Advanced General Relativity and Quantum Field Theory in Curved Spacetimes

Lecture Notes

Stefano Liberati

SISSA, IFPU, INFN

January 2020

Preface

The following notes are made by students of the course of "Advanced General Relativity and Quantum Field Theory in Curved Spacetimes", which was held at the International School of Advanced Studies (SISSA) of Trieste (Italy) in the year 2017 by professor Stefano Liberati.

Being the course directed to PhD students, this work and the notes therein are aimed to interested readers that already have basic knowledge of Special Relativity, General Relativity, Quantum Mechanics and Quantum Field Theory; however, where possible the authors have included all the definitions and concept necessary to understand most of the topics presented.

The course is based on different textbooks and papers; in particular, the first part, about Advanced General Relativity, is based on:

- "General Relativity" by R. Wald [1]
- "Spacetime and Geometry" by S. Carroll [2]
- "A Relativist Toolkit" by E. Poisson [3]
- "Gravitation" by T. Padmanabhan [4]

while the second part, regarding Quantum Field Theory in Curved Spacetime, is based on

- "Quantum Fields in Curved Space" by N. C. Birrell and P. C. W. Davies [5]
- "Vacuum Effects in strong fields" by A. A. Grib, S. G. Mamayev and V. M. Mostepanenko [6]
- "Introduction to Quantum Effects in Gravity", by V. F Mukhanov and S. Winitzki [7]

Where necessary, some other details could have been taken from paper and reviews in the standard literature, that will be listed where needed and in the Bibliography.

Every possible mistake present in these notes are due to misunderstandings and imprecisions of which only the authors of the following works must be considered responsible.

Credits

- Andrea Oddo (Chapters 1, 2, 3, 6, 7, Appendix A; revision)
- Giovanni Tambalo (Chapters 8, 9, 10, 11)
- Lumen Boco (Chapters 4, 5)

Also thanks to Giovanni Tricella and Paolo Campeti for useful scientific discussions.

Contents

Part I

Advanced General Relativity

1 Foundations of Relativity

1.1 Definitions and postulates

We often refer to spacetime as a metaphysical entity, but in physics this would never be the case; therefore, our first task is to define both space and time in an operative fashion. We can replace the concepts of time with duration measured by a clock and space with distance measured with a rod. This is also because duration and distance are experimental quantities, that are directly meaningful, and that don't need any interpretation whatsoever. With this, the properties of space and time are now directly related to the respective properties of rods and clocks. This gives us an operative definition of spacetime, rather than a metaphysical definition.

We can then define an **event**. We will assume this to be a primitive notion in physics, in the same way a point is a primitive notion in geometry. In particular, an event is something that happens for a very small period time Δt in a very small region of space Δx; in this way, we can define our spacetime as the collection of all possible events: given a generic event $\mathcal{E}$, a spacetime $\mathcal{M}$ is

$$\mathcal{M} \equiv \bigcup \mathcal{E}. \tag{1.1}$$

We can postulate that every measurement is either local or reducible to local, and in this way we can introduce the notion of an **observer**. We can imagine to place an observer in each point of space, and hence define a set of observers $\mathcal{O}_i$ in our spacetime $\mathcal{M}$, that measure in different locations of spacetime. All these observers have trajectories in spacetime, and therefore we can define the set of these worldlines as a **congruence** of observers, one for each point in $\mathcal{M}$, in the sense that the worldlines of all the observers never cross. The set of these worldines is $\mathcal{U} \subset \mathcal{M}$.

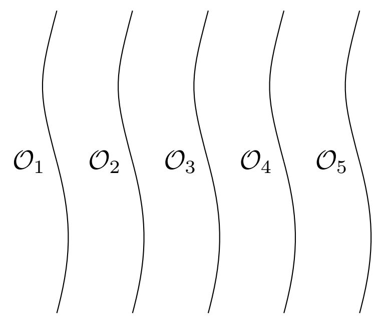


Figure 1.1: Pictorial representation of a congruence of worldlines, each one representing one observer (or a point in $\mathcal{M}$).

One can go from an observer to another by performing a boost from the frame of the first observer to the frame of the second observer through a coordinate transformation. We want now to relate observations in different frames, and therefore we want to agree on a system of rulers and clocks.

Once we agree on a recipe to build a clock (e.g. quartz oscilation, period of radiation from atomic transitions, etc...), we can give a clock to each observer in the set, thus introducing the concept of local time.

To do kinematics, we need to relate measures made by different observers, so we need to synchronize clocks. In order to do this, we want a way to relate measures of different observers: we introduce a function $t : \mathcal{U} \to \mathbb{R}$, such that $t(\mathcal{O}_i)$ gives the proper time on that worldline. Synchronizing the clocks means to characterize the difference in time, or the lapse of time, between two events: $\Delta t = t(\mathcal{E}_2) - t(\mathcal{E}_1)$. Events $\mathcal{E}_1$ and $\mathcal{E}_2$ are said to be simultaneous if $t(\mathcal{E}_1) = t(\mathcal{E}_2)$.

We can now separate space from spacetime and define it in the following way: given a time t_0 we can define the space S_{t_0} at t_0 as the set of events such that their time is t_0

$$S_{t_0} = \{\mathcal{E} \,|\, t(\mathcal{E}) = t_0\}\,, \tag{1.2}$$

that is, a slice at some t_0 that we can call "space". Note that there is no independent notion of space from the system of observers. We can also define co-local events as events that happen on the same worldline of an observer.

If an experiment takes place in a sufficiently small region of spacetime, then we can introduce a postulate: there is always a system of observers, along with a procedure of synchronization and choices of length units, so that the distance Δx that exists between two observers does not depend on time, and this distance respects the euclidean geometry axioms. In practice, it is always possible to find a region of spacetime that is locally euclidean.

In general we can choose time arbitrarily, but not all choices are equivalent: we naturally prefer to have a simple function defining time, for example a time function such that the equation of motion is

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2} = 0; \tag{1.3}$$

a change of the time function implies the introduction of t', such that $t = f(t')$, and therefore

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t'^2} = \frac{\mathrm{d}^2 f}{\mathrm{d}t'^2}\frac{\mathrm{d}\vec{x}}{\mathrm{d}t}; \tag{1.4}$$

now, if we want to find another function in which the equation of motion is still valid, we need to transform time in such a way that $t'(t) = at + b$, where a and b are constants. With this function, corresponding to a shift of the origin and a change of units, the law describing the equation of motion is conserved.[1]

So, locally space is euclidean; this is telling us that space is homogeneous and isotropic: there is no privileged point or preferred direction. For what time is concerned, we choose our clocks in a way that time is homogeneous: every point in time is not preferred and laws don't change if we perform measurements at different times. There is no isotropy in time, however, since time is not reversible.

A system of observers in which $\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = 0$ is called an **inertial reference frame**. At this point, we introduce a second postulate: given an inertial reference frame, another frame moving with uniform rectilinear motion with respect to the first one is again inertial. So, given an inertial reference frame $\mathcal{K}$, another reference frame $\mathcal{K}'$ is inertial if the velocity $\mathbf{v}$ between the two frames is constant. This is the kinematical formulation of **relativity principle**. There is also a stronger, dynamical, formulation of the relativity principle: physical laws in any reference frame are unchanged, and dynamics appear the same in all inertial frames.

Now we want to add the condition of **precausality**: changing the reference frame doesn't change the order of co-local events.

We now have all the ingredients to derive Special Relativity axiomatically, without knowing anything about the speed of light.

[1]Note also that changing units should not change physics: there should be a way to formulate physical laws so that they are evidently unit independent.

1.2 Axiomatic derivation of Special Relativity

There is a theorem by von Ignatowski (1911), where it is showed that with all previous assumptions, being

- Relativity principle
- Spatial and temporal homogeneity
- Spatial isotropy
- Precausality

we can derive Special Relativity. Let us introduce two reference frames, $\mathcal{K}(O,x,y,z,t)$ and $\bar{\mathcal{K}}(\bar{O},\bar{x},\bar{y},\bar{z},\bar{t})$ the former moving with a velocity $\mathbf{v}$ with respect to the latter. We can always set $t=x=y=z=0$ if $\bar{t}=\bar{x}=\bar{y}=\bar{z}=0$ without loss of generality. We want to find coordinate transformations between the two systems, x^μ and $\bar{x}^\mu$; in particular: $\bar{x}^\mu = f^\mu(t,x,y,z,\mathbf{v})$.

Homogeneity implies that the time or space intervals measured in $\bar{\mathcal{K}}$ can depend only on the corresponding intervals in $\mathcal{K}$, not on the precise instants or positions in which the measurements are performed.

We can write $\mathrm{d}\bar{x}^\mu = \frac{\partial f^\mu}{\partial x^\nu}\mathrm{d}x^\nu$, and due to homogeneity we also have $\frac{\partial f^\mu}{\partial x^\nu} = \text{const}$ since it cannot depend on x; therefore f^μ must be a linear function of the coordinates x^μ. For simplicity, we can assume $\mathbf{v}$ to be along one of the spatial directions, let's say x: $\mathbf{v}=(v,0,0)$. We can then write $\bar{t} = A(v)t + B(v)x$ and $\bar{x} = C(v)t + D(v)x$.

A, B, C and D cannot be all independent. In fact, consider the motion of the $\mathcal{K}$ origin, O, w.r.t. $\bar{\mathcal{K}}$: we are comoving with $\bar{\mathcal{K}}$, therefore at $\bar{t}=0$, O is in $\bar{x}=0$. Later O will be at $\bar{x}_O = v\bar{t}$. But $x_O = 0$, therefore $\bar{t} = A(v)t$, $\bar{x}_O = C(v)t$, therefore $\bar{x}_O = \frac{C(v)}{A(v)}\bar{t}$, and then

$$\boxed{C(v) = vA(v)}. \tag{1.5}$$

Conversely, $\bar{O}$ moves in $\mathcal{K}$ with some velocity $\bar{v}$; so $x_{\bar{O}} = \bar{v}t$, $\bar{x}_{\bar{O}} = 0$, therefore $C(v)t + D(v)x_{\bar{O}} = 0$, and then

$$\boxed{C(v) = -\bar{v}D(v)}. \tag{1.6}$$

By putting together the two previous boxed relations 1.5 and 1.6, we can find

$$D(v) = -\frac{C(v)}{\bar{v}} = -\frac{v}{\bar{v}}A(v). \tag{1.7}$$

Then, the transformation law can be written in matrix form as

$$\begin{pmatrix} \bar{t} \\ \bar{x} \end{pmatrix} = \gamma(v)\begin{pmatrix} 1 & \xi(v) \\ v & -\frac{v}{\bar{v}} \end{pmatrix}\begin{pmatrix} t \\ x \end{pmatrix}, \tag{1.8}$$

where we defined $\xi(v) \equiv \frac{B(v)}{A(v)}$ and renamed $A(v) \equiv \gamma(v)$.

Given the relativity principle, we only know that if $\bar{v} = f(v)$, then $v = f(\bar{v})$. However, we can use the isotropy of space to ask for invariance after inversion of the spatial axis and velocity. (Reciprocity principle) [8, 9]. In order to see this, let us start by defining the overall transformation as $\Lambda(v)$

$$\begin{pmatrix} \bar{t} \\ \bar{x} \end{pmatrix} = \Lambda(v)\begin{pmatrix} t \\ x \end{pmatrix}, \tag{1.9}$$

Note that for $v=0$, we must have that the transformation matrix reduces to the identity, $\Lambda(0) = \mathbb{1}$, therefore we know that $\xi(0) = 0$ and $\gamma(0) = 1$.

The above statement about isotropy implies that we should get the same transformation between (t,x) and $(\bar{t},\bar{x})$ coordinates if we invert the direction of the the x and $\bar{x}$ axes (so that also $\bar{v} \to -\bar{v}$

and $v \to -v$). This in turns implies that $\gamma(-v) = \gamma(v)$, $\xi(-v) = -\xi(v)$ and that $-\bar{v} = f(-v)$ as well. However, we also know that $\bar{v} = f(v)$, hence we conclude that $f(-v) = -f(v)$. Evidently we must then have that $f = \pm I$ with the $+$ option unphysical as it would lead to the conclusion that $\bar{x} = -x$ for the $v = 0$ limit of our transformation. Hence we get in the end $\bar{v} = -v$, and therefore our transformation law becomes

$$\begin{pmatrix} \bar{t} \\ \bar{x} \end{pmatrix} = \gamma(v) \begin{pmatrix} 1 & \xi(v) \\ v & 1 \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix}. \tag{1.10}$$

Now, due to the relativity principle, $\Lambda^{-1}(v) = \Lambda(\bar{v})$, such that

$$\begin{pmatrix} t \\ x \end{pmatrix} = \Lambda(\bar{v}) \begin{pmatrix} \bar{t} \\ \bar{x} \end{pmatrix}. \tag{1.11}$$

But $\Lambda(\bar{v}) = \Lambda(-v)$, therefore $\Lambda(v)\Lambda(-v) = \Lambda(v)\Lambda^{-1}(v) = \mathbb{1}$, where the inverse transformation matrix is

$$\Lambda^{-1}(v) = \frac{1}{\Delta} \begin{pmatrix} \gamma(v) & \gamma(v)\xi(v) \\ -v\gamma(v) & \gamma(v) \end{pmatrix}, \tag{1.12}$$

where $\Delta = \det \Lambda = \gamma^2 - v\gamma^2\xi(v)$. Now, by imposing that $\Lambda(v)\Lambda(-v) = \mathbb{1}$ (and using that $\gamma(v)$ is even and $\xi(v)$ is odd under $v \to -v$) we find $\gamma^2 - v\gamma^2\xi(v) = 1$. This implies that $\Delta = 1$ and, that γ is (discarding the negative solution as it would imply $\gamma(0) = -1$),

$$\gamma(v) = \frac{1}{\sqrt{1 - v\xi(v)}}, \tag{1.13}$$

So that in the end

$$\Lambda(v) = \frac{1}{\sqrt{1 - v\xi(v)}} \begin{pmatrix} 1 & \xi(v) \\ v & 1 \end{pmatrix}. \tag{1.14}$$

We can now use the relativity principle to impose a group structure. Introduce a third reference frame $\bar{\bar{K}}$ which moves with constant velocity $\mathbf{u}$ with respect to K and $\bar{\mathbf{u}}$ with respect to $\bar{K}$. Again, without loss of generality, let us reduce the problem to the case in which $\mathbf{u} = (u, 0, 0)$ and $\bar{\mathbf{u}} = (\bar{u}, 0, 0)$. Therefore, $\bar{u}$ is a function of both v and u: $\bar{u} \equiv \Phi(u, v)$. Now, by simply applying the transformation laws between couples of reference frames, we can write

$$\begin{pmatrix} \bar{t} \\ \bar{x} \end{pmatrix} = \Lambda(\bar{u}) \begin{pmatrix} \bar{\bar{t}} \\ \bar{\bar{x}} \end{pmatrix} \qquad \begin{pmatrix} \bar{t} \\ \bar{x} \end{pmatrix} = \Lambda(v) \begin{pmatrix} t \\ x \end{pmatrix}, \tag{1.15}$$

which in turn simply implies

$$\begin{pmatrix} \bar{t} \\ \bar{x} \end{pmatrix} = \Lambda^{-1}(v)\Lambda(\bar{u}) \begin{pmatrix} \bar{\bar{t}} \\ \bar{\bar{x}} \end{pmatrix} = \Lambda(u) \begin{pmatrix} \bar{\bar{t}} \\ \bar{\bar{x}} \end{pmatrix}. \tag{1.16}$$

Therefore, $\Lambda^{-1}(v)\Lambda(\bar{u}) = \Lambda(u)$, $\Lambda(\bar{u}) = \Lambda(u)\Lambda(v)$, and we have found the composition of transformations:

$$\boxed{\Lambda(\Phi(u, v)) = \Lambda(u)\Lambda(v)}, \tag{1.17}$$

The above relation (together with the statement that if $u*$ is such that $\bar{u} = \Phi(u^*, v) = 0$, then $\Lambda(v)\Lambda(u^*) = \Lambda(0) = \mathbb{1}$) implies that our transformations form a group.

From the previous boxed relation, we can also see that

$$\gamma(u)\gamma(v) \begin{pmatrix} 1 + u\xi(v) & \xi(u) + \xi(v) \\ v + u & 1 + v\xi(u) \end{pmatrix} = \gamma(\Phi) \begin{pmatrix} 1 & \xi(\Phi) \\ \Phi & 1 \end{pmatrix}, \tag{1.18}$$

therefore

$$\gamma(\Phi) = \gamma(u)\gamma(v)\left(1 + u\xi(v)\right) = \gamma(u)\gamma(v)\left(1 + v\xi(u)\right), \tag{1.19}$$

which implies

$$u\xi(v) = v\xi(u). \tag{1.20}$$

This equation has two possible solutions. We could have $\xi(u) = \xi(v) = 0$: in this case,

$$\Lambda(v) = \gamma(v) \begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix} \qquad \gamma(v) = \frac{1}{\sqrt{1 - v\xi(v)}} = 1, \tag{1.21}$$

and we find the Galilean transformations.

Another solution is given by:

$$\frac{u}{v} = \frac{\xi(u)}{\xi(v)}; \tag{1.22}$$

this implies $\xi(u) \propto u$, $\xi(u) = \alpha u$, and therefore

$$\Lambda(v) = \frac{1}{\sqrt{1 - \alpha v^2}} \begin{pmatrix} 1 & \alpha v \\ v & 1 \end{pmatrix}. \tag{1.23}$$

The composition law implies $\gamma(u)\gamma(v)(u+v) = \gamma(\Phi)\Phi$; by squaring both sides of the equation, we find

$$\gamma^2(u)\gamma^2(v)(u+v)^2 = \gamma^2(\Phi)\Phi^2 = \frac{1}{1 - \alpha\Phi^2}\Phi^2; \tag{1.24}$$

solving for Φ gives

$$\Phi(u, v) = \frac{u+v}{1 + \alpha uv}, \tag{1.25}$$

with $\alpha > 0$. It is evident that α has the dimension of the inverse of the square of a velocity; also, note, that if $\alpha < 0$, it would be possible to get a composite velocity equal to infinity even if the two velocities are finite. We shall hence write $\alpha \equiv \frac{1}{C^2}$, where C is a constant. Therefore,

$$\Lambda(v) = \frac{1}{\sqrt{1 - \frac{v^2}{C^2}}} \begin{pmatrix} 1 & \frac{v}{C^2} \\ v & 1 \end{pmatrix} \qquad \Phi(u, v) = \frac{u+v}{1 + \frac{uv}{C^2}}; \tag{1.26}$$

if either u or v is equal to C, we find $\Phi(u, v) = C$.

In order to conclude our analysis we now need to impose the precausality condition:

$$\frac{\partial \bar{t}}{\partial t} > 0, \tag{1.27}$$

which gives $\gamma(v) > 0$ for any v, $\alpha > 0$, $\alpha v^2 < 1$ and, in conclusion, $v^2 < C^2$. So basically we have recovered the Lorentz transformation and the relativistic composition law of velocities modulo the fact that we have an undetermined velocity constant C appearing in our equations.

$$\begin{aligned} \bar{t} &= \frac{1}{\sqrt{1 - \frac{v^2}{C^2}}} \left(t + \frac{v}{C^2} x\right) \\ \bar{x} &= \frac{1}{\sqrt{1 - \frac{v^2}{C^2}}} (vt + x) \end{aligned}$$

Experimentally, it is found that C is actually the speed of light, $C = c$. But in general, the value of C depends on the physics behind the clocks and rods used to construct our spacetime: for example, for a bat using sonar to probe its environment the corresponding C would naturally be the speed of sound. (see eg. A. Trautman in Postepy Fizyki 45, 1 (1994) (in Polish). http://www.fuw.edu.pl/ amt/BATPOL.pdf).

What happens if we break some of the assumptions we made? Relaxing the relativity principle implies the existence of a privileged frame, and this brings to all kinds of aether theories; breaking isotropy [10, 11] of space implies finding a new symmetry group of transformations that depends both on positions and velocities (Finsler geometry); breaking homogeneity means that spacetime locally is no longer euclidean, we are losing the operative meaning of our definitions of space and time, and this could require again for the geometry to be Finsler; breaking precausality is in general the more exotic case.

For more details on von Ignatowski theorem look at references [12, 13, 14].

1.3 Equivalence Principles

Now, spacetime is locally Minkowski, so now we can go a step beyond and see what are the principles that are governing physics.

A crucial role is played by the equivalence principle; actually there are a lot of equivalence principles we can formulate. For a review on Equivalence Principles, see [15].

Newton Equivalence Principle (NEP) This states that in the newtonian limit the inertial mass is equal to the gravitational mass, $m_i = m_g$. According to Galileo, masses shouldn't enter in the laws of free fall; in particular, take two different objects at any distance and let them fall; now join the two objects: they fall as one object, but in the same way as the two distinct objects. Therefore, the mass shouldn't be involved in the laws of free fall.

This principle makes sense only in theories of gravity that admit a Newtonian limit. If $\Phi(x(t), t)$ is the Newtonian potential, then $m_i \ddot{x}(t) = -m_g \nabla\Phi(x(t), t)$, and thus $m_i \equiv m_g$, since from Newton's third law we cannot distinguish between active and passive gravitational masses.

A possible test of this principle can be performed using pendula with different bobs of different weight and nature but same length, that oscillate with the same period.

Weak Equivalence Principle (WEP) This is the universality of free fall: the motion under gravitational forces of bodies does not depend on their weights or compositions.

More specifically, we want to consider bodies that are test particles and that are not self-gravitating: a test particle is a particle that is not modifying the gravitational field in which it is falling in (so it has no backreaction on the gravitational field); the self-gravitating nature of an object, instead, is connected to its compactness σ, or the ratio between its gravitational energy and inertial energy

$$\sigma \equiv \frac{G_N m_g^2}{r} \frac{1}{m_i c^2}, \tag{1.28}$$

where r is the scale dimension of the object. If NEP holds, $m_i = m_g = m$, and therefore

$$\sigma = \frac{G_N m}{c^2 r}; \tag{1.29}$$

when $\sigma \ll 1$, the body is not self-gravitating.

Note that being a test particle depends on what background one considers, e.g. the Moon is a test particle for the Sun, but it is not a test particle for the Earth.

Tests of this principle are, for example, torsion balances *à la* Eötvos.

Gravitational Weak Equivalence Principle (GWEP) This is exactly like the WEP plus the self-gravity condition. In practice: does a test particle follow the same geodesic whether it has self-gravity or not? Does having self-gravity make any difference in the motion of the object? If not, the GWEP holds. This principle turns out to be true only in theories of gravity with second order field equations: General Relativity (with and without cosmological constant), Nordstrom gravity and Lanczos-Lovelock theories for $D > 4$, [16].

A test for this equivalence principle is the Lunar Laser Ranging experiment.

Einstein Equivalence Principle (EEP) This states that fundamental non-gravitational test physics is not affected locally at any point of spacetime by the presence of gravitational field: you can completely geometrise the action of the gravitational field. Why is it necessary to talk about fundamental non-gravitational test physics? Because, for example, we want to avoid composite systems, e.g. two particles. Given two particles at a distance ℓ, we would have

$$\frac{\ddot{\ell}}{\ell} \propto \frac{\partial^2 \Phi}{\partial x^i \partial x^j} \sim \tau_{ij},$$

where τ_{ij} is the tidal field, that is zero in flat spacetime but different from zero in the presence of gravitational fields.

This principle can be reformulated *à la* Clifford Will: in particular,

$$\mathrm{EEP} = \mathrm{WEP} + \mathrm{LLI} + \mathrm{LPI},$$

where WEP is the Weak Equivalence Principle, LLI stands for Local Lorentz Invariance and LPI stands for Local Position Invariance. Basically, with the WEP you can describe gravity geometrically (you can define an affine connection); with the LLI, it can be shown that local physical laws obey Special Relativity, i.e. you can "undo" gravity and go in a free fall (local inertial frame) where non-gravitational physics obey Special Relativity; LPI allows you to do this in every point of spacetime.

This equivalence principle is at the heart of metric theories of gravity.

A test of this principle could be to test the constancy of the fine-structure constant α, or of mass ratios (m_p/m_e).

Strong Equivalence Principle (SEP) Essentially, this is the extension of the EEP to gravitational physics: all fundamental test physics is not affected locally by the presence of the gravitational field.

À la Will, SEP = GWEP + LLI + LPI: the SEP requires the GWEP, the universality of free-fall extended to self-gravitating bodies, and the LLI and LPI; therefore, it requires that $G_N = \text{const}$, and this has been tested up to 10% of confidence.

This principle selects two theories of gravity: General Relativity and the Nordstrom gravity, which is characterised by an equation, which is

$$R = 24\pi G_N T; \tag{1.30}$$

however this Nordstrom gravity does not predict the deviation of light and has been therefore ruled out.

All these equivalence principles are related. In particular, the WEP implies the NEP, but the converse is not generally true, but only if masses enter in the equations of motion via their ratio, m_i/m_g.

Also, the GWEP implies the WEP in the limit for $\sigma = 0$. Note that if the GNEP (which is the NEP for self-gravitating bodies) fails, then also the GWEP fails.

Finally, the SEP implies both the GWEP and the EEP separately.

However, there are still some unsolved relations, the Schiff's conjecture and the Gravitational Schiff's conjecture. Respectively, it is still unknown if the WEP implies the EEP or whether the GWEP (with the possible addition of the EEP) implies the SEP.

2 Differential Geometry

2.1 Manifolds, Vectors, Tensors

The WEP implies that gravity is actually due to the geometry of spacetime. This pushes us to introduce a connection, $\Gamma^{\rho}_{\mu\nu}$, and to promote the derivative operator ∂_μ to a covariant derivative operator ∇_μ.

The EEP implies that, locally, spacetime is flat, a euclidean plane, $g_{\mu\nu} = \eta_{\mu\nu}$, and therefore we can compute the connection as a function of the metric g. Thus, spacetime must have the structure of a differential manifold.

We can define a **manifold** as a set of points $\mathcal{M}$ so that, in each point of $\mathcal{M}$, there is an open neighborhood which has a continuous one-to-one map ϕ onto an open set of $\mathbb{R}^n$ for some n, meaning $\phi : \mathcal{M} \to \mathbb{R}^n$.

In poor-man words, $\mathcal{M}$ might be a complicated space with non-trivial topology and curvature, however locally it has to look like an n-dimensional euclidean space $\mathbb{R}^n$.

The map ϕ is generally $\mathcal{C}^p$, and in this case the manifold is said to be $\mathcal{C}^p$ differentiable; if $p \to \infty$, then the manifold $\mathcal{M}$ is called smooth. Differentiability is an important property, since it allows to introduce tensor-like structures on the manifold.

2.1.1 Vectors

As a matter of fact, we can now define elementary objects on the manifold, like **vectors**. We can imagine each vector as located at some point $p \in \mathcal{M}$. Conversely, all the vectors v at a point $p \in \mathcal{M}$ form a tangent space T_p. In general, the vector v is independent on the choice of coordinates, therefore it should be possible to describe T_p without introducing any coordinate dependence.

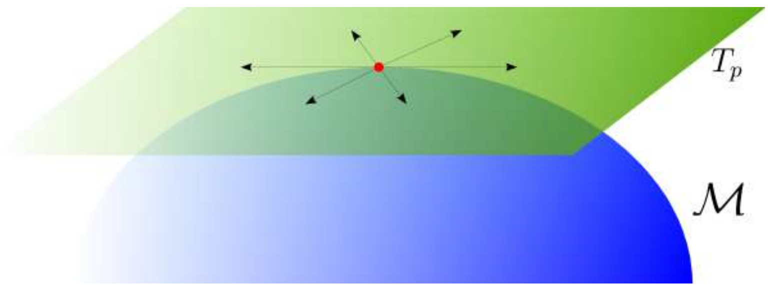


Figure 2.1: Pictorial representation of the tangent space T_p to the point $p \in \mathcal{M}$

We could consider all the parametrized curves through p, $\gamma : \mathbb{R} \to \mathcal{M}$ and their tangent vectors, but this would be coordinate dependent, since we would write $\gamma = x^\mu(\lambda)$ in some coordinate system.

Let us try to consider, instead, $\mathcal{F}$: the space of all the smooth functions on $\mathcal{M}$, meaning that $f \in \mathcal{F}$ if $f : \mathcal{M} \to \mathbb{R}$ and if it is $\mathcal{C}^\infty$. Then, we notice that each curve through p defines an operator on $\mathcal{F}$, the directional derivative, which maps $f \to \left.\frac{\mathrm{d}f}{\mathrm{d}\lambda}\right|_p$, where λ is an arbitrary parameter for the curve through p.

So, the tangent space at p, T_p, can be identified with all the directional derivative operators along curves through p, meaning that $\frac{\mathrm{d}}{\mathrm{d}\lambda}$ maps f in $\frac{\mathrm{d}f}{\mathrm{d}\lambda}$. It can be shown that T_p defines a vector space with the same dimension as $\mathcal{M}$, that the directional derivative acts linearly, and that it satisfy the Leibniz rule [2, p. 63].

Now, given a coordinate system $\{x^\mu\}$ in some open neighborhood $\mathcal{U}$ of p, these coordinates provide an obvious set of n-directional derivatives at p, being this the set of partial derivatives at p, $\{\partial_\mu\}$:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} f = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\partial f}{\partial x^\mu} = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\partial_\mu f, \tag{2.1}$$

and therefore $\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$ are the components of the vector $\frac{\mathrm{d}}{\mathrm{d}\lambda}$ in the coordinate basis $\{x^\mu\}$.

Now, we can write the vector by expanding it in components through the choice of this natural basis: $v = v^\mu \partial_\mu$, where v^μ are the components and ∂_μ is the basis of vectors.

Since the vector must be independent from the coordinate set, under a coordinate transformation $x^\mu \to x'^\mu$, the vector must behave as

$$v^\mu \partial_\mu = v = v^{\mu'}\partial_{\mu'} = v^{\mu'}\frac{\partial x^\mu}{\partial x^{\mu'}}\partial_\mu, \tag{2.2}$$

therefore, the components of the vector transform as

$$v^{\mu'} = v^\mu \frac{\partial x^{\mu'}}{\partial x^\mu}. \tag{2.3}$$

Since a vector at a point can be thought as a directional derivative along a path through that point, then a vector field defines a map from smooth functions to smooth functions over the manifold by taking a derivative at each point. So, given two vector fields X and Y one can define their commutator as

$$[X, Y](f) = X(Y(f)) - Y(X(f)), \tag{2.4}$$

which, in coordinate basis can be written as

$$[X, Y]^\mu = X^\lambda \partial_\lambda Y^\mu - Y^\lambda \partial_\lambda X^\mu; \tag{2.5}$$

notice that the commutator is linear and obeys Leibniz rule; also, this is coordinate independent in spite of the dependence on the partial derivative.

2.1.2 Dual vectors

Then, we can define a **dual vector** through the definition of cotangent space: a cotangent space $T_p^\star$ is the set of linear maps $\omega : T_p \to \mathbb{R}$, such that $\omega(v) \in \mathbb{R}$. Then, the map ω is a dual vector, $\omega \in T_p^\star$.

We can then define $\mathrm{d}f \in T_p^\star$ in such a way that $\mathrm{d}f \frac{\mathrm{d}}{\mathrm{d}\lambda} \equiv \frac{\mathrm{d}f}{\mathrm{d}\lambda}$, where $\frac{\mathrm{d}}{\mathrm{d}\lambda} \in T_p$.

If $\omega \in T_p^\star$, the natural choice of basis, in this case, is $\mathrm{d}x^\mu$, therefore we can write

$$\omega = \omega_\mu \, \mathrm{d}x^\mu; \tag{2.6}$$

as before, we want this object to be independent from the coordinate system, therefore, under coordinate transformation, the components ω_μ must transform as

$$\omega_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}}\omega_\mu. \tag{2.7}$$

The basis of vectors and dual vectors, also, are defined in such a way that

$$\mathrm{d}x^{\mu}\partial_{\nu} = \frac{\partial x^{\mu}}{\partial x^{\nu}} = \delta^{\mu}{}_{\nu}, \tag{2.8}$$

where $\delta^{\mu}{}_{\nu}$ is the Kronecker delta.

If we apply a vector v to a dual vector ω, we get

$$\omega_{\mu}\,\mathrm{d}x^{\mu}v^{\nu}\partial_{\nu} = \omega_{\mu}v^{\nu}\delta^{\mu}{}_{\nu} = (\omega \cdot v) \in \mathbb{R}, \tag{2.9}$$

which is an inner product.

Now that we have vectors and dual vectors, we can also generate **tensors**, as multilinear maps from a collection of k vectors and l dual vectors; thus, a generic tensor of rank (k,l) is given by

$$T = T^{\mu_1\ldots\mu_k}{}_{\nu_1\ldots\nu_l}\partial_{\mu_1}\otimes\cdots\otimes\partial_{\mu_k}\otimes\mathrm{d}x^{\nu_1}\otimes\cdots\otimes\mathrm{d}x^{\nu_l}, \tag{2.10}$$

where the symbol $\otimes$ indicates a tensor product. With this, similarly as we did before, we can write the transformation of the components of T under a coordinate transformation as

$$T^{\mu'_1\ldots\mu'_k}{}_{\nu'_1\ldots\nu'_l} = \frac{\partial x^{\mu'_1}}{\partial x^{\mu_1}}\cdots\frac{\partial x^{\mu'_k}}{\partial x^{\mu_k}}\frac{\partial x^{\nu'_1}}{\partial x^{\nu_1}}\cdots\frac{\partial x^{\nu'_l}}{\partial x^{\nu_l}}T^{\mu_1\ldots\mu_k}{}_{\nu_1\ldots\nu_l}. \tag{2.11}$$

Notice that, with this representation, vectors are $(1,0)$ tensors, dual vectors are $(0,1)$ tensors, and scalars are $(0,0)$ tensors.

Now, we can introduce the **metric tensor**, which is a symmetric, non-degenerate, rank $(0,2)$ tensor g defined by its relation to the line element

$$\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^{\mu}\,\mathrm{d}x^{\nu}, \tag{2.12}$$

and such that $\det g_{\mu\nu} \neq 0$. The metric provides the notions of causality and light-cones. From the metric g, we can also introduce the proper time (by setting all the spatial terms to zero), the proper length, and introduce the concept of shortest distance between two points. This is sometimes called **geodesic**, however beware that actually geodesics are straightest lines, that can be different from shortest lines if the connection is different from the Christoffel symbol of the metric.

Moreover, the metric tensor plays the role that the gravitational potential played in newtonian dynamics; it also tells us how to compute inner products:

$$(u,v) = g_{\mu\nu}u^{\mu}v^{\nu}; \tag{2.13}$$

in metric theories of gravity, the metric tensor also determines the curvature through the connection (see below).

2.2 Tensor Densities

Tensor densities, differently from tensors, are objects that fail to transform as tensors, modulo the Jacobian of the coordinate transformation taken into account

$$J = \left|\frac{\partial x^{\mu'}}{\partial x^{\mu}}\right|. \tag{2.14}$$

An example is given by the Levi-Civita symbol:

$$\tilde{\epsilon}_{\mu_1\ldots\mu_n} = \begin{cases} +1 & \text{even permutations of the indices} \\ -1 & \text{odd permutations of the indices} \\ 0 & \text{otherwise} \end{cases}. \tag{2.15}$$

Now, this object gives only $0, \pm 1$, therefore it cannot transform as a tensor through an arbitrary coordinate transformation. In general, we get

$$\tilde{\epsilon}_{\mu_1' \dots \mu_n'} = \left| \frac{\partial x^{\mu'}}{\partial x^{\mu}} \right|^{+1} \tilde{\epsilon}_{\mu_1 \dots \mu_n} \frac{\partial x^{\mu_1}}{\partial x^{\mu_1'}} \cdots \frac{\partial x^{\mu_n}}{\partial x^{\mu_n'}}, \tag{2.16}$$

and the Jacobian is the object that spoils the tensor nature of the Levi-Civita symbol. The exponent of the Jacobian is called *weight*.

Let us define determinant of the metric as

$$g \equiv \det g = |g|, \tag{2.17}$$

and this, under a coordinate transformation, transforms as

$$\det g(x') = \left| \frac{\partial x^{\mu'}}{\partial x^{\mu}} \right|^{-2} \det g(x), \tag{2.18}$$

with a weight equal to -2. Therefore, we can transform the Levi-Civita symbol into a tensor by defining

$$\epsilon_{\mu_1 \dots \mu_n} = \sqrt{|g|} \tilde{\epsilon}_{\mu_1 \dots \mu_n}. \tag{2.19}$$

2.3 Differential forms

A natural generalization of dual vectors is that of differential forms. A differential differential p-form is a $(0, p)$ tensor which is completely anti-symmetric.
Scalars are in this sense 0-forms, dual vectors are 1-forms and the Levi-Civita tensor is a four-form. The space of all the p-forms over a manifold is often denoted as $\Lambda^p(\mathcal{M})$. Noticeable it is possible to show that on a n-dimensional Manifold there are only $\frac{n!}{p!(n-p)!}$ linearly independent p-forms. This implies that for $n = 4$ there are at a point

- 4 one-form
- 6 two-forms
- 4 three-forms
- 1 four-form (the Levi-Civita tensor)

As we shall see the existence and uniqueness of the Levi-Civita tensor is important for an unambiguous definition of the volume element.

