paradigm and entropy of black holes in the euclidean action approach. *Physical Review D*, 84(6), September 2011.

[13] Richard H. Price and Kip S. Thorne. Membrane viewpoint on black holes: Properties and evolution of the stretched horizon. *Phys. Rev. D*, 33:915–941, Feb 1986.

[14] Gerard 't Hooft. The black hole interpretation of string theory. *Nucl. Phys. B*, 335:138–154, 1990.

[15] S.D. Mathur. The fuzzball proposal for black holes: an elementary review. *Fortschritte der Physik*, 53(7–8):793–827, June 2005.

[16] Aaron J. Amsel, Donald Marolf, and Amitabh Virmani. Physical process first law for bifurcate Killing horizons. *Phys. Rev. D*, 77(2):024011, January 2008.

[17] Dawood Kothawala, Sudipta Sarkar, and T. Padmanabhan. Einstein's equations as a thermodynamic identity: The cases of stationary axisymmetric horizons and evolving spherically symmetric horizons. *Physics Letters B*, 652(5-6):338–342, September 2007.

[18] Raphael Bousso. The holographic principle. *Reviews of Modern Physics*, 74(3):825–874, August 2002.

[2] https://samreay.github.io/ChainConsumer/

[3] https://bitbucket.org/cgnieder/exsheets/

[19] Jacob D. Bekenstein. Universal upper bound on the entropy-to-energy ratio for bounded systems. *Phys. Rev. D*, 23:287–298, Jan 1981.

[20] Leonard Susskind. The world as a hologram. *Journal of Mathematical Physics*, 36(11):6377–6396, November 1995.

[21] Raphael Bousso. A covariant entropy conjecture. *Journal of High Energy Physics*, 1999(7):004, July 1999.

[22] David Silva Pereira, João Ferraz, Francisco S. N. Lobo, and José Pedro Mimoso. Thermodynamics of the Primordial Universe. *Entropy*, 26(11):947, November 2024.

[23] Robert M. Wald. The Thermodynamics of Black Holes. *Living Reviews in Relativity*, 4(1):6, December 2001.

[24] Gerard 't Hooft. Dimensional reduction in quantum gravity. In *Salamfest 1993:0284-296*, pages 0284–296, 1993.

[25] Sean M. Carroll and Ashmeet Singh. Quantum mereology: Factorizing Hilbert space into subsystems with quasiclassical dynamics. *Phys. Rev. A*, 103(2):022213, February 2021.

[26] Paolo Zanardi, Daniel A. Lidar, and Seth Lloyd. Quantum Tensor Product Structures are Observable Induced. *Phys. Rev. Lett.*, 92(6):060402, February 2004.

[27] Federico Piazza. Glimmers of a Pre-geometric Perspective. *Foundations of Physics*, 40(3):239–266, March 2010.

[28] Jordan S. Cotler, Geoffrey R. Penington, and Daniel H. Ranard. Locality from the Spectrum. *Communications in Mathematical Physics*, 368(3):1267–1296, June 2019.

[29] Nicolas Loizeau, Flaviano Morone, and Dries Sels. Unveiling order from chaos by approximate 2-localization of random matrices. *Proceedings of the National Academy of Science*, 120(39):e2308006120, September 2023.

[30] Nicolas Loizeau and Dries Sels. Quantum mereology and subsystems from the spectrum. *arXiv e-prints*, page arXiv:2409.01391, September 2024.

[31] Paolo Zanardi, Emanuel Dallas, Faidon Andreadakis, and Seth Lloyd. Operational Quantum Mereology and Minimal Scrambling. *Quantum*, 8:1406, July 2024.

[32] Arsalan Adil, Manuel S. Rudolph, Andrew Arrasmith, Zoë Holmes, Andreas Albrecht, and Andrew Sornborger. A Search for Classical Subsystems in Quantum Worlds. *arXiv e-prints*, page arXiv:2403.10895, March 2024.

[33] Mark Srednicki. Entropy and area. *Phys. Rev. Lett.*, 71(5):666–669, August 1993.

[34] Luca Bombelli, Rabinder K. Koul, Joohan Lee, and Rafael D. Sorkin. Quantum source of entropy for black holes. *Phys. Rev. D*, 34(2):373–383, July 1986.