Given a p-form A and a q-form B we can get a (P+q)-form by a completely anti-symmetrised tensor product: the wedge product.

$$(A \wedge B)_{\mu_1 \dots \mu_{(p+q)}} = \frac{(p+q)!}{p! q!} A_{[\mu_1 \dots \mu_p} B_{\mu_{p+1} \dots \mu_{p+q}]} \tag{2.20}$$

which for the case of two vectors A_μ and B_ν, is simply

$$(A \wedge B)_{\mu\nu} = 2 A_{[\mu} B_{\nu]}. \tag{2.21}$$

The notation with the indices inside square brackets implies anti-symmetrization with respect to the indices inside the brackets. In particular, one has

$$T_{[\mu_1 \dots \mu_n]} = \frac{1}{n!} \sum_{\pi} \delta_\pi T_{\mu_{\pi(1)} \dots \mu_{\pi(n)}},$$

where the sum is taken over the permutations π of $1, \dots, n$ and δ_π is $+1$ for even permutations and -1 for odd permutations. For example, for a two indices object, one has

$$T_{[\mu\nu]} = \frac{1}{2}\left(T_{\mu\nu} - T_{\nu\mu}\right).$$

Similarly for the symmetrization:

$$T_{(\mu_1\ldots\mu_n)} = \frac{1}{n!}\sum_{\pi} T_{\mu_{\pi(1)}\ldots\mu_{\pi(n)}},$$

and in the case of two indices, one again has

$$T_{(\mu\nu)} = \frac{1}{2}\left(T_{\mu\nu} + T_{\nu\mu}\right).$$

We can also define the exterior derivative d mapping a p-form into a (p+1)-form. This is basically equal to the wedge product of a p-form, say A, and the one-form ∂:

$$(\mathrm{d}A)_{\mu_1\ldots\mu_{(p+1)}} = (p+1)\partial_{[\mu_1}A_{\mu_2\ldots\mu_{p+1}]} \tag{2.22}$$

and the simplest example of it is the gradient of a scalar field $(\mathrm{d}\phi)_\mu = \partial_\mu\phi$. It might seem strange that an operation involving the simple partial derivative turns a tensor into a tensor. This is due to the intrinsic antisymmetric structure of the exterior derivative.

Note that the Leibniz rule it is slightly modified for the exterior derivative. If ω is a p-form and η is a q-form, then

$$\mathrm{d}(\omega\wedge\eta) = (\mathrm{d}\omega)\wedge\eta + (-1)^p\omega\wedge(\mathrm{d}\eta) \tag{2.23}$$

2.4 The Volume element

We previously noticed the important of the existence and uniqueness of a n-form in a n-dimensional Manifold, the levi-Civita tensor. This is indeed criucial so to define volume elements (n-dimensional infinitesimal regions which can be represented as a parallelepiped defined by an ordered set of n vectors stemming from the same point in spacetime) in an integral like

$$\int \mathrm{d}x^\mu, \tag{2.24}$$

Let us see how.

We want to defy an oriented volume element (if two vectors are interchanged we should get a volume with same magnitude but opposite sign, or the volume should vanish when two of the vectors are collinear) so the volume element should indeed be replaced by a wedge-product,

$$d^nx = \mathrm{d}x^0\wedge\cdots\wedge\mathrm{d}x^{n-1}, \tag{2.25}$$

The problem is that the expression on the r.h.s of (2.25) is not a an n-form as it might seem. Indeed it is a tensor density: while it is true that given two functions on the manifold say f and g, then $\mathrm{d}f\wedge\mathrm{d}g$ is a coordinate independent two-form, this is not the case if the functions one choses are the coordinates themselves as in our case. One can check this explicitly by making a coordinate change after rewriting the element as

$$\mathrm{d}x^0\wedge\cdots\wedge\mathrm{d}x^{n-1} = \frac{1}{n!}\tilde{\epsilon}_{\mu_1\ldots\mu_n}\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_n}$$

Again the fact that the volume element naively transforms as a density can be remedied by applying $|g|$ with the right weight

$$\sqrt{|g|}\,d^nx = \sqrt{|g|}\,\mathrm{d}x^0\wedge\cdots\wedge\mathrm{d}x^{n-1}, \tag{2.26}$$

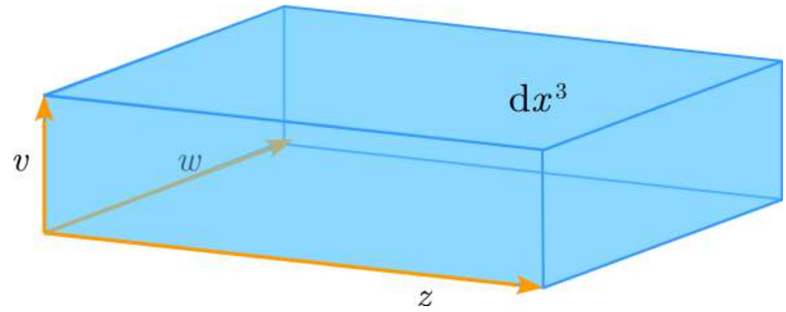


Figure 2.2: Pictorial representation of a 3D volume element.

and indeed, by doing this, we recover a proper n-form which moreover can be easily recognised by nothing else than the Levi-Civita tensor in coordinate basis

$$\begin{aligned}\epsilon &= \epsilon_{\mu_1\dots\mu_n}\,\mathrm{d}x^{\mu_1}\otimes\cdots\otimes\mathrm{d}x^{\mu_n} = \frac{1}{n!}\epsilon_{\mu_1\dots\mu_n}\,\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_n} = \\ &= \frac{1}{n!}\sqrt{|g|}\tilde{\epsilon}_{\mu_1\dots\mu_n}\,\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_n} = \sqrt{|g|}\,\mathrm{d}x^0\dots\mathrm{d}x^{n-1}\equiv\sqrt{|g|}\,\mathrm{d}^n x.\end{aligned} \tag{2.27}$$

2.5 Curvature

We can now start to deal with curvature; a first way to introduce curvature is to seek for a generalization of partial derivatives ∂_μ that: allows for a rank (k,l) tensor to become a rank $(k,l+1)$ tensor, that acts linearly, and that obeys the Leibniz rule.

This actually implies the existence of an affine connection. In metric theories of gravity, the metric enters in the definition of the connection ${\Gamma^\sigma}_{\mu\nu}$:

$${\Gamma^\sigma}_{\mu\nu} = \frac{1}{2}g^{\sigma\rho}\left(\partial_\mu g_{\rho\nu} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}\right); \tag{2.28}$$

the connection defined through the metric tensor is called the **Christoffel symbol**, and this will enter (as we will shortly see) in the computation of the Riemann tensor.

With this, it is natural to define a **covariant derivative** ∇_μ so that it is coordinate invariant, and in fact what we find is:

$$\nabla_\mu v^\nu = \partial_\mu v^\nu + {\Gamma^\nu}_{\mu\lambda}v^\lambda \qquad\qquad \nabla_\mu\omega_\nu = \partial_\mu\omega_\nu - {\Gamma^\lambda}_{\mu\nu}\omega_\lambda. \tag{2.29}$$

Now, $\nabla_\mu v^\nu$ is a tensor; if we change coordinates, also $\hat{\nabla}_\mu v^\nu$ must be a tensor; therefore, also their difference must be a tensor, and it can be shown that

$$\nabla_\mu v^\nu - \hat{\nabla}_\mu v^\nu = \left({\Gamma^\lambda}_{\mu\nu} - {\hat{\Gamma}^\lambda}_{\mu\nu}\right)v^\nu. \tag{2.30}$$

Now, we can form a different connection also by permuting the two lower indices; therefore, in general, we could have ${\Gamma^\lambda}_{\mu\nu} \neq {\Gamma^\lambda}_{\nu\mu}$. We then define the **torsion** as a rank $(1,2)$ tensor which is the difference of the two connections:

$${T^\lambda}_{\mu\nu} \equiv {\Gamma^\lambda}_{\mu\nu} - {\Gamma^\lambda}_{\nu\mu} = 2{\Gamma^\lambda}_{[\mu\nu]}. \tag{2.31}$$

In particular, if the connection is metric compatible[1], it inherits the symmetry of the metric, and therefore one has that the torsion ${T^\lambda}_{\mu\nu} = 0$. Notice that, while the torsion is a tensor, the connection is not a tensor, as it does not transform as a tensor.

[1]Saying that the connection is metric compatible means that it is calculated through the metric, and therefore this connection is the Christoffel symbol of the metric.

In general, assuming the EEP, one has that the torsion is zero, therefore in General Relativity the torsion is automatically zero and also

$$\nabla_\rho g_{\mu\nu} = 0. \tag{2.32}$$

We can define the divergence of a vector as

$$\operatorname{div} v \equiv \nabla_\mu v^\mu = \partial_\mu v^\mu + \Gamma^\mu{}_{\mu\lambda} v^\lambda. \tag{2.33}$$

It can be proven that the following relation holds true:

$$\Gamma^\mu{}_{\mu\lambda} = \frac{1}{\sqrt{|g|}} \partial_\lambda \sqrt{|g|}, \tag{2.34}$$

therefore the divergence of a vector reduces to

$$\operatorname{div} v = \frac{1}{\sqrt{|g|}} \partial_\mu \left(\sqrt{|g|} v^\mu \right). \tag{2.35}$$

We can also introduce the D'Alembert operator, defined as

$$\Box \phi = \nabla_\mu \nabla^\mu \phi, \tag{2.36}$$

and it can be proven similarly that[2]

$$\Box \phi = \frac{1}{\sqrt{g}} \partial_\mu \left(\sqrt{g} g^{\mu\nu} \partial_\nu \right) \phi. \tag{2.37}$$

In curved spacetime, one can also generalize the **Stokes' theorem**:

$$\int_\Sigma \nabla_\mu v^\mu \sqrt{g}\, \mathrm{d}^n x = \int_{\partial\Sigma} n_\mu v^\mu \sqrt{g}\, \mathrm{d}^{n-1} x, \tag{2.38}$$

where n_μ is the vector orthogonal to the hypersurface $\partial\Sigma$, which is the boundary of the region Σ.

We can also define the exterior derivative:

$$\mathrm{d}A \equiv \nabla_{[\mu} A_{\nu]} = \partial_{[\mu} A_{\nu]}, \tag{2.39}$$

and this is independent from the connection. Another quantity independent from the connection is

$$[x, y] = x^\nu \nabla_\nu y^\mu - y^\nu \nabla_\nu x^\mu. \tag{2.40}$$

Now we want to apply the commutator to two covariant derivatives in order to find the **Riemann tensor**. The Riemann tensor can be seen geometrically as the failure of parallel transporting a vector in a closed curve. What one can find is that

$$[\nabla_\mu, \nabla_\nu] v^\rho = \left(\partial_\mu \Gamma^\rho{}_{\nu\sigma} - \partial_\nu \Gamma^\rho{}_{\mu\sigma} + \Gamma^\rho{}_{\mu\lambda} \Gamma^\lambda{}_{\nu\sigma} - \Gamma^\rho{}_{\sigma\lambda} \Gamma^\lambda{}_{\mu\sigma} \right) v^\sigma \equiv R^\rho{}_{\sigma\mu\nu} v^\sigma. \tag{2.41}$$

However, one can find that, in the case in which the connection is not symmetric, another term emerges from the commutator of the two covariant derivatives:

$$[\nabla_\mu, \nabla_\nu] v^\rho = R^\rho{}_{\sigma\mu\nu} v^\sigma - 2\Gamma^\lambda{}_{[\mu\nu]} \nabla_\lambda v^\rho; \tag{2.42}$$

basically, if there is torsion, it is like the curve does not close, as if there is a defect in spacetime.

For the Riemann tensor a number of properties holds true:

1. $R_{\rho\sigma\mu\nu} = -R_{\sigma\rho\mu\nu}$; this holds true only if the connection is Levi–Civita one.

[2] From now on, we will denote the absolute value of the determinant of the metric simply as g.

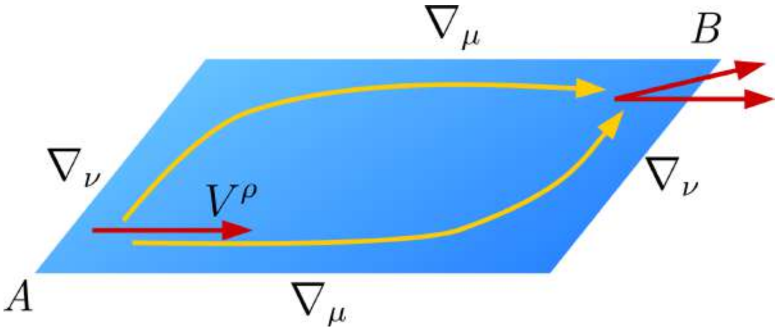


Figure 2.3: Pictorial representation of the failure of parallel transporting a vector in a closed curve, giving rise to the Riemann tensor.

2. $R_{\rho\sigma\mu\nu} = -R_{\rho\sigma\nu\mu}$; this is always true by construction.
3. $R_{\rho\sigma\mu\nu} = R_{\mu\nu\rho\sigma}$; this is true only if the connection is Levi–Civita one.
4. $R_{\rho\sigma\mu\nu} + R_{\rho\nu\sigma\mu} + R_{\rho\mu\nu\sigma} = 0$; this is true only if the torsion is zero.
5. Bianchi identity: $\nabla_{[\lambda} R_{\rho\sigma]\mu\nu} = 0$; this is true only if the torsion is zero, otherwise $\nabla_{[\lambda} R_{\rho\sigma]\mu\nu} = T^{\alpha}_{[\lambda\rho} R_{\sigma]\alpha\mu\nu}$ [17].

Note that the Riemann tensor has $n^2/12(n^2-1)$ independent components, i.e. twenty in four dimensions.

We can now define the **Ricci tensor** as the contraction of the Riemann tensor with respect to the first and third indices, $R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu}$. Why exactly those indices?

Contracting the first index with the second gives zero if the connection is symmetric, $R'_{\mu\nu} = R^{\rho}{}_{\rho\mu\nu} = 0$. However, if the connection is not metric compatible, the new Ricci tensor is not zero:

$$R'_{\mu\nu} = -\partial_\nu \Gamma^{\alpha}{}_{\mu\alpha} + \partial_\mu \Gamma^{\alpha}{}_{\nu\alpha}. \tag{2.43}$$

Nonetheless, if the connection is not the Christoffel symbol of the metric but it is symmetric, then we have $R'_{\mu\nu} = 2R_{[\mu\nu]}$.

One can also define, at this point, the **Ricci scalar** as the trace of the Ricci tensor

$$R \equiv g^{\mu\nu} R_{\mu\nu}; \tag{2.44}$$

notice that, if we used the alternative Ricci tensor to compute the Ricci scalar, we would have found $R' = 0$, therefore R is unique even in metric-affine gravity.

At this point, the **Einstein tensor** can be introduced:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R. \tag{2.45}$$

An important property that holds true for the Einstein tensor is $\nabla^\mu G_{\mu\nu} = 0$, which is the contracted Bianchi identity.

From the Riemann tensor, we can extract its completely traceless part, that has the name of **Weyl tensor**; in $n+1$ dimensions, this becomes:

$$C_{\rho\sigma\mu\nu} = R_{\rho\sigma\mu\nu} - \frac{2}{n-2}\left(g_{\rho[\mu} R_{\nu]\sigma} - g_{\sigma[\mu} R_{\nu]\rho}\right) + \frac{2}{(n-1)(n-2)} g_{\rho[\mu} g_{\nu]\sigma} R, \tag{2.46}$$

and if the connection is metric compatible, then the following properties of the Weyl tensor hold true:

$$C_{\rho\sigma\mu\nu} = C_{[\rho\sigma][\mu\nu]} \qquad C_{\rho\sigma\mu\nu} = C_{\mu\nu\rho\sigma} \qquad C_{\rho[\sigma\mu\nu]} = 0. \tag{2.47}$$

One can also show that the Weyl tensor is invariant under conformal transformations, therefore doesn't convey informations about volume, but rather it conveys informations about perturbations, like gravitational waves, and about how the shape of a body changes due to tidal forces. In GR, in vacuum solutions, the Weyl tensor is also the only non-zero part of the Riemann tensor.

2.6 Intrinsic and Extrinsic curvature

There are two different relevant concepts of curvature: the **intrinsic curvature** and the **extrinsic curvature**.

- The Intrinsic curvature is an intrinsic property of the manifold, and therefore one does not need to embed a manifold in a higher dimensional manifold in order to describe its intrinsic curvature.
- The Extrinsic curvature is a property of a submanifold embedded in a higher dimensional manifold. So it does depend on the embedding as the latter induces a metric on the submanifold.

2.6.1 Example: Intrinsic-Extrinsic (Gaussian) curvature of two surfaces

For example, we can characterize the curvature of a two-sphere without embedding it in 3-dimensional space: given the measured area of the sphere A, its radius a will be given by

$$a = \sqrt{\frac{A}{4\pi}}. \tag{2.48}$$

If we take the line element on the two-sphere

$$\mathrm{d}s^2 = a^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2 \right), \tag{2.49}$$

one finds that $R_{\theta\phi\theta\phi} = a^2 \sin^2\theta$, $R_{\phi\phi} = \sin^2\theta$, $R_{\theta\theta} = 1$, and therefore

$$R = \frac{2}{a^2}, \tag{2.50}$$

which is twice the gaussian curvature of the sphere. So, note that the intrinsic curvature is $R = 2K_G$, where $K_G = \kappa_1\kappa_2$ is the gaussian curvature and $\kappa_{1,2}$ are the so called principal curvatures of the surface which can be written as the inverses of the radii of curvature in the two principal directions on the surface $\kappa_i = 1/a_i$. However, the extrinsic curvature turns out to be $K_{Ext} = \kappa_1 + \kappa_2$ [4, p. 232]. So for example if one takes a flat sheet of paper and rolls it in one direction with radius R, it will appear to have made a cylinder in three dimensional space. The extrinsic curvature will then be non-zero $K_{Ext} = 1/R + 0 = 1/R$ but the intrinsic (Gaussian) curvature will still be zero $K_G = 0 \cdot 1/R = 0$.

> Note that if you do not change the metric, you cannot change the gaussian curvature: this is due to the *Theorema Egregium* by Gauss, the gaussian curvature of a surface is unchanged when the surface is bent but not stretched. This is also why the surface of the Earth cannot be mapped on a plane without stretching or cutting.
>
> This is also connected to a somewhat more common experience: eating pizza.
>
> When on the plate, each slice of a pizza is flat, meaning that it has a zero gaussian curvature. Picking it up, the slice curves downward, but in the perpendicular direction it still has zero curvature. This means that, when we introduce a curvature along the direction of the crust, we will force a zero gaussian curvature along the radius of the slice, which allows us to eat comfortably the pizza slice!

2.6.2 Hypersurfaces and their Extrinsic curvature

All of this can be better formalized through the definition of hypersurfaces. A **hypersurface** is an $(n-1)$-submanifold Σ of an n-dimensional manifold $\mathcal{M}$.

We can define an hypersurface by setting a function of the coordinates $f(x)$ to a constant value f^* on Σ; we can then define a vector $\zeta^\mu \equiv g^{\mu\nu}\nabla_\nu f(x)$. This vector is automatically orthogonal to the hypersurface: take $l_\mu \in T_p(\Sigma)$, then

$$\zeta^\mu l_\mu = l^\nu \nabla_\nu f(x) = 0, \tag{2.51}$$

since on the hypersurface the function is constant; so, ζ^μ is the normal vector to Σ. This vector is defined modulo some function, since given a function $h(x)$, $h(x)\zeta^\mu$ is still orthogonal to Σ.

Now, by definition, if ζ^μ is timelike, then Σ is spacelike; viceversa, if ζ^μ is spacelike, then Σ is timelike. Finally, if ζ^μ is null, then $\Sigma = \mathcal{N}$ is null.

We can normalize the orthogonal vector by defining n^μ as

$$n^\mu = \varepsilon \frac{\zeta^\mu}{|\zeta^\nu \zeta_\nu|^{1/2}}, \tag{2.52}$$

where ε is either $+1, -1$ if Σ is spacelike or timelike respectively.

The null case is less trivial, since a null vector orthogonal to a null hypersurface is also parallel to the hypersurface (its product with itself is zero). We can then take a parametrized family of curves $x^\mu(\lambda) \in \mathcal{N}$ and write $\zeta^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$ i.e. as the tangent vector to these curves; it can be shown that these curves are indeed (null) geodesics (to show this one just needs to show that $\zeta^\nu \nabla_\nu \zeta^\mu \propto \zeta^\mu$ [18]. Now, by definition, the null geodesics $x^\mu(\lambda)$ with λ being an affine parameter, for which the tangent vectors $\mathrm{d}x^\mu / \mathrm{d}\lambda$ are normal to a null hypersurface $\mathcal{N}$, are the "generators" of $\mathcal{N}$.

Now we want to define the extrinsic curvature of hypersurfaces. We can take an hypersurface Σ in a manifold $\mathcal{M}$; suppose Σ is spacelike, so that its normal vector n^μ is timelike. Take the congruence of all the timelike geodesics, so at each point n^μ is the normal vector to Σ. Associate an affine parameter τ to the vector field and construct a basis $y^1, \dots y^{n-1}$ on Σ: then we can construct an adapted set of coordinates $(\tau, y^1, \dots, y^{n-1})$ such that

$$g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\,\mathrm{d}\tau^2 + \gamma_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j. \tag{2.53}$$

The induced metric γ_{ij} has dimension $n-1$, and of course we can define the curvature associated to this induced metric. However, we would like to be able to define the extrinsic curvature as a property in n dimensions. For this reason, we construct the **projector** $P_{\mu\nu}$ as

$$P_{\mu\nu} = g_{\mu\nu} - \sigma n_\mu n_\nu, \tag{2.54}$$

where $\sigma = n^\mu n_\mu$ is the norm of the normal vector.

The projector is sometimes called the first fundamental form of the hypersurface; it has a number of properties, for example, it projects any vector of the manifold $\mathcal{M}$ on the spacelike hypersurface Σ, as it can be checked by observing that any vector after the action of the projector is orthogonal to the normal vector to the hypersurface

$$(P_{\mu\nu} v^\mu)\, n^\nu = 0 \quad \forall v^\mu \in \mathcal{M}; \tag{2.55}$$

moreover, the projector acting on two vectors on Σ will behave like the metric:

$$P_{\mu\nu} v^\mu w^\nu = v \cdot w \quad \forall v^\mu, w^\nu \in \Sigma; \tag{2.56}$$

finally, the projector tensor is idempotent, $P^n = P$, or alternatively

$$P^\mu_\lambda P^\lambda_\nu = P^\mu_\nu. \tag{2.57}$$

The extrinsic curvature is now defined as a rank $(0, 2)$ tensor

$$K_{\mu\nu} \equiv \frac{1}{2} \mathcal{L}_n P_{\mu\nu}, \tag{2.58}$$

where the symbol $\mathcal{L}_n$ represents the Lie derivative along n, and it is the evolution of the object on which the derivative is applied along n. We will discuss this Lie derivative better in the following section.

We can take the covariant derivative of an object on the hypersurface Σ by taking projections of the whole object:

$$\hat{\nabla}_\sigma X^\mu{}_\nu = P^\alpha{}_\sigma P^\mu{}_\beta P^\gamma{}_\nu \nabla_\alpha X^\beta{}_\gamma. \tag{2.59}$$

Then we can define a Riemann tensor on the hypersurface as

$$\left[\hat{\nabla}_\mu, \hat{\nabla}_\nu\right] v^\rho \equiv \hat{R}^\rho{}_{\sigma\mu\nu}, \tag{2.60}$$

which will be related to the Riemann tensor on the n-dimensional manifold by a suitable contraction of the indices with appropriate projector operators

$$\hat{R}^\rho{}_{\sigma\mu\nu} = P^\rho{}_\alpha P^\beta{}_\sigma P^\gamma{}_\mu P^\delta{}_\nu R^\alpha{}_{\beta\gamma\delta}; \tag{2.61}$$

this is the so called Gauss-Codazzi equation; there is also the Codazzi equation that relates the Ricci tensor to the extrinsic curvature:

$$\hat{\nabla}_{[\mu} K^\mu_{\nu]} = \frac{1}{2} P^\sigma{}_\nu R_{\rho\sigma} n^\rho. \tag{2.62}$$

2.7 Lie Derivative

The Lie derivative is a primitive connection-independent notion of derivative, and it is different from the covariant derivative. In particular, it is the derivative with respect to the flow of a vector field. Consider a congruence of curves $x^\mu(s)$; we can define the set of vectors tangent to these curves as $X(x^\mu(s)) = \frac{\mathrm{d}x^\mu}{\mathrm{d}s}$. Consider an object of any rank T. At point P on the curve, we will have generally $T(P)$. We can then drag T from P to the point Q along the curve and obtain $T'(Q)$. We can then define the Lie derivative of T along the curve defined by X as

$$\mathcal{L}_X T = \lim_{s\to 0} \frac{T(Q) - T'(Q)}{s}, \tag{2.63}$$

where $T(Q)$ is just the same object $T(P)$, but at Q. Basically, the idea behind the Lie derivative is

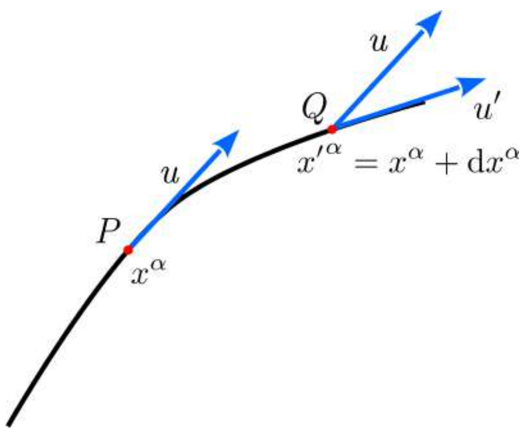


Figure 2.4: Pictorial representation of the comparison of a dragged vector w.r.t. the vector taken at the same point.

that we want to use the u vector field flow to drag objects of any form and compare with their dragged value with their actual value at some point.

In order to understand the physical meaning of this derivative, we can restrict our definition to a curve with tangent $u^\alpha = \frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda}$, along which we want to calculate the Lie derivative of the vector field A^α.

The Lie-dragged vector $A'^\alpha(x')$ can be determined as

$$A'^\alpha(x') = \frac{\partial x'^\alpha}{\partial x^\beta} A^\beta(x) = \frac{\partial}{\partial x^\beta} \left(x^\alpha + u^\alpha \,\mathrm{d}\lambda\right) A^\beta(x) = \left(\delta^\alpha_\beta + \partial_\beta u^\alpha \,\mathrm{d}\lambda\right) A^\beta(x), \tag{2.64}$$

and therefore

$$A'^{\alpha}(Q) = A^{\alpha}(P) + \partial_{\beta} u^{\alpha} A^{\beta}(P)\,\mathrm{d}\lambda; \tag{2.65}$$

conversely, by performing a Taylor expansion, we can write

$$A^{\alpha}(Q) = A^{\alpha}(P) + u^{\beta}\partial_{\beta} A^{\alpha}(P)\,\mathrm{d}\lambda. \tag{2.66}$$

Therefore, the Lie derivative of the vector field A^{α} along the curve of tangent u^{α} is

$$\mathcal{L}_u A^{\alpha} = \partial_{\beta} A^{\alpha} u^{\beta} - \partial_{\beta} u^{\alpha} A^{\beta}. \tag{2.67}$$

One can then generalize this to objects of any rank; for a scalar function $f(x)$, one has

$$\mathcal{L}_u f = u^{\alpha}\partial_{\alpha} f; \tag{2.68}$$

for a $(2,0)$ rank tensor $T^{\mu\nu}$ one has

$$\mathcal{L}_u T^{\mu\nu} = u^{\sigma}\partial_{\sigma} T^{\mu\nu} - T^{\mu\sigma}\partial_{\sigma} u^{\nu} - T^{\sigma\nu}\partial_{\sigma} u^{\mu}; \tag{2.69}$$

for a generic (k,l) rank tensor, we can write

$$\mathcal{L}_x T^{\mu_1,\dots,\mu_k}{}_{\nu_1,\dots,\nu_l} = x^{\sigma}\partial_{\sigma} T^{\mu_1,\dots,\mu_k}{}_{\nu_1,\dots,\nu_l} - \sum_i \left(\partial_{\rho} x^{\mu_i}\right) T^{\mu_1,\dots,\rho,\dots\mu_k}{}_{\nu_1,\dots,\nu_l} + \sum_j \left(\partial_{\nu_j} x^{\rho}\right) T^{\mu_1,\dots,\mu_k}{}_{\nu_1,\dots,\rho,\dots,\nu_l}. \tag{2.70}$$

The Lie derivative has several properties:

1. it is coordinate independent, and it maps a rank (k,l) tensor into a rank (k,l) tensor;
2. it is linear: $\mathcal{L}_X(aT + bS) = a\mathcal{L}_X T + b\mathcal{L}_X S$, with constant a and b;
3. it obeys the Leibniz rule: $\mathcal{L}_X\left(T \otimes S\right) = \mathcal{L}_X(T) \otimes S + T \otimes \mathcal{L}_X(S)$.

We can prove the first property, according to which the Lie derivative is coordinate invariant. Basically, we want to prove that $(\mathcal{L}_X T)' = \mathcal{L}_X T'$.

Assume to have a parameter s such that we have a congruence of curves $x^{\mu}(s)$. Now, take s as one of the coordinates, in a way that we have an adapted system of coordinates. In this way, we have

$$x = \frac{\partial}{\partial s} = x^{\alpha}\partial_{\alpha}. \tag{2.71}$$

Now we want to write $\mathcal{L}_X T = \frac{\partial}{\partial s} T'$; specifically, we get (in order to not get a too messy notation, we use indices creatively)

$$(\mathcal{L}_X T)'^{\mu\dots}{}_{\nu\dots} \tag{2.72}$$

which proves the property.

Now we can determine explicitly the extrinsic curvature

$$K_{\mu\nu} = \frac{1}{2}\mathcal{L}_u(g_{\mu\nu} + u_{\mu}u_{\nu}) = \frac{1}{2}\left(u^{\sigma}\cancel{\nabla_{\sigma} g_{\mu\nu}} + \nabla_{\mu} u^{\sigma} g_{\sigma\nu} + \nabla_{\nu} u^{\sigma} g_{\mu\sigma}\right) = \frac{1}{2}\left(\nabla_{\mu} u_{\nu} + \nabla_{\nu} u_{\mu}\right) = \nabla_{\mu} u_{\nu}, \tag{2.73}$$

where we have used $\mathcal{L}_u u = 0$ and substituted partial derivatives with covariant derivatives and we have used the fact that $u_{\nu} = \nabla_{\nu} f$, and thus $\nabla_{\mu} u_{\nu}$ is symmetric. Therefore we have $K_{\mu\nu} = \nabla_{\mu} u_{\nu}$ and $K = K^{\mu}{}_{\mu} = \nabla^{\mu} u_{\mu}$. Note that if the congruence is not geodesic, then the quantity $a^{\mu} = u^{\nu}\nabla_{\nu} u^{\mu} \neq 0$, and therefore $K_{\mu\nu} = \nabla_{\mu} u_{\nu} - \sigma n_{\mu} a_{\nu}$

2.8 Killing Vectors

Consider now the case in which $\mathcal{L}_X g_{\mu\nu} = 0$; we can easily see from the steps in (2.73) that in this case one gets

$$\nabla_\mu x_\nu + \nabla_\nu x_\mu = 0. \tag{2.74}$$

Therefore, if we Lie-drag the metric along x, and we find that the metric is invariant under this drag, than we get that x defines a symmetry of spacetime, an **isometry**. In this case, the vector x is called a **Killing vector** and equation (2.74) is called the **Killing equation**.

How many symmetries do we expect to find in a four-dimensional spacetime?

If a spacetime possesses a maximal number of isometries, the spacetime is called **maximally symmetric**. An example is given by Minkowski spacetime with the Poincaré group: translations plus boosts and rotations.

In n-dimensions, if $\mathcal{M}$ has the maximum number of isometries, then $\mathcal{M}$ is maximally symmetric. A property of such spacetimes is that they have constant curvature, and thus constant Ricci scalar.

Consider for example the **Friedmann-Robertson-Walker metric**. It can be obtained using the Copernican principle. Assuming that space is homogeneous and isotropic, we can write that $\mathcal{M} = \mathbb{R} \times \Sigma^3$, where Σ^3 is a maximally symmetric 3-surface. Now there are only three constant curvature 3-surfaces: flat, hyperbolic and spherical. So, the metric will be of the form

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + R(t)\,\mathrm{d}\sigma^2.$$

We can put a space-independent scale factor in front of the three-metric $\mathrm{d}\sigma^2$ which, since we want a maximally-symmetric 3-surface, must be of the form

$$\mathrm{d}\sigma^2 = \gamma_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j.$$

We can write an element that embodies all the three possibilities for the constant curvature:

$$\mathrm{d}\sigma^2 = \frac{\mathrm{d}\bar{r}^2}{1-\kappa\bar{r}^2} + \bar{r}^2\,\mathrm{d}\Omega^2,$$

where $\kappa = 0, \pm 1$; if $\kappa = 0$, we have the 3-plane, if $\kappa = +1$ we have the 3-sphere, and if $\kappa = -1$ we have the constant-curvature hyperbolic 3-space.

One can then verify that the 3-dimensional Ricci scalar $R^{(3)} = \text{const}$ and that $R_{ij} = 2\kappa\gamma_{ij}$.

Sometimes, we can find a Killing vector with the blink of an eye: assume to have an adapted system of coordinates; then, if the metric is independent on one of these coordinates, we have a Killing vector associated to that one coordinate.

Imagine we have a coordinate σ^* such that $\partial_{\sigma^*} g_{\mu\nu} = 0$; then $\frac{\partial}{\partial\sigma^*}$ is a Killing vector. As an example, consider the Schwarzschild metric

$$\mathrm{d}s^2 = -\left(1 - \frac{2M}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{2M}{r}\right)^{-1}\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2, \tag{2.75}$$

and we immediately realize that the metric is independent both on t and ϕ. Therefore both $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial\phi}$ are Killing vectors.

Of course, this trick can not be complete: at most there are 10 Killing vectors but only 4 coordinates, so by studying only the coordinates from which the metric is independent does not give us all the Killing vectors.

A useful way to write a Killing vector is as follows: for example, take the Killing vector $\frac{\partial}{\partial\phi}$; then, the components of this Killing vector are $k^\mu = (0,0,0,1) = \delta^\mu_\phi$; in general, if the metric is independent on the coordinate σ^*, we can write the components of the Killing vector relative to σ^* as

$$k^\mu = \delta^\mu_{\sigma^*}, \tag{2.76}$$

and it is easy to check that the Killing vector itself then becomes

$$k = k^\mu \partial_\mu = \delta^\mu_{\sigma^*} \partial_\mu = \partial_{\sigma^*}. \tag{2.77}$$

However, in general you cannot find a coordinate system in which all the Killing vectors are simultaneously of the form ∂_{σ^*}.

Imagine to be in spacetime with $1+1$ dimension:

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + \mathrm{d}x^2; \tag{2.78}$$

we have two evident isometries, which are the translations along t and x, and therefore two Killing vectors $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial x}$. Let us perform these transformations:

$$\begin{cases} t = -\lambda \sinh \eta \\ x = -\lambda \cosh \eta \end{cases} ; \tag{2.79}$$

with this, the line element can be simply rewritten as

$$\mathrm{d}s^2 = -\lambda^2\,\mathrm{d}\eta^2 + \mathrm{d}\lambda^2. \tag{2.80}$$

Then, we can see that the metric does not depend on the coordinate η, therefore we have that $\frac{\partial}{\partial \eta}$ is a Killing vector. This Killing vector is simply

$$\frac{\partial}{\partial \eta} = x\frac{\partial}{\partial t} + t\frac{\partial}{\partial x}, \tag{2.81}$$

and this Killing vector is a generic boost, so this is the Killing vector associated to the boost of flat spacetime. What is the norm of this Killing vector? It is easy to check that

$$\left|\left|\frac{\partial}{\partial \eta}\right|\right| = g_{\mu\nu}\delta^\mu_\eta \delta^\nu_\eta = g_{\eta\eta} = -\lambda^2 < 0, \tag{2.82}$$

and therefore this Killing vector is timelike. This is an important comment, since it is possible that the norm of a Killing vector is spacelike, and in that case, if we want to "see" the isometry defined by it, we would have to follow a spacelike geodesic. Also, it is possible that the norm of a Killing vector depends on the coordinates of spacetime, and therefore the behaviour of the Killing vector could change.