[35] Tao Han and Scott Willenbrock. Scale of quantum gravity. *Physics Letters B*, 616(3-4):215–220, June 2005.

[36] Gia Dvali and Michele Redi. Black hole bound on the number of species and quantum gravity at CERN LHC. *Phys. Rev. D*, 77(4):045027, February 2008.

[37] G. Dvali. Black holes and large N species solution to the hierarchy problem. *Fortschritte der Physik*, 58(6):528–536, June 2010.

[38] G. Dvali, C. Gomez, and D. Lüst. Black hole quantum mechanics in the presence of species. *Fortschritte der Physik*, 61(7-8):768–778, July 2013.

[39] A. Castellano, A. Herráez, and L. E. Ibáñez. The emergence proposal in quantum gravity and the species scale. *Journal of High Energy Physics*, 2023(6):47, June 2023.

[40] ChunJun Cao, Sean M. Carroll, and Spyridon Michalakis. Space from Hilbert space: Recovering geometry from bulk entanglement. *Phys. Rev. D*, 95(2):024031, January 2017.

[41] ChunJun Cao and Sean M. Carroll. Bulk entanglement gravity without a boundary: Towards finding Einstein's equation in Hilbert space. *Phys. Rev. D*, 97(8):086003, April 2018.

[42] Juan Maldacena. The Large-N Limit of Superconformal Field Theories and Supergravity. *International Journal of Theoretical Physics*, 38:1113–1133, January 1999.

[43] S. S. Gubser, I. R. Klebanov, and A. M. Polyakov. Gauge theory correlators from non-critical string theory. *Physics Letters B*, 428(1-2):105–114, May 1998.

[44] Edward Witten. Anti-de Sitter space and holography. *Advances in Theoretical and Mathematical Physics*, 2:253–291, January 1998.

[45] L. Susskind and Edward Witten. The Holographic Bound in Anti-de Sitter Space. *arXiv e-prints*, pages hep–th/9805114, May 1998.

[46] Thomas Faulkner, Monica Guica, Thomas Hartman, Robert C. Myers, and Mark Van Raamsdonk. Gravitation from entanglement in holographic CFTs. *Journal of High Energy Physics*, 2014:51, March 2014.

[47] Shinsei Ryu and Tadashi Takayanagi. Holographic Derivation of Entanglement Entropy from the anti de Sitter Space/Conformal Field Theory Correspondence. *Phys. Rev. Lett.*, 96(18):181602, May 2006.

[48] Daniel Harlow. The Ryu-Takayanagi Formula from Quantum Error Correction. *Communications in Mathematical Physics*, 354(3):865–912, September 2017.

[49] Ahmed Almheiri, Xi Dong, and Daniel Harlow. Bulk locality and quantum error correction in AdS/CFT. *Journal of High Energy Physics*, 2015:163, April 2015.

[50] Oliver Friedrich, ChunJun Cao, Sean M. Carroll, Gong Cheng, and Ashmeet Singh. Holographic phenomenology via overlapping degrees of freedom. *Classical and Quantum Gravity*, 41(19):195003, October 2024.

[51] Rui Chao, Ben W. Reichardt, Chris Sutherland, and Thomas Vidick. Overlapping qubits. *arXiv e-prints*, page arXiv:1701.01062, January 2017.

[52] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[53] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with scipy. *Nature*, 585(7825):357–362, September 2020.

[54] Fredrik Johansson et al. *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.18)*, December 2013. `http://mpmath.org/`.

[55] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[56] Oliver Friedrich, Ashmeet Singh, and Olivier Doré. Toolkit for scalar fields in universes with finite-dimensional Hilbert space. *Classical and Quantum Gravity*, 39(23):235012, December 2022.

[57] E. Kh. Akhmedov. Vacuum energy and relativistic invariance. *arXiv e-prints*, pages hep–th/0204048, April 2002.

[58] Jérôme Martin. Everything you always wanted to know about the cosmological constant problem (but were afraid to ask). *Comptes Rendus Physique*, 13(6-7):566–665, July 2012.

[59] Alberto Güijosa and David A. Lowe. New twist on the dS/CFT correspondence. *Phys. Rev. D*, 69(10):106008, May 2004.

[60] Ning Bao, Sean M. Carroll, and Ashmeet Singh. The Hilbert space of quantum gravity is locally finite-dimensional. *International Journal of Modern Physics D*, 26(12):1743013, January 2017.

[61] Samir D. Mathur. Three puzzles in cosmology. *International Journal of Modern Physics D*, 29(14):2030013, January 2020.

[62] F. Mandl and G. Shaw. *Quantum Field Theory, Revised Edition.* 1994.

[63] Palash B. Pal. Dirac, Majorana, and Weyl fermions. *American Journal of Physics*, 79(5):485–498, May 2011.

A The scalar field as a collection of oscillators

We closely follow the notation of [56], see also [4, 5]. The action of a massive scalar field in Minkowski space is given by

$$S = \frac{1}{2}\int \mathrm{d}t\mathrm{d}^3x \, \left[\dot{\phi}^2 - (\boldsymbol{\nabla}\phi)^2 - m^2\phi^2\right]$$
$$= \frac{1}{2}\int \frac{\mathrm{d}t\mathrm{d}^3k}{(2\pi)^3} \, \left[|\dot{\phi}_{\boldsymbol{k}}|^2 - (|\boldsymbol{k}|^2 + m^2)|\phi_{\boldsymbol{k}}|^2\right] \, , \tag{A.1}$$

where in the second line we moved to Fourier space, with $\boldsymbol{k}$ labeling the Fourier modes. Expressing the Fourier transform of the field in terms of real and imaginary parts, $\phi_{\boldsymbol{k}} = A_{\boldsymbol{k}} + iB_{\boldsymbol{k}}$, we must have $A_{\boldsymbol{k}} = A_{-\boldsymbol{k}}$ and $B_{\boldsymbol{k}} = -B_{-\boldsymbol{k}}$ because ϕ is real. This allows us to define a new field

$$q_{\boldsymbol{k}} = \sqrt{2}\begin{cases} A_{\boldsymbol{k}} \text{ for } k_1 \leq 0 \\ \\ B_{\boldsymbol{k}} \text{ for } k_1 > 0 \end{cases} \tag{A.2}$$

such that the action can be re-written as [4]

$$S = \int \frac{\mathrm{d}t\mathrm{d}^3k}{(2\pi)^3} \, \left[\frac{1}{2}\dot{q}_{\boldsymbol{k}}^2 - \frac{|\boldsymbol{k}|^2 + m^2}{2} q_{\boldsymbol{k}}^2\right] \, . \tag{A.3}$$

To make it explicit that this can be interpreted as a collection of harmonic oscillators, let us restrict the field ϕ to a finite box of side length ℓ_{box}, imposing periodic boundary conditions. This modifies the action to

$$S_{\mathrm{box}} = \int \mathrm{d}t \, \frac{1}{\ell_{\mathrm{box}}^3}\sum_{\boldsymbol{k}} \left[\frac{1}{2}\dot{q}_{\boldsymbol{k}}^2 - \frac{|\boldsymbol{k}|^2 + m^2}{2} q_{\boldsymbol{k}}^2\right] \equiv \int \mathrm{d}t \; L_{\mathrm{box}}\left(\{q_{\boldsymbol{k}}\}, \{\dot{q}_{\boldsymbol{k}}\}, t\right) \, , \tag{A.4}$$

where we have replaced $\mathrm{d}^3k \to \Delta k^3 = (2\pi/\ell_{\mathrm{box}})^3$ and the sum is over all $\boldsymbol{k} = (k_1, k_2, k_3)$ with $k_i \in \{2\pi n/\ell_{\mathrm{box}} \mid n \in \mathbb{Z}\}$. The second equality serves as a definition of the Lagrangian L_{box} of the discretized field. It is literally the Lagrangian of a set of harmonic oscillators with masses $1/\ell_{\mathrm{box}}^3$ and frequencies $\sqrt{|\boldsymbol{k}|^2 + m^2}$. The corresponding Hamiltonian is given by