By definition, we have

$$(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu)\xi_\rho = R_\rho{}^\sigma{}_{\mu\nu}\xi_\sigma = R_{\mu\nu\rho}{}^\sigma \xi_\sigma, \tag{2.83}$$

but using the Killing equation, $\nabla_\mu \xi_\rho = -\nabla_\rho \xi_\mu$, we can also write

$$(\nabla_\mu \nabla_\nu \xi_\rho + \nabla_\nu \nabla_\rho \xi_\mu) = R_{\mu\nu\rho}{}^\sigma \xi_\sigma. \tag{2.84}$$

We can write this same equation by permuting cyclically the indices (μ, ν, ρ), and in particular, if we add the equation with (μ, ν, ρ) to the equation with (ν, ρ, μ), and subtract the one with (ρ, μ, ν), we get [1, p. 442], $2\nabla_\nu \nabla_\rho \xi_\mu = -2R_{\mu\nu\rho}{}^\sigma \xi_\sigma$, i.e. that for any Killing vector one has

$$\nabla_\mu \nabla_\nu \xi_\rho = -R_{\nu\rho\mu}{}^\sigma \xi_\sigma \quad \Longrightarrow \quad \nabla_\mu \nabla_\rho \xi_\nu = R_{\nu\rho\mu\sigma}\xi^\sigma. \tag{2.85}$$

which contracted alos implies

$$\nabla_\mu \nabla_\rho \xi^\mu = R_{\rho\sigma}\xi^\sigma. \tag{2.86}$$

2.9 Conserved quantities

Due to Noether's Theorem, given a symmetry, there is a conserved quantity. Let us write the action for a point-like particle:

$$S = \frac{1}{2} \int g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu \, \mathrm{d}\lambda; \tag{2.87}$$

extremizing the action, $\delta S = 0$, gives

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \left(g_{\mu\alpha} \dot{x}^\mu \right) - \frac{1}{2} \partial_\alpha g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu = 0; \tag{2.88}$$

it can be proven that this is equivalent to the usual geodesic equation with a metric compatible Γ.

Now, assume that there exists a coordinate $\hat{\alpha}$ such that $\partial_{\hat{\alpha}} g_{\mu\nu} = 0$; then, the previous equation reduces to

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \left(g_{\mu\hat{\alpha}} \dot{x}^\mu \right) = 0, \tag{2.89}$$

and therefore $g_{\mu\hat{\alpha}} \dot{x}^\mu$ is constant along the geodesic parametrized by λ. The Killing vector relative to this symmetry is $k^\mu = \delta^\mu_{\hat{\alpha}}$; we can also define $\dot{x}^\mu \equiv u^\mu$ as the four-velocity along the geodesic, the tangent vector to the curve. Therefore

$$g_{\mu\hat{\alpha}} \dot{x}^\mu = g_{\mu\nu} k^\mu u^\nu = (k \cdot u) = \text{const.}, \tag{2.90}$$

and this is the conserved quantity, written in a covariant way.

We can now introduce a notion of conserved energy if we have a globally defined timelike Killing vector (we have a conserved momentum if the Killing vector is spacelike): take the stress-energy tensor of the matter,

$$T_{\mu\nu} = -\frac{2}{\sqrt{g}} \frac{\delta S}{\delta g^{\mu\nu}}; \tag{2.91}$$

we can also define a current $J^\mu = k_\nu T^{\mu\nu}$, where k_ν is a timelike Killing field (for example, in a stationary spacetime). The covariant derivative of this current is

$$\nabla_\mu J^\mu = (\nabla_\mu k_\nu) T^{\mu\nu} + k_\nu (\nabla_\mu T^{\mu\nu}) = 0, \tag{2.92}$$

since the first term is zero by the Killing equation, while the second term is zero due to the conservation of the stress-energy tensor. Since $\nabla_\mu J^\mu = \text{const}$, we can apply Stokes' theorem; we define the energy as

$$E \equiv \int_\Sigma J^\mu n_\mu \sqrt{\gamma} \, \mathrm{d}^3 x; \tag{2.93}$$

since $\nabla_\mu J^\mu = 0$, we have

$$\int_{\mathcal{M}} \nabla_\mu J^\mu \sqrt{g} \, \mathrm{d}^4 x = 0, \tag{2.94}$$

and by Stokes' theorem we have

$$0 = \int_{\partial\mathcal{M}} J^\mu n_\mu \sqrt{\gamma} \, \mathrm{d}^3 x = \int_{\Sigma_2} \mathrm{d}^3 x \sqrt{\gamma} J^\mu n_\mu - \int_{\Sigma_1} \mathrm{d}^3 x \sqrt{\gamma} J^\mu n_\mu, \tag{2.95}$$

where we assumed that the fields go to zero sufficiently rapidly in order to neglect the spatial infinity boundary ι^0 of $\partial\mathcal{M}$. Since the choice of Σ_1 and Σ_2 is arbitrary, we immediately see that the energy E is a conserved quantity of our stationary spacetime.

Consider the previous example of a point-like particle; the energy of this particle is then given by $E = -p \cdot k$; now p is always timelike, but in Schwarzschild we can verify that for $r > 2M$ the Killing vector $k = \frac{\partial}{\partial t}$ is timelike, but inside $r < 2M$ the same Killing vector is spacelike. Therefore, $E > 0$ outside the horizon, but $E < 0$ inside $r < 2M$. Therefore, in Schwarzschild, it seems that the energy of a particle falling in the black hole changes sign, and thus for an observer at infinity it has negative energy: it seems that the particle brings energy away from the black hole. This is connected to the ergoregion, a region of spacetime where there can be negative-energy states. We will delve more into details in the following.

Consider a maximally symmetric spacetime in n dimensions; what is the maximum number of isometries? Take $\mathbb{R}^n$: it will have n independent translation and n rotations, but not all of these rotations are independent. The number of independent rotations is

$$N_{\text{rot}} = \frac{n(n-1)}{2}; \tag{2.96}$$

the total number of isometries, then, is

$$N_{\text{tot}} = n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}; \tag{2.97}$$

in $n = 4$ we then have at most 10 isometries; in $n = 3$ we have at most 6 isometries.

In 4 dimensions, there are only three maximally symmetric spacetimes: Minkowski, de Sitter and anti-de Sitter. Also, in a maximally symmetric spacetime there is a very simple formula for the Riemann tensor:

$$R_{\rho\sigma\mu\nu} = \frac{R}{n(n-1)} \left(g_{\rho\mu} g_{\sigma\nu} - g_{\rho\nu} g_{\sigma\mu} \right), \tag{2.98}$$

where we remember that R is constant for maximally symmetric spacetimes.

3 Kinematics

3.1 Geodesic Deviation Equation

The basic concept of General Relativity is that, in presence of gravity, geodesics would feel a kind of tidal effect, such that they would start to diverge or converge. Indeed, as we saw the WEP implies that the gravitational field in a small region of size L can be exchanged with a suitably accelerating frame only if $l \ll L$ where L is the typical scale over which the gravitational field is varying. Actually, if ϕ is the gravitational potential, the gravitational acceleration is $g \approx \partial\phi$ and $L^{-2} \approx \partial^2\phi/\phi$, so one has to require $l \ll L \to [l^2\partial^2\phi]/\phi \ll 1$.

If the above condition it is not satisfied the effects of curvature will be observable by looking e.g. at the relative acceleration between two geodesics. To be concrete, let us imagine we have a flat two dimensional plane. If two geodesics cross at a point on this plane they will then diverge with a *constant* angle. However, if the same two geodesics happen to lie on a two-sphere such angle will not be constant (they will actually cross again). So the effect of curvature clearly resides in the variation of the rate of change of the two geodesic relative position.

Indeed, we can see this simply in the context of the motion of a set of non-relativistic particle in Newtonian gravity [4, p. 197]. Imagine the set of their trajectories is parametrised as $x^\mu(t,n)$ where t is the absolute Newtonian time and n labels the trajectory. Then the separation between two nearby trajectories is described by a vector $n \equiv (\partial x^\mu/\partial n)$. As we said, the effect of curvature is not in the rate of change of this vector but in its non-constancy. This, given two nearby trajectories, is

$$\frac{\partial^2 n^\mu}{\partial t^2} = \frac{\partial^2}{\partial t^2}\left(\frac{\partial x^\mu}{\partial n}\right) = \frac{\partial}{\partial n}\left(\frac{\partial^2 x^\mu}{\partial t^2}\right) = n^\nu \frac{\partial}{\partial x^\nu}\left(-\frac{\partial \phi}{\partial x^\mu}\right) = -n^\nu \frac{\partial^2\phi}{\partial x^\nu \partial x^\mu}\,. \tag{3.1}$$

We see then that the effect depends on the second derivative of the gravitational potential rather (while the gravitational force is proportional to the first derivative). It is precisely this quantity that measures the effect of curvature and that cannot be made vanishing by moving to a freely falling frame.

Let us now study the same phenomenon in full relativistic gravity. Define a set of geodesics $\gamma_s(t)$ that are parametrized by some parameter s that varies from geodesics to geodesics, while t is the affine parameter along the geodesics themselves. The collection of these curves define a smooth 2-dimensional surface in spacetime, since we can assume that the parameter s changes continuously. We can now define two vectors, T^μ and S^μ: $T^\mu \equiv \frac{\partial x^\mu}{\partial t}$ is the tangent vector to the geodesic, while $S^\mu \equiv \frac{\partial x^\mu}{\partial s}$ is the deviation vector that goes from a geodesic to the other.

We are interested in understanding how geodesics move with respect to the others as t varies. We can then define a vector describing the rate of change of the distance between two nearby geodesics V^μ, as the variation of S^μ with respect to the parameter t

$$V^\mu = \frac{\partial S^\mu}{\partial t} = (\nabla_T S)^\mu = T^\sigma \nabla_\sigma S^\mu, \tag{3.2}$$

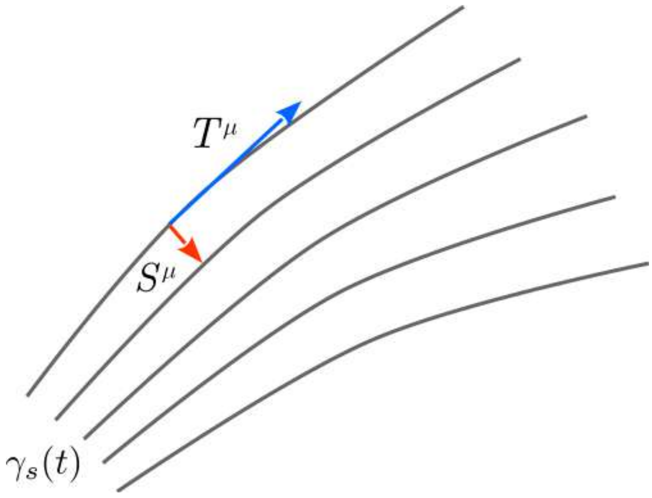


Figure 3.1: Pictorial representation of one parameter family of geodesics $\gamma_s(t)$ with its tangent and deviation vectors.

and a "relative acceleration" of geodesics A^μ, given by the variation of V^μ in t

$$A^\mu = \frac{\partial V^\mu}{\partial t} = (\nabla_T V)^\mu = T^\sigma \nabla_\sigma V^\mu. \tag{3.3}$$

Note that the acceleration along a given curve is a different concept, and it would be $a^\mu = T^\sigma \nabla_\sigma T^\mu$, which is by definition zero for an affinely parametrized geodesic. Moreover, the fact that T and S are basis vectors adapted to our coordinate system, implies that the their commutator vanishes, $[S, T] = 0$ (one vector is Lie dragged by the other); this, by the definition of the commutator, translates as

$$S^\lambda \nabla_\lambda T^\mu = T^\lambda \nabla_\lambda S^\mu. \tag{3.4}$$

Now, let us compute explicitly A^μ:

$$\begin{aligned} A^\mu &= T^\rho \nabla_\rho V^\mu = T^\rho \nabla_\rho \left(T^\lambda \nabla_\lambda S^\mu\right) \overset{(3.4)}{=} T^\rho \nabla_\rho \left(S^\lambda \nabla_\lambda T^\mu\right) = \\ &= \left(T^\rho \nabla_\rho S^\lambda\right)(\nabla_\lambda T^\mu) + T^\rho S^\lambda (\nabla_\rho \nabla_\lambda T^\mu) = \\ &= \left(T^\rho \nabla_\rho S^\lambda\right)(\nabla_\lambda T^\mu) + T^\rho S^\lambda (\nabla_\lambda \nabla_\rho T^\mu) + T^\rho S^\lambda R^\mu{}_{\nu\rho\lambda} T^\nu = \\ &= \left(T^\rho \nabla_\rho S^\lambda\right)(\nabla_\lambda T^\mu) + S^\lambda \nabla_\lambda (T^\rho \nabla_\rho T^\mu) - S^\lambda (\nabla_\lambda T^\rho)(\nabla_\rho T^\mu) + R^\mu{}_{\nu\rho\lambda} T^\rho T^\nu S^\lambda. \end{aligned} \tag{3.5}$$

Up to dummy indices, the first and third term in the last line are one the opposite of the other; the second term is zero because we are considering affinely parametrized geodesics. Therefore, we obtain as a final result

$$A^\mu = R^\mu{}_{\nu\rho\lambda} T^\rho T^\nu S^\lambda. \tag{3.6}$$

Note that one can derive the same result more simply using index-free notation. We have defined $S = \partial_s$ and $T = \partial_t$, and so that the commutator is $[S, T] = \nabla_S T - \nabla_T S = 0$. Now, the acceleration is given simply by

$$A = \nabla_T \nabla_T S = \nabla_T \nabla_S T = (\nabla_T \nabla_S - \nabla_S \nabla_T) T = R^a{}_{bcd} T^b T^c S^d \partial_a.$$

The **Hawking effect** can be described heuristically as the breaking apart of virtual pairs close to a black hole. The Riemann tensor near the horizon is (in Schwarzschild coordinates) of the order $\frac{2M}{r^3} \sim \frac{1}{M^2}$ ($r_{BH} = 2M$ for a Schwarzschild black hole), therefore this is the strength of the the tidal forces at the horizon responsible for the breaking of the virtual pairs. We will further expand on this later on in this lectures, however this simple observation makes it clear that the Hawking effect should be expected to be more intense for small black holes and negligible for large ones.

More practically, given the dependence written above, one can conclude that an extended object near the horizon will be stretched toward the center of the black hole, with a tidal force which will be bigger the smaller is the mass of the black hole. This effect is generically known as ***spaghettification***. The tensile force acting on a uniform rod of length ℓ in the radial direction, mass m, and distance from the BH R, oriented radially, the tensile force is

$$F_T = \frac{G_N M_{BH} m}{4R^3}\ell \tag{3.7}$$

For black holes with masses of order $1-100 M_\odot$, the *spaghettification* is such to be lethal to humans. For black holes, above $10^4 - 10^5\, M_\odot$ the tensile force should be weak enough for allowing horizon crossing (but of course it will depends on the size and strength e.g. of your spaceship). Exercise: compute the minimal mass of a black hole for which a 20 meters long spaceship weighting $15{*}10^3$kg and with a tensile strength of 10,000 N can cross unharmed the horizon. (Answer $M \approx 10^5\, M_\odot$).

3.2 Raychaudhuri Equations

We said that the geodesics deviation equation is actually the more direct way to look at gravity. This is however a a special case of as it keeps characterises only the behaviour of a one-parameter family of geodesics. In general, one is instead interested in keeping track of the action of gravity on a multi-dimensional set of neighbouring geodesics, a congruence, such as those describing the trajectory of a bundle of photons or of a cloud of massive particles. The set of equations encapsulating such behaviour are the so called **Raychaudhuri equations**.

3.2.1 Time-like geodesics congruence

Consider a set of particles moving through space; now we want to describe the motion of this whole set of particles, more than every single particle. For concreteness let us analize first the case of massive particles moving along time-like geodesics. Generalising the previous concept, take a congruence of multiparameter timelike geodesics $\gamma_{s_i}(\tau)$; for every geodesic we define a normalized tangent vector $u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$, such that $u^\mu u_\mu = -1$, in a way that the geodesics are affinely parametrized, $u^\mu \nabla_\mu u^\nu = 0$.

Take an adapted system of coordinates $(\tau, s_1, \ldots, s_{n-1})$, such that $[S_i, u] = 0$. We want now to consider the variation of S_i^μ along the geodesic:

$$\frac{\mathrm{d}S^\mu}{\mathrm{d}\tau} \equiv u^\nu \nabla_\nu S^\mu = S^\nu \nabla_\nu u^\mu \equiv B^\mu{}_\nu S^\nu, \tag{3.8}$$

where we have defined the tensor $B^\mu{}_\nu$ as the measure of the failure of S^μ to being parallely transported along the geodesic. We want to evaluate $B^\mu{}_\nu$ along the congruence.

Now, $B^\mu{}_\nu \equiv \nabla_\nu u^\mu$, indeed lives on the subspace orthogonal to u^μ; in fact

$$u^\mu B_{\mu\nu} = u^\mu \nabla_\nu u_\mu = \frac{\nabla_\nu (u^\mu u_\mu)}{2} = \frac{\nabla_\nu(-1)}{2} = 0. \tag{3.9}$$

Conversely, $u^\nu B_{\mu\nu} = u^\nu \nabla_\nu u_\mu = 0$, since we are on an affinely parametrized geodesic.

Since $B_{\mu\nu}$ belongs to the subspace orthogonal to u^μ, it can be generally expanded in a part containing the trace, a traceless symmetric part, and an antisymmetric part, each of them belonging to the same orthogonal subspace:

$$B_{\mu\nu} = \frac{\theta}{3} P_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu}. \tag{3.10}$$

We again the projector operator is $P^\mu{}_\nu = \delta^\mu_\nu + u^\mu u_\nu$ for a timelike normal vector. and we have introduced the trace of $B_{\mu\nu}$,

$$\theta \equiv P^{\mu\nu} B_{\mu\nu}, \tag{3.11}$$

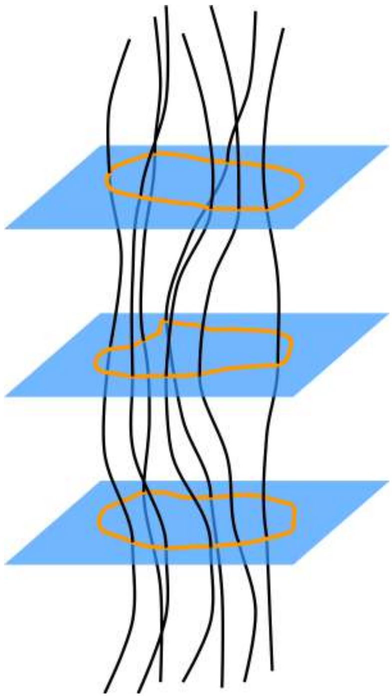


Figure 3.2: Pictorial representation of a geodesic congruence.

which is called the **expansion** of the congruence, and it tells us how much our cloud of particle expands or contracts isotropically. In the above equation $\sigma_{\mu\nu}$ is he symmetric traceless part of $B_{\mu\nu}$,

$$\sigma_{\mu\nu} = B_{(\mu\nu)} - \frac{\theta}{3} P_{\mu\nu}, \tag{3.12}$$

and it is called the **shear**, which tells us how things get elongated or squashed. Finally, we had also to introduce the antisymmetric part of $B_{\mu\nu}$,

$$\omega_{\mu\nu} = B_{[\mu\nu]} \tag{3.13}$$

which is called the **twist**, and describes rotation of the particle of the cloud as they move forward along the congruence. Note that whenever a geodesic congruence is hypersurface orthogonal (i.e its tangent vector can be written as the gradient of a scalar function) then it must have $\omega_{\mu\nu} = 0$ (see e.g. [4, p. 226]).

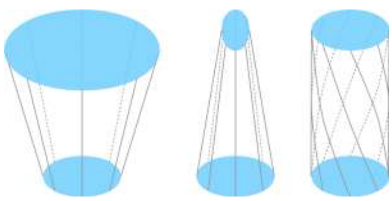


Figure 3.3: Pictorial representation of the congreunce evolution associated to expansion, shear and twist.

In order to derive the Raychaudhuri equations, we start by studying the variation of the tensor

$B_{\mu\nu}$ with respect to τ:

$$\begin{aligned}\frac{\mathrm{d}B_{\mu\nu}}{\mathrm{d}\tau} &= u^\rho \nabla_\rho B_{\mu\nu} = u^\rho \nabla_\rho \nabla_\nu u_\mu = u^\rho \nabla_\nu \nabla_\rho u_\mu + u^\rho R^\lambda{}_{\mu\nu\rho} u_\lambda = \\ &= \nabla_\nu \cancel{(u^\rho \nabla_\rho u_\mu)} - (\nabla_\nu u^\rho)(\nabla_\rho u_\mu) + u^\rho R_{\lambda\mu\nu\rho} u^\lambda = \\ &= -B^\rho{}_\nu B_{\mu\rho} + R_{\lambda\mu\nu\rho} u^\lambda u^\rho = \\ &= -B^\rho{}_\nu B_{\mu\rho} - R_{\mu\lambda\nu\rho} u^\lambda u^\rho.\end{aligned} \tag{3.14}$$

To derive the **Raychaudhuri equation for the expansion**, we can simply take the trace of the previous relation. By doing this, we get

$$\begin{aligned}\frac{\mathrm{d}\theta}{\mathrm{d}\tau} &= -B^\rho{}_\nu B^\nu{}_\rho - R^\nu{}_{\lambda\nu\rho} u^\lambda u^\rho \\ &= -\left(\frac{\theta^2}{9} P^\rho{}_\nu P^\nu{}_\rho - \frac{\theta}{3} P^\rho{}_\nu \sigma^\nu{}_\rho - \frac{\theta}{3} P^\rho{}_\nu \omega^\nu{}_\rho + \sigma^\rho{}_\nu \sigma^\nu{}_\rho + \omega^\rho{}_\nu \omega^\nu{}_\rho\right) - R_{\lambda\rho} u^\lambda u^\rho = \\ &= -\frac{\theta^2}{3} - \sigma_{\mu\nu}\sigma^{\mu\nu} + \omega_{\mu\nu}\omega^{\mu\nu} - R_{\mu\nu} u^\mu u^\nu.\end{aligned} \tag{3.15}$$

If we take now the traceless symmetric part, we get the **Raychaudhuri equation for the shear** [3]

$$\frac{\mathrm{d}\sigma_{\mu\nu}}{\mathrm{d}\tau} = -\frac{2}{3}\theta\sigma_{\mu\nu} - \sigma_{\mu\alpha}\sigma^\alpha{}_\nu - \omega_{\mu\rho}\omega^\rho{}_\nu + \frac{1}{3}P_{\mu\nu}\left(\sigma_{\alpha\beta}\sigma^{\alpha\beta} - \omega_{\alpha\beta}\omega^{\alpha\beta}\right) + C_{\alpha\nu\mu\beta}u^\alpha u^\beta + \frac{1}{2}\bar{R}_{\mu\nu}, \tag{3.16}$$

where $C_{\alpha\nu\mu\beta}$ is the Weyl tensor, here related to the deformation of space due to gravitational waves, and $\bar{R}_{\mu\nu}$ is the trace-free Ricci tensor projected onto the three-dimensional subspace [3]

$$\bar{R}_{\mu\nu} = P^\alpha_\mu P^\beta_\nu R_{\alpha\beta} - \frac{1}{3}P_{\mu\nu}P^{\alpha\beta}R_{\alpha\beta}. \tag{3.17}$$

Taking the antisymmetric part now gives us the **Raychaudhuri equation for the twist** [3]

$$\frac{\mathrm{d}\omega_{\mu\nu}}{\mathrm{d}\tau} = -\frac{2}{3}\theta\omega_{\mu\nu} + \sigma^\alpha{}_\mu\omega_{\nu\alpha} - \sigma^\alpha{}_\nu\omega_{\mu\alpha}. \tag{3.18}$$

In general, if the congruence is not made of geodesics, an extra term $-a_\mu u_\nu$ appears in $B_{\mu\nu}$, while in $\frac{\mathrm{d}\theta}{\mathrm{d}\tau}$ will have to add a term $\nabla_\mu a^\mu$, where $a^\mu = \frac{\mathrm{d}u^\mu}{\mathrm{d}\tau}$. If the congruence, however, is made of geodesics that are not affinely parametrized, generally we have $u^\nu\nabla_\nu u^\mu = \kappa u^\mu$ and in the expansion equation we have to add a term $-\kappa\theta$.

3.2.2 Null geodesics congruence

If we instead consider a null congruence, we have slight differences. Our timelike u^μ is now substituted by a null k^μ. As a consequence the space orthogonal to the congruence is no more three dimensional but rather bi-dimensional (if spacetime has four dimensions). As a way to see this, consider the four-dimensional Minkowski metric in spherical coordinates:

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + \,\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2.$$

We can define two types of null coordinates, $u \equiv t - r$ and $v \equiv t + r$, such that null vectors have either u or v constant; with these two new coordinates, the metric becomes

$$\mathrm{d}s^2 = -\,\mathrm{d}u\,\mathrm{d}v + r^2(u, v)\,\mathrm{d}\Omega^2.$$

If we take a constant null surface at e.g. $u = \textit{constant}$, the $\mathrm{d}u\,\mathrm{d}v$ part of the line element goes to zero, and one can immediately see that the orthogonal subspace is thus two-dimensional. The problem is that we cannot define a metric on such subspace e.g. via the usual projector $g+\sigma kk$ as this would not have a zero contraction with the tangent vector to the congruence (and in any case $\sigma = k^\mu k_\mu = 0$).

We can however define another, auxiliary, null vector l^μ, which (in some frame) points in the opposite spatial direction to k^μ and is normalized so that $l^\mu l_\mu = 0$ and $l^\mu k_\mu = -1$. We also demand that this auxiliary vector is parallel-transported, $k^\mu \nabla_\mu l^\nu = 0$ (which is compatible with the previous conditions as the inner product is preserved by parallel-transport). [For example, in flat spacetime if $k_\alpha = -\partial_\alpha u$ in the local inertial frame, one can take $l_\alpha = -(1/2)\partial_\alpha v$.] The problem with this "trick" is that this auxiliary vector is by no means unique (as pointing in the opposite spatial direction is frame-dependent as we said). But let us proceed for the moment and see what we get.

The projector operator for a subspace which is orthogonal to both k^μ and l^μ is then

$$Q_{\mu\nu} = g_{\mu\nu} + k_\mu l_\nu + k_\nu l_\mu. \tag{3.19}$$

Now, all the relevant information we need are encoded in the projected version of $B^\mu{}_\nu$:

$$\hat{B}^\mu{}_\nu = Q^\mu{}_\alpha Q^\beta{}_\nu B^\alpha{}_\beta, \tag{3.20}$$

where we have basically projected every index; with this, decomposing in all the components, we have

$$\hat{B}_{\mu\nu} = \frac{1}{2}\hat{\theta} Q_{\mu\nu} + \hat{\sigma}_{\mu\nu} + \hat{\omega}_{\mu\nu}. \tag{3.21}$$

It should be noted that now the equations will not depend on l^μ; in order to realize this, consider the expansion

$$\hat{\theta} = Q^{\mu\nu}\hat{B}_{\mu\nu} = \hat{B}^\mu{}_\mu = g^{\mu\nu}B_{\mu\nu} = \theta, \tag{3.22}$$

where we used that fact that the projector Q is idempotent, and that any contraction of k with B is zero. So, we see that in fact the expansion, that in principle could have had a dependence on l^μ, does not depend on l^μ.

This is promising, however if we would compute explicitly $\hat{\sigma}$ and $\hat{\omega}$, we would find that they depend on the choice of l^μ! Nonetheless, the Raychaudhuri equation for the expansion has still the familiar form

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{\theta^2}{2} - \hat{\sigma}_{\mu\nu}\hat{\sigma}^{\mu\nu} + \hat{\omega}_{\mu\nu}\hat{\omega}^{\mu\nu} - R_{\mu\nu}k^\mu k^\nu \quad (+\kappa\theta)\,, \tag{3.23}$$

and the dependence on l^μ cancels in the squares of $\hat{\sigma}$ and $\hat{\omega}$ (while the projector tensors drop out of the Ricci term when one takes the trace). Hence, we can conclude that in the end our equation is safely frame independent.

One might wonder why in the above equation the expansion squared term has a two at denominator rather than the factor 3 we found for time-like geodesics congruence. The point is that the factor at the denominator represent the dimension of the subspace orthogonal to u^μ, and when u^μ is timelike, the orthogonal subspace has dimension 3 in a four dimensional spacetime. However, as we saw, the subspace orthogonal to a null vector has dimension 2.

The expansion that we have defined here can be seen exactly as the variation of the volume of test particles in $n+1$ dimensions [4, p. 227]:

$$\theta = \frac{1}{\delta V^{(n)}}\frac{\delta V^{(n)}}{\delta\tau};$$

in four spacetime dimensions, the expansion of a timelike congruence is then

$$\theta^{(4)}_{\text{timelike}} = \frac{1}{\delta V^{(3)}}\frac{\delta V^{(3)}}{\delta\tau};$$

for a null congruence, instead we have

$$\theta^{(4)}_{\text{null}} = \frac{1}{\delta A}\frac{\delta A}{\delta\tau}.$$

3.3 Energy Conditions

While we just saw that the Raychaudhuri equation is completely geometric in its nature, the presence of the term proportional to the Ricci tensor clearly suggest that it can be used together with the Einstein equations (or their generalizations) so to tell us how the geometry behaves on the base of conditions imposed on the matter stress-energy tensor.

Indeed, Einstein gravity is a particularly simple theory where the Ricci tensor is directly related to the matter Stress Energy Tensor (and this is not the case for more general theories of gravity, as we shall see later on)

$$R_{\mu\nu} = 8\pi G_N \left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right).$$

Hence it natural to substitute the Ricci tensor for the SET and depending on the nature of the congruence introduce some reasonable conditions for predicting the behaviour of matter as described by the Raychaudhuri equations. Let us see them in detail.

1. **Null Energy Condition** (NEC): we define this in a null Raychaudhuri equation; what we have is a term $R_{\mu\nu}k^\mu k^\nu$. Now, through Einstein's equation, we replace this by $T_{\mu\nu}k^\mu k^\nu$ (the part with $T g_{\mu\nu}$ is zero, since k^μ is null). The NEC then states that

$$T_{\mu\nu}k^\mu k^\nu \geq 0 \quad \forall k^\mu \text{ null}.$$

   If we consider a perfect fluid, $T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu}$, the NEC implies $\rho + p \geq 0$.

2. **Weak Energy Condition** (WEC): this is sometimes used for constraining the behaviour of a congruence of timelike geodesics; the WEC states that

$$T_{\mu\nu}v^\mu v^\nu \geq 0 \quad \forall v^\mu \text{ timelike}.$$

   For a perfect fluid, the WEC implies both $\rho + p \geq 0$ and $\rho \geq 0$.

3. **Dominant Energy Condition** (DEC): the DEC requires the WEC and that the object $T_{\mu\nu}v^\nu$ (which is the flux measured by an observer along the congruence) is not spacelike for any timelike v^μ. For a perfect fluid the DEC implies $\rho \geq 0$ and $p \in [-\rho, \rho]$. In order to give a physical interpretation, we can say that in this case the locally measured energy density is always positive for any timelike observer, and the energy flux si timelike or null; therefore, there are no superluminal fluxes.

4. **Strong Energy Condition** (SEC): this is the natural one to use for a timelike congruence as derives directly from using the above written simple relation the Ricci-SET in GR. It requires that

$$\left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right) v^\mu v^\nu \geq 0 \quad \forall v^\mu \text{ timelike}.$$

   For a perfect fluid the SEC takes the form: $\rho + p \geq 0$ and $\rho + 3p \geq 0$. The SEC implies the NEC but not the WEC, and the DEC does not imply the SEC.

As a practical application, consider the copernican principle which, as we have already seen, translates into the FRW metric. Take a timelike congruence. The tangent to this congruence, u^μ, can be written as the gradient of a scalar, $\nabla^\mu \phi$ (the cosmic time). Therefore, we can set the twist to zero [4, p. 226], while by isotropy we can set the shear to zero.

Consider now a small spherical volume, $V = \frac{4}{3}\pi R^3$, where $R = R(t) = a(t)R_0$. The expansion is

$$\theta = \lim_{\delta V \to 0} \frac{1}{V}\frac{\delta V}{\delta \tau} = \frac{1}{\frac{4}{3}\pi R^3}\frac{4}{3}\pi 3R^2 \dot{R} = 3\frac{\dot{a}}{a} = 3H.$$

Then, the Raychaudhuri equation for the expansion becomes

$$\dot{\theta} = -\frac{\theta^2}{3} - R_{\mu\nu}u^\mu u^\nu;$$

Without loss of generality, one can choose u^μ in a way that it is purely timelike; then the Raychaudhuri equation becomes

$$3\dot{H} = -3H^2 - R_{00} \implies 3\frac{\ddot{a}}{a} = -R_{00}.$$

Until now, we have not used Einstein's equations (and per se the Raychaudhuri equations are valid in **any** theory of gravity, being purely geometrical in nature) but we can do that now in order to determine R_{00}. So in General relativity

$$R_{00} = 8\pi G_N \left(T_{00} - \frac{1}{2}Tg_{00}\right),$$

which when used in the Raychaudhuri ends up provinding us with nothing else than the well known second Friedmann equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{3}\left(\rho + 3p\right),$$

In particular, if for the cosmological fluid $\rho + 3p > 0$, then $\ddot{a}/a < 0$, and the expansion of the Universe decelerates; however, for $\rho + 3p < 0$, i.e. if the SEC is violated, then $\ddot{a}/a > 0$, and the expansion of the Universe accelerates. Current observations indeed seem to imply that *if GR holds at cosmological scales* then the current energy-matter composition of the Universe is dominated by a SEC-violating component. This is what we call dark energy.

4 Variational Principle

4.1 Lagrangian Formulation of General Relativity

In the previous sections, we have seen a series of topics concerning kinematics, like the Raychaudhuri equations; now we want to start studying the dynamics connecting gravity to the matter distribution through the Einstein's field equations:

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{4.1}$$

We will provide a formal derivation of the above equations using the variational principle: starting from an action, we will vary it with respect to our dynamical field, arriving to the equations of motion for that field. In our case the field is the metric, and we want to find a suitable Lagrangian density in order to have second order equations of motion for it.

First of all we want to find a theory that satisfies the strong equivalence principle (universality of free fall even for self gravitating bodies, local Lorentz invariance and local position invariance for gravitational experiments). Usually, the Lagrangians we write depend on the fields and their first derivatives, but, since for the strong equivalence principle we can always find a reference frame in which the first derivatives of the metric vanish, we have to go to the next order and write a Lagrangian depending also from the second derivatives of the metric (quite reasonably so, given that the latter encodes the curvature of spacetime):

$$\mathscr{L}_{\text{grav}} = \mathscr{L}_{\text{grav}}(g, \partial g, \partial^2 g). \tag{4.2}$$

However, notice that the second derivatives dependence should be at most linear, otherwise by integrating by parts we would end up with higher than the second order equations of motion.

Any non trivial tensor that can be made from the metric and its first and second derivatives is a combination of the metric itself and the Riemann tensor (which takes into account all the possible curvatures), and from these two we can build up only one possible scalar that is the Ricci scalar. We saw in Chapter 2 that this is a good choice also if we are in non-metric theories, given that the Ricci scalar is unique even if the Γ are not the Christoffel symbols.

Thus the action for gravity can be written as:

$$S_{\text{EH}} = \frac{1}{\kappa} \int \mathrm{d}^4x \sqrt{-g} R \tag{4.3}$$

where $\kappa = 16\pi G$ is the appropriate factor (which has to be constant for SEP to hold) fixed by the requirement to recover standard non-relativistic gravity in the Newtonian limit. In the following, we will sometime omit this for conciseness.

The action we have just written is called the Einstein–Hilbert action. We can also write a total action

$$S_{\text{tot}} = S_{\text{EH}} + S_{\text{m}}, \tag{4.4}$$

where $S_{\rm m}$ is the action for the matter fields

$$S_{\rm m} = \int \mathrm{d}^4x\sqrt{-g}\mathscr{L}\left(\phi^i, \nabla_\mu\phi^i\right). \tag{4.5}$$

Notice that we are not allowing non-minimal gravity-matter couplings of the form $f(\phi)R$, since these are equivalent to have a dynamical gravitational constant $G(x)$ depending on the spacetime position, as this would violate the requirement of local position invariance of all the interaction dictated by the SEP. Similarly, we do not consider more general derivative couplings that would lead to even more general classes of scalar tensor theories (more on this later).

The fact that the gravitational action contains up to second order derivatives of the metric, implies that we cannot use straightforwardly the usual Euler-Lagrange equations. We can however still resort to a minimal action principle and impose $\delta S_{\rm tot} = 0$. [1] By performing the variation of the action with respect to $g_{\mu\nu}$ and imposing it to be 0, we shall get the gravitational equations of motion.

Actually, it is more convenient to perform the variation with respect to the inverse metric $g^{\mu\nu}$ given that $R = g^{\mu\nu}R_{\mu\nu}$. It can be shown that the following relation holds [2]

$$\delta g_{\mu\nu} = -g_{\mu\rho}g_{\nu\sigma}\delta g^{\rho\sigma}. \tag{4.6}$$

Therefore, apart from some metric factors, the variation with respect to the metric or the inverse metric leads to the same result.

Now, if we vary the Einstein–Hilbert action we get (for concreteness in what follows explicitly adopt a $(-,+,+,+)$ signature)

$$\delta S_{\rm EH} = \frac{1}{k}\delta\int \mathrm{d}^4x\sqrt{-g}g^{\mu\nu}R_{\mu\nu} = \frac{1}{k}\int \mathrm{d}^4x\left[\sqrt{-g}g^{\mu\nu}\delta R_{\mu\nu} + \sqrt{-g}R_{\mu\nu}\delta g^{\mu\nu} + R\delta\sqrt{-g}\right] \tag{4.7}$$

Since the functional derivative of a functional S with respect to a generic field is defined as:

$$\delta S = \int \mathrm{d}^4x\sqrt{-g}\frac{\delta S}{\delta\phi^i}\delta\phi^i \tag{4.8}$$

we would like to arrive to a form in which the term $\delta g^{\mu\nu}$ is factorized. The second term in Equation (4.7) is already in the right form, however for the other two we need some manipulations. Let us start from the third one. The following property holds for a square matrix:

$$\ln\left(\det(A)\right) = \mathrm{tr}(\ln(A)) \quad\Longrightarrow\quad \frac{1}{\det(A)}\delta(\det(A)) = \mathrm{tr}(A^{-1}\delta A), \tag{4.9}$$

so that

$$\frac{1}{g}\delta g = g^{\mu\nu}\delta g_{\mu\nu} \quad\Longrightarrow\quad \delta g = -g(g_{\rho\sigma}\delta g^{\rho\sigma}) \quad\Longrightarrow\quad \delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}g_{\mu\nu}\delta g^{\mu\nu}, \tag{4.10}$$

where in the second step we have used Equation (4.6).

Finally, the first term in (4.7) can be shown (see the details below) to be a boundary term and therefore one would be tempted to say that this term does not give any contribution to the variation of the action. We will see in the following that this is not exactly the case, but for the moment we can decide to neglect this term and postpone its discussion for a while.

---

[1] For a Lagrangian function of i fields and their derivatives one can of course use generalised EL-equations in the form

$$\frac{\delta^{EL}}{\delta\phi^i}L(\phi^i, \partial\phi^i, \partial^2\phi^i, \dots) = 0 \quad \text{where} \quad \frac{\delta^{EL}}{\delta\phi^i} = \frac{\partial}{\partial\phi^i} - \partial_\mu\frac{\partial}{\partial(\partial_\mu\phi^i)} + \partial_\mu\partial_\nu\frac{\partial^2}{\partial(\partial_\mu\partial_\nu\phi^i)} - \dots$$

but this is not necessarily more efficient than doing the direct variation of the action.

[2] This can be check starting from the identity $g^{\mu\lambda}g_{\lambda\nu} = \delta^\mu_\nu$, the variation then implies $\delta g^{\mu\lambda}g_{\lambda\nu} + g^{\mu\lambda}\delta g_{\lambda\nu} = 0$ which of course can be rewritten has $\delta g_{\lambda'\nu} = -g_{\mu\lambda'}g_{\lambda\nu}\delta g^{\mu\lambda}$.

This said, using only the second and third term in Equation (4.7) and using the relation (4.10), we immediately get the Einstein's equations. To see this clearly we remind the definition of the stress energy tensor:

$$T_{\mu\nu} \equiv -\frac{2}{\sqrt{-g}}\frac{\delta S_{\mathrm{m}}}{\delta g^{\mu\nu}}. \tag{4.11}$$

So we have

$$\frac{\delta S}{\delta g^{\mu\nu}} = \frac{\delta S_{\mathrm{EH}}}{\delta g^{\mu\nu}} + \frac{\delta S_{\mathrm{m}}}{\delta g^{\mu\nu}} = 0 \tag{4.12}$$

where

$$\frac{\delta S_{\mathrm{EH}}}{\delta g^{\mu\nu}} = \frac{\sqrt{-g}}{16\pi G}\left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right) \tag{4.13}$$

and

$$\frac{\delta S_{\mathrm{m}}}{\delta g^{\mu\nu}} = -\frac{1}{2}\sqrt{-g}T_{\mu\nu} \tag{4.14}$$

Equation (4.12) expresses the Einstein's equations.

Let us now try to motivate the choice of that definition of the stress-energy tensor.

Consider a scalar field with an action

$$S_\phi = \int \mathrm{d}^4x\sqrt{-g}\left[-\frac{1}{2}g^{\mu\nu}(\nabla_\mu\phi)(\nabla_\nu\phi) - V(\phi)\right]; \tag{4.15}$$

if we vary it with respect to the inverse metric and apply the definition we used above, Equation (4.11), we get

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S_\phi}{\delta g^{\mu\nu}} = \nabla_\mu\phi\nabla_\nu\phi - \frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\nabla_\rho\phi\nabla_\sigma\phi - g_{\mu\nu}V(\phi), \tag{4.16}$$

and in flat spacetime this reduces exactly to the expected stress energy tensor for a scalar field.

We can also notice that the quantity defined in Equation (4.11) is a $(0,2)$ rank tensor, with the dimension of an energy density, symmetric and covariantly conserved, since we can prove that it satisfies the relation

$$\nabla^\mu T_{\mu\nu} = 0,$$

where the covariant derivative is referred to the metric appearing in equation (4.11). This conservation would be related to the Bianchi identity of the Einstein tensor.

Now we can go back to the discussion of the first term in (4.7). In order to be more accurate, we have to do the variation of the Ricci tensor. It can be proven that (see e.g. [2])

$$\delta R_{\mu\nu} = \nabla_\alpha(\delta\Gamma^\alpha{}_{\mu\nu}) - \nabla_\nu(\delta\Gamma^\alpha{}_{\alpha\mu}). \tag{4.17}$$

This equation provides the generic variation of the Ricci tensor and it is always true, independently on the theory of gravity considered. Since we are in a metric theory, the affine connection is provided by the Christoffel symbols, so that

$$\delta\Gamma^\alpha{}_{\mu\nu} = -\frac{1}{2}\left[g_{\lambda\mu}\nabla_\nu(\delta g^{\lambda\sigma}) + g_{\lambda\nu}\nabla_\mu(\delta g^{\lambda\sigma}) - g_{\mu\alpha}g_{\nu\beta}\nabla^\sigma(\delta g^{\alpha\beta})\right] \tag{4.18}$$

This can be seen as the difference of two connections and hence it is a tensor and one can apply to it the covariant derivative in the standard way:

$$\nabla_\rho(\delta\Gamma^\alpha{}_{\mu\nu}) = \partial_\rho(\delta\Gamma^\alpha{}_{\mu\nu}) + \Gamma^\alpha{}_{\rho\sigma}\delta\Gamma^\sigma{}_{\mu\nu} - \Gamma^\sigma{}_{\rho\mu}\delta\Gamma^\alpha{}_{\sigma\nu} - \Gamma^\sigma{}_{\rho\nu}\delta\Gamma^\alpha{}_{\mu\sigma}\,. \tag{4.19}$$

So, putting everything in (4.17) gives in the end

$$\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}g^{\mu\nu}\delta R_{\mu\nu} = \int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}\nabla_\mu\nabla_\nu(-\delta g^{\mu\nu} + g^{\mu\nu}g_{\alpha\beta}\delta g^{\alpha\beta}) = \int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}\nabla_\mu v^\mu, \tag{4.20}$$

where the integral is performed over the whole manifold $\mathcal{M}$ and in the last step we have defined

$$v^\mu \equiv \nabla_\nu(-\delta g^{\mu\nu} + g^{\mu\nu} g_{\alpha\beta} \delta g^{\alpha\beta}). \tag{4.21}$$

Since we have the divergence of a vector, we can use Stokes' theorem to rewrite the integral as

$$\int_{\mathcal{M}} \mathrm{d}^4x \sqrt{-g}\, \nabla_\mu v^\mu = \oint_{\partial\mathcal{M}} \sigma\, \mathrm{d}^3x \sqrt{-h}\, v^\mu n_\mu \tag{4.22}$$

where the integral is performed over the boundary of the manifold $\partial\mathcal{M}$, we have introduced the metric h induced on such a boundary and again $\sigma = n^\rho n_\rho$. Usually, when we do variations of a field, we keep it fixed at the boundary and thus integrals of this type vanish. However, in this case we do not have only terms with the variation of the metric appears, but as well terms involving the variation of the derivative of the metric, which can be generically non-zero. At this point, we can choose two different approaches to solve this problem: we can either introduce a counter term in the action that cancels this first term when we do the variation, or we recognise that the problematic terms are introduced by the assumption that the connections are metric compatible. Let us consider below both paths.

4.1.1 Gibbons-Hawking-York counterterm

Let us analyse the product:

$$n_\mu v^\mu = n_\mu(-\nabla_\alpha \delta g^{\mu\alpha} + g_{\alpha\beta} \nabla^\mu \delta g^{\alpha\beta}) = n_\mu g_{\alpha\beta}(\nabla^\mu \delta g^{\alpha\beta} - \nabla^\beta \delta g^{\mu\alpha}) = n_\mu P_{\alpha\beta}(\nabla^\mu \delta g^{\alpha\beta} - \nabla^\beta \delta g^{\mu\alpha}) \tag{4.23}$$

where in the last passage we have substituted the projection operator $P_{\alpha\beta} = g_{\alpha\beta} - \sigma n_\alpha n_\beta$ in place of the metric, since the second term of the projection operator applied to the parentheses vanishes.

Now, we can immediately see that the term $P_{\alpha\beta} \nabla^\beta \delta g^{\mu\alpha}$ vanishes. Indeed, this terms corresponds to take the covariant derivative projected on the hypersurface; this gives 0 because the field is held fixed at the boundary in a variational principle and hence δg vanishes on $\partial\mathcal{M}$. So the integral in (4.20) becomes

$$\sigma \int_{\partial\mathcal{M}} \mathrm{d}^3x \sqrt{-h}\, (n_\mu P_{\alpha\beta} \nabla^\mu \delta g^{\alpha\beta}). \tag{4.24}$$

We can now prove that the term in parentheses is the variation of the trace of the extrinsic curvature of the boundary. The trace of the extrinsic curvature is given by

$$K = \nabla_\alpha n^\alpha = P^\alpha{}_\beta \nabla_\alpha n^\beta, \tag{4.25}$$

where in the last step we introduced the projection operator because the directional derivative of n along n is zero. If we now vary this quantity, we get (note that $\delta P_{\alpha\beta} = 0$ on the boundary)

$$\delta K = \frac{1}{2} n^\gamma P^\alpha{}_\beta g^{\beta\delta} \left[\nabla_\alpha(\delta g_{\gamma\delta}) + \nabla_\gamma(\delta g_{\alpha\delta}) - \nabla_\delta(\delta g_{\alpha\gamma})\right] = \frac{1}{2} n^\gamma P^\alpha{}_\beta g^{\beta\delta} \nabla_\gamma(\delta g_{\alpha\delta}) = \frac{1}{2} n^\gamma P^{\alpha\delta} \nabla_\gamma(\delta g_{\alpha\delta}); \tag{4.26}$$

where we used twice that the projected derivative applied on the variation of the metric gives zero.

So this shows that the term in equation (4.24) can be written as

$$2\sigma \int_{\partial\mathcal{M}} \mathrm{d}^3x \sqrt{-h}\, \delta K. \tag{4.27}$$

It is clear that in order to recover the standard Einstein equations, we need to add to the original gravitational action a counter term, the so called **Gibbons-Hawking-York term** (GHY) and redefine the gravitational action as

$$S_{\text{grav}} = S_{\text{EH}} + S_{\text{GHY}} = \frac{1}{16\pi G} \int_{\mathcal{M}} \mathrm{d}^4x \sqrt{-g}\, R - \frac{2\sigma}{16\pi G} \int_{\partial\mathcal{M}} \mathrm{d}^3x \sqrt{-h}\, K. \tag{4.28}$$

Indeed, if we perform the variation of this action, the additional term perfectly cancels the term containing the variation of the Ricci tensor. Note that, in principle, when we do the variation, we should have also the variation with respect to $\sqrt{-h}$; still, since the full metric is fixed at the boundary and its variation is zero, also the projected metric has null variation. In this way we can obtain again the Einstein's equations.

Thanks to this counternterm, we obtain the correct result for what concerns Einstein's equations; however, if we are interested in the numerical value of the action (e.g. in a path integral approach), we still have to face some issues. As an example, consider a Minkowski spacetime with a boundary of cylindrical shape (Figure 4.1). The Einstein–Hilbert part of the action is zero, since the spacetime is flat. The boundary is made of two circular surfaces, respectively at $t_1 = \text{const}$ and $t_2 = \text{const}$, where the extrinsic curvature is zero, and a cylindrical surface at constant radius R with an induced metric $h_{ij} = -\,\mathrm{d}t^2 + R^2\,\mathrm{d}\Omega^2$. The vector orthogonal to this surface is $n_\alpha = \partial_\alpha r$ and the trace of the extrinsic curvature can be computed to be $K = \nabla_\alpha n^\alpha = \frac{2}{R}$. It is also easy to see that $\sqrt{|h|} = R^2\,|\sin(\theta)|$ and so the GHY term turns out to be proportional to $S_{\text{GHY}} \propto R(t_2 - t_1)$, which diverges as $R \to \infty$. Thus the action diverges.

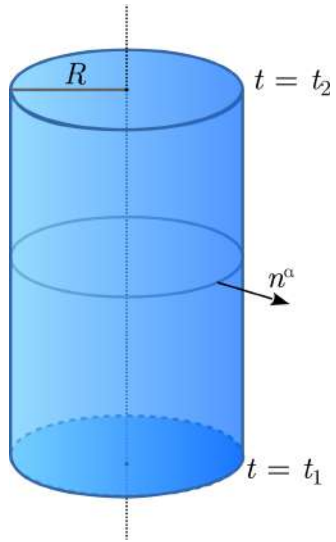


Figure 4.1: Minkowski spacetime with a boundary cylindrical shape.

The way out is to regularize this in some way. First of all we remember that, even if the example above is in flat spacetime, the action for gravity is written in a completely general way and it is valid also when there is matter and the spacetime is curved. Now let us proceed to regularize the GHY term as follows:

$$S_{\text{GHY}} = \frac{2\sigma}{16\pi G}\int_{\partial\mathcal{M}} \mathrm{d}^3x\sqrt{-h}(K - K_0). \tag{4.29}$$

Let us try to explain the meaning of this regularization. K_0 is the extrinsic curvature of the boundary $\partial\mathcal{M}$ computed in flat spacetime. Even if this term diverges, neither K nor K_0 represent actual physical quantities, the real physical quantity is the difference between the surface term in our spacetime with respect to the case in which there is no gravity; such a difference is finite. In the case of Minkowski spacetime we analysed before, $K = K_0$ and the GHY term goes to zero; in the case of curved spacetime, however, the difference will be different from zero, but it will still be finite. This case is similar to the Casimir effect, where we do a subtraction of the zero point energy to obtain a finite effect.

To be fair, the definition of K_0 we gave is not precise: it is the extrinsic curvature of the boundary in the *maximally symmetric limit* of the solution we are considering. For example, if we considered the Schwarzschild solution, the K_0 term of course would be the extrinsic curvature of the boundary in flat spacetime, but if we considered a Schwarzschild-de Sitter solution, it would be the extrinsic curvature computed in de Sitter.

We conclude by stressing that, since K_0 is a constant, its variation with respect to the metric gives zero, and therefore it does not affect the equation of motion for the metric. Hence, we recover the Einstein's equations.

4.1.2 Schrödinger action

Let us now try to understand the reason why the boundary term appears. Previously, we wrote the Lagrangian for gravity as: $\mathscr{L}_{\text{grav}}(g, \partial g, \partial^2 g)$. The origin of the problem is in the second derivatives of the metric, since they give rise to terms containing the variation of the derivative of the metric once we integrate by parts. Indeed, one can write

$$\sqrt{-g}R = \sqrt{-g}\, Q_a{}^{bcd} R^a{}_{bcd} \tag{4.30}$$

where $Q_a{}^{bcd} = \frac{1}{2}(\delta_a{}^c g^{bd} - \delta_a{}^d g^{bc})$. Writing the Riemann tensor explicitly we get [4, p. 242]

$$\sqrt{-g}R = 2\sqrt{-g} Q_a{}^{bcd} \Gamma^a{}_{dk} \Gamma^k{}_{bc} + 2\partial_c[\sqrt{-g} Q_a{}^{bcd} \Gamma^a{}_{bd}] \equiv \sqrt{-g}\mathscr{L}_{\text{quadratic}} + \mathscr{L}_{\text{surface}}. \tag{4.31}$$

The $\mathscr{L}_{\text{quadratic}}$ part contains terms like gg or $(\partial g)^2$, and from there we can recover Einstein's equations by performing the metric variation, this is sometimes called the Schrödinger action; the surface part $\mathscr{L}_{\text{surface}} = \partial_c\left[\sqrt{-g}V^c\right]$, instead, has terms containing second derivatives of the metric and it is related to the boundary term we have seen before. One could think that this term is exactly equal to the GHY term, but this is true only if the chosen boundary $\partial\mathcal{M}$ has some specific coordinate σ^* that remains constant on it. Therefore, on this boundary one has to have $x^{\sigma^*} = \text{const}$ and $g^{\mu\sigma^*} = 0$.

Remarkably, it can be noticed [4, p. 655] that $\mathscr{L}_{\text{bulk}} \equiv \sqrt{-g}\mathscr{L}_{\text{quadratic}}$ and $\mathscr{L}_{\text{surface}}$ are not independent in GR but rather related in a simple way as (in $D > 2$ dimensions)

$$\mathscr{L}_{\text{surface}} = -\frac{1}{[(D/2) - 1]}\partial_i \left(g_{ab} \frac{\partial \mathscr{L}_{\text{bulk}}}{\partial(\partial_i g_{ab})} \right). \tag{4.32}$$

This "holographic" relation is typical of GR and of a class of theories in higher dimension still endowed with secondo order field equations (the so called Lanczos-Lovelock theories) which we shall meet in the next chapter.

4.1.3 Palatini variation

Palatini variation is another way to avoid the problems given by the boundary term in the gravitational action and to derive the Einstein's equations. The core of the Palatini variation is to not assume a priori that the connection is provided by the Christoffel symbols. In fact, problems started to arise with the boundary term after equation (4.17), when we assumed that the connection is provided by the Christoffel symbols. The only assumptions we do in this case are:

1. the connection is symmetric (no torsion), $\Gamma^\rho_{[\mu\nu]} = 0$;

2. matter couples to the metric and not to the connection, $S_{\text{m}} = S_{\text{m}}(g, \phi^i, \nabla\phi^i)$.

The Ricci tensor will no longer depend on the metric but only on the connection: $R_{\mu\nu} = R_{\mu\nu}(\Gamma)$. Thus, when we do the total variation of the gravitational action we get:

$$\delta S_{\text{grav}} = \frac{\delta S_{\text{grav}}}{\delta g^{\mu\nu}}\delta g^{\mu\nu} + \frac{\delta S_{\text{grav}}}{\delta\Gamma}\delta\Gamma. \tag{4.33}$$

From the first term, together with the variation of the matter action, we can directly rederive the Einstein's equations. The second term will lead us to the equation of motion for the connection. In particular, we have

$$\frac{\delta S}{\delta\Gamma} = \int \mathrm{d}^4x\sqrt{|g|}g^{\mu\nu}\frac{\delta R_{\mu\nu}}{\delta\Gamma} = \int \mathrm{d}^4x\sqrt{|g|}g^{\mu\nu}\left[\overset{\Gamma}{\nabla}_\lambda\left(\delta\Gamma^\lambda{}_{\mu\nu}\right) - \overset{\Gamma}{\nabla}_\nu\left(\delta\Gamma^\lambda{}_{\lambda\mu}\right)\right], \tag{4.34}$$

where we have used equation (4.17). The symbol $\overset{\Gamma}{\nabla}$ indicates that now the covariant derivatives are made from the connection, which are no more functions of the metric (we are not in a metric compatible theory). Thus we cannot safely bring the term $\sqrt{|g|}g^{\mu\nu}$ inside the covariant derivative; we have to use the Leibniz's rule to do it. Specifically, we get

$$\frac{\delta S}{\delta \Gamma} = \int \mathrm{d}^4x \left[\overset{\Gamma}{\nabla}_\lambda (\sqrt{|g|}g^{\mu\nu}\delta\Gamma^\lambda{}_{\mu\nu}) - \overset{\Gamma}{\nabla}_\lambda (\sqrt{|g|}g^{\mu\nu})\delta\Gamma^\lambda{}_{\mu\nu} - \overset{\Gamma}{\nabla}_\nu (\sqrt{|g|}g^{\mu\nu}\delta\Gamma^\lambda{}_{\lambda\mu}) + \overset{\Gamma}{\nabla}_\nu (\sqrt{|g|}g^{\mu\nu})\delta\Gamma^\lambda{}_{\lambda\mu} \right]. \tag{4.35}$$

The first and third terms inside the square brackets are the divergence of a vector therefore, for Stokes' theorem, we can take the integral on the boundary of those terms, multiplied by a directional vector. Fortunately, in this case Γ is an independent field, and it can be kept fixed at the boundary; hence the two terms vanish. What remains is

$$\frac{\delta S}{\delta \Gamma} = \int \mathrm{d}^4x \left[\delta_\lambda{}^\sigma \overset{\Gamma}{\nabla}_\nu (\sqrt{|g|}g^{\mu\nu}) - \overset{\Gamma}{\nabla}_\lambda (\sqrt{|g|}g^{\mu\sigma}) \right] \delta\Gamma^\lambda{}_{\mu\sigma}. \tag{4.36}$$

Since the connection is symmetric, also the $\delta\Gamma^\lambda{}_{\mu\sigma}$ are symmetric in the exchange of μ and σ; therefore, it is sufficient to set to zero the symmetric part of the expression in the square brackets (in μ and σ) in order to extremize the action. Through this requirement we can recover the equations of motion for the connections: $\overset{\Gamma}{\nabla} \sqrt{|g|} = 0$ and $\overset{\Gamma}{\nabla} g^{\mu\nu} = 0$, so the connection is metric compatible.
(**Exercise**: show this. Hint: consider first the trace and then the symmetric traceless part of the above term.

This is an elegant way to recover Einstein's equations: *a priori* we do not consider a metric compatible theory, and we vary with respect to the metric and the connections as independent fields; the fact that the connection is exactly provided by the Christoffel symbols of the metric is given by the dynamics. This is not general; in 4 dimensions it happens only for GR. The fact that the Palatini variation gives the same answer as the metric variation is a property only of theories with two degrees of freedom: those associated a massless spin 2 particle, the graviton. If we have a theory with extra degrees of freedom, this can always be made explicit by adoption a representation of the latter with extra dynamical fields which we shall then need to set them constant at the boundary.

Before we move one, let us comment on the step that led us to (4.36): in equation (4.35) we set to zero the first and third terms applying the Stokes theorem even if we had no more $\sqrt{|g|}$ in the volume element. However, in those terms we are performing the covariant derivative of a *tensor density*, and not of a tensor, since $\sqrt{|g|}$ is inside the derivative, and this turns out to be equivalent to apply a simple partial derivative. Indeed, when we perform a covariant derivative of a tensor density we have something like:

$$\nabla_\rho(A^{\cdots}{}_{\dots}) = \partial_\rho(A^{\cdots}{}_{\dots}) + (\text{standard tensor terms}) - A^{\cdots}{}_{\dots}\Gamma^\sigma{}_{\sigma\rho}, \tag{4.37}$$

where we have left out the indices of A and the standard terms are the usual Christoffel terms (see e.g. [19, p. 69] for an example on the metric). Therefore, in our case we have

$$\nabla_\rho V^\rho = \partial_\rho V^\rho + \Gamma^\rho{}_{\rho\lambda}V^\lambda - \Gamma^\sigma{}_{\sigma\rho}V^\rho = \partial_\rho V^\rho, \tag{4.38}$$

the covariant derivative becomes the partial derivative, the Stokes theorem in curved space becomes the usual one, and we can apply it in a integral with just d^4x without any problem.

5 Alternative Theories of Gravity

In this Chapter, we will briefly review some of the most relevant alternative theories of gravity. While their investigation started rather early (e.g. Brans–Dicke theory is dating back to 1961), it is undoubted that the renewed in extension/alternatives to GR was strongly propelled by the need to explain in recent times cosmological observations, and in particular Dark Energy. Bear in mind that this Chapter is not exhaustive nor excessively detailed, its purpose is just to give the reader an overview, a glimpse of how it is possible to formulate gravity theories different from General Relativity.

5.1 Holding SEP: GR in $D \neq 4$

If one wanted to create alternative theories of gravity, the first idea that could come to mind is to change the spacetime dimension $D = d + 1$ (where d represents the space dimension). Let us give a quick excursus on the main ideas and issues of such theories. We can divide them in two classes: $D < 4$ and $D > 4$.

5.1.1 $D < 4$

In the case of $D = 2$ we have that the Einstein tensor identically vanishes: $G_{\mu\nu} = 0$ in GR. The theory trivialises: the Riemann tensor has only one independent component (remember that such number is $D^2(D^2 - 1)/12$ in D dimensions) which is indeed the Ricci scalar. Furthermore the integral of the Ricci scalar is nothing else than a topological invariant (the Euler characteristic of the manifold, more later). In such theories it is usually introduced a dilaton scalar field and the Lagrangian is written as: $\mathscr{L} = \phi R + X(\phi) + U(\phi)$, where $X(\phi)$ is the kinetic term of the scalar field and $U(\phi)$ the potential. Also one can still derive a non-trivial 2D gravity action by the so called "dimensional reduction": e.g. in spherical symmetry one could integrate out the angular dependence so to get an effective 2D Lagrangian.

More interesting is the case of $D = 3$. In fact, we can write the action: $S_{\text{grav}} = \frac{1}{16\pi G} \int d^3x \sqrt{-g} R$ and we arrive to the Einstein equations: $R_{ij} - \frac{1}{2} g_{ij} R = 8\pi G T_{ij}$. However in three dimensions the Weyl tensor vanishes, so the curvature is completely encoded in the Ricci tensor and you cannot have gravitational waves. We cannot have structure formation either, unless we introduce a negative cosmological constant, $\Lambda = -\frac{1}{l^2}$.

The BTZ black hole

A notable example of solution of the Einstein equations in 3 spacetime dimensions is the one of a rotating black hole embedded in an anti-de Sitter background, the so called BTZ (Bañados–Teitelboim–

Zanelli) black hole

$$ds^2 = -F(r)dt^2 + \frac{dr^2}{F(r)} + r^2(d\varphi - \Omega dt)^2, \tag{5.1}$$

where $\Omega(r) = -\frac{J}{2r^2}$ is the angular velocity and $F(r) = -M + \frac{r^2}{l^2} + \frac{J}{4r^2}$ (with M and J being the mass and the angular momentum of the black hole with $|J| \leq M$). In this case we have two horizons,

$$r_\pm = \frac{Ml^2}{2}\left[1 \pm \left(1 - \left(1 - \frac{J}{Ml}\right)^2\right)\right]^{1/2}. \tag{5.2}$$

Since the BTZ solution is a vacuum solution with negative cosmological constant it is not surprising that is locally, in the limit $J \to 0$, like an AdS one with the extra constraint $\phi = \phi + 2\pi$. I.e. the BTZ BH is locally AdS but with a different global topology.

5.1.2 $D > 4$

Let us move to the more interesting case, $D > 4$. We can have two kind of theories, compactified extra dimensions and large extra dimensions.

In the first case, we imagine that there are compactified extra dimensions with a typical radius R. Since the energy scales as $E \propto \frac{1}{R}$, having a compactified dimension means that, if the radius of the extra dimension is small enough, we will not be able to probe it, since we would be probing it at too low energies, and we would never realize that our spacetime has e.g. five dimension and topology $\mathcal{R} \times \Sigma^3 \times S^1$ However we are doing a theory of gravity, and this means that this structure cannot be a background structure, but rather should have dynamics. So we have to promote the radius of this extra dimension to a field $R(x)$, and its dynamics is hard to study and implies to deal with technical issues.

An alternative point of view stems from the following considerations. In $D = 4$ spacetime dimensions we can write the Newtonian potential as

$$V^{(4)} = \frac{G_N M}{r} = \frac{M}{M_P^{(4)2}}\frac{1}{r}, \tag{5.3}$$

where in the last step $M_{\mathrm{P}}^{(4)}$ is the Planck mass in 4 dimensions and we have written G_N as the square of the inverse of the Planck mass. We can try to generalize it in D dimensions. From a simple power-counting of the usual Einstein–Hilbert action, we can find that the dimensionality of the Newton constant in D spacetime dimensions in mass unit is $\left[G_N^{(D)}\right] = 2 - D$ such that $G_N^{(D)} \propto 1/M_{\mathrm{P}}^{D-2}$; also, the Newton potential in D dimension scales as r^{3-D}, therefore,

$$V^{(D)} = \frac{G_N^{(D)} M}{r^{D-3}} \sim \frac{M}{\left(M_P^{(D)}\right)^{D-2}}\frac{1}{r^{D-3}} = \frac{M}{\left(M_P^{(D)}\right)^{D-2}}\frac{1}{R^{D-4}}\frac{1}{r}, \tag{5.4}$$

where in the last step we introduced a characteristic scale of the extra dimensions R, and we rewrote the potential in order to see explicitly the $1/r$ dependence experienced in the $D = 4$ case. Basically we are trying to interpret the Newton potential (Equation 5.3) that we see in four dimensions as described in Equation 5.4. So, equating the two potentials, we get:

$$\left(M_P^{(D)}\right)^2 = \frac{\left(M_P^{(4)}\right)^2}{\left(M_P^{(D)} R\right)^{D-4}}. \tag{5.5}$$

It is a well known problem in theoretical physics that the coupling associated to gravity is very different with respect to the coupling of the other Standard Model interactions (hierarchy problem).

However, the above reasoning shows that with increasing number of dimensions D, with a characteristic length which is large enough, we can reduce the value of the Planck mass, in order to make it comparable to the couplings of other forces. One can than conjecture that our Universe is actually a 4-dimensional brane embedded in a higher-dimensional manifold, where the extra dimensions are accessible only to gravity. Then, the standard model forces are stuck on the brane, while gravity is the only interaction that can "communicate" with the other dimensions and has a characteristic scale $M_P^(D)$ similar to those of the standard model.

However, all these theories are partially ruled out by experiments at LHC that show no signals predicted by them (micro black holes from TeV collisions). Also the observation of GW170817 with its electromagnetic counterpart, tells us that $d_L^{(\mathrm{GW})} \simeq d_L^{(\mathrm{EM})}$ pointing towards $D = 4$ [20].

5.2 Holding SEP: Lanczos-Lovelock theories

We saw that the SEP can be seen as a combination of the GWEP, LLI and LPI where the latter is applied also to gravitational experiments. It can be shown [16] that GWEP substantially requires that in vacuum a theory does not admit extra dynamical degrees of freed apart from the graviton. In four dimension GR is the only theory (with spin-2 gravitons) satisfying the SEP. [1]

Nordström Gravity: We specified that GR is the only theory with spin-2 gravitons satisfying SEP in four dimensions because indeed one can have a scalar theory of gravity which satisfies the same principle. In 1913 such a theory was proposed by Gunnar Nordström as a way to have a relativistic generalisation of the Poisson equation. After several iterations with Einstein and the crucial contribution of a brilliant graduate student of Hendrik Lorentz, Adriaan Fokker, the theory settled in its modern form $R = 24\pi G_N T_{matter}$ (the so called Einstein–Fokker equation) and $C_{abcd} = 0$ where R is the Ricci scalar, T_{matter} is the trace of the SET and C_{abcd} is the Weyl tensor. The theory is a scalar theory of gravitation because the condition on the Weyl implies that all the solutions are conformally flat metrics $g_{\mu\nu} = \phi^2(x)\eta_{\mu\nu}$. The theory was superseded by GR which among other things crucially predicts correct factor of light deflection by gravitating bodies (angle which is trivially zero in Nortström gravity due to the constraint of conformally flat solutions). See e.g. [25] for more details.

Lanczos-Lovelock theories are alternative gravity theories that have an equation of motion of the second order in the metric. In order to have this feature, we expect the Lagrangian to be at most linear in the second derivatives of the metric, which is the same feature we had in the Lagrangian of the previous Chapter. Hence the Lagrangian should contain the curvature tensor at most linearly. However, in such theories we can construct Lagrangians with even higher powers of the curvature tensor, but in a special combination that allows us to eliminate all higher order terms in the field equations. Let us see how to generalize the Lagrangian to include terms with higher power of the curvature. We saw that we can rewrite the Einstein–Hilbert Lagrangian as

$$\mathscr{L}_{\mathrm{EH}} = R = Q_\alpha{}^{\beta\gamma\delta} R^\alpha{}_{\beta\gamma\delta}, \tag{5.6}$$

where we have defined

$$Q_\alpha{}^{\beta\gamma\delta} \equiv \frac{1}{2}(\delta^\gamma_\alpha g^{\beta\delta} - \delta^\delta_\alpha g^{\beta\gamma}). \tag{5.7}$$

This tensor is constructed only from the metric, has the same symmetries of the Riemann tensor and it satisfies the property $\nabla^\mu Q_\mu^{\beta\gamma\delta} = 0$. It can also be proven that this is the only tensor constructed from the metric alone that has such properties.

[1] One can argue that also unimodular gravity [21] — which is like GR with the additional requirement $\sqrt{-g} \equiv 1$ – satisfy the SEP. However, this theory can be rather seen as a generalisation of general relativity, in which the cosmological constant appears as a single additional variable (not a field), but is then seen to be a constant of the motion. Indeed, if one tries to reconstruct GR as a nonlinear field theory of gravitons over Minkowski spacetime (see e.g. [22] but also [23]), one can get both GR and Unimodular gravity making different assumptions of the way diffeomorphism invariance is implemented, see e.g. [24].

Actually let us write the above expression as

$$\mathscr{L}_{\rm EH} = \delta^{\gamma\delta}_{\alpha\beta} R^{\alpha\beta}_{\gamma\delta}\,, \tag{5.8}$$

where the object

$$\delta^{\gamma\delta}_{\alpha\beta} \equiv \frac{1}{2}(\delta^{\gamma}_{\alpha}\delta^{\delta}_{\beta} - \delta^{\delta}_{\alpha}\delta^{\gamma}_{\beta}) \tag{5.9}$$

is known as the alternating determinant, or alternating tensor. In general, we have

$$\delta^{ab...j}_{mq...n} \equiv \frac{1}{n!}\delta^{m}_{[a}\delta^{q}_{b}...\delta^{n}_{j]}, \tag{5.10}$$

that is the determinant of a matrix of deltas, constructed as follows,

$$\det\begin{pmatrix} \delta^m_a & \dots & \delta^n_a \\ \dots & \dots & \dots \\ \delta^m_j & \dots & \delta^n_j \end{pmatrix}. \tag{5.11}$$

This procedure to write the Einstein–Hilbert lagrangian by the contraction of the Riemann tensor with the divergence-free tensor Q leads to second order field equations [4]. To go beyond GR we would like to maintain the same construction, but generalizing Q. We will allow it to be a linear function of the curvature tensor. Hence, now Q will have the same symmetries of the Riemann tensor, it will be divergence-free, and it will be a function of $g^{\alpha\beta}$ and $R^{\alpha}_{\beta\gamma\delta}$. Q can be built as:

$$Q^{\alpha\beta\gamma\delta} = R^{\alpha\beta\gamma\delta} - G^{\alpha\gamma}g^{\beta\delta} + G^{\beta\gamma}g^{\alpha\delta} + R^{\alpha\delta}g^{\beta\gamma} - R^{\beta\delta}g^{\alpha\gamma}; \tag{5.12}$$

Such a form has again the right symmetries. This leads to a new Lagrangian, given by the contraction between Q and the Riemann tensor, called the Gauss–Bonnet Lagrangian:

$$\mathscr{L}_{\rm GB} = QR = \frac{1}{2}[R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta} - 4R^{\alpha\beta}R_{\alpha\beta} + R^2] \tag{5.13}$$

which will cancel on the terms of order higher than the second in the equations of motion. In 4 dimensions, the variation of this quantity is a pure divergence term and hence does not contribute to the equations of motion. Furthermore, its integral on the four volume is equal to the **Euler characteristic** of the manifold, which is a number characterising the topology of the considered manifold: a topological invariant. However, in a higher number of dimensions, this is not equal to the Euler characteristics anymore, and therefore it becomes relevant.