$$H_{\mathrm{box}}\left(\{q_{\boldsymbol{k}}\}, \{p_{\boldsymbol{k}}\}, t\right) = \sum_{\boldsymbol{k}} \left[\frac{\ell_{\mathrm{box}}^3}{2} \, p_{\boldsymbol{k}}^2 + \frac{1}{\ell_{\mathrm{box}}^3}\frac{(|\boldsymbol{k}|^2 + m^2)}{2} q_{\boldsymbol{k}}^2\right] \, , \tag{A.5}$$

where we introduced the conjugate momenta $p_{\boldsymbol{k}} = \partial L_{\mathrm{box}}/\partial \dot{q}_{\boldsymbol{k}}$. To obtain the quantum theory of this field one would usually promote $q_{\boldsymbol{k}}$ and $p_{\boldsymbol{k}}$ to conjugate Hermitian operators satisfying the Heisenberg commutation relations

$$[\hat{q}_{\boldsymbol{k}}, \hat{p}_{\boldsymbol{k}'}] = i\delta_{\boldsymbol{k},\boldsymbol{k}'} \tag{A.6}$$

such that the Hamiltonian operator of the field becomes

$$\hat{H}(t) = \sum_{\boldsymbol{k}} \left[\frac{\ell_{\mathrm{box}}^3}{2} \, \hat{p}_{\boldsymbol{k}}^2 + \frac{1}{\ell_{\mathrm{box}}^3}\frac{(|\boldsymbol{k}|^2 + m^2)}{2} \hat{q}_{\boldsymbol{k}}^2\right] \, . \tag{A.7}$$

which has the minimum eigenvalue

$$\lambda_{\mathrm{min}}\left[\hat{H}(t)\right] = \sum_{|\boldsymbol{k}|<k_{\mathrm{max}}} \frac{\sqrt{|\boldsymbol{k}|^2 + m^2}}{2} \, . \tag{A.8}$$

Here we have introduced an ultra-violet cut-off $k_{\max}$ in order to regularise this otherwise divergent expression. To obtain the vacuum energy density of the field we need to divide this eigenvalue by the volume of the box, i.e.

$$\begin{aligned}\epsilon_{\text{vac}} &= \frac{1}{\ell_{\text{box}}^3} \sum_{|\boldsymbol{k}|<k_{\max}} \frac{\sqrt{|\boldsymbol{k}|^2 + m^2}}{2} \\ &\approx \int_{|\boldsymbol{k}|<k_{\max}} \frac{\mathrm{d}^3 k}{(2\pi)^3} \frac{\sqrt{|\boldsymbol{k}|^2 + m^2}}{2} \\ &\approx \int_{|\boldsymbol{k}|<k_{\max}} 2\pi \frac{\mathrm{d}k}{(2\pi)^3} k^3 \quad \text{for} \quad k_{\max} \gg m \\ &= \frac{1}{4} \frac{k_{\max}^4}{(2\pi)^2} \ . \end{aligned} \tag{A.9}$$

The sharp cut-off we used in the above expressions has been criticized because it breaks Lorentz symmetry [57, 58]. There are however reasons to believe that the breaking of Lorentz symmetry is physical [59–61]. What's more relevant to us: the sharp cutoff does not sufficiently regularize quantum fields to make them consistent with holography.

B The Weyl field as a collection of qubits

This section is taken from [50] and serves as a reminder of the Weyl field, which is the simples Fermionic quantum field. We also make explicit in which sense the Weyl field can be decomposed into a collection of qubits.

B.1 Weyl field basics

The (left-handed) Weyl spinor is a two-component field ψ with the Lagrangian

$$\mathcal{L} = i\psi^\dagger \sigma^\mu \partial_\mu \psi \ , \tag{B.1}$$

where $\sigma^0 \equiv \mathbf{1}$ and σ^i are e.g. the Pauli matrices (we are following here the notation of [62, 63]). The above Lagrangian leads to the equations of motion