As it was done with the Einstein–Hilbert Lagrangian, also this term can be rewritten in a more compact form:

$$\mathscr{L}_{\rm EH} = \delta^{1\,3}_{2\,4}R^{2\,4}_{1\,3}, \tag{5.14a}$$

$$\mathscr{L}_{\rm GB} = \delta^{1\,3\,5\,7}_{2\,4\,6\,8}R^{2\,4}_{1\,3}R^{6\,8}_{5\,7}. \tag{5.14b}$$

Now we can easily extend it to higher orders in the Lagrangian; for example, at third order we have

$$\mathscr{L}^{(3)} = \delta^{1\,3\,5\,7\,9\,11}_{2\,4\,6\,8\,10\,12}R^{2\,4}_{1\,3}R^{6\,8}_{5\,7}R^{10\,12}_{9\,11}, \tag{5.15}$$

and in general at the m-th order we have

$$\mathscr{L}^{(m)} = \delta^{1\,3\,5\,7...2k-1}_{2\,4\,6\,8...2k}R^{2\,4}_{1\,3}R^{6\,8}_{5\,7}\dots R^{2k-2\;2k}_{2k-3\;2k-1}, \tag{5.16}$$

where $k = 2m$.

How we see explicitly that these Lagrangian are all leading to second order field equations? First of all note that the splitting we showed for the EH action in a bulk action (quadratic in the connections) and a surface part did not make explicit use of the form of Q^{abcd}, just of its properties.

$$\sqrt{-g}Q_a{}^{bcd}R^a_{bcd} = 2\sqrt{-g}Q_a{}^{bcd}\Gamma^a{}_{dk}\Gamma^k{}_{bc} + 2\partial_c[\sqrt{-g}Q_a{}^{bcd}\Gamma^a{}_{bd}] \equiv \sqrt{-g}\mathscr{L}_{\rm quadratic} + \mathscr{L}_{\rm surface}. \tag{5.17}$$

Alternative one can see that each $\mathscr{L}^{(m)}$ is by construction a homogeneous function of the Riemann tensor of degree m.

$$L^{(m)} = \frac{1}{m}\left(\frac{\partial L^{(m)}}{\partial R^a{}_{bcd}}\right) R^a{}_{bcd} \equiv \frac{1}{m} P_a{}^{bcd} R^a{}_{bcd} \tag{5.18}$$

where evidently $P^{abcd} = mQ^{abcd}$ and hence, $\nabla_a P^{abcd} = 0 = \nabla_a Q^{abcd} = 0$.

The above form of the Lagrangian allows to show [4] that the EOM take in this case a simple form

$$\left(P_b{}^{cde} R_{acde} - \frac{1}{2} g_{ab} L^{(m)}\right) = \frac{1}{2} T_{ab} \tag{5.19}$$

which does not involve more than the second derivatives of the metric (albeit for $m \geq 2$ will be non linear in the latters).

So in the end, the most general Lanczos–Lovelock (LL) Lagrangian is:

$$\mathscr{L}_{\text{LL}} = \sum_{m=1}^{k} \mathscr{L}^{(m)}. \tag{5.20}$$

Because of the antisymmetry of the deltas, we have that for $2m > D$ the term $\mathscr{L}^{(m)}$ vanishes, so we have a finite sum of different terms depending on the dimensionality of our manifold. If $2m = D$ the term does not vanish, but it is a topological invariant (the Euler characteristic of that D-dimensional manifold) and does not affect the equations of motion, just like the Gauss–Bonnet term in $D = 4$. Only terms for which $D \geq 2m + 1$ introduce a non trivial dynamics.

Finally, one might wonder how in $D = 4$ dimension one could theoretically prefer GR over say its LL counterpart given that SEP (at least in its formulation has GWEP+LLI+LPI) is satisfied by both. Interestingly it is still true that these theories differs in fundamental aspects such as gravitational waves propagation (which in LL generically does not happen on the same background geometry and seems connected to a problem of well-posedness of the Cauchy problem [26, 27]) or critical collapse [28]. The fact that Lanczos–Lovelock theories comply with the GWEP albeit physically different from GR, seems to indicate that there is potential room for improvement in the selection rules given by the equivalence principles, perhaps by integrating them with some restrictions on the kind of self-interaction of the gravitational degrees of freedom.

5.2.1 More about the Euler characteristic

The Euler characteristic is a topological invariant originally introduced for describing the topology of polyhedra surfaces. It can be calculated according to the formula $\chi = V - E + F$ where V, E, and F are respectively the numbers of vertices (corners), edges and faces in the given polyhedron. Remarkably, it is easy to see that any convex polyhedron's surface has Euler characteristic $\chi = 2$, which indeed is also the Euler characteristic of a sphere.

More generally still, for any topological space, we can define the n-th Betti number b_n as the rank of the n-th singular homology group. The Euler characteristic can then be defined as the alternating sum $\chi = b_0 - b_1 + b_2 - b_3 + \cdots$. This quantity is well-defined if the Betti numbers are all finite and if they are zero beyond a certain index n_*.[2]

The Euler characteristic of a closed orientable surface can also be calculated from its genus g (the number of tori in a connected sum decomposition of the surface; intuitively, the number of "handles") as $\chi = 2 - 2g$ as well as from the Gauss–Bonnet integral as shown before.

[2]Homology itself was developed as a way to analyse and classify manifolds according to their cycles: closed loops (or more generally submanifolds) that can be drawn on a given n dimensional manifold but not continuously deformed into each other. Informally, the k-th Betti number refers to the number of k-dimensional holes on a topological surface. The first few Betti numbers have the following definitions for 0-dimensional, 1-dimensional, and 2-dimensional simplicial complexes: b_0 is the number of connected components, b_1 is the number of one-dimensional or "circular" holes, b_2 is the number of two-dimensional "voids" or "cavities". Thus, for example, a torus has one connected surface component so $b_0 = 1$, two "circular" holes (one equatorial and one meridional) so $b_1 = 2$, and a single cavity enclosed within the surface so $b_2 = 1$. Which give an Euler characteristic equal to zero.

Figure 5.1: Illustration of a $g = 3$ surface, hence with $\chi = -4$. (Image from en.wikipedia.org)

A funny application of these concepts is that of soccer balls. These are spheres covered by stitching together pentagons and hexagons. If we call P the number of Pentagons and H that of hexagons then

Figure 5.2: Soccer ball and its pentagon-hexagon traditional tesselation. (Image from en.wikipedia.org)

the number of faces is $F = P + H$, the number of vertexes is $V = (5P + 6H)/3$ (because at each vertex three patches meet: two hexagons and one pentagon), and finally $E = (5P + 6H)/2$ edges (because each edge is shared by two patches). This implies that $V - E + F = P/6$ but we also know that $\chi_{sphere} = 2 = V - E + F$ so we can deduce that any soccer ball, no matter how big it will be will always have 12 pentagonal patches!

5.3 Relaxing SEP: $F(R)$ Theories as a first example of higher curvature gravity

If one decides to move away from SEP and fall back on requiring just the EEP then a whole set of metric theories of gravity is explorable. A first obvious generalisation is to consider GR as the low energy EFT action of a more complicated theory with a hierarchy of higher curvature terms. A simple step in this direction would be to take the Gauss–Bonnet action with arbitrary coefficients (as all its term have the same mass dimension). However, this is generally a dangerous path: the fourth order equations of motions show that there are in general eight dynamical degrees of freedom that in a suitable representation (choice of the field variables) can be associated to a massless graviton (carrying two degrees of freedom), a scalar (carrying one degree of freedom) and a massive spin-2 particle (carrying five degree of freedom) where the latter is a ghost and hence is associated to an instability.

While it is still possible to construct sane higher curvature theories involving terms in Riemann and Ricci a very convenient framework to start with is to revert to theories which generalise GR by changing the action from $R \to F(R)$, where usually $F(R)$ is taken to be an analytical function. The gravitational action then becomes

$$S_{\text{grav}} = \frac{1}{16\pi G}\left[\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}F(R) - 2\sigma\int_{\partial\mathcal{M}} \mathrm{d}^3x\sqrt{-h}F'(R)K\right] \tag{5.21}$$

where $F'(R)$ is the derivative of $F(R)$ with respect to the Ricci scalar R. The second term, where K is the trace of the extrinsic curvature of the manifold, is the analogous of the GHY term and we need this term in order to cancel the part $F'(R)g^{\mu\nu}\delta R_{\mu\nu}$ that one obtains when we perform the variation of the first term.

However, from the variation of the second term we get two contributions: a term like $F'(R)\delta K$ which cancels the term involving the variation of the Ricci shown above, and a term like $K\delta F'(R)$, that will give the variation of the Ricci scalar. Hence, even if we started from a metric theory of gravity without any new field, we would end up with a boundary term where $F'(R)$ seems to be an independent field, which needs to be fixed at the boundary.

Indeed if we write the equations of motion obtainable in this way we get

$$F'(R)R_{\mu\nu} - \frac{1}{2}F(R)g_{\mu\nu} - \nabla_\mu\nabla_\nu F'(R) + g_{\mu\nu}\Box F'(R) = 8\pi G T_{\mu\nu}. \tag{5.22}$$

As a sanity check, please note that if we put $F(R) = R$, the previous relation reduces to the Einstein equations. These are fourth order partial differential equations. The first part is similar to the Einstein tensor, but the other two terms have no analogue in General Relativity. We can see that a $\Box F'(R)$ appears, with a box operator in general implying a field propagation. We can also take the trace of this expression and get

$$F'(R)R - 2F(R) + 3\Box F'(R) = 8\pi G T. \tag{5.23}$$

This is a differential equation, so, while in General Relativity R and T were algebraically related, here we have a bigger set of solutions. Let us consider now a maximally symmetric solution, $R = \text{const}$. If we consider the case $T = 0$, in General Relativity (in absence of a cosmological constant) we obtain uniquely $R = 0$ (Minkowski spacetime); however, in this case, you can have also other solutions to the equation (5.23), as $R = C \neq 0$, with $R_{\mu\nu} = \frac{C}{4}g_{\mu\nu}$. This is interesting in a context related to dark energy: as a matter of fact, depending on the choice of $F'(R)$, we can have that the solution is de Sitter or anti-de Sitter without introducing a cosmological constant which would violate the SEC, since it would be readily provided by the theory without the need of exotic matter.

5.3.1 Palatini variation

We saw that in General Relativity, Palatini variation of the Einstein–Hilbert action gives the Einstein equations, and tells us that the connection is provided by the Christoffel symbols just from the dynamics, without assuming it *a priori*. We now consider Palatini variation for $F(R)$ theories, which means that we are constructing a metric-affine $F(R)$ theory, where the connection, again, is not fixed to be the Christoffel symbols, but has its own dynamics. We can write the Palatini action,

$$S_{\text{Pal}} = \frac{1}{16\pi G}\int \mathrm{d}^4x\sqrt{-g}F(\mathcal{R}) + S_{\text{m}}(g, \psi_i) \tag{5.24}$$

where with $\mathcal{R}$ we mean the Ricci scalar obtained considering symmetric connections that are not *a priori* the Christoffel symbols. If we vary with respect to the metric and with respect to the connection, we respectively obtain the field equations

$$F'(\mathcal{R})\mathcal{R}_{\mu\nu} - \frac{1}{2}g_{\mu\nu}F(\mathcal{R}) = 8\pi G T_{\mu\nu}, \tag{5.25a}$$

$$\overset{\Gamma}{\nabla}\left(\sqrt{-g}F'(\mathcal{R})g^{\mu\nu}\right) = 0, \tag{5.25b}$$

where, again, the covariant derivative is made from the connection. In analogy with what we did in Palatini variation with the Einstein–Hilbert action, we see that the equation (5.25b) is forcing the connection to be the Christoffel symbols not of the metric $g_{\mu\nu}$, but of a conformally related metric $h_{\mu\nu}$ defined as

$$\sqrt{-h}h_{\mu\nu} = \sqrt{-g}F'(\mathcal{R})g_{\mu\nu}. \tag{5.26}$$

Is it possible to eliminate the dependence of this new metric h and write the equations only in terms of g? In principle yes. Let us write

$$\Gamma^\lambda{}_{\mu\nu} = \Gamma^\lambda{}_{\mu\nu}(h, \partial h) = \Gamma^\lambda{}_{\mu\nu}(g, \partial g, F'(\mathcal{R}), \partial F'(\mathcal{R})); \tag{5.27}$$

the connection can thus be defined in terms of g, $F'(R)$ and their derivatives, and we have the relation

$$\mathcal{R} = \mathcal{R}(\Gamma(h)) = \mathcal{R}(\Gamma(g, F')). \tag{5.28}$$

If we now take the trace of the field equation, (5.25a), we obtain

$$F'(\mathcal{R})\mathcal{R} - 2F(\mathcal{R}) = 8\pi G T, \tag{5.29}$$

that tells us that $\mathcal{R}$ is *algebraically* related to T (in analogy with the Einstein equations), hence $\mathcal{R} = \mathcal{R}(T, F')$. But in F' there is only the dependence from g and F' again, as we can see from (5.28). Therefore we can in principle eliminate the independent connections from the field equations and express them only in terms of the metric and the matter fields, once the shape of the function F is known.

It can also be proven that the following relations between the Ricci scalars in the two different conformally related metrics holds,

$$\mathcal{R} = R + \frac{3}{2F'(\mathcal{R})^2}(\nabla_\mu F'(\mathcal{R}))(\nabla^\mu F'(\mathcal{R})) + \frac{3}{F'(\mathcal{R})}\Box F'(\mathcal{R}), \tag{5.30}$$

which leads us to write the dynamical equation in the form:

$$\begin{aligned} G_{\mu\nu} = \frac{8\pi G}{F'(\mathcal{R})}T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\left(\mathcal{R} - \frac{F(\mathcal{R})}{F'(\mathcal{R})}\right) + \frac{1}{F'(\mathcal{R})}(\nabla_\mu\nabla_\nu - g_{\mu\nu}\Box)F'(\mathcal{R}) - \\ - \frac{3}{2}\frac{1}{F'(\mathcal{R})^2}\left[(\nabla_\mu F'(\mathcal{R}))(\nabla_\nu F'(\mathcal{R})) - \frac{1}{2}g_{\mu\nu}(\nabla F'(\mathcal{R}))^2\right] \end{aligned} \tag{5.31}$$

Knowing $F(\mathcal{R})$ and the root of (5.29) $\mathcal{R} = \mathcal{R}(T)$ we have completely eliminated the dependence on Γ and everything is determined: all the terms are given as a function of g and F'. These are similar to the Einstein equations, but with a big source term. Again, if $F(\mathcal{R}) = \mathcal{R}$ we recover General Relativity.

However the source term contains derivatives of the stress energy tensor. This can lead to a problem: if we have, for example, a star with a sharp boundary (a crust), the source term could diverge; so the Palatini variation for $F(R)$ theories does not work for discontinuous distributions of matter.

5.4 Relaxing SEP: Generalized Brans–Dicke theories

In Brans–Dicke theories [29] we relax the SEP by promoting the Newton constant to be a field that varies in spacetime (the term "generalised" that we use in the title of this section is due to the fact that originally such theories did not include a potential term). In order to preserve background independence of the theory we shall of course need to give a dynamics to such field by introducing a kinetic term and a in general a potential as well. The action is of the form:

$$S_{\text{grav}} = \frac{1}{16\pi G_0}\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}\left[\phi R - \frac{\omega}{\phi}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - V(\phi)\right]; \tag{5.32}$$

as evident from the first term, G_0 is a book-keeping constant, such that Newton's constant is promoted to a field $G_N = \frac{G_0}{\phi}$ that is a function of the spacetime point. The second term is a kinetic term that gives the dynamics to the new field ϕ and the coupling depends on spacetime through ϕ; the third term is the potential.

The second term provides a long range interaction; using the Cassini probe, it has been proved that the value of ω can be well constrained, $\omega > 40000$. This in principle could be a good result because with an high ω we have that the value of the field tends to be constant and it freezes-out (the field becomes ultra-heavy and it is difficult to put it in motion); so for $\omega \to \infty$ we recover General Relativity and the Newton constant appears to be constant. However, we can still make the theory compatible with solar system tests of the GWEP if in the potential term we introduce an effective mass to the scalar field. This leads to a Yukawa potential term, so that an heavy field cannot effectively propagate: in this way, the problematic long range interaction is turned into short range, and therefore the tests like the Cassini probe ones cannot constrain efficiently Brans–Dicke theories.

5.4.1 $F(R)$ theories as generalized Brans–Dicke theories

The aim of this Subsection is to show that sometimes it is possible to map two theories of gravity one in the other, even if in principle they appear to be very different. This can be explicitly seen for $F(R)$ theories, which were proposed in 1970 and discovered to be equal to special generalized Brans–Dicke theories in 2006.

Metric $F(R)$ Let us consider again *metric* $F(R)$ theories (the connection is provided by the Christoffel symbols of the metric). We have seen above that $F'(R)$ is like an extra degree of freedom (its D'Alembert operator appears in the fields equations, Equation (5.23)). We use the *auxiliary field method* to make this explicit. Let us rewrite the main part of the gravitational action in another form,

$$S_{\text{grav}} = \frac{1}{16\pi G}\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}F(R) = \frac{1}{16\pi G}\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}[F(\chi) + F'(\chi)(R-\chi)], \tag{5.33}$$

where we have introduced the auxiliary field χ. If we perform the variation with respect to χ, we obtain $\chi = R$ if the condition $F''(\chi) \neq 0$ holds. Therefore, imposing this last condition, by performing the variation of the new action with respect to g, we obtain again the same equations of motion: dynamically speaking, we have changed nothing. Let us now introduce a scalar field $\phi \equiv F'(\chi)$, and define a potential $V(\phi) = \chi(\phi)\phi - F(\chi(\phi))$; the action becomes

$$S_{\text{grav}} = \frac{1}{16\pi G}\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}[\phi R - V(\phi)]. \tag{5.34}$$

This equation makes explicit an extra degree of freedom ϕ that first was encoded in the function F. Comparing (5.34) with (5.32), we see that metric $F(R)$ theories with $F''(R) \neq 0$ are generalized Brans–Dicke theories with $\omega = 0$ and the potential $V(\phi) = \chi(\phi)\phi - F(\chi(\phi))$.

Palatini $F(R)$ If we instead start from the Palatini action and apply the auxiliary field method, we obtain an action that can be written as

$$S_{\text{Pal}} = \frac{1}{16\pi G}\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}\left[\phi R + \frac{3}{2\phi}\nabla^\mu\phi\nabla_\mu\phi - V(\phi)\right] + S_{\text{m}}(g,\psi_i), \tag{5.35}$$

where R is the Ricci scalar associated to the metric, expressed as a function of $\mathcal{R}$ using Equation (5.30). This is a Brans–Dicke theory with $\omega = -\frac{3}{2}$, which can be proven to be singular. Indeed, in general, Brans–Dicke theories have the property that, by rescaling the metric by a conformal factor that is a function of ϕ, we can always get back to the Einstein gravity, with ϕ now coupled to the matter. For $\omega = -\frac{3}{2}$ this cannot be done. See also [30, 31].

Just to complete the picture of $F(R)$ theories we show a picture that illustrates the link between the different $F(R)$ theories and the fact that they are mappable in particular cases of Brans–Dicke theories.

5.5 Relaxing SEP: Scalar-Tensor Theories

Let us generalize the Brans–Dicke idea: we write down the most general scalar-tensor action compatible with the Einstein equivalence principle,

$$S_{\text{ST}} = \int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}\left[\frac{A(\phi)R}{16\pi G} - \frac{B(\phi)}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - V(\phi)\right] + S_{\text{m}}\left(e^{2\alpha(\phi)}g,\psi_i\right), \tag{5.36}$$

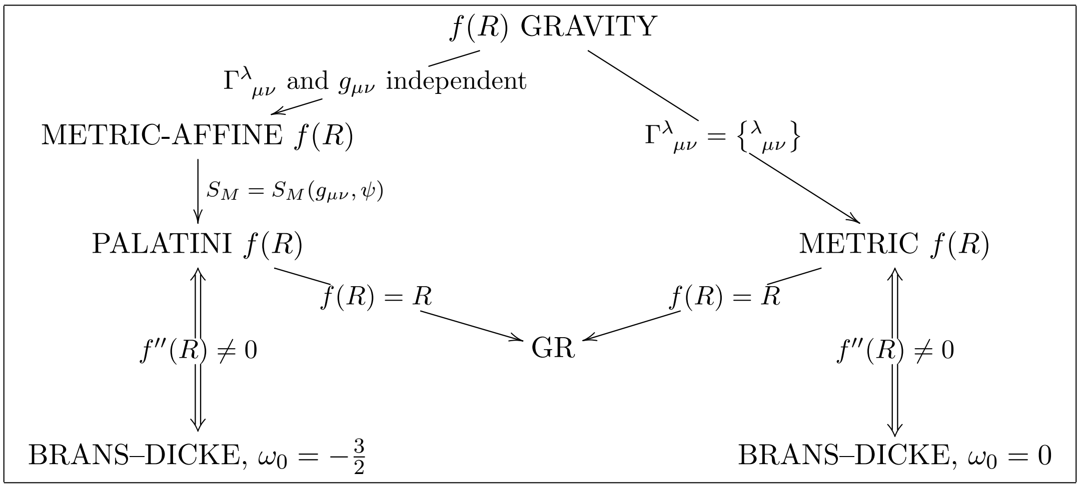


Table 5.1: Classification of $f(R)$ theories of gravity and equivalent Brans–Dicke theories (from [32]).

where we are allowing the scalar field to mediate the interaction between the metric and the matter fields (no minimal coupling). Using the symmetries of the action, we can represent such theories in two different but (classically) equivalent representations, called frames: the Einstein and the Jordan frame. This will show once again that even if two gravity theories looks different they could have the same physical content. As a matter of fact, we can always rescale the field ϕ without generating new terms. Moreover we can always perform a conformal transformation (the action is invariant under conformal symmetry) and rescale the metric of a function of ϕ (conformal symmetry). These two transformations allow us to eliminate 2 of the 4 unknown functions in the action (A, B, α, V).

5.5.1 Einstein frame

If we set $A = B = 1$ we get

$$S_{\rm ST} = \int_{\mathcal{M}} \mathrm{d}^4x \sqrt{-\tilde{g}} \left[\frac{\tilde{R}}{16\pi G} - \frac{1}{2}\tilde{g}^{\mu\nu}\tilde{\nabla}_\mu\phi\tilde{\nabla}_\nu\phi - V(\phi) \right] + S_{\rm m}\left(e^{2\alpha(\phi)}\tilde{g}, \psi_i\right) \tag{5.37}$$

This is the Einstein frame representation: $\tilde{g}$ is the Einstein metric, and in the action we have an Einstein–Hilbert part plus a standard kinetic and a potential term for the field ϕ. However, we do not recover General Relativity, since the coupling of the matter fields with gravity is mediated by the scalar field.[3]

5.5.2 Jordan frame

We can also set $\alpha = 0$, $A = \phi$. In this case, we get

$$S_{\rm ST} = \int_{\mathcal{M}} \mathrm{d}^4x \sqrt{-g} \left[\frac{\phi R}{16\pi G} - \frac{B(\phi)}{2} g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - V(\phi) \right] + S_{\rm m}(g, \psi_i). \tag{5.38}$$

The gravity part is more complicated since there is a ϕR coupling. The matter fields are coupled to the Jordan metric in a minimal way, so the matter follows the geodesics of the Jordan metric, not of the Einstein one. In fact, in the Einstein representation it is coupled to the Einstein metric through the scalar field; the matter is never really free.

[3] Of course, we still have the freedom to redefine our matter fields with a conformal transformation so that to reabsorb the $e^{2\alpha(\phi)}$. However our new fields, say $\tilde{\psi}_i = \Omega^{s_i}(\phi)\psi_i$ (where s_i its the i-field conformal weight) will be dependent on $\phi(x)$, $\tilde{\psi}_i = \tilde{\psi}_i(\phi)$. This is tantamount to say that one does trade this simplification with spacetime position dependent masses and coupling constants (which will still make the theory different from GR).

5.5.3 Horndeski theory

Now we can ask: is the action we wrote, depending on A, B, α, and V, the most general action that can give second order field equations? The answer to this question is actually no: there is a more general action, called Horndeski action [33].

Its general form is given by

$$\mathscr{L} = \sum_{i=2}^{5} \mathscr{L}_i(\phi, X), \tag{5.39}$$

where we introduced $X = -\partial^\mu\phi\partial_\mu\phi/2$ and the terms apearing in the sum are defined by:

$$\mathscr{L}_2 = K(\phi, X) \tag{5.40a}$$

$$\mathscr{L}_3 = -G_3(\phi, X)\Box\phi \tag{5.40b}$$

$$\mathscr{L}_4 = G_4(\phi, X)R + G_{4,X}\left[(\Box\phi)^2 - (\nabla_\mu\nabla_\nu\phi)(\nabla^\mu\nabla^\nu)\right] \tag{5.40c}$$

$$\mathscr{L}_5 = G_5(\phi, X)G_{\mu\nu}(\nabla^\mu\nabla^\nu\phi) - \frac{1}{6}G_{5,X}\left[(\Box\phi)^2 - 3(\Box\phi)(\nabla_\mu\nabla_\nu\phi)(\nabla^\mu\nabla^\nu\phi) + 2(\nabla^\mu\nabla_\alpha\phi)(\nabla^\alpha\nabla_\beta\phi)(\nabla^\beta\nabla_\mu\phi)\right] \tag{5.40d}$$

Here, K and the G_i's are functions of ϕ and the kinetic term X, and $G_{i,X} = \partial G_i/\partial X$.

Horndeski theory admits an Einstein frame under some restrictions and is achieved via a disformal transformation $\bar{g}_{\mu\nu} = A(\phi)g_{\mu\nu} + B(\phi)\nabla_\mu\phi\nabla_\nu\phi$.

5.5.4 DHOST Theories

So far we blindly required for the modified theories to have second-order equations of motion. This requirement is imposed for two reasons: first we want to fix the number of degrees-of-freedom of the theory (higher-order equations need more initial conditions) and second we want to avoid instabilities in the action. Indeed, in general higher-order equations imply the presence of additional degrees-of-freedom which are *ghost-unstable* (meaning with the wrong sign for their kinetic term). For example, a theorem guarantees that generically this sort of instabilities appears in the case of equations of motion with higher than second order derivatives in time, the so called *Ostrogradsky instability*.[4]

However, it was suggested that — despite the appearance of higher order time derivatives — it is still possible to avoid the Ostrogradsky ghost if the equations of motion can be recast into a second order form by field redefinitions. This idea was firstly applied to Horndeski showing that it is possible to add two more terms with one arbitrary function of the scalar field and its kinetic term entailing third order derivatives which however do not lead to an Ostrogadsky instability. These terms and the corresponding theory are called "beyond Horndeski".

However, it was realised in recent years that even more generic Lagrangians with higher-order equations can be stable and propagate the desired number of degrees-of-freedom, provided some degeneracy conditions are imposed on the theory. Such scalar-tensor theories go under the name of DHOST (Degenerate Higher Order Scalar-Tensor theories) and further generalise the Horndeski theory but still propagate one scalar and one massless graviton.

We can elucidate the degeneracy-mechanism behind DHOST with an example from classical mechanics (for more details see e.g. [35]). Consider the Lagrangian for the variables $\phi(t)$ and $h(t)$ defined as

$$\mathscr{L} = \frac{a}{2}\ddot{\phi}^2 + b\,\ddot{\phi}\dot{h} + \frac{c}{2}\dot{h}^2 + \frac{1}{2}\dot{\phi}^2 - V(\phi, h)\;, \tag{5.41}$$

[4]Note that Ostrogadsky theorem does not straightforwardly apply to non-local EOMs (here meaning theories with infinite higher derivatives): e.g. if one has a $f(\Box)\phi = 0$ equation of motion. (Note however, that no general theorem guarantees the viability of any non-local theory, see e.g. [34]). Also, we shall see in what follows that constrained/degenerate Lagrangians can evade the theorem, this is for example evidently the case for $F(R)$ theories which have only an extra scalar d.o.g. in addition to the graviton.

where a, b and c are generic coefficients and $V(\phi, h)$ is a generic potential. The equations of motion are

$$\begin{cases} a\ddddot{\phi} + b\dddot{h} - \ddot{\phi} = \partial_\phi V(\phi, h) \\ -c\ddot{h} - b\dddot{\phi} = \partial_h V(\phi, h) \end{cases} . \tag{5.42}$$

These equations are fourth-order for $a \neq 0$, third-order for $a = 0$ and $b \neq 0$ and second-order for $a = 0 = b$. In total we need 6 initial conditions to find a solution, meaning we have 3 d.o.f. in general. Despite the higher-order equations we can still reduce the number of d.o.f. by imposing a condition for the coefficients. To show this, it is best to introduce an auxiliary variable Q and a Lagrange-multiplier μ to put the system in a more familiar form:

$$\mathscr{L} = \frac{a}{2}\dot{Q}^2 + b\,\dot{Q}\dot{h} + \frac{c}{2}\dot{h}^2 + \frac{1}{2}Q^2 - V(\phi, h) - \mu(Q - \dot{\phi}) . \tag{5.43}$$

It is easy to see that this Lagrangian is equivalent to the previous one. At this point we can recast the first three pieces in terms of the vector $\boldsymbol{X}^T = (Q, h)$ as follows

$$\mathscr{L} = \frac{1}{2}\dot{\boldsymbol{X}}^T \mathcal{M}\,\dot{\boldsymbol{X}} + \frac{1}{2}Q^2 - V(\phi, h) - \mu(Q - \dot{\phi}) , \quad \mathcal{M} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} . \tag{5.44}$$

From this equation we see that for $\det \mathcal{M} \neq 0$ we have 6 d.o.f., whereas for $\det \mathcal{M} = 0$ the kinetic term acquires a zero-eigenvalue and the system only has 2 degrees-of-freedom. This is indeed the degeneracy condition we were looking for; degenerate theories have parameters a, b and c satisfying it. Furthermore one can also show that the resulting Hamiltonian is bounded-from-below, indicating that indeed we avoid the presence of the Ostrogradsky ghost.

The same principle can be applied to scalar-tensor theories. Such setup can be made analogous to this example by identifying $\phi(t)$ with the scalar field $\phi(x)$ and $h(t)$ with the metric $g_{\mu\nu}(x)$. The most general DHOST theory is given by [36]

$$\begin{aligned} \mathscr{L} =& f_0(X, \phi) + f_1(X, \phi)\Box\phi + f_2(X, \phi)R + f_3(X, \phi)G_{\mu\nu}\phi^{\mu\nu} + \\ & C_{(2)}^{\mu\nu\rho\sigma}\phi_{\mu\nu}\phi_{\rho\sigma} + C_{(3)}^{\mu\nu\rho\sigma\alpha\beta}\phi_{\mu\nu}\phi_{\rho\sigma}\phi_{\alpha\beta} , \end{aligned} \tag{5.45}$$

where for convenience we define $\phi_\mu = \nabla_\mu\phi$ and the tensor $C_{(2)}^{\mu\nu\rho\sigma}$ can be decomposed as

$$C_{(2)}^{\mu\nu\rho\sigma}\phi_{\mu\nu}\phi_{\rho\sigma} = \sum_{A=1}^{5} a_A(X, \phi)\mathscr{L}_A^{(2)} , \tag{5.46}$$

with

$$\begin{aligned} &\mathscr{L}_1^{(2)} = \phi_{\mu\nu}\phi^{\mu\nu} , \quad \mathscr{L}_2^{(2)} = (\Box\phi)^2 , \quad \mathscr{L}_3^{(2)} = (\Box\phi)\phi^\mu\phi_{\mu\nu}\phi^\nu , \\ &\mathscr{L}_4^{(2)} = \phi^\mu\phi_{\mu\rho}\phi_\nu , \quad \mathscr{L}_5^{(2)} = (\phi^\mu\phi_{\mu\nu}\phi^\nu)^2 , \end{aligned} \tag{5.47}$$

and $C_{(3)}^{\mu\nu\rho\sigma\alpha\beta}$ can be similarly decomposed as

$$C_{(3)}^{\mu\nu\rho\sigma\alpha\beta}\phi_{\mu\nu}\phi_{\rho\sigma}\phi_{\alpha\beta} = \sum_{A=1}^{10} b_A(X, \phi)\mathscr{L}_A^{(3)} , \tag{5.48}$$

with $\mathscr{L}_A^{(3)}$ this time given by operators cubic in the second derivatives of ϕ (since it contains many terms that are not essential for our discussion we do not write it explicitly).

For the theory (5.45), the degeneracy condition can be achieved by imposing some restrictions on the functions f_2, f_3 and a_A and b_A. If such condition is imposed the theory propagates one scalar and one graviton and avoid the Ostrogradsky instability. Of course, appropriate choices for these functions will give the Horndeski theory seen in the previous paragraph. We can imagine these DHOST theories as the largest (so far) class of scalar-tensor theories propagating one scalar, as depicted in the diagram of Figure 5.3.

At this point one could be puzzled about the statement that higher-derivatives in the Lagrangian lead to instabilities. Indeed, from the point of view of Effective Field Theories (EFTs) we know that if we restrict to low energies and we integrate out some heavy field of mass M, we are going to generate higher-derivative operators in the Lagrangian, containing schematically $\sim \frac{\partial^{2n}}{M^{2n}}\phi$. Furthermore in general these operators do not have the structure appearing in Horndeski or DHOST theories, and hence lead to higher-order equations (and additional d.o.f.). Usually in this context one does not worry about Ostrogradsky instabilities, but why? The reason is that, as one can easily check, the "additional" d.o.f. induced by these operators have masses $\sim M$, and are therefore outside of the EFT. Whenever these operator become relevant (for instance when we go to energies close to M), the EFT breaks down and we need to consider the full theory. Moreover, these operators are in general sub-dominant compared to other effective operators with a lower number of derivative and are therefore usually neglected. On the other hand, here we are considering theories where higher-derivative operators can dominate the dynamics (e.g. in strong-field regimes around Black Holes or in cosmology at late times), and in order to make sense of them we need to impose the absence of instabilities. From this discussion we can also realize that it is difficult to interpret such theories as standard EFTs where we integrate out weakly coupled heavy fields. It is indeed an open problem to find possible UV completions for these scalar-tensor theories.

5.6 Going further

This is far from an exhaustive list of generalized theories. Most noticeable we have not discussed metric-affine theories of gravity such as Einstein–Cartan [37] (a generalization of GR including non propagating torsion), Metric affine $F(R)$ [38] or Lorentz breaking theories of gravity such as Einstein–Aether [39] (in vacuum equivalent to GR plus a timelike vector setting a preferred frame) or Hořava–Lifshitz gravity [40] (GR plus an hypersurface orthogonal vector which hence can be written as the gradient of a scalar, the Kronon, providing a preferred foliation). We leave this to further study for the interested student.

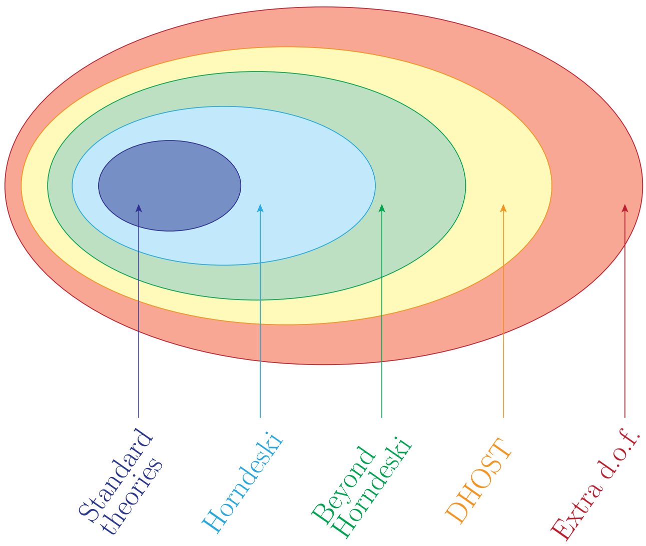


Figure 5.3: Schematic diagram of the various scalar theories discussed.

6 Global Methods

6.1 Carter–Penrose Diagrams

Now we want to introduce global methods to discuss spacetime solutions and their dynamics, and it would be useful if we could "see" the global structure of a spacetime, including infinities. Basically, imagine that we want to describe the whole spacetime in such a way that we can draw it completely on a piece of paper, and therefore we want to find a way to draw spatial and time infinities. The starting point of this task is the light cone.

We can generally define one kind of spatial infinity, ι^0, and two kind of time infinities: a future time infinity ι^+ and a past time infinity, ι^-. Also, we can define two types of null infinities, reachable by following the light-cone surface: we then have future null infinity $\mathscr{I}^+$ and past null infinity $\mathscr{I}^-$. In total, we have 5 notions of infinity.

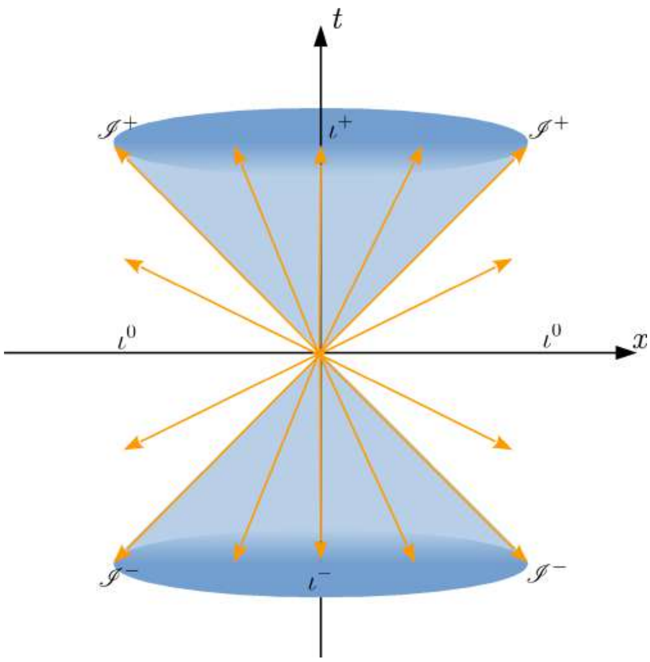


Figure 6.1: Representation of the different types of inifinity with respect to the light-cone.

The idea behind the so called Carter–Penrose diagrams is to start from this causal structure and draw a diagram that conveys it in a compact form by bringing the infinities to a finite distance and exposing it by a suitable conformal rescaling of the metric.

Conformal transformations: Let $(\mathcal{M}, g)$ be a spacetime, and $\Omega : \mathcal{M} \to \mathbb{R}$ a nowhere-vanishing, smooth function on $\mathcal{M}$. This is called a conformal transformation, a scale transformation of the metric $g \to \tilde{g} = \Omega^2 g$; this leaves invariant the angles between vectors, since

$$\cos\tilde{\theta} = \frac{(\tilde{x}, \tilde{y})}{\|\tilde{x}\| \, \|\tilde{y}\|} = \frac{\Omega^2 (x, y)}{\Omega^2 \, \|x\| \, \|y\|} = \frac{(x, y)}{\|x\| \, \|y\|} = \cos\theta, \tag{6.1}$$

but obviously not the norm of vectors. However, null vectors remain null:

$$g(x, x) = 0 \Rightarrow \tilde{g}(x, x) = 0 \tag{6.2}$$

In this way, we have introduced transformations that preserve the causal structure of spacetime.

6.1.1 Minkowski spacetime Carter–Penrose diagram

For example, let us take Minkowski spacetime; in spherical coordinates, the line element is

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + \,\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2. \tag{6.3}$$

As a first step, we reintroduce the null coordinates $u \equiv t - r$ and $v \equiv t + r$, that are such that u is constant for outgoing photons and v is constant for ingoing photons; since $r \geq 0$, then $v \geq u$. With this, the metric becomes

$$\eta = -\,\mathrm{d}u\,\mathrm{d}v + \frac{(v-u)^2}{4}\,\mathrm{d}\Omega^2. \tag{6.4}$$

We now want to map infinities to a finite distance. We can do so by introducing suitable coordinates of the form, $\mathcal{U} \equiv \arctan u$, $\mathcal{V} = \arctan v$, which map the range $-\infty < u, v < +\infty$, to the finite range $-\pi/2 < \mathcal{U}, \mathcal{V} < \pi/2$.

By inverting the above coordinate transformation we also have $u = \tan\mathcal{U}$, $v = \tan\mathcal{V}$, $\mathrm{d}u = (1+u^2)\,\mathrm{d}\mathcal{U}$, $\mathrm{d}v = (1+v^2)\,\mathrm{d}\mathcal{V}$, and thus

$$\eta = -(1+u^2)(1+v^2)\,\mathrm{d}\mathcal{U}\,\mathrm{d}\mathcal{V} + \frac{1}{4}(v-u)^2\,\mathrm{d}\Omega^2. \tag{6.5}$$

From this, we can change again coordinates by defining $\tau \equiv \mathcal{V} + \mathcal{U}$, $R \equiv \mathcal{V} - \mathcal{U}$. We can see that $\tau \pm R$ are limited to the real interval $(-\pi, \pi)$. Moreover, from $v \geq u$ follows that $\mathcal{V} \geq \mathcal{U}$ and hence $R \geq 0$. The metric then becomes

$$\eta = \frac{1}{4}\frac{1}{\cos^2\mathcal{U}\cos^2\mathcal{V}}\left[-\,\mathrm{d}\tau^2 + \,\mathrm{d}R^2 + \sin^2 R\,\mathrm{d}\Omega^2\right]. \tag{6.6}$$

We can make now two observations:

- In the above metric element R behaves as an angular variable.
- The Minkowski metric in this coordinates conformal appears to be conformal — with conformal factor $\Omega^2 = \frac{1}{4}\frac{1}{\cos^2\mathcal{U}\cos^2\mathcal{V}}$ — to the line element in square brackets, which is indeed (with the range restrictions we have inherited) a portion of and Einstein static Universe which is homogeneous and isotropic, and has a structure like $\mathbb{R} \times S^3$.
- we have now singularities at $\mathcal{U} = \frac{\pi}{2}$ and $\mathcal{V} = \frac{\pi}{2}$

The above points suggest an easy way forward: we define a fictitious metric $\tilde{\eta} = \Omega^{-2}\eta$ through a conformal transformation, so to consider a portion of Einstein static Universe. Although distances measured with the metric $\tilde{\eta}$ will differ (by a possibly an infinite factor) from those measured with the η metric, the causal sructure will be the same in both metrics, with the advantage that the fictitious metric $\tilde{\eta}$ is well behaved at the values of $(\mathcal{U}\mathcal{V})$ correspondings to the asymptotic regions of η.

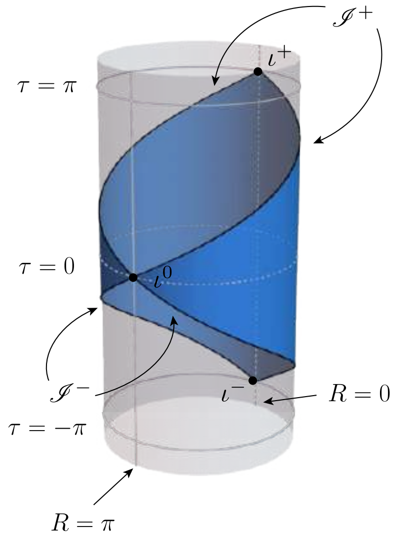


Figure 6.2: Spacetime diagram of Einstein static Universe.

If we draw this, by suppressing angular coordinates θ and φ, and carefully taking into account the ranges of our variables, we obtain what can be seen in Figure 6.2.

Therefore, each point of the cylinder is a 2-sphere of area $4\pi \sin^2 R$, timelike geodesics start at ι^- and end at ι^+, null geodesics start at $\mathscr{I}^-$ and end at $\mathscr{I}^+$, while spacelike geodesics start and end at ι^0.

If we cut this at $R = \pi$ and unfold it, the shaded area will be the Carter-Penrose diagram of Minkowski, Figure 6.3.

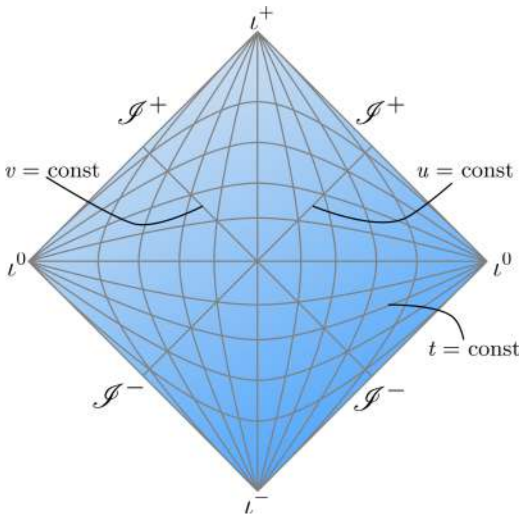


Figure 6.3: Carter-Penrose diagram of Minkowski spacetime.

The above example shows the basic procedure behind a causal diagram:

- Expose the causal structure by introducing a suitable set of null coordinates.
- Do a coordinate change that maps infinities to a finite range (normally using the *arctan* function or something equivalent).
- Use a conformal transformation to get rid of annoying conformal factors and analyse the causal structure of a simpler fictitious spacetime.

Let us then apply this technique now to a more interesting spacetime.

6.1.2 Schwarzschild spacetime Carter–Penrose diagram

One can perform a similar computation with Schwarzschild metric:

$$g = -\left(1 - \frac{2M}{r}\right) \mathrm{d}t^2 + \left(1 - \frac{2M}{r}\right)^{-1} \mathrm{d}r^2 + r^2\, \mathrm{d}\Omega^2. \tag{6.7}$$

First of all we want to introduce null coordinates. In order to find them in Schwarzschild spacetime, let us study the behaviour of null rauys. One can set $\mathrm{d}s^2 = 0$ and $\theta, \phi = \text{const.}$; then we immediately find

$$\left(\frac{\mathrm{d}t}{\mathrm{d}r}\right)^2 = \left(\frac{r}{r - 2M}\right)^2, \tag{6.8}$$

from which we find the following relation between the time and radial coordinates

$$\pm\, t = r + 2M \ln\left(\frac{r - 2M}{2M}\right) + \text{const.}; \tag{6.9}$$

we can now define the so-called Regge–Wheeler coordinate, or tortoise coordinate

$$r^* \equiv r + 2M \ln\left(\frac{r - 2M}{2M}\right), \tag{6.10}$$

and then by performing a similar transformation as the one we made before,

$$\begin{cases} u = t - r^* \\ v = t + r^*, \end{cases} \tag{6.11}$$

which are Schwarzschild null coordinates, Schwarzschild metric becomes

$$g = -\left(1 - \frac{2M}{r}\right) \mathrm{d}u\, \mathrm{d}v + r^2\, \mathrm{d}\Omega^2, \tag{6.12}$$

that still looks singular at $r = 2M$ where $r^* \to -\infty$ and consequently $u \to +\infty$ and $v \to -\infty$.

Eddington–Finkelstein coordinates: The just introduced null coordinates for Schwarzschild can be used to define a very useful set of coordinates the Eddington–Finkelstein ones. More precisely we can write the metric in the Ingoing Eddington–Finkelstein coordinates (v, r)

$$g = -\left(1 - \frac{2M}{r}\right) \mathrm{d}v^2 + 2\, \mathrm{d}v\, \mathrm{d}r + r^2\, \mathrm{d}\Omega^2, \tag{6.13}$$

for which ingoing and outgoing radial light rays have trajectories respectively given by

$$\frac{\mathrm{d}v}{\mathrm{d}r} = 0 \quad \text{and} \quad \frac{\mathrm{d}v}{\mathrm{d}r} = \frac{2}{1 - 2M/r}, \tag{6.14}$$

Alternatively one can have the so called Outgoing Eddington–Finkelstein coordinates (u, r)

$$g = -\left(1 - \frac{2M}{r}\right) \mathrm{d}u^2 + 2\, \mathrm{d}u\, \mathrm{d}r + r^2\, \mathrm{d}\Omega^2, \tag{6.15}$$

for which outgoing and ingoing radial light rays have trajectories respectively given by

$$\frac{\mathrm{d}u}{\mathrm{d}r} = 0 \quad \text{and} \quad \frac{\mathrm{d}u}{\mathrm{d}r} = -\frac{2}{1 - 2M/r}, \tag{6.16}$$

See also Box 31.2 of [41].

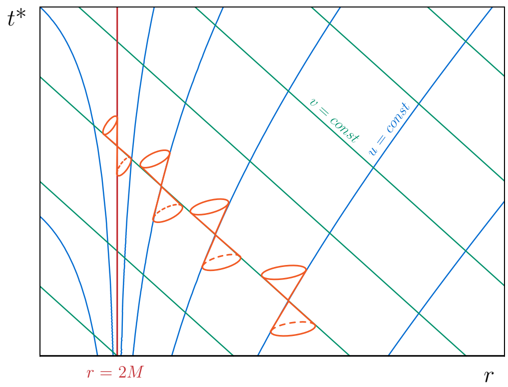


Figure 6.4: Eddington–Finkelstein plot showing the tilting of light cones close to the horizon of a black hole (with the two angular dimensions suppressed). Please note that rays of constant v being ingoing null ones are normally plotted at 45-degrees slant, just as they would be in flat spacetime. This is equivalent to say that the vertical axis is defined w.r.t. the time $t^* \equiv v - r = t + 2M \ln |r/2M - 1|$.

The apparent singularity at $r = 2M$ can be avoided by another change of coordinates. Let us then introduce, $U = -\mathrm{e}^{-u/4M}$ and $V = e^{v/4M}$ with implied ranges $-\infty < U < 0$ and $0 < V < +\infty$, so that the metric becomes

$$g = -\frac{32M^3}{r}\mathrm{e}^{-r/2M}\,\mathrm{d}U\,\mathrm{d}V + r^2(u,v)\,\mathrm{d}\Omega^2; \tag{6.17}$$

These new coordinates are called Kruskal–Szekeres null coordinates. We can immediately see that, in these coordinates, the metric is no longer singular in $r = 2M$ or $U, V = 0$, not surprisingly given that we already know that the singularity of the metric at the horizon is just a coordinate artefact. So nothing forbids us from doing an analytical continuation and consider the whole range $-\infty < U, V < +\infty$.

Finally, one can then again perform another coordinate transformation to get to Kruskal–Szekeres cartesian coordinates: $T \equiv \frac{V+U}{2}$, $X \equiv \frac{V-U}{2}$, and the metric becomes

$$\mathrm{d}s^2 = -\frac{32M^3}{r}\mathrm{e}^{-r/2M}\left(-\,\mathrm{d}T^2 + \mathrm{d}X^2\right) + r^2\,\mathrm{d}\Omega^2. \tag{6.18}$$

If we plot this metric, we get a Kruskal diagram, Figure 6.5.

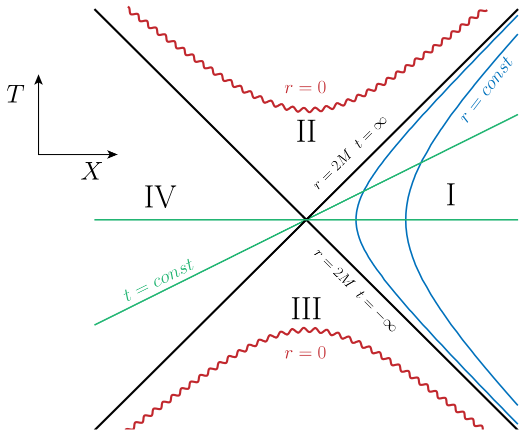


Figure 6.5: Kruskal diagram of a Schwarzschild black hole.

However, this diagram is still not capturing the asymptotic behaviour of the geometry. For that we shall need to introduce again a coordinate transformation based on the *arctan* function. Let us then introduce a new set of coordinates

$$\mathcal{U} = \arctan\left(\frac{U}{\sqrt{M}}\right) \qquad \mathcal{V} = \arctan\left(\frac{V}{\sqrt{2M}}\right), \tag{6.19}$$

we have coordinates that are, as before, in a finite range, $-\frac{\pi}{2} < \mathcal{V} < \frac{\pi}{2}$ and $-\frac{\pi}{2} < \mathcal{U} < \frac{\pi}{2}$. If we perform one final coordinate transformation

$$\bar{T} = \frac{\mathcal{V} + \mathcal{U}}{2} \qquad \bar{X} = \frac{\mathcal{V} - \mathcal{U}}{2}, \tag{6.20}$$

We get a metric that after a suitable conformal rescaling (**Exercise**: compute the conformal factor) leads to the the conformal diagram shown in Figure 6.6.

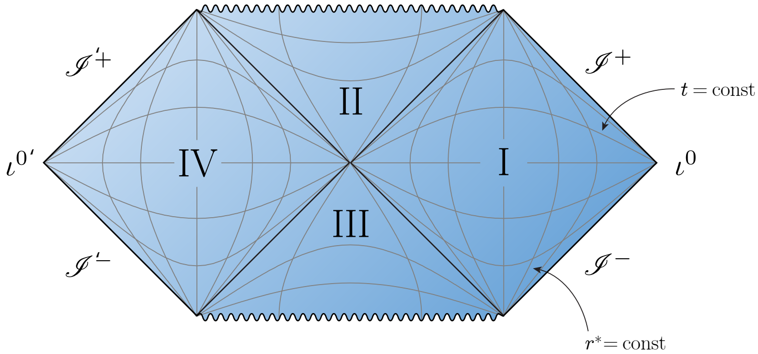


Figure 6.6: Carter–Penrose diagram of the maximal analytical extension of a Schwarzschild spacetime. This is sometime said to describe an eternal black hole.

Again, here each point is actually a 2-sphere. Differently from Minkowski, however, now $\mathscr{I}^{\pm}$, $\iota^{\pm}$, and ι^{0} are all made of disconnected pieces; moreover, not all null geodesics end on $\mathscr{I}^{+}$ or begin from $\mathscr{I}^{-}$, and this can actually give us a hint on the notion of the complex causal structure associated to a black hole.

Here we have four different regions: region I and region IV are asymptotically flat regions, which represent the "outside" of a black hole, and thus here $\mathscr{I}^{\pm} \cup \iota^{0}$ is identical to what we find in Minkowki. Here we defined an asymptotically flat spacetime as a spacetime with some asymptotic structure of Minkowski.

This actually implies that Schwarzschild metric possesses the same Killing vectors of Minkowski at infinity, which will be important for allowing to define a mass M and an angular momentum J on spacelike hypersurfaces.

Region II is the standard black hole region, with the wavy line representing the black hole singularity of Schwarzschild spacetime in $r = 0$; notice that the only way to exit region II is either ending into the singularity in $r = 0$ or escaping to region I or IV by travelling faster than light. Also notice that the singularity is spacelike.

Region III is a region called white hole, which is the temporal inverse of a black hole: if in a black hole nothing can exit and everything must remain inside, in a white hole nothing can enter and everything must leave.

Remarkably, it is possible to create an analogue of a white hole in a sink [42]! This is connected to **shallow water waves**: perturbations that propagate in a fluid under the influence of gravity in a shallow water basin. Indeed it is possible to show that these perturbations propagate at long wavelengths with a wave equation o a curved geometry determined by the fluid low [43].

$$\mathrm{d}s^2 = \frac{1}{c^2}\left[-(c^2 - v_{\mathrm{B}}^{\|2})\,\mathrm{d}t^2 - 2\mathbf{v}_{\mathrm{B}}^{\|} \cdot \mathbf{dx}\,\mathrm{d}t + \mathbf{dx} \cdot \mathbf{dx}\right],$$

where $c \equiv \sqrt{g h_{\mathrm{B}}}$ where g is the gravitational acceleration and h_{B} is the depth of the flow. So the speed of these waves depends on the height of the basin in which they are flowing. Hence, in a shallow water regime, these perturbations are slow. Let's now consider the experiment where we open out kitchen tap over a flat surface. The water forms a column that spreads radially as it meets the plate and forms a circle centred around the column. This circle marks the transition from supersonic to subsonic flow for the gravity waves. On the circle the water has the same speed as the waves in that fluid. If we create perturbations in the water outside the circle so they attempt to enter the circle, we would see that they would actually bounce off it, in the same way in which it is impossible to enter a white hole [42].

Is this maximally analytic extension realistic? The Carter-Penrose diagram of an eternal black hole only reflects the fact that, *per sé*, General Relativity solutions are time reversible, so if there is a black hole, there is also a white hole. However, initial conditions break time reversal, for example in the case in which a black hole is generated by a collapse of star. If we have a collapse, the Carter–Penrose diagram of Schwarzschild spacetime becomes like the one shown in Figure 6.7

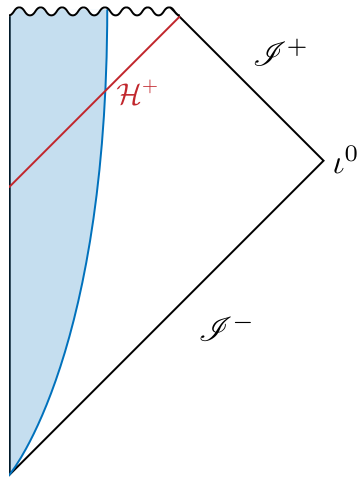


Figure 6.7: Carter-Penrose diagram of a collapsing star generating a Schwarzschild black hole.

We see that in this case the white hole is no more present; nonetheless, we can also see that the causal structure of black holes from collapse at late times would be undistinguishable from that of eternal black holes (this is rigorously proven in [44]).

6.2 Black Holes: Singularities and Event Horizon

Black holes are, in general, objects that possess horizons and singularities. Let us start with the latter. What is a singularity? Most often is identified as a set of events where the metric has some pathology, or the curvature becomes infinity. However, the metric is a solution of our field equations and we cannot assume it a priori. Hence an event in spacetime has a physical meaning only if (M, g) is well defined at such event. This in turns implies that a singularity cannot be a set of events in spacetime but has to be rather characterised as a set of missing events in (M, g), i.e. where spacetime itself is not defined.

6.2.1 Singularities

More formally,*an inextendible spacetime has a singularity if there is at least one incomplete geodesic.*

Here, with **inextendible spacetime** we mean a spacetime which literally misses a point: for example, if we take Minkowski spacetime and we trivially remove a point, that is **not** considered inextendible. Therefore, we define an inextendible spacetime as a spacetime which is *not isometric to a proper subset of another spacetime*.

An **incomplete geodesic** is a geodesic in which it is possible to reach the end of such geodesic at a *finite* value of its affine parameter. I.e. a geodesic from which we cannot extend our geodesic beyond a *finite* value of the affine parameter.

Note that albeit not completely bullet proof (there can be special cases when the first derivatives of the metric are OK but the second derivatives blow up, and hence the geodesic are Ok but the curvature is not) it is by far preferable to other definitions used in the past.

- Look for a singular behaviour in the metric: in that case we could have simply a coordinate artefact. For example consider the metric $g = (4|x|)^{-1}dx^2 + dy^2$ which looks singular in $x = 0$. However, is enough to intruduce the coordinate $X = \pm\sqrt{|x|}$ to see that $g = dX^2 + dy^2$ i.e. that the metric is just that of the Euclidean plane.

- Look for some curvature scalar blowing up. Unfortunately even if a scalar does blow up one can still have a Riemann well behaved everywere except at infinity. Alternatively, one can have well behaved scalars but a ill behaved Riemann.

- Also one can have curvature singularities without having a ill behaved Riemann tensor. For example if we consider e a flat disk from which we excise a slice and glue the opposite side together we still get a metric $g = \rho^2\,\mathrm{d}\theta^2 + \mathrm{d}\rho^2$ but with $\theta \notin [0, 2\pi]$ which corresponds to have a conical singularity in $\rho = 0$ even if Riemann is zero everywhere else.

We can classify singularities in 3 different types: we could have future spacelike singularities (like the singularity in $r = 0$ of a Schwarzschild black hole or the Big Crunch), past spacelike singularities (like the physical singularity of a white hole, or the Big Bang) and timelike singularities (that are present, for example, in rotating or charged black holes, or naked singularities).

6.2.2 Horizons

Now we want to define the notion of horizon. In order to do that, we need to define some concepts.

If at each point of a spacetime $(\mathcal{M}, g)$ we can locally distinguish a future and a past lightcone, then we say that this spacetime is **time orientable**.

In this case, there must be a continuous and everywhere non-zero, globally defined, timelike vector field v^μ. A spacetime is **space orientable** if there is a three-form $\epsilon_{[\mu\nu\lambda]} \neq 0$ everywhere and continuous, and if $v^\mu\epsilon_{[\mu\nu\lambda]} = 0$ for any v^μ timelike [45].

A spacetime is **orientable** if and only if there is a continuous, everywhere non-null, globally defined 4-form $\epsilon_{[\mu\nu\lambda\rho]}$, and this is the volume form $\mathrm{d}x^0 \wedge \cdots \wedge x^3$.

If a spacetime is both time orientable and space orientable, then $\epsilon_{[\mu\nu\rho\sigma]} = v_{[\mu}\epsilon_{\nu\rho\sigma]}$, so saying that a spacetime is both time orientable and space orientable is equivalent to say that it is orientable.

We can define chiral fermions in a spacetime if and only if $\epsilon_{\mu\nu\lambda\rho}$ exists. In fact, $\gamma_5\psi = \pm\psi$ and

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = i\frac{\mathrm{sign}([abcd])}{4!}\gamma^a\gamma^b\gamma^c\gamma^d \tag{6.21}$$

in Minkowski, while in a curved spacetime it becomes

$$\gamma_5 = i\frac{\epsilon_{\mu\nu\lambda\rho}}{4!}\mathbf{e}^\mu_a\mathbf{e}^\nu_b\mathbf{e}^\lambda_c\mathbf{e}^\rho_d\gamma^a\gamma^b\gamma^c\gamma^d \tag{6.22}$$

The **chronological future** of a point p, $I^+(p)$ is the set of points q of the manifold $\mathcal{M}$ such that there exists a future directed timelike curve γ with $\gamma(0) = p$ and $\gamma(1) = q$ (with an appropriate affine parametrization). Similarly, the **chronological past** of a point p, $I^-(p)$ is the set of points q of the manifold $\mathcal{M}$ such that there exists a past directed timelike curve γ with $\gamma(0) = p$ and $\gamma(1) = q$.

We can also define the **causal future** of a point p, denoted as $J^+(p)$ as the set of points q of the manifold $\mathcal{M}$ such that there exists a future directed causal curve γ with $\gamma(0) = p$ and $\gamma(1) = q$. Similarly, the **causal past** of a point p, denoted as $J^-(p)$ is the set of points q of the manifold $\mathcal{M}$ such that there exists a past directed causal curve γ with $\gamma(0) = p$ and $\gamma(1) = q$.

We can also define the chronological and causal pasts and futures of a subset of the manifold, $\mathcal{U} \subset \mathcal{M}$, in a very intuitive way, as

$$I^\pm(\mathcal{U}) = \bigcup_{p\in\mathcal{U}} I^\pm(p) \qquad J^\pm(\mathcal{U}) = \bigcup_{p\in\mathcal{U}} J^\pm(p). \tag{6.23}$$

With these definitions in mind, we can define a **black hole** region $\mathcal{B}_{\rm bh}$ as:

$$\mathcal{B}_{\rm bh} \equiv \mathcal{M} - J^-(\mathscr{I}^+), \tag{6.24}$$

where we include the asymptotes in $\mathcal{M}$. This definition characterises the fact that there is a portion of events in our spacetime which will be never causally connected with $\mathscr{I}^+$. We can also define the **future event horizon** as the border of this region [18]

$$\mathcal{H}^+ \equiv \dot{J}^-(\mathscr{I}^+); \tag{6.25}$$

where we have introduced the border of the topological closure of the causal past of a set of points U as $\dot{\bar{J}}^-(\mathcal{U}) = \bar{J}^-(\mathcal{U}) - J^-(\mathcal{U})$ with $\bar{J}^-(\mathcal{U})$ being the topological closure of the set (the intersection of all closed sets containing $J^-(\mathcal{U})$) and $J^-(\mathcal{U})$ being the union of all the open sets containing $J^-(\mathcal{U})$, which in the case of an open set coincides with the set itself. Note that the subtraction implicit in the definition $\dot{\bar{J}}^-(\mathscr{I}^+)$ is important as it gets rid of ι^0 and $\mathscr{I}^-$ — contained in $J^-(\mathscr{I}^+)$ — which otherwise we would also be included in the event horizon.

Of course, all of this can also be extended to a white hole:

$$\mathcal{B}_{\rm wh} \equiv \mathcal{M} - J^+(\mathscr{I}^-);$$

and analogously define a **past event horizon**

$$\mathcal{H}^- \equiv \dot{J}^+(\mathscr{I}^-).$$

Notice that we have defined the event horizon of a black hole and the black hole region by using future null infinity, $\mathscr{I}^+$. This means that, in order to properly define the event horizon, we need to know *a priori* what is the infinite future evolution of our spacetime, and in particular what is the $\mathscr{I}^+$ of our manifold $\mathcal{M}$. The event horizon defined in this way is, therefore, a *teleological* concept, but we shall see that there are also other notions of horizon that are much more down-to-earth and effective when we want to describe physical processes and interactions with black holes.

Some final remarks. Event horizons are null hypersurfaces. Therefore they can be seen as a collection of null geodesics which are the generators of the null hypersurface. Also we can always find a null vector l^μ which is affinely parametrized on the horizon such that $l^\mu\nabla_\mu l^\nu|_{\mathcal{H}^+} = 0$.

No two points on the horizon can be timelike separated, it will be always null. Locally, this is obvious for the very fact that the horizon is a null hypersurface. Globally, it can be proven by contradiction: take two points, p and q, on the event horizon $\mathcal{H}^+$. Suppose now that the curve γ connecting q to p is timelike so that $p \in J^-(q)$. Consider two points: p' close to p slightly inside $\mathcal{H}^+$ and q' close to q slightly outside. The timelike curve between p and q could then be deformed to a nearby timelike curve between p' and q' with $q' \in J^-(\mathscr{I}^+)$ and $p' \notin J^-(\mathscr{I}^+)$. However, this also implies that $p' \in J^-(q')$, but since q' is outside the horizon, $J^-(q') \subset J^-(\mathscr{I}^+)$, and therefore $p' \in J^-(\mathscr{I}^+)$. Hence, we get a contradiction.

There is also a theorem (due to Penrose) according to which the generators of the event horizon can have a past end points but no future end points: a null light ray can become a generator of the

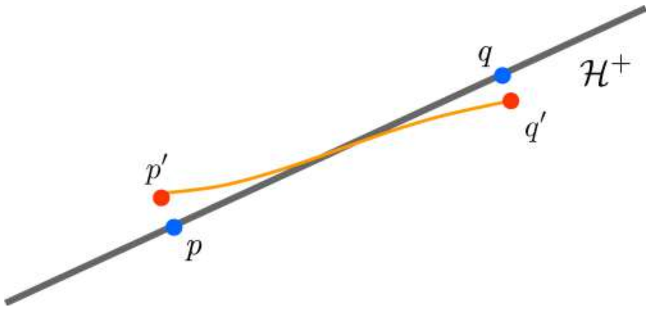


Figure 6.8: Illustration of the demonstration that two points on the horizon cannot be connected by a timelike curve.

horizon, but the contrary is not possible. Accordingly given appropriate energy conditions, horizons cannot bifurcate and black holes cannot split. Note that black hole evaporation by Hawking radiation does imply that there can be a future end-point; however, in this case, all of the energy conditions are violated.

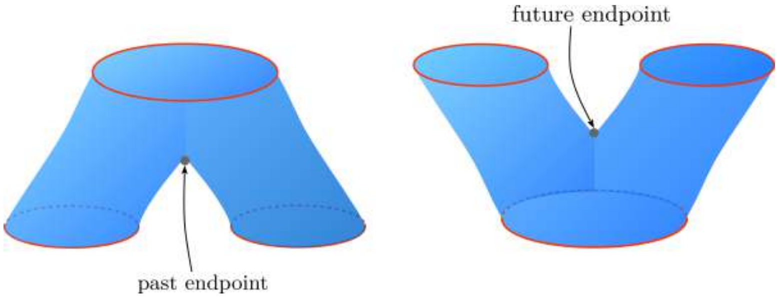


Figure 6.9: Illustration of the implication of past and future end points for the horizon generators.

6.3 Killing Horizons and Surface Gravity

Let ξ^a be a Killing vector field in region of spacetime which contains a null hypersurface $\mathcal{S}$. If $\xi^a \perp \mathcal{S}$, then $\mathcal{S}$ is said to be a Killing horizon.

A theorem by S. W. Hawking states that in GR: *the event horizon of a stationary asymptotically flat spacetime is a* ***Killing horizon***[1]. Note:

- A stationary spacetime is a spacetime which admits a globally defined, non vanishing, Killing vector which is timelike near infinity.
- A static spacetime is a stationary spacetime with a Killing vector which is also hypersurface orthogonal, i.e. irrotational.

Hawking theorem implies that for stationary black holes it is possible to find a Killing vector orthogonal to the horizon. But we also know that there is also another null vector along the surface, whose orbits are the generators of the null hypersurface: i.e. there must be on $\mathcal{H}^+$ a vector l^μ which is both null and affinely parametrized, so $l^\mu \nabla_\mu l^\nu|_{\mathcal{H}^+} = 0$.

Since ξ and l are both null vectors, we can write $\xi = f(x) l$, and also $\xi^\mu \nabla_\mu \xi^\nu = \kappa \xi^\nu|_{\mathcal{H}^+}$, since we cannot assume *a priori* that they are both affinely parametrized. The variable κ is associated to the failure of the geodesic to be affinely parametrized, and it is called **surface gravity** of the Killing horizon. We can easily derive κ by considering on the horizon

$$\xi^\mu \nabla_\mu \xi^\nu = f l^\mu \nabla_\mu (f l^\nu) = f (l^\mu \partial_\mu f) l^\nu + f^2 \cancel{l^\mu \nabla_\mu l^\nu} = f l^\mu \partial_\mu \ln|f| f l^\nu = \xi^\mu \partial_\mu \ln|f| \xi^\nu \equiv \kappa \xi^\nu, \quad (6.26)$$

[1] The converse is not true: a Killing horizon does not need to be an event horizon

so that

$$\kappa = \xi^\mu \partial_\mu \ln|f|\,. \tag{6.27}$$

In the case that the surface gravity is zero, the black hole is said to be extremal.

We can now apply the above definition to the surface gravity of a Schwarzschild black hole. Consider the Killing vector $\xi = \frac{\partial}{\partial t}$, with components $\xi^\mu = \delta^\mu_0$. This is also the Killing vector orthogonal to the Killing horizon. However, this is not the generator of the Killing horizon. Indeed one can check that it is $l = \frac{\partial}{\partial V}$ which satisfies $l^\mu \nabla_\mu l^\nu|_{\mathcal{H}^+} = 0$ as indeed U and V coincide with the affine parameters along ingoing and outgoing null geodesics respectively (alternatively one can directly proof this see e.g. [18] page 29).

The Killing vector $\xi = \frac{\partial}{\partial t}$ can be written in terms of the exponential null coordinates U and V as

$$\xi = \frac{\partial}{\partial t} = \frac{\partial V}{\partial t}\frac{\partial}{\partial V} + \frac{\partial U}{\partial t}\frac{\partial}{\partial U}; \tag{6.28}$$

in these coordinates, the horizon is located in $U = 0$ (indeed, $U \propto \mathrm{e}^{-u/4M}$, and at $r = 2M$, the tortoise coordinate $r^* \to -\infty$, which means that $u = (t - r^*) \to +\infty$ and $U = 0$). Then at the horizon we have

$$\frac{\partial}{\partial t} = \frac{\partial V}{\partial t}\frac{\partial}{\partial V}, \tag{6.29}$$

and since $V = \mathrm{e}^{v/4M} = \mathrm{e}^{(t+r^*)/4M}$, we get

$$\xi = \frac{\partial}{\partial t} = \frac{V}{4M}\frac{\partial}{\partial V}. \tag{6.30}$$

This is tantamount to say that $\xi = fl$, with $f = \frac{V}{4M}$ and $l = \frac{\partial}{\partial V}$. So surface gravity is

$$\kappa = \xi^\mu \partial_\mu \ln|f| = \frac{V}{4M}\frac{\partial}{\partial V}\ln\left|\frac{V}{4M}\right| = \frac{V}{4M}\frac{1}{V} = \frac{1}{4M}, \tag{6.31}$$

What is its physical meaning? Why is this relevant?