$$\sigma^\mu \partial_\mu \psi = 0 \ . \tag{B.2}$$

A general solution to these equations can be expressed as

$$\psi(\boldsymbol{x}, t) = \int \tfrac{\mathrm{d}^3 p}{(2\pi)^3 E_p} \ \left\{ a_{\boldsymbol{p}}(t) u(\boldsymbol{p}) \ e^{i\boldsymbol{p}\boldsymbol{x}} + b_{\boldsymbol{p}}^*(t) u(\boldsymbol{p}) \ e^{-i\boldsymbol{p}\boldsymbol{x}} \right\} \ , \tag{B.3}$$

where the time evolution of the coefficients $a_{\boldsymbol{p}}$ and $b_{\boldsymbol{p}}^*$ is given by

$$a_{\boldsymbol{p}}(t) = a_{\boldsymbol{p},0} \ e^{-iE_p t} \ , \ b_{\boldsymbol{p}}(t)^* = b_{\boldsymbol{p},0}^* \ e^{iE_p t} \tag{B.4}$$

with $E_p \equiv |\boldsymbol{p}|$, and where $u(\boldsymbol{p})$ are eigenvectors of the matrix $\sigma^j p_j$ with eigenvalues $+E_p$ [4] ,

$$\sigma^j p_j \cdot u(\boldsymbol{p}) = +E_p \ u(\boldsymbol{p}) \ . \tag{B.5}$$

[4] Note that $u(-\boldsymbol{p})$ is then an eigenvector with eigenvalue $-E_p$. This is why only a single family of functions $u(\boldsymbol{p})$ appears in the expansion of Equation B.3.

Note that in the following we will keep the time dependence of $a_{\boldsymbol{p}}$ and $b^*_{\boldsymbol{p}}$ implicit in our notation. The reason is that this dependence will deviate from Equation B.4 once we consider overlapping degrees of freedom.

Normalising the $u(\boldsymbol{p})$ such that

$$u(\boldsymbol{p})^\dagger \cdot u(\boldsymbol{p}) = E_p \tag{B.6}$$

ensures that they are orthogonal wrt. the Lorentz invariant momentum space measure

$$\mathrm{d}\tilde{p} := \frac{\mathrm{d}^3 p}{(2\pi)^3\ E_p}\ . \tag{B.7}$$

Up to an irrelevant phase factor we can e.g. choose $u(\boldsymbol{p})$ as [63]

$$u(\boldsymbol{p}) = \sqrt{E_p}\begin{pmatrix} e^{-i\phi}\sin\frac{\theta}{2} \\ \cos\frac{\theta}{2}\end{pmatrix}\ , \tag{B.8}$$

where $\boldsymbol{p} = (p\sin\theta\cos\phi, p\sin\theta\sin\phi, p\cos\theta)^T$.

In the quantum version of the above field theory we consider the operator valued field

$$\hat{\psi}(\boldsymbol{x}, t) = \int \mathrm{d}\tilde{p}\ \left\{\hat{a}_{\boldsymbol{p}}(t)u(\boldsymbol{p})\ e^{i\boldsymbol{p}\boldsymbol{x}} + \hat{b}_{\boldsymbol{p}}(t)^\dagger u(\boldsymbol{p})\ e^{-i\boldsymbol{p}\boldsymbol{x}}\right\}\ , \tag{B.9}$$

where the operator $\hat{b}^\dagger_{\boldsymbol{p}}$ can be thought of as creating an anti-spinor of momentum $\boldsymbol{p}$ while $\hat{a}_{\boldsymbol{p}}$ is destroying a spinor with momentum $\boldsymbol{p}$. At equal times these operators satisfy the anti-commutation relations

$$0 = \{\hat{a}_{\boldsymbol{p}}, \hat{a}_{\boldsymbol{q}}\} = \{\hat{b}_{\boldsymbol{p}}, \hat{b}_{\boldsymbol{q}}\} = \{\hat{a}_{\boldsymbol{p}}, \hat{b}_{\boldsymbol{q}}\} = \{\hat{a}_{\boldsymbol{p}}, \hat{b}^\dagger_{\boldsymbol{q}}\} \tag{B.10}$$

$$\{\hat{a}_{\boldsymbol{p}}, \hat{a}^\dagger_{\boldsymbol{q}}\} = \{\hat{b}_{\boldsymbol{p}}, \hat{b}^\dagger_{\boldsymbol{q}}\} = (2\pi)^3 E_p\ \delta_D(\boldsymbol{p} - \boldsymbol{q})\ . \tag{B.11}$$

This ensures that the field operators satisfy the equal-time anti-commutation relation

$$\{\hat{\psi}(\boldsymbol{x}), i\hat{\psi}(\boldsymbol{y})^\dagger\} \tag{B.12}$$

$$= i\int \mathrm{d}\tilde{p}\mathrm{d}\tilde{q}\ \left[\{\hat{a}_{\boldsymbol{p}}, a^\dagger_{\boldsymbol{q}}\}\ u(\boldsymbol{p})u(\boldsymbol{q})^\dagger + \{\hat{b}^\dagger_{-\boldsymbol{p}}, b_{-\boldsymbol{q}}\}\ u(-\boldsymbol{p})u(-\boldsymbol{q})^\dagger\right] e^{i\boldsymbol{p}\boldsymbol{x} - i\boldsymbol{q}\boldsymbol{y}}$$

$$= i\int \frac{\mathrm{d}^3 p}{(2\pi)^3 E_p}\ \left[u(\boldsymbol{p})u(\boldsymbol{p})^\dagger + u(-\boldsymbol{p})u(-\boldsymbol{p})^\dagger\right] e^{i\boldsymbol{p}(\boldsymbol{x}-\boldsymbol{y})}$$

$$= i\ \mathbf{1}_{\mathrm{2D}}\ \delta_D(\boldsymbol{x} - \boldsymbol{y})\ , \tag{B.13}$$

as is needed because $i\hat{\psi}^\dagger$ is the conjugate momentum of the field $\hat{\psi}$.

B.2 Decomposition into qubits

To make it explicit that the above field can be considered as a collection of qubits, let us constrict $\psi(\boldsymbol{x})$ to a box of finite size L . This means that we have to perform the substitutions

$$\mathrm{d}^3 p \to \frac{(2\pi)^3}{L^3}\ ,\ \delta_D(\boldsymbol{p} - \boldsymbol{q}) \to \frac{L^3}{(2\pi)^3}\delta_{\boldsymbol{p},\boldsymbol{q}}\ , \tag{B.14}$$

and replace integrals with sums over the discrete grid $\boldsymbol{p} \in \{\ \frac{2\pi}{L}(n_1, n_2, n_3)\ |\ n_i \in \mathbb{Z}\ \}$. For convenience, we will also consider redefined mode operators

$$\hat{c}_{\boldsymbol{p}} = \frac{\hat{a}_{\boldsymbol{p}}}{(E_p V)^{\frac{1}{2}}}\ ,\ \hat{d}_{\boldsymbol{p}} = \frac{\hat{b}_{\boldsymbol{p}}}{(E_p V)^{\frac{1}{2}}}\ , \tag{B.15}$$

such that the new operators satisfy the anti-commutation relations

$$0 = \{\hat{c}_{\boldsymbol{p}}, \hat{c}_{\boldsymbol{q}}\} = \{\hat{d}_{\boldsymbol{p}}, \hat{d}_{\boldsymbol{q}}\} = \{\hat{c}_{\boldsymbol{p}}, \hat{d}_{\boldsymbol{q}}\} = \{\hat{c}_{\boldsymbol{p}}, \hat{d}^\dagger_{\boldsymbol{q}}\} \tag{B.16}$$

$$\{\hat{c}_{\boldsymbol{p}}, \hat{c}^\dagger_{\boldsymbol{q}}\} = \{\hat{d}_{\boldsymbol{p}}, \hat{d}^\dagger_{\boldsymbol{q}}\} = \delta_{\boldsymbol{p},\boldsymbol{q}} \ . \tag{B.17}$$

Our field can now be decomposed in terms of these operators as

$$\hat{\psi}(\boldsymbol{x}, t) = \sum_{\boldsymbol{p}} \frac{1}{(E_p V)^{\frac{1}{2}}} \left\{ \hat{c}_{\boldsymbol{p}}(t)\ u(\boldsymbol{p})\ e^{i\boldsymbol{p}\boldsymbol{x}} + \hat{d}_{\boldsymbol{p}}(t)^\dagger\ u(\boldsymbol{p})\ e^{-i\boldsymbol{p}\boldsymbol{x}} \right\} \ . \tag{B.18}$$