Let us start again with $\xi^\mu \nabla_\mu \xi^\nu = \kappa_{\rm in}\xi^\nu|_{\mathcal{H}^+}$, where the subscript stands for "inaffinity". Now, introduce a point splitting so that we can compute the same quantity in the coincidence limit of a point p that lies on the horizon and a point q slightly outside of it

$$\kappa^2_{\rm in} = \lim_{q\to p}\left[-\frac{(\xi^\mu\nabla_\mu\xi^\sigma)(\xi^\nu\nabla_\nu\xi_\sigma)}{\xi^\rho\xi_\rho}\right]_q. \tag{6.32}$$

This formula is useful because it allows a physical interpretation of the surface gravity. Let us start by recognising that the normalized acceleration along the orbits of ξ^μ can be written as

$$a^\sigma \equiv \frac{(\xi^\mu\nabla_\mu\xi^\sigma)}{-\xi^\nu\xi_\nu}, \tag{6.33}$$

and that we can make a scalar out of it by $a \equiv (a^\sigma a_\sigma)^{1/2}$. We also know the redshift factor is just $\chi \equiv (-\xi^\mu\xi_\mu)^{1/2}$. Therefore, we can see that the surface gravity in this limiting procedure is given by

$$\kappa_{\rm in} = \lim_{q\to p}(\chi a), \tag{6.34}$$

and we notice that basically the surface gravity is the acceleration per unit mass needed to remain static and hovering just over the horizon as measured from infinity (since there is a redshift factor). If we lower an object attached with a rope, this is the acceleration we have to give the rope in order to keep the object hovering above the horizon. At the horizon, the acceleration $a \to \infty$ while $\chi \to 0$, so that their product remains finite.

It can be proven that κ is constant on the horizon $\mathcal{H}^+$ (we shall show this explicitly later). Also, if $\mathcal{H}^+$ is the event horizon of a Killing vector ξ, it is also the event horizon of a vector $\bar{\xi} = C\xi$: in this case, we would have $\bar{\kappa} = C\kappa$, and this means that the normalization of the surface gravity is arbitrary. However, in general the normalization is set in such a way that at infinity the Killing vector ξ becomes exactly the one of Minkowski spacetime, that has norm -1. We would be left then with an arbitrary choice on the sign of the Killing vector, but this can be resolved by the condition that the Killing vector is future-directed.

6.3.1 Surface gravity: alternative definitions

In general, there are many other definitions of surface gravity, and they coincide with the ones given here only in a stationary spacetime and when we are considering Killing horizons (see e.g. [46] for a more detailed review).

For example, we can introduce the **peeling surface gravity**, κ_{peeling}: consider a generic spherically symmetric metric (not necessarily stationary)

$$\mathrm{d}s^2 = -\mathrm{e}^{-2\phi(r,t)}\left(1 - \frac{2M(r,t)}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{2M(r,t)}{r}\right)^{-1}\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2. \tag{6.35}$$

We will define an evolving horizon as the place where $g_{00} = 0$. This is not exactly an event horizon, but it is a dynamical horizon r_H such that

$$\frac{2M(r_H, t)}{r_H} = 1. \tag{6.36}$$

Radial null geodesics in this metric will satisfy the relation

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \pm\mathrm{e}^{-\phi}\left(1 - \frac{2M}{r}\right). \tag{6.37}$$

Now, consider a geodesic infinitesimally close to the horizon, and expands around it. We get

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \pm e^{-\phi}\left(\frac{1 - 2M'}{r_H}\right)(r(t) - r_H(t)) + \mathcal{O}(r - r_H)^2; \tag{6.38}$$

we define κ_{peeling} as the coefficient

$$\kappa_{\text{peeling}} = \frac{1}{2}e^{-\phi}\left(\frac{1 - 2M'}{r_H}\right), \tag{6.39}$$

and therefore $\frac{\mathrm{d}r}{\mathrm{d}t} = \pm 2\kappa_{\text{p}}(t)\,(r(t) - r_H(t)) + \dots$. One can now get that the distance of the geodesic from the horizon as a function of time is

$$|r - r_H|(t) \simeq |r - r_H|\,(t_0)\exp\left[2\int \kappa_{\text{p}}(t)\,\mathrm{d}t\right]. \tag{6.40}$$

Similarly for two nearby null geodesics $r_1(t), r_2(t)$ one gets

$$|r_1 - r_2|(t) \simeq |r_1 - r_2|\,(t_0)\exp\left[2\int \kappa_{\text{p}}(t)\,\mathrm{d}t\right]. \tag{6.41}$$

In Schwarzschild limit, $\phi \to 0$ and $M(r,t) = \mathit{const} = M$, so that $\kappa_{\text{p}} = \frac{1}{4M} = \kappa_{\text{in}}$. More generally, two notion of surface gravity always coincides in stationary situations, when the horzion is also a Killing horizon. However, κ_{p} better captures how null geodesics "peel" away from the horizon. If we plot $u = \text{const}$ and $v = \text{const}$ lines in a (t^*, r) plane, we would see that the $u = \text{const}$ lines would peel away exponentially from the horizon in a way ruled by this factor.

This is also related to Hawking radiation: one has this exponential peeling, ruled by κ_{p}, both inside and outside the horizon. In the already discussed heuristic picture for Hawking radiation (see the section of geodesic deviation equation) a virtual pair forms at the two sides of the horizon, it will break apart with a law similar to the previous equation, and therefore one particle will end in the black hole, while the other one will escape to infinity. The above mentioned peeling null rays on the two sides of the horizon are those followed by these Hawking pairs. This is why when we have an evaporating black hole, its temperature depends really on the peeling surface gravity and the latter will coincides with κ_{in} only in stationary spacetimes (see e.g. [47]).

Another definition for surface gravity we can give is the so-called **normal surface gravity**, κ_{normal},

$$\nabla^a \left(\xi^b \xi_b\right) \equiv -2\kappa_{\mathrm{normal}} \xi^a. \tag{6.42}$$

Indeed, we know that $\mathcal{H}^+$ is null and that ξ^μ is normal to $\mathcal{H}^+$, but being a null vector, we can write $(\xi^\mu \xi_\mu)\,|_{\mathcal{H}^+} = \mathrm{const} = 0$, so that this norm can be taken as the function defining the horizon hypersurface. Now this implies that if we consider the vector $\nabla^\nu\left(\xi^\mu \xi_\mu\right)$ this will be orthogonal to $\mathcal{H}^+$ as ξ^ν is. Consequently they should be proportional to each other

$$\nabla^\nu \left(\xi^\mu \xi_\mu\right) \propto \xi^\nu. \tag{6.43}$$

with the proportionality factor defining the normal surface gravity as above.

Again also this notion of surface gravity is equivalent to the inaffinity surface gravity when dealing with stationary black holes, i.e. with Killing horizons. This can be seen using the Killing equation: in fact, we can write

$$-\kappa_{\mathrm{normal}} \xi^a \equiv \frac{1}{2} \nabla^a \left(\xi^b \xi_b\right) = \xi_b \nabla^a \xi^b = -\xi_b \nabla^b \xi^a \equiv -\kappa_{\mathrm{in}} \xi^a, \tag{6.44}$$

I.e. $\kappa_{\mathrm{normal}} = \kappa_{\mathrm{in}}$.

Another definition of surface gravity is the **generator surface gravity** κ_{g}:

$$\kappa_{\mathrm{g}}^2 = -\frac{1}{2} \left(\nabla^{[a} \xi^{b]}\right) \left(\nabla_{[a} \xi_{b]}\right). \tag{6.45}$$

If the Killing equation holds, this is still equivalent to κ_{in}; in order to prove this (**Exercise**), one has to use the Frobenius theorem, that holds whenever we have a hypersurface orthogonal vector and assures that $\xi_{[\mu} \nabla_\nu \xi_{\rho]}|_{\mathcal{H}} = 0$.

6.4 Cauchy Horizon

There are situation in which one would like to deal with the evolution of spacetimes which are neither static nor stationary. For example if we want to describe black holes mergers, evolution, and so on. In order to do that, we need to introduce some important concepts needed to be able to describe the evolution of the spacetime from initial conditions: i.e. one needs to solve a Cauchy problem. In General Relativity, solving a Cauchy problem requires that we have a good hypersurface on which our initial conditions can be provided. This hypersurface is called **Cauchy hypersurface**.

In order to define these hypersurfaces formally, we need some preliminary concepts. First of all, we need to define an **acronal set**. *An acronal set is a subset of the manifold $\mathcal{S} \subset \mathcal{M}$ such that, for any two points p, $q \in \mathcal{S}$ we have that $q \notin I^{\pm}(p)$*; in practice, this means that there cannot be a timelike curve that links them, and therefore $I^{\pm}(\mathcal{S}) \cap \mathcal{S} = \varnothing$.

Another concept is that of **domain of dependence**. *The future domain of dependence is defined as the set of points p such that every past inextensible causal curve intersects $\mathcal{S}$.*
Here, with *inextensible* we mean that there are no future or past endpoints: a past/future endpoint is a point p such that, for any neighbourhood $\mathcal{O}(p)$ of p there is a t_0 so that for any $(t < t_0)$ or

$(t > t_0)$ respectively, one has $\lambda(t) \in \mathcal{O}(p)$. (Note also that this definition is different from the case of incomplete geodesics for which there are no $t > t_0 \neq \infty$.) Similarly, *the past domain of dependence is the set of points q such that every future inextensible causal curve intersects $\mathcal{S}$.*

We denote the future and past domain of dependence of a point p as $D^{\pm}(p)$, and similarly for a submanifold $\mathcal{S}$, $D^{\pm}(\mathcal{S})$. We then define a **Cauchy hypersurface** as an acronal, non-compact surface Σ for which $D(\Sigma) \equiv D^{-}(\Sigma) \cup D^{+}(\Sigma)$ coincides with the whole manifold, $D(\Sigma) = \mathcal{M}$.[2] Given some initial data on a Cauchy hypersurface, one can evolve backward and forward the whole manifold. A spacetime that admits a Cauchy hypersurface is called **globally hyperbolic**.

An example of a globally hyperbolic spacetime is given by the spacetime of a collapsing star, since we can construct Cauchy hypersurfaces covering the whole manifold.

There might be limits to the Cauchy evolution: these are called **Cauchy horizons**. If a spacetime is not globally hyperbolic, it happens when D^+ or D^- has a boundary; there is no surface we can find such that D^+ and D^- fill the whole spacetime.

There are two types of Cauchy horizons: there are Cauchy horizons from timelike singularities and Cauchy horizons from time machines (or causality violations, or closed timelike curves).

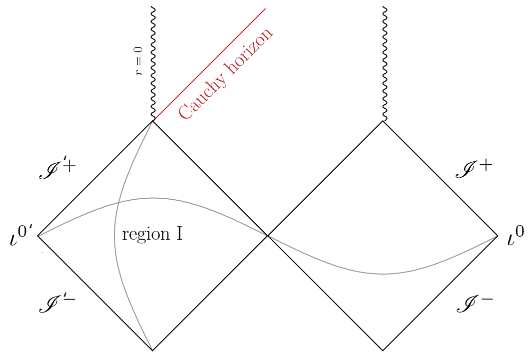


Figure 6.10: Carter-Penrose diagram of the Reissner-Nordström spacetime, highlighting the Cauchy horizon.

An example of Cauchy horizon from timelike singularities can be found, for example, in the Reissner-Nordström metric, which is the metric around a static and stationary black hole with mass M and electric charge Q; this is given by simply substituting in Schwarzschild metric

$$\left(1 - \frac{2M}{r}\right) \longrightarrow \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right); \tag{6.46}$$

the Cauchy horizon in this case is the one visible in Figure 6.10 for the non extremal case, $Q^2 < M^2$ (the extremal RN black hole being the solution with $Q^2 = M^2$ and the naked singularity case being the one with $Q^2 > M^2$). It is clear that the presence of the timelike singularity prevents us to evolve from a set of initial data given on the drawn hypersurface to the region beyond the Cauchy horizon.

A simple example of a Cauchy horizon from causality violation can be given for a $1+1$ geometry of a cylindrical spacetime with $(t, x) = (t, x+1)$ with metric

$$\mathrm{d}s^2 = -\cos(\lambda)\,\mathrm{d}t^2 - \sin(\lambda)\,(\,\mathrm{d}t\,\mathrm{d}x + \,\mathrm{d}x\,\mathrm{d}t) + \cos(\lambda)\,\mathrm{d}x^2; \tag{6.47}$$

here, $\lambda = \cot^{-1}(t)$, and thus $\lambda(t = -\infty) = 0$ and $\lambda(t = +\infty) = \pi$; the structure of the spacetime is then similar to the one visible in Figure 6.11. In this case, the Cauchy horizon is called a chronological horizon.

Finally, let us comment that for a future Cauchy horizon the entire infinite history of the external spacetime in region I, and its counterpart on the right side of figure 6.11, is in its causal past, i.e.

[2] There is also a weaker concept of Cauchy surface namely that of a **partial Cauchy surface** which for for a spacetime $\mathcal{M}$ is a hypersurface which no causal curve intersects more than once.

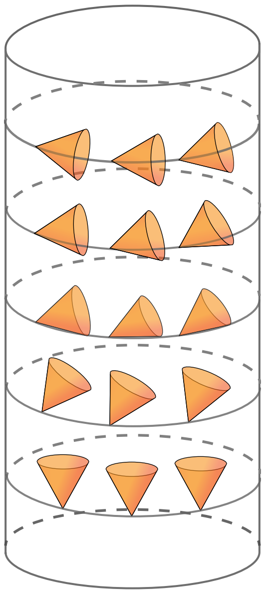


Figure 6.11: Causal structure of a causality-violating spacetime. The Cauchy horizon is the point where the light-cones tilt enough to permit the existence of a closed null curve.

visible, so signals from I must undergo an infinite blueshift as they approach the Cauchy horizon. For this reason, Cauchy horizons are generally considered to be unstable under classical perturbations no matter how small. So, for any physically realistic collapse, the Cauchy horizon is often seen as a singular null hypersurface for which new physics beyond GR is needed. Let us just add for completeness that this issue is not completely settle as recently it was claimed that in spacetimes with a cosmological constant stable Cauchy horizons might exist (see [48] and follow up literature) so violating the Cosmic censorship conjecture in its strong form.

The Cosmic Censorship Conjecture: After the realisation that singularities are a generic prediction of GR (see singularity theorems later on in this lecture notes), it was understood that if they could be observed to outside observers then they could mark a sharp breakdown of the predictive power of physics (at least the classical one). For this reason was conjectured by Penrose that singularities should always be screened by horizons (non naked singularities). This concept has later on evolved in a distinction between a weak cosmic censorship and a strong cosmic censorship conjectures to take also int account the breakdown of Cauchy evolution introduced by Cauchy horizons

Weak Cosmic censorship conjecture : No singularity can be visible from future null infinity. I.e. for generic initial data, the maximal Cauchy development possesses a complete future null infinity.

Strong Cosmic censorship conjecture : General relativity is a deterministic theory, in the same sense that classical mechanics is a deterministic theory. In other words, physical solutions of the theory should always be globally hyperbolic spacetimes, which implies that the maximal Cauchy development of generic compact or asymptotically flat initial data is locally inextendible as a regular Lorentzian manifold.

It is evident that solutions with Cauchy horizons violate the strong cosmic censorship while they are compatible with the weak one.

6.5 Cosmological Horizon

Sometimes the term Hubble horizon is used in cosmology, however what it is then defined as a horizon is not really a horizon in the causal sense. For a more detailed discussion see [49, 50].

Consider the FRW metric

$$\mathrm{d}s^2 = -c^2\,\mathrm{d}t^2 + R^2(t)\left[\mathrm{d}\chi^2 + S_\kappa^2(\chi)\,\mathrm{d}\Omega^2\right], \quad \text{where } S_\kappa(\chi) = \begin{cases} \sin\chi & k=1 \\ \chi & k=0 \\ \sinh\chi & k=-1. \end{cases} \tag{6.48}$$

In general, we can write the proper distance of an object as a function of time as $D(t) = R(t)\chi(z)$. If we differentiate with respect to proper time we get that the velocity with which an object, observed at redshift z now, gets far from us is

$$V(t,z) = \dot{D}(t) = \dot{R}(t)\chi(z) + R(t)\dot{\chi}(z), \tag{6.49}$$

where the first term is the recession velocity V_{rec} and the second term is the peculiar velocity V_{pec}.[3]

Even neglecting peculiar motion ($\dot{\chi} = 0$), one can easily see that the recession velocity e.g. of a galaxy from us is a time dependent quantity as the expansion of the universe varies in time. The current recession velocity of a galaxy observed now at $t = t_0$ will then be $V = \dot{R}_0\chi(z)$.

The Hubble sphere is defined as the radius when the velocity of recession is equal to c; therefore, neglecting the peculiar velocity for the moment we get $\chi(z) = c/\dot{R}(t)$ and hence the Hubble radius is

$$D_H = \frac{c}{H}; \tag{6.50}$$

where $H(t) = \dot{D}(t)/D(t)$. This is not an horizon, it is the speed of light sphere (SLS) radius and corresponds to a redshift of roughly $z_{SLS} = 1.46$ in ΛCDM; therefore, since we can see beyond this distance, this is definitely not an horizon.

For photons, we can write $c\,\mathrm{d}t = R(t)\,\mathrm{d}\chi$, and from this we find a peculiar velocity $v_\mathrm{p} = R\dot{\chi} = c$. This also implies that the total velocity for photons travelling towards us is $V = V_{rec} - c$ which is still positive (away from us) if $V_{rec} > c$, i.e. if they are outside our Hubble sphere. However, the Hubble sphere expands in time (e.g. in a matter dominated universe $D_H = 3ct/2$) and so it can overtake these receding incoming photons that will then find themselves in a region with $V_{rec} < c$ and hence will be able to get to us.

Also, with photons, we can define the so called particle horizon, which again is not an event horizon. This is defined in comoving distance as

$$\chi_{\mathrm{PH}}(t) = c\int_0^t \frac{\mathrm{d}t'}{R(t')}, \tag{6.51}$$

and this tells us that the coordinate position of a particle after some time t from when it was emitted goes like the integral above. This defines the region that can be causally connected at time t after the Big Bang. In physical distance, this is of course $D_{\mathrm{PH}}(t) = R(t)\chi_{\mathrm{PH}}(t)$.

The event horizon tells us the boundary of the region that can be causally connected from the start to the end of the Universe; therefore, the only event horizon we can define in cosmology is

$$\chi_{\mathrm{EH}} = c\int_0^\infty \frac{\mathrm{d}t'}{R(t')}. \tag{6.52}$$

Now, if our Universe were asymptotically de Sitter, we would get asymptotically $R(t) = \exp\sqrt{\Lambda/3}t$, so we would find that this horizon would be finite and equal to the size of the de Sitter Universe, which scales with the cosmological constant as $D_{EH} = c/H = c\sqrt{3/\Lambda}$ and coincides in this special case with the Hubble sphere.

[3] Note that we can always "trade" redshift for time if we know the universe content, i.e. $R(t)$, via the relation $(1+z) = R_0/R(t)$.

6.6 Local black hole horizons

As we discussed above the characterization of a black hole horizon as event horizon is very straightforward and tightly linked to the causal structure of spacetime. We shalls ee later on that it plays a crucial role in the definition of the four laws of black hole mechanics. It is however, as we also stressed, an intrinsically non-local definition which is also teleological in nature: defining the EH of a black hole requires the full knowledge of the spacetime in the infinite future. Hence, per se, these are not useful concepts when dealing e.g. with dynamical processes or astrophysical observation (see e.g. [51]). For this reason alternative notions of a horizon, more intrinsic and local, have been proposed. But in order to understand them we shall need first to add a few notions to our toolkit.

6.6.1 Trapped Surfaces and Trapped Regions

We now want to introduce now the concept of trapped surfaces. Consider Figure 6.12 were we have time in the vertical axis and radius and one angle on a $t = constant$ slice (i.e. we have suppressed the extra space dimension). There are two possible null rays that can be emitted from each point of the

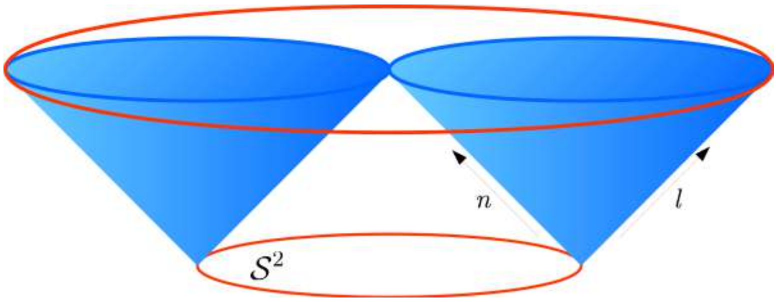


Figure 6.12:

initial spacelike 2-surface $\mathcal{S}^2$: ingoing light rays, along vector n, and outgoing light rays, along vector l, with l and n null vectors. Being both l and n null vectors, they are both orthogonal to $\mathcal{S}^2$.

In Minkowski, all the ingoing light rays will converge in a point and form a kind of cone with a"bowl", made out of outgoing light rays, around it (Figure 6.13). We define F^4 the chronological future of $\mathcal{S}^2$, $\mathcal{F}^4 \equiv I^+(S^2)$ and $\mathcal{B}^3$ the null hypersurface bounding it $\mathcal{B}^3 = \partial\mathcal{F}^4$. It is easy to see that this is a $\mathcal{C}^0$ submanifold of $\mathcal{M}$ which is also acronal.

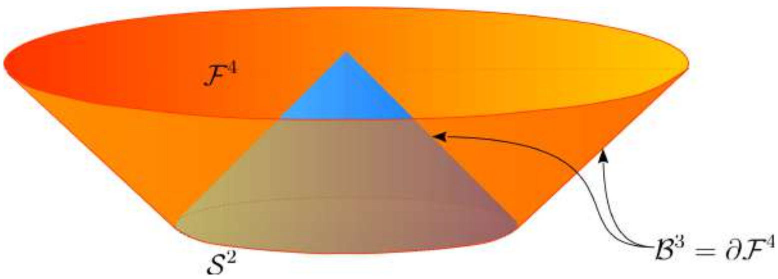


Figure 6.13:

Now, let us introduce gravity. In an Eddington–Finkelstein plot, we can see that light-cones are straight far away from the horizon and begin to tilt as they get close to the horizon, Figure 6.4. This notion can be used to define the notion of a no-escape, a trapped, region without resorting to non-local concepts like in the case of the event horizon. Indeed, time-like observers are forced to travel inside their own light cones and then if in a gravitational collapse they become enough tilted, after some time the region $\mathcal{S}^2$ will be causally connected with a smaller and smaller region (see Figure 6.14).

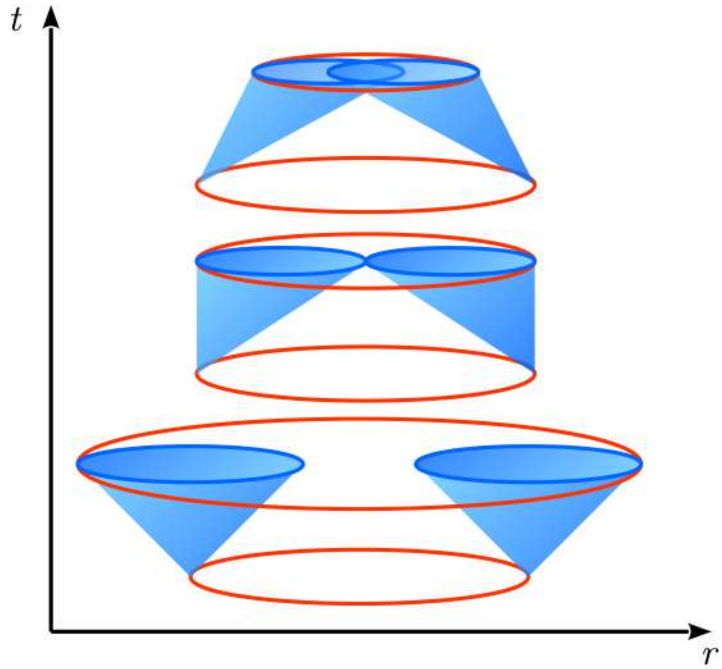


Figure 6.14: Representation of how gravity can tilt the light cones and lead to the formation of a trapped surface.

This is the idea behind a trapped region. Now we want to give a more rigorous definition. Remember that, for a null congruence, the expansion can be written as

$$\theta = \lim_{\delta A \to 0} \frac{1}{\delta A} \frac{\delta A}{\delta \lambda}. \tag{6.53}$$

It is easy to see that the expansion of the ingoing light rays is always negative, $\theta_n < 0$, even in Minkowsky. In order to characterise the behaviour of a trapped region we can however notice that in this case also the expansion of the outgoing null congruence, θ_l, is negative, $\theta_l < 0$.

Hence, we can finally give the following definition: a **trapped surface** is a $\mathcal{T}^2$, closed[4], spacelike, 2-surface such that both ingoing and outgoing generators are converging; this means that the expansion of both ingoing and outgoing congruences orthogonal to the 2-surface are negative: $\theta_n < 0$, $\theta_l < 0$. The limit case between having a trapped surface and a non-trapped surface, with $\theta_l = 0$, is called a **marginally trapped surface**.

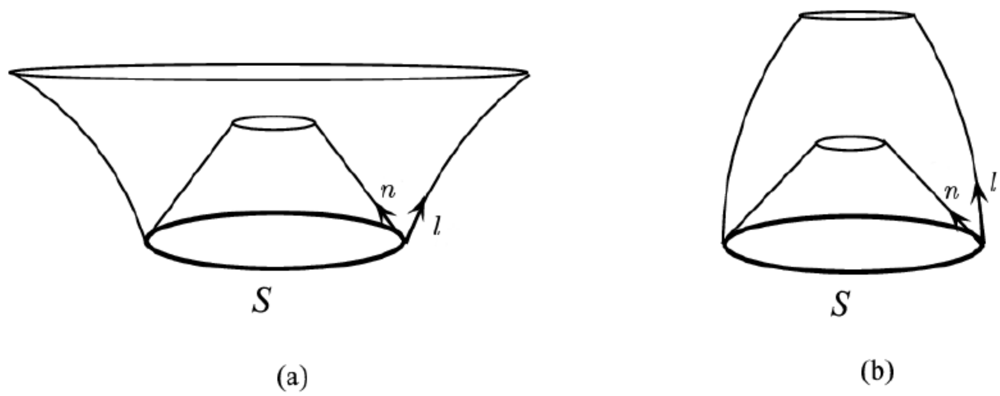


Figure 6.15: (a) Untrapped region; (b) future trapped region (from [52]).

[4] A closed manifold is a manifold which is compact and boundary-less.

A **trapped region** on a 3-D hypersurface Σ is the set of all the points in Σ through which a trapped surface passes. In general, we could have disconnected trapped regions, therefore we can define a **total trapped region** on a Cauchy hypersurface as the closure of the union of all the trapped region $\mathcal{T}_i$ on Σ.

6.6.2 Apparent Horizon

Now we introduce a new concept of horizon: an **apparent horizon** $\mathcal{A}$ is the boundary $\dot{\mathscr{F}}$ of a total trapped region on Σ. Therefore, $\mathcal{A}$ is an *outer marginally trapped surface* with $\theta_{\text{out}} = 0$. For a stationary black hole, the apparent horizon coincides with the event horizon, $\mathcal{A} = \mathcal{H}$. During a collapse, however, the apparent horizon is included inside the event horizon, $\mathcal{A} \subset \mathcal{H}$.

The idea is then to regard the apparent horizon as the instantaneous surface of a black hole. Notice that, differently from the concept of event horizon, where we need to know our $J^-(\mathscr{I}^+)$ in order to be able to define it, the apparent horizon is a local notion, and as such in principle observable. Basically, the event horizon is drawn by the light rays entering at 45°, bended by gravity, that then stick to the same radius. The apparent horizon, however, has another shape, following the one of $\mathcal{B}^3$ as it can be seen in the following Figure 6.16.

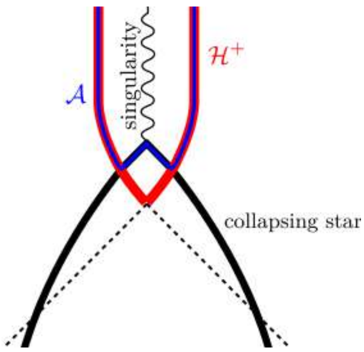


Figure 6.16: Formation of an event horizon (red) and an apparent horizon (blue and red) during a stellar collapse. The dashed lines represent the incoming light rays of the future horizon generators. (2 space dimensions are suppressed).

In practice, the difference between apparent and event horizon can be summarized as follows: the apparent horizon is the boundary of a region through which *at some time* anything can't escape; the event horizon is the boundary of a region through which *at any time* anything can't escape.

6.6.3 Trapping Horizon, Dynamical Horizon, Isolated Horizon

The concept of apparent horizon has recently become a little inadequate, since in this case the notion of a trapped region relies on the choice of Σ: of course one can calculate the expansion θ_l of $\mathcal{T}^2$ knowing only the intrinsic 3-metric q_{ab} and the extrinsic curvature K_{ab} of Σ and one does not need to evolve these quantities away from Σ even locally. In this sense the notion is local to M. However, this locality is achieved at the price of restricting $\mathcal{T}^2$ to lie in Σ and to strongly depends on its choice. However, a foliation of the spacetime in Σ surfaces can be made in different ways. So the existence of a trapped region, not just its shape, can depend on the choice of foliation. This is not so desirable, and in modern times there has been an evolution of the concept of horizon (see [53, 54] for an in depth discussion).

Instead of giving a slicing, we can focus on the structure of the geodesics stemming from the horizon. A **future outer trapping horizon** (FOTH) is a smooth 3-submanifold $\underline{\mathcal{H}}$ of spacetime $\mathcal{M}$ foliated by closed, 2-surfaces $\mathcal{S}$ such that:

1. the expansion of one future-directed null normal to the foliation l^a is zero, $\theta_l = 0$;

2. the expansion of the other future-directed null normal to the surface is negative, $\theta_n < 0$;

3. the Lie derivative along n of the expansion of l is negative, $\mathcal{L}_n \theta_l < 0$.

The first condition is analogous to having *marginally trapped surface*; the second condition tells us that we are considering a *black hole* instead of a white hole; the third condition implies the notion of an *outer* surface.

Consider now the Raychaudhuri equation supplemented by the condition of hypersurface orthogonality of l

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{\theta^2}{2} - |\sigma|^2 + \cancel{|\omega|^2} - R_{ab} l^a l^b; \tag{6.54}$$

using the Einstein equation, $R_{ab} l^a l^b = T_{ab} l^a l^b = 0$, if $\sigma = 0$ (no GW) and initially $\theta = 0$, then also $\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = 0$, which means that this remains null. However, if the black hole is accreting the horizon will be growing, and therefore will be spacelike.

Let us see that an accreting horizon will be locally spacelike [53]. Denote by V a a vector field which is tangential to $\underline{\mathcal{H}}$, everywhere orthogonal to the foliation by marginally trapped surfaces and preserves this foliation. We can take it as a linear combination of the two null vectors $V = l - fn$ for some constant f. Let us normalize l and n so that $l \cdot n = -1$. This implies that $|V|^2 = 2f$ and hence $\underline{\mathcal{H}}$ is respectively, space-like, null or time-like, depending on whether f is positive, zero or negative. Now, even during an accretion event V lies by definition always on $\underline{\mathcal{H}}$ which is also defined by the requirement that $\theta_l = 0$. Hence $\mathcal{L}_V \theta(l) = 0$ which implies $\mathcal{L}_l \theta(l) = f \mathcal{L}_n \theta(l)$. But (6.54) tell us that in absence of exotic matter $\mathcal{L}_l \theta(l) \leq 0$. At the same, θ_l becomes negative as one moves along n to the interior of the marginally trapped surfaces, whence $\mathcal{L}_n \theta(l) < 0$ also. So we conclude that f is non-negative, i.e. that the time-like case is ruled out. Finally, if the flux of non-exotic energy across $\underline{\mathcal{H}}$ is non-zero on some leafs of the foliation of $\underline{\mathcal{H}}$, the right side of (6.54) cannot vanish identically on that leaf. Thus, f will be strictly positive somewhere on each of these leafs, and the horizon will be space-like there.

The definition of a FOTH is not completely intrinsic to $\underline{\mathcal{H}}$: we need to know the flow of a congruence along the other congruence; if we relax this third condition, retaining the rest of the FOTH definition, we have a **dynamical horizon** which is a smooth, three-dimensional, *space-like* $\underline{\mathcal{H}}$ of spacetime $\mathcal{M}$. Basically a dynamical horizon is a space-like 3-manifold which is foliated by closed, marginally trapped 2-surfaces and whose definition is now foliation independent, and intrinsic to $\underline{\mathcal{H}}$ (while in FOTH the condition 3 requires knowledge of the geometry in the n direction away from $\underline{\mathcal{H}}$) and they can be present even in spacetime which do not admit an event horizon (e.g. also in spatially compact spacetimes). Note however, that this is payed at the price to have a a less stringent definition. Indeed, in time dependent situations, if the dominant energy condition holds and the space-time is asymptotically predictable, it is easy to see that dynamical horizons will lie inside the event horizon. However, in the interior of an expanding event horizon, there may be many dynamical horizons. Nonetheless, under fairly general conditions one can associate with each evolving black hole an outer-most or canonical dynamical horizon and it is then natural to focus just on this canonical one.

Finally we can also have an **isolated horizon**, which have the local structure of an event horizon. More precisely an **isolated horizon**, is a 3D submanifold Δ which satisfy

1. Δ is topologically $S^2 \times R$ and null

2. along any null normal field tangent to Δ the outgoing expansion rate θ_l vanishes (the horizon area is constant in time)

3. All field equations holds on Δ and the SET T_{ab} on Δ is such that $-T^a{}_b l^b$ is a future directed causal vector (it is not spacelike) for any null norm l^a

4. the commutator $[\mathcal{L}_l, D_a] V^b = 0$ for any V^b tangent to Δ where D_a is the induced derivative operator on Δ.

A trapping horizon for which there are no flows, no gravitational waves, and so on, becomes an isolated horizon.

Finally, note that, FOTHs, dynamical horizons, isolated horizons are intrinsic "spacetime notions", defined quasi-locally. They are not defined relatively to special space-like surfaces, as in the case of apparent horizons, nor do they require to know the future null infinity of spacetime, as in the case of event horizons. I.e., being quasi-local, they are not teleological.

6.7 Singularity Theorems

Singularity theorems represented a crucial step in our understanding of gravity. In the early sixties of the XX century indeed it was far from clear if the singularity chacracterzing stationary black hole solutions could not be avoided in realistic stellar collapse scenarios for example due to large deviation from spherical or axis symmetry. The first application of topological methods to gravity by Penrose in 1965 was indeed a breakthrough that cannot be overstated and which lead in the following to a new way of proceeding at the base of the subsequent revolution that culminated in the seventies with the discovery of black hole thermodynamics.

As we said, the first of the singularity theorems is due to Penrose [55, 56], in layman words it states that a singularity is the unavoidable endpoint of any gravitational collapse if one does not have exotic matter. To prove his theorem, Penrose starts by assuming classical General Relativity, that energy is always positive, and that there is a star that starts collapsing, all relatively general assumptions. He proves and deduces that if the star collapses too much, then the collapse becomes unstoppable, that black holes are common and that they contain a singularity.

Another type of singularity, however, is represented by the one preceding a Big Bang scenario and also in this case singularity theorems adapted to cosmology were developed [57, 58]; according to these theorems, singularities in cosmological settings arise at the start of the Big Bang scenario, only by assuming that the energy is always positive and that the Universe is expanding right now. From this it can be deduced that there is a singularity in our past and that time had actually a beginning.

Let us focus on the singularity theorem by Penrose. During a collapse, is a singularity inevitable? If the collapse is not isotropic, can a singularity be avoided? All this problems are tackled and solved by the singularity theorems. Penrose used topological tools in General Relativity.

Recapping, our hypothesis will include the fact that we use classical General Relativity, that we assume appropriate energy conditions, and that we have stellar collapse, which means that we will eventually have a trapped region $\mathcal{T}^2$. The thesis is that a singularity forms.

Penrose's Singularity theorem. Assume the following:

1. we have a trapped surface, and there is a maximum expansion associated to this trapped surface, $\theta_{\max} = \theta_0 < 0$;

2. the NEC condition holds: $T_{ab}k^a k^b \geq 0$;

3. General Relativity holds: in this way, we can write

$$R_{ab}k^a k^b = 8\pi \left(T_{ab} - \frac{1}{2}T g_{ab}\right) k^a k^b = 8\pi T_{ab}k^a k^b;$$

4. the manifold $\mathcal{M}$ is a globally hyperbolic spacetime, therefore $\mathcal{M} = \mathbb{R} \times \Sigma^3$, and therefore no Cauchy horizon forms.

Then, a singularity must form.

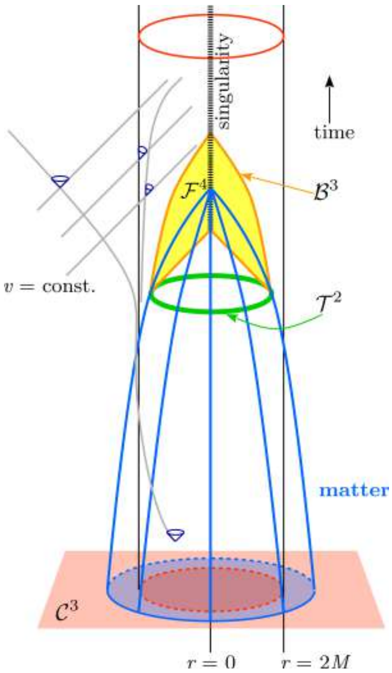


Figure 6.17: A schematic depiction of a spherically symmetric collapse with the essential elements of the singularity theorem demonstration.