Usually, each of the grid points $\boldsymbol{p}$ in the above sum represents a 4-dimensional Hilbert space factor

$$\mathcal{H}_{\boldsymbol{p}} = \mathcal{H}^c_{\boldsymbol{p}} \otimes^{\mathrm{JW}} \mathcal{H}^d_{\boldsymbol{p}} \tag{B.19}$$

and the total Hilbert space is the tensor product over all these factors,

$$\mathcal{H} = \bigotimes_{\boldsymbol{p}}^{\mathrm{JW}} \mathcal{H}_{\boldsymbol{p}} \ , \tag{B.20}$$

where the superscript "JW" again indicates that operators in the individual Hilbert spaces need to be embedded into the product space via Jordan-Wigner-factors in order for them to anti-commute (as opposed to commute). The $\hat{c}_{\boldsymbol{p}}$, $\hat{d}_{\boldsymbol{p}}$ and their Hermitian conjugates act non-trivially only on the factors $\mathcal{H}^c_{\boldsymbol{p}}$ and $\mathcal{H}^d_{\boldsymbol{p}}$ respectively. On the factor $\mathcal{H}^c_{\boldsymbol{p}}$ (and similarly for $\mathcal{H}^d_{\boldsymbol{p}}$) we can define

$$\hat{\sigma}^c_{x,\boldsymbol{p}} = \hat{c}_{\boldsymbol{p}} + \hat{c}^\dagger_{\boldsymbol{p}} \tag{B.21}$$

$$\hat{\sigma}^c_{y,\boldsymbol{p}} = i\left(\hat{c}_{\boldsymbol{p}} - \hat{c}^\dagger_{\boldsymbol{p}}\right) \tag{B.22}$$

$$\hat{\sigma}^c_{z,\boldsymbol{p}} = -i\hat{\sigma}^c_{x,\boldsymbol{p}}\hat{\sigma}^c_{y,\boldsymbol{p}} = 2\hat{c}^\dagger_{\boldsymbol{p}}\hat{c}_{\boldsymbol{p}} - 1 \ . \tag{B.23}$$

These operators constitute a Pauli algebra on the qubit Hilbert space $\mathcal{H}^c_{\boldsymbol{p}}$. Note however, that the labels x, y, z are simply notation, and not meant to indicate directions in physical space. The Hamiltonian of the field can be expressed in terms of these operators as

$$\begin{aligned} \hat{H} &= \textstyle\sum_{\boldsymbol{p}} E_p \left\{ \left(\hat{c}^\dagger_{\boldsymbol{p}}\hat{c}_{\boldsymbol{p}} - \frac{1}{2}\right) + \left(\hat{d}^\dagger_{\boldsymbol{p}}\hat{d}_{\boldsymbol{p}} - \frac{1}{2}\right) \right\} \\ &= \textstyle\sum_{\boldsymbol{p}} \frac{E_p}{2} \left\{ \hat{\sigma}^c_{z,\boldsymbol{p}} + \hat{\sigma}^d_{z,\boldsymbol{p}} \right\} \ . \end{aligned} \tag{B.24}$$

So our field behaves like a set of non-interacting spins in a $\boldsymbol{p}$-dependent magnetic field. Of course, the occupation of different Fourier modes $\boldsymbol{p}$ of the Weyl field does not measure the state of any actual spins, but rather the existence or non-existence of particles with momentum $\boldsymbol{p}$.

C Penrose diagrams

D Horizon definitions

E Flow across a surface

$$T^{ab} = \rho\, u^a u^b \tag{E.1}$$

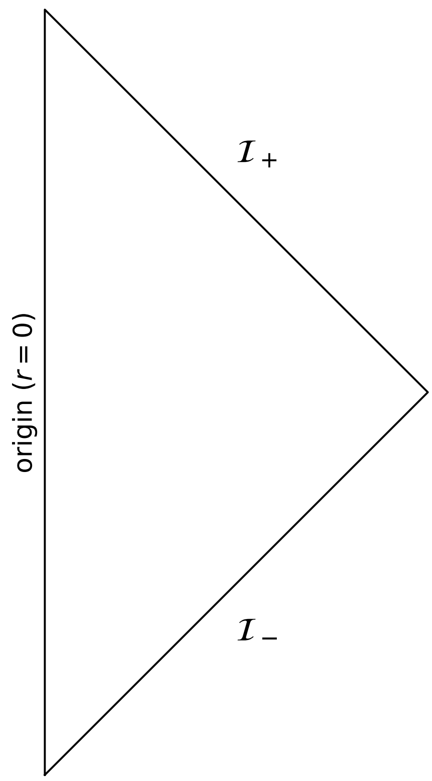

Figure 17: Penrose diagram for Minkowski space

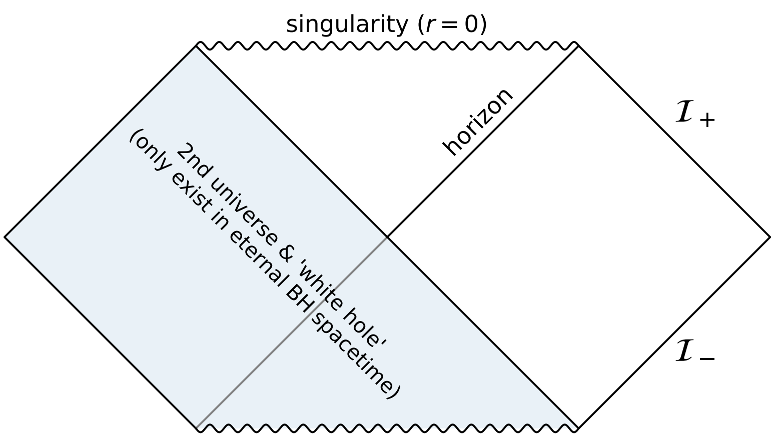

Figure 18: Penrose diagram for the Schwarzschild blackhole

$$\eta^{ab} = \mathrm{diag}(-1, 1, 1, 1) \tag{E.2}$$

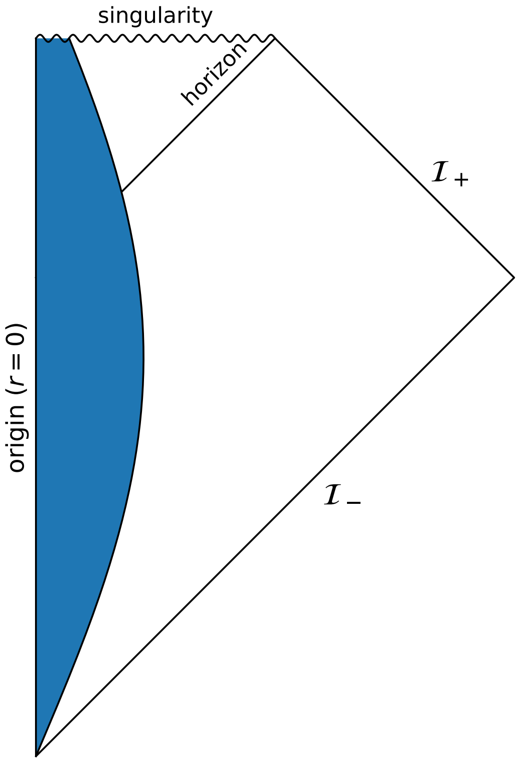

Figure 19: Penrose diagram for a real black hole, forming as a result of the collapse of a massive object (blue shaded region)

$$(u^a) = \gamma(c, \boldsymbol{v}) \tag{E.3}$$
$$(u_a) = \gamma(-c, \boldsymbol{v}) \tag{E.4}$$

$$(\xi^a) = (1, 0, 0, 0) \tag{E.5}$$
$$(\xi_a) = (-1, 0, 0, 0) \tag{E.6}$$

$$(n^a) = (0, 1, 0, 0) \tag{E.7}$$
$$(n_a) = (0, 1, 0, 0) \tag{E.8}$$

$$T^{ab}\xi_a n_b = T^{01}\xi_0 n_1 \tag{E.9}$$
$$= -\rho u^0 u^1 = -\rho\gamma^2 c v_1 \ . \tag{E.10}$$