Proof: we have $\mathcal{F}^4 = I^+(\mathcal{T}^2)$, $B^3 = \partial F^4 = \dot{I}^+(\mathcal{T}^2)$ [5]. In a stellar, spherically symmetric, collapse there will be a stage in which all the congruences will converge in a point Ω.

Consider the Raychaudhuri equation

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{\theta^2}{2} - R_{ab}k^a k^b; \tag{6.55}$$

using the NEC, we can then write

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} \leq -\frac{\theta^2}{2}, \tag{6.56}$$

and in the best case scenario, this becomes

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{\theta^2}{2}. \tag{6.57}$$

By solving this differential equation, we simply get

$$\theta^{-1} = \theta_0^{-1} + \frac{\lambda}{2}, \tag{6.58}$$

and evidently θ diverges to infinity as λ approaches $-2\theta_0^{-1}$. Therefore, in a finite λ, the expansion reaches $-\infty$, and thus all congruences converge to the point Ω (this is strictly true for a spherically symmetric collapse, in general there could be a extended region where the focussing happens). The question now becomes: is Ω part of $\mathcal{M}$?

Assume that $\Omega \in \mathcal{M}$, which means that Ω is not a singularity. Then $\mathcal{B}^3$ is compact and boundaryless, being generated by a compact system of closed segments, and therefore is a closed submanifold of $\mathcal{M}$.[6]

[5]Note, we adopt here [55] notation but one can alternatively find in textbooks $\mathcal{F}^4 = J^+(\mathcal{T}^2)$, $B^3 = \partial F^4 = \dot{J}^+(\mathcal{T}^2)$.

[6]By a compact manifold we mean that it is topologically compact. A sphere, torous or a disk are topologically compact. A closed manifold is a manifold which is compact but also boundariless. A disk is compact, and it is a closed subset of the plane, but it is not a closed manifold as it has a boundary in the Euclidean plane.

Now, consider a timelike vector field t^μ. It can be shown that such a congruence of timelike curves must exists in a spacetime with a Cauchy hypersurface. Since one and only one curve of the congruence passes through each point of space, and since a timelike curve cannot intersect a null surface such as $\mathcal{B}^3$ more than once (i.e. $\mathcal{B}^3$ is acronal), each integral curve of t^μ will pinch $\mathcal{B}^3$ only once and the same will be true for Σ^3 (which is also acronal).

So t^μ induces a continuous one to one map, a homeomorphism, from points on $\mathcal{B}^3$ to points on Σ^3. Let us call $\mathcal{S}^3$, the image of $\mathcal{B}^3$ generated via this map on Σ^3 (notice that we are assuming that there are no causality violations, since we assumed $\mathcal{M} = \mathbb{R} \times \Sigma^3$). In particular, given that we are dealing with a homeomorphism, this image of B^3 over Σ^3 must be compact.

However, Σ^3 is non-compact, and since $\mathcal{S}^3$ is compact, we have that $\mathcal{S}^3 \subset \Sigma^3$, and therefore $\mathcal{S}^3$ must have a boundary in Σ^3. But if this is the case, $\mathcal{S}^3$ is not boundary-less, and even if it is compact, it is not closed.

But, $\mathcal{S}^3$ is homeomorphic to $\mathcal{B}^3$, which is closed by assumption, so $\mathcal{S}^3$ cannot have a boundary. Therefore we have reached a contradiction, which means that $\Omega \notin \mathcal{M}$, and thus Ω is a singularity. ■

The above theorem was later extended to singularity theorems in cosmology (see e.g. [58]).

Wald provides us with a similar and alternative proof of this theorem.

In particular, assume that $\mathcal{B}^3$ is compact and boundary-less, because of geodesic completeness; $\mathcal{S}^3$ is an image with a one-to-one homeomorphic map of $\mathcal{B}^3$ on Σ^3.

$\mathcal{B}^3$ is compact, hence $\mathcal{S}^3$ is compact; but it is also closed on Σ^3, since $\mathcal{B}^3$ is compact and boundary-less, therefore $\mathcal{S}^3$ must be too. $\mathcal{S}^3$ must be a closed subset of Σ^3. But it is also true that on each point of $\mathcal{B}^3$ there is an open neighbourhood homeomorphic to an open ball. But this must be true also for $\mathcal{S}^3$, because the two are homeomorphic; but this implies that $\mathcal{S}^3$, as a subset of Σ^3, is also open. The only way in which $\mathcal{S}^3$ can be both closed and open on Σ^3 is that $\mathcal{S}^3$ is the same set of Σ^3 (Wald assumes that the spacetime is connected, and therefore only $\mathcal{M}$ and $\varnothing$ are both open and closed). Therefore the image of $\mathcal{B}^3$ must be the whole Cauchy hypersurface.

However $\mathcal{S}^3$ is also supposed to be compact, but Σ^3 is not compact by definition, so it is not possible that $\mathcal{S}^3$ coincides with Σ^3. Then, $\mathcal{S}^3 \subset \Sigma^3$, but then $\mathcal{B}^3$ cannot be compact and boundary-less.

Penrose's singularity theorem can be seen as morally similar to Gödel incompleteness theorems in mathematics, in the sense that it tells us that **General Relativity is an incomplete theory**. This does not necessarily mean that the theory must be quantized, but just that we have a theory, General Relativity, that predicts its own demise, since singularities form behind the event horizon of black holes, and we have very strong, albeit non conclusive, evidences of the existence of General Relativity black holes, with the detection of gravitational waves emitted by sources compatible with black hole binaries (e.g. [59]), and with the image of the black hole M87*, published in April 2019 by the Event Horizon Telescope collaboration, [60].

Note that a more general theorem one was proposed by Hawking and Penrose later on in 1970. According to this theorem, the spacetime $\mathcal{M}$ necessarily contains incomplete timelike or null geodesics which cannot be continued, provided the following conditions are satisfied:

1. The spacetime contains no Closed Timelike Curves (which implies that the spacetime has no problems causality-wise, and therefore there are no Cauchy Horizons)

2. For any timelike vector u^μ, $R_{\mu\nu}u^\mu u^\nu \geq 0$ (meaning that the Strong Energy Conditions applies)

3. For each timelike or null geodesics with tangential vector u^μ there is a point in which $u_{[\alpha}R_{\beta]\gamma\delta[\epsilon}u_{\rho]}u^\gamma u^\delta \neq 0$ (spacetime is of general type with special symmetries)

4. A trapped surface exists.

This theorem applies also when a trapped surface is generated e.g. in a closed universes, where there are no non-compact Σ^3, so that the Penrose theorem can be straightforwardly be applied.

Can the singularity be resolved by quantum gravity? Of course yes, this is after all what we do expect from a quantum theory of gravitation. It is interesting to not that if one also requires this resolution to respect the strong cosmic censorship conjecture a finite number of possibilities are left (see e.g. [61] and reference therein).

7 From Black Hole Mechanics to Black Hole Thermodynamics

Birkhoff's theorem states that any spherically symmetric vacuum solution is static, which effectively implies that it must be Schwarzschild (independently if it is describing the geometry of a black hole or that of the exterior of a spherical object). A generalisation of this theorem to the Einstein–Maxwell system shows that the only spherically symmetric solution is Reissner–Nordström.

However, if the geometry is not spherically symmetric, we have no analogue of Birkhoff's theorem; for example, around the Earth we do not know a priori what is the metric. In principle, we can still write the metric in a weak field approximation – where the metric contains terms like $(1-2\Phi)$ – and essentially expand the gravitational potential Φ in spherical harmonics

$$\Phi = \sum_{j=1}^{\infty}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{D_{jl}^{m}}{r^{j}}Y_{l}^{m}(\theta,\varphi). \tag{7.1}$$

For a general potential, we shall have to determine all these coefficients; however, black holes are simple as they obey the **No Hair Theorem**: if a black hole forms after a sufficient long time all higher multipoles are radiated away, only the lower multipole coefficients will be relevant.

In particular a theorem, due to Carter–Robinson (plus Hawking and Wald), states: If $(\mathcal{M}, g)$ is an asymptotically-flat, stationary, vacuum spacetime that is non-singular on and outside an event horizon, then $(\mathcal{M}, g)$ is a member of the two-parameter family of axisymmetric black hole solution [1]. The parameters being the mass M an the angular momentum J

Consider gravitational waves: the conservation of the SET implies the conservation of mass, linear and angular momentum so the emission happens from the quadrupole moment up. Basically, this tells us that the mass and the angular momentum of the black hole cannot be the source of gravitational waves and are therefore conserved once the black hole is stationary. During the formation of a black hole (in a general collapse or in a merging event), mass and angular momentum are partially shed away (not just by matter loss but also by gravitational waves that carry away both energy and angular momentum), but once the black hole forms they becomes the conserved charges characterising its structure.

What about electric and magnetic fields? The electric field is conserved, since the charge Q is conserved. According to the membrane paradigm [62], the charge sticks on the horizon which from outside can be seems as a thin membrane made up by the infinite overlapping layers of matter (a stretched horizon) which the outside observer can never see crossing the horizon. This matter can carry charge which then becomes for the ouside observer a black hole charge. The magnetic field is

[1] An asymptotically flat spacetime is axisymmetric if there exists a Killing vector field ψ (the "axial" Killing vector field) that is spacelike near ∞ and for which all orbits are closed.

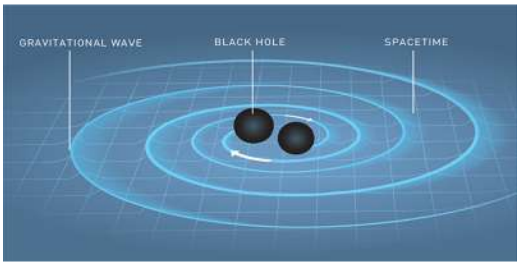


Figure 7.1: Artist impression of the gravitational waves emission by a black hole merger. Credits (Nobel prize pages)

another story: the magnetic monopole P (if it exists) can be attached to the stretched horizon of the black hole as the electric charge, but the magnetic field $\mathbf{B}$ itself is different; it cannot exist in a black hole surrounded by vacuum, as nothing would prevent fields lines from being radiated away as closed loops in the collapse. Hoverer, for a black hole surrounded by a disk of matter, a torus (what sometimes it unceremoniously called a "dirty black hole") field lines can attach on the stretched horizon on one side and the matter on the other. Something very relevant for astrophysical phenomena like the Blandford–Znajek process.

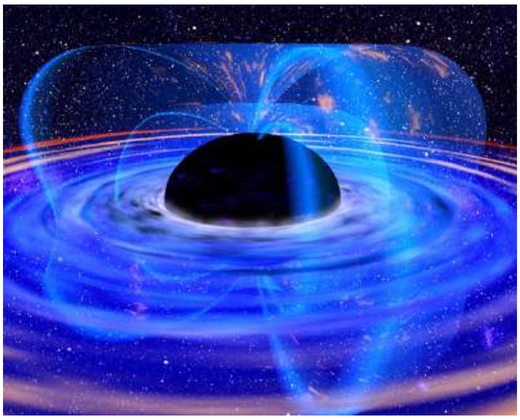


Figure 7.2: Artist impression of the magnetic field lines from the stretched horizon to the accretion disk. Credits Dana Berry (NASA)

7.1 Rotating Black Holes

The generic metric describing a rotating, charged black hole in a vacuum is the **Kerr–Newman metric** and it was found only in 1965, just two years after Roy Kerr had provided the metric for a rotating black hole which is the same as below with $e = 0$:

$$\begin{aligned} \mathrm{d}s^2 = -\left(\frac{\Delta - a^2\sin^2\theta}{\Sigma}\right)\mathrm{d}t^2 - 2a\sin^2\theta\left(\frac{r^2+a^2-\Delta}{\Sigma}\right)\mathrm{d}t\,\mathrm{d}\varphi + \\ + \left[\frac{(r^2+a^2)^2 - \Delta a^2\sin^2\theta}{\Sigma}\right]\sin^2\theta\,\mathrm{d}\varphi^2 + \frac{\Sigma}{\Delta}\,\mathrm{d}r^2 + \Sigma\,\mathrm{d}\theta^2, \end{aligned} \tag{7.2}$$

where

$$\Sigma \equiv r^2 + a^2\cos^2\theta,$$

$$\Delta \equiv r^2 - 2Mr + a^2 + e^2,$$

$$e^2 \equiv Q^2 + P^2 \quad \text{and} \quad a \equiv \frac{J}{M},$$

Note that in astrophysics it is often used a dimensionless spin parameter $\bar{a} = a/M = J/M^2$ (which is often presented simply as a so to confuse relativists!).

The above solution is an electrovac (vacuum apart from electromagnetic field) solution of the Einstein equations. The associated electromagnetic one form potential non-vanishing components are

$$A_t = \frac{Qr - Pa\cos\theta}{\Sigma} \quad \text{and} \quad A_\varphi = \frac{-Qar\sin^2\theta + P(r^2+a^2)\cos\theta}{\Sigma}$$

This is an axisymmetric stationary but not static solution (e.g. you can check it is not symmetric under time reversal due to the $\mathrm{d}t\,\mathrm{d}\phi$ term), it admits a globally defined, non-vanishing, timelike Killing vector outside of the horizon; it is also asymptotically flat, since at infinity this reduces to Minkowski. The coordinate we used are called Boyer–Lindquist coordinates.

Setting $a = Q = P = 0$, we recover the Schwarzschild solution; setting $a = 0$ we recover the Reissner–Nordström solution; setting $P = Q = 0$ we recover the standard Kerr solution for a rotating black hole. All of them are Ricci flat solutions. It is sometimes argued that Kerr–Newman black holes cannot exist in nature: an electrically charged black hole can selectively attract charge of the opposite sign, and this can neutralize the electric charge of the black hole. Therefore, Kerr–Newman black holes should not exist (even though this is still disputed) and astrophysical black holes should be Kerr ones. From now one we shall focus on this solution then, albeit we shall provide sometimes the generalisation of important results to the full Kerr-Newman case.

7.1.1 A ring Singularity

Let us try to see if there are singularities of the metric.

- In $\theta = 0$ (i.e. on the axis of rotation of the solution) we have a singularity but this is obviously the standard singularity at the poles of our polar coordinate set: changing coordinates we can easily eliminate it.

- Another interesting point is at $\Sigma = 0$. In this case, there are two possible solutions; at $r = 0$ and $\theta = \frac{\pi}{2}$ we could have a problem. What is the shape of this singularity? It looks like a point, but only in the equator

Let us try to analyse this in another set of coordinates, the Kerr–Schild coordinates, in which the metric acquires the form $g_{\mu\nu} = \eta_{\mu\nu} + f(r,z)k_\mu k_\nu$, where k_μ is a null vector. The coordinate transformation is defined as

$$\begin{cases} x + iy = (r+ia)\sin\theta\exp\left[i\int\left(\mathrm{d}\varphi + \frac{a}{\Delta}\,\mathrm{d}r\right)\right] \\ z = r\cos\theta \\ \hat{t} = \int\left(\mathrm{d}t + \frac{r^2+a^2}{\Delta}\,\mathrm{d}r\right) - r \end{cases} \tag{7.3}$$

with the radius given implicitly by the equation

$$r^4 - \left(x^2+y^2+z^2-a^2\right)r^2 - a^2z^2 = 0.$$

In these coordinates the Kerr metric has the form

$$\mathrm{d}s^2 = -\,\mathrm{d}\hat{t}^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + \frac{2Mr^3}{r^4+a^2z^2}\left[\mathrm{d}\hat{t} + \frac{r}{r^2+a^2}(x\,\mathrm{d}x + y\,\mathrm{d}y) - \frac{a}{r^2+a^2}(x\,\mathrm{d}y - y\,\mathrm{d}x) + \frac{z\,\mathrm{d}z}{r}\right]^2, \tag{7.4}$$

With this line element one can recognize that

$$f(r,z) = \frac{2Mr^3}{r^4+a^2z^2} \quad \text{and} \quad (k_{\hat{t}}, k_x, k_y, k_z) = \left(1, \frac{rx+ay}{r^2+a^2}, \frac{ry-ax}{r^2+a^2}, \frac{z}{r}\right). \tag{7.5}$$

Let us start noticing that in the limit $M \to 0$ we recover Minkowski as expected (this is true in any coordinate system but in the Kerr–Schild one is straightforward to recognise the left over metric). Moreover, in this coordinate systems it is evident that surfaces of constant r are cofocal ellipsoids, indeed from the coordinates definitions it is easy to see that

$$x^2 + y^2 = (r^2 + a^2)\sin^2\theta \quad \text{and} \quad z^2 = r^2\cos^2\theta \tag{7.6}$$

so that

$$x^2 + y^2 = (r^2 + a^2)\left(1 - \frac{z^2}{r^2}\right) \Rightarrow \frac{x^2 + y^2}{r^2 + a^2} + \frac{z^2}{r^2} = 1, \tag{7.7}$$

We can easily see from (7.6) that as $r \to 0$ these surfaces degenerate to a disk: at $r = 0$ we have $z = 0$ and $x^2 + y^2 \leq a^2$ (as $\sin^2\theta \leq 1$). At the singular point $r = 0$, $\theta = \frac{\pi}{2}$, we get

$$x^2 + y^2 = a^2, \tag{7.8}$$

which is the equation of a circle: the singular point is actually resolved in these coordinates into a ring on the equatorial plane. An investigation of the causal structure of Kerr shows that this is indeed a timelike singularity, we hence expect to find an associated Cauchy horizon.

7.1.2 Horizons

Consider now the singularity at $\Delta = 0$: this implies (in the case of astrophysical black holes, $e = 0$)

$$r^2 - 2Mr + a^2 = 0, \tag{7.9}$$

which admits solutions

$$r_\pm = M \pm \sqrt{M^2 - a^2}. \tag{7.10}$$

At $a = 0$, we get $r_+ = 2M$, so we recover the event horizon of a Schwarzschild black hole. However, differently from the Schwarzschild case, here we have in general two horizons. In particular, we have some special cases.

If $a^2 > M^2$, we would have no horizon, since $r_\pm$ would acquire an imaginary part; however, we would still have the singularity at $\Sigma = 0$; this means we would have a **naked singularity**. According to the Cosmic Censorship Conjecture, naked singularities should not exist in the Universe: any singularity should be protected by an event horizon. In case a naked singularity existed, we would have strong effects of quantum gravity exposed to us, and even worse, because we would have serious problems with causality.

Another possibility is given by $a^2 = M^2$: then we would have $r_+ = r_- = M$, and this is called an extremal black hole, the maximally rotating black hole. They are also considered unphysical: in astrophysics there is the so-called Thorne limit, $a \lesssim 0.998M$ [2], that is close to the extremal limit. Extremal black holes have also zero surface gravity and, therefore, have a zero Hawking temperature; black hole thermodynamics, we shall see later on, implies via its third law that it should be impossible to turn a non-extremal black hole into an extremal one in a finite number of thermodynamical steps.

Consider, then, the standard Kerr black hole case $a^2 < M^2$. We now want to see if the singularity at $r_\pm$ is really a coordinate artefact. We can introduce so-called Kerr coordinates, horizon penetrating coordinates, similar to ingoing Eddington–Finkelstein coordinates, defined w.r.t. BL coordinates as (see Se.g. Box 33.2 of [41]. See also [63] for an extensive pedagogical introduction to the Kerr geometry).

$$\begin{cases} \mathrm{d}v = \mathrm{d}t + \frac{r^2 - a^2}{\Delta}\,\mathrm{d}r \\ \mathrm{d}\tilde{\varphi} = \mathrm{d}\varphi + \frac{a}{\Delta}\,\mathrm{d}r, \end{cases} \tag{7.11}$$

[2] Thorne in 1974 recognized that thin discs radiate and as a result, the black hole spin would be upper bounded because the black hole would preferentially swallow negative angular momentum photons emitted by the accretion flow. A number of studies since then have suggested that this limit may not hold (or that the limiting spin may be higher) if accretion takes place via a geometrically thick flow.

and with these we can rewrite the Kerr metric as

$$\begin{aligned}\mathrm{d}s^2 = &-\frac{(\Delta - a^2\sin^2\theta)}{\Sigma}\,\mathrm{d}v^2 + 2\,\mathrm{d}v\,\mathrm{d}r - 2\theta\sin^2\theta\frac{r^2+a^2-\Delta}{\Sigma}\,\mathrm{d}v\,\mathrm{d}\tilde{\varphi} + \\ &- 2a\sin^2\theta\,\mathrm{d}\tilde{\varphi}\,\mathrm{d}v + \frac{(r^2+a^2)^2 - \Delta a^2\sin^2\theta}{\Sigma}\sin^2\theta\,\mathrm{d}\tilde{\varphi}^2 + \Sigma\,\mathrm{d}\theta^2,\end{aligned} \tag{7.12}$$

which is perfectly fine at $\Delta = 0$. Note also that for $a \to 0$ gives the Schwarzschild metric in ingoing $(v, r, \theta, \tilde{\varphi})$ EF coordinates

$$\mathrm{d}s^2 = -\left(1 - \frac{2M}{r}\right)\mathrm{d}v^2 + 2\,\mathrm{d}v\,\mathrm{d}r + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\tilde{\varphi}^2). \tag{7.13}$$

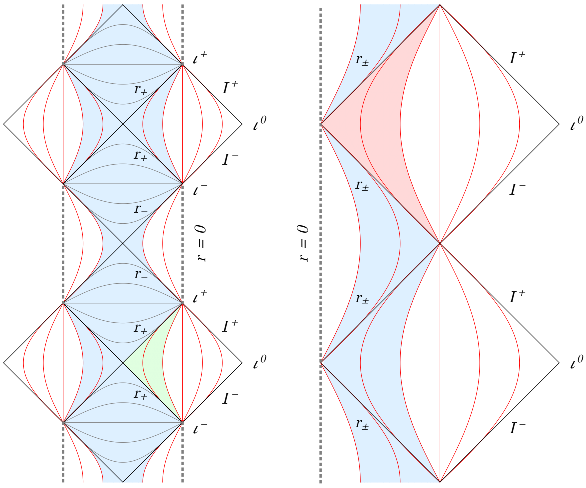


Figure 7.3: Carter–Penrose diagrams of non-extremal (left) and extremal (right) Kerr black hole (from [64]). For further details about the conformal structure of Kerr, see also [58, p. 165].

7.1.3 Killing vectors

Consider now the symmetries of the metric: this admits two self-evident Killing vectors, $\xi \equiv \frac{\partial}{\partial t}$ and $\psi \equiv \frac{\partial}{\partial \varphi}$, since the metric is independent from t and φ. The norm of ξ at $r = r_+$ is given by

$$|\xi|^2_{r=r_+} = g_{00}|_{r_+} = \frac{a^2\sin^2\theta}{r_+^2 + a^2\cos^2\theta} \neq 0; \tag{7.14}$$

however, the surfaces at both $r = r_\pm$ are null surfaces, but since ξ is not a null Killing vector field at the horizon, ξ is not a generator of the horizons. Moreover, we see that at the horizon ξ is spacelike, while it is timelike at infinity (Kerr is asymptotically Minkowski), which means that there must be a region where ξ becomes null. Finally, since ψ is spacelike spacelike at r_+, therefore neither ψ can be a generator of the horizon.

We can nonetheless introduce a Killing vector $\Xi_\pm$ which is a linear combination of ξ and ψ which is null on r_+:

$$\Xi_\pm \equiv \xi + \frac{a}{r_\pm^2 + a^2}\psi = \frac{\partial}{\partial t} + \Omega_H^\pm \frac{\partial}{\partial \varphi}. \tag{7.15}$$

One can easily check that $|\Xi_\pm|^2_{r_\pm} = 0$; also, it can be proven that the surface gravity of the Kerr black hole for both the horizons is given by

$$\kappa_\pm = \frac{r_\pm - r_\mp}{2(r_\pm^2 + a^2)}. \tag{7.16}$$

The Ω_H defined above, $\Omega_H \equiv \frac{a}{r_\pm^2+a^2}$, is the angular velocity of the horizon. In fact, we have

$$\Xi^\mu \nabla_\mu(\varphi - \Omega_H t) = (\xi^\mu \partial_\mu + \Omega_H \psi^\mu \partial_\mu)(\varphi - \Omega_H t) = \frac{\partial \varphi}{\partial t} - \Omega_H \frac{\partial t}{\partial t} + \Omega_H \frac{\partial \varphi}{\partial \varphi} - \Omega_H^2 \frac{\partial t}{\partial \varphi} = 0, \tag{7.17}$$

which means that $\varphi - \Omega_H t = \text{const}$ along the flow defined by Ξ^μ and $\varphi = \Omega_H t$ are orbits of Ξ. We deduce that particles following these orbits rotate with angular velocity Ω_H w.r.t. static observers at infinity (whose worldlines are instead orbits of ξ). Since null generators of the EH follows the orbits of Ξ_+ then we can say that the EH rotates w.r.t. to asymptotic observers at infinity with angular velocity Ω_H

7.2 Ergosphere and Penrose Process

Let us now find the region where $|\xi|^2 = 0$; this happens where

$$-\left(1 - \frac{2Mr}{r^2 + a^2\cos^2\theta}\right) = 0 \qquad \Longrightarrow \qquad r^2 + a^2\cos^2\theta - 2Mr = 0, \tag{7.18}$$

which has solution $r_{\mathrm{E}} = M \pm \sqrt{M^2 - a^2\cos^2\theta}$; since we are interested into values of r outside r_+, we consider only the solution with the plus sign, and therefore we find that the radius of the region in which ξ becomes spacelike, called **ergosphere**, is

$$r_{\mathrm{E}} = M + \sqrt{M^2 - a^2\cos^2\theta}. \tag{7.19}$$

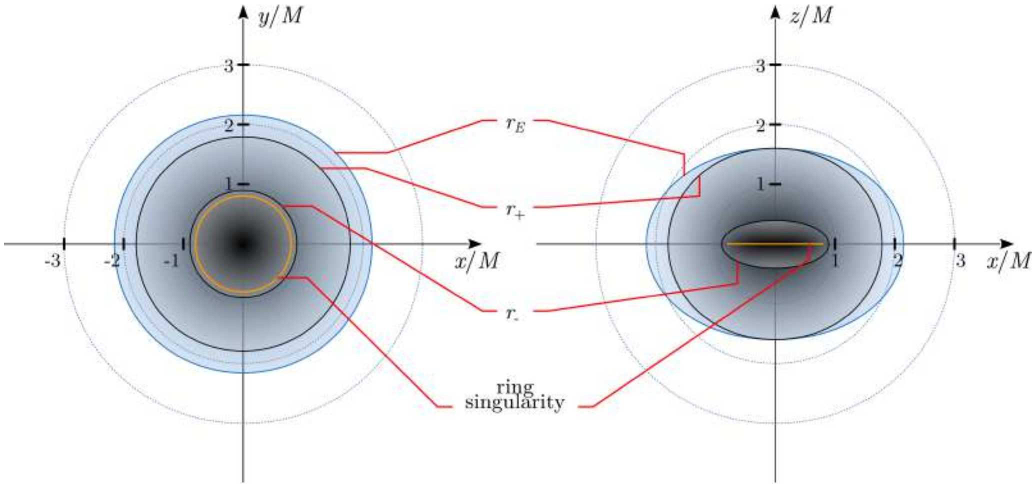


Figure 7.4: Representation of the structure of a Kerr black hole with $a = 0.8M$ in Kerr–Schild coordinates at constant time; the left portion represents the slice at $z = 0$ (face-on, top view), while the right portion represents the slice at $y = 0$ (edge-on, side view); r_E is the border of the outer ergoregion, r_+ is the outer horizon, r_- is the inner horizon. In orange, the ring singularity of Kerr. Notice that, in the side view, both r_+ and r_- are ellipses in Kerr–Schild coordinates, while r_E is not.

The fact that ξ becomes spacelike means that, inside the ergosphere, one cannot remain still: usually, we follow trajectories along $\frac{\partial}{\partial t}$, but we can do that only if $\frac{\partial}{\partial t}$ is timelike. Therefore, inside the ergosphere, we would be forced to rotate with the black hole, we would be dragged by the spacetime. Notwithstanding the strange nature of the surface at r_{E}, this is not an horizon, since it is always possible to spiral out of the ergosphere.

If we have a Killing vector and we contract it with the four-momentum, we find a constant along geodesics, $p^\mu \chi_\mu = \text{const}$. If $\chi = \xi$, this constant is the energy

$$E = -p^\mu \xi_\mu; \tag{7.20}$$

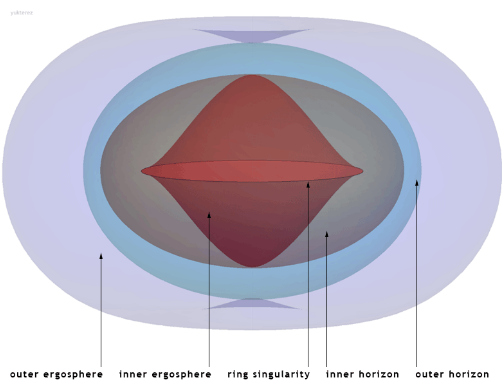


Figure 7.5: Location of the horizons, ergospheres and the ring singularity of the Kerr spacetime in Cartesian Kerr-Schild coordinates. From wikipedia.org

if $\chi = \psi$, the constant conserved along the geodesic motion is the angular momentum

$$L_\phi = p^\mu \psi_\mu. \tag{7.21}$$

Now, since ξ is spacelike inside the ergosphere, it is possible to have negative energy states in this region, $E < 0$. Now we can realize that, since negative energy states exist outside the horizon, we can sort of "mine out" energy from the black hole. This is called **Penrose process**.

Consider a material particle with energy E_0 approaching a Kerr black hole. Once it goes inside the ergoshpere, suppose that this particle splits into two particles, one of energy E_1 escaping the black hole and one of energy E_2 going to the black hole. Is it possible to have a situation in which $E_1 > E_0$? We have

$$E_2 = -p_2 \cdot \xi|_{r<r_{\mathrm{E}}}, \qquad E_1 = -p_1 \cdot \xi > 0, \qquad E_0 > 0. \tag{7.22}$$

The fact that ξ becomes spacelike inside the ergoregion implies that for $r < r_{\mathrm{E}}$ we can choose orbits where $E_2 < 0$; in this case we have that, once the particle goes inside the black hole, we have $E_{\mathrm{BH}} \to E_{\mathrm{BH}} - |E_2|$, and therefore $E_1 > E_0$. How is this possible? Who "pays the bill"?

Consider the Killing vector generating the horizon contracted with the momentum of the infalling particle at $\mathcal{H}^+$. Being Ξ timelike outside the horizon one has

$$\begin{aligned} -p_2 \cdot \Xi &\geq 0 \\ &= -p \cdot \xi - \Omega_H p \cdot \psi = E_\xi - \Omega_H L_\psi, \end{aligned} \tag{7.23}$$

and therefore $L_\psi \leq \frac{E_\xi}{\Omega_H}$. Hence, if and when the Penrose process is realized, we have $E_\xi < 0$, which in turn implies $L_\psi < 0$, and therefore the infalling particle carries both negative energy and negative angular momentum.

Now, suppose that the infalling particle makes the black hole change its mass from M to $M + \delta M$ and its angular momentum from J to $J + \delta J$. We have $\delta M = E_2 < 0$ and $\delta J = L_2 < 0$. Also, $\delta J \leq \frac{\delta M}{\Omega_H}$, and if we write explicitly the angular momentum of the black hole

$$\Omega_H = \frac{J}{2M^3 + 2M\sqrt{M^4 - J^2}}, \tag{7.24}$$

we immediately get

$$\delta J \leq 2M\left(M^2 + \sqrt{M^4 - J^2}\right)\frac{\delta M}{J}. \tag{7.25}$$

This must be true for each step of the Penrose process, and can be rearranged as

$$\delta\left(M^2 + \sqrt{M^4 - J^2}\right) \geq 0; \tag{7.26}$$

we can then define the **irreducible mass** of a rotating black hole as

$$M_{\rm irr}^2 \equiv \frac{1}{2}\left(M^2 + \sqrt{M^4 - J^2}\right). \tag{7.27}$$

We immediately see that if $J \to 0$, the irreducible mass reduces to the mass M of the black hole; also, by inverting the relation, we can see that

$$M^2 = M_{\rm irr}^2 + \frac{J^2}{4M_{\rm irr}} \geq M_{\rm irr}^2. \tag{7.28}$$

Therefore, the Penrose process continues as long as there is angular momentum, and when this is no longer the case, the black hole mass equates the irreducible mass (from here the name "irreducible").

The maximum energy we can extract using the Penrose process is therefore $E_{\rm max} = M_0 - M_{\rm irr}$; therefore, we can determine the efficiency of the Penrose process:

$$\frac{M_0 - M_{\rm irr}}{M_0} = 1 - \frac{M_{\rm irr}}{M_0} = 1 - \left(\frac{M_0^2 + \sqrt{M_0^4 - J^2}}{2M_0^2}\right)^{1/2} \overset{a=1}{=} 1 - \frac{1}{\sqrt{2}} \simeq 29\%. \tag{7.29}$$

Let us now try to determine the area of a black hole; this is given by

$$A_{\rm BH} = \int_{r_+} \sqrt{g_{\theta\theta} g_{\varphi\varphi}}\, \mathrm{d}\theta\, \mathrm{d}\varphi = 4\pi\left(r_+^2 + a^2\right) = 16\pi M_{\rm irr}^2; \tag{7.30}$$

therefore, since we know that $\delta M_{\rm irr} \geq 0$, we also know that $\delta A_{\rm BH} \geq 0$. Hence, the Penrose process can go on as long as the black hole area keeps on increasing.

This is a special case of a theorem by Hawking – known as the **Area Law** – that states that, if matter satisfies the WEC, in any physical process, $\delta A_{\rm BH} \geq 0$. And this is in analogy with thermodynamics: energy extraction, in thermodynamics, is related to entropy which, in any physical process, cannot decrease, while here it is the area of a black hole that cannot decrease. We will delve deeper into this concept in later on.

7.2.1 Superradiance

Is there an analogue of the Penrose process in a field-theoretical context? The answer is still yes, and in this case it is called **superradiance** [65]. We send a wave toward a black hole, and this wave is partly transmitted with a coefficient I and partly reflected with a coefficient R. What happens is that $|R|^2 > |I|^2$ and the wave gets amplified. If we have a scalar field wave of equation $\phi = \phi_0 \cos(\omega t - m\varphi)$, superradiance is realised when

$$0 < \omega < m\Omega_H \ \ (+e\phi_H). \tag{7.31}$$

where in parenthesis we have added also the contribution in the case a Kerr–Neuman black hole The average power loss is, then

$$P = \frac{1}{2}\phi_0^2 A_{\rm BH}\omega\left(\omega - m\Omega_H\right). \tag{7.32}$$

Note that, like the Penrose process, superradiance has to be there in order not to decrease the area of a black hole in the above described scattering process. Indeed, the existence of superradiance can also be derived from Hawking's Area Theorem (see e.g. page 329 [1]).

Another thing to keep in mind is that fermions do not superradiate [66]: this is a consequence of the fermions SET violating the WEC which is a basic assumption necessary to demonstrate the Area theorem. However, superradiance applies to scalar fields, electromagnetic field and tensor fields, like gravitational waves.

A final remarks. Is there an ergoregion in a Schwarzschild geometry? Of course there is, but it is inside the event horizon. In fact, we already saw that, inside the event horizon of a Schwarzschild black hole, the Killing vector associated to time translations becomes spacelike. Also, notice that in the limit for $J \to 0$, the radius of the ergosphere r_{E} approaches the radius of the event horizon of a Schwarzschild black hole. Therefore, the negative energy states are also present in Schwarzschild geometry, however they cannot be extracted from behind the horizon, at least via classical processes. This is another hint suggesting to look into quantum effects in curved spacetime.

7.3 Black Hole Thermodynamics

In 1970, Wheeler posed a peculiar question: what happens if we throw something into a black hole? Is the second law of thermodynamics violated? How do the laws of thermodynamics work in presence of a black hole?

In 1971-72, we have the realization of the Penrose process, Christodoulou and Ruffini introduce the concept of irreducible mass and Hawking formulates the Area Law.

In 1973, Bekenstein was the first to make the conjecture that there was a connection between entropy of a black hole and area of a black hole, as well as between the surface gravity of the black hole and the temperature of the black hole.

All this led to the formulation of a generalized second law, which basically states that

$$\delta\left(S_{\text{out}} + \eta\frac{A_{\text{BH}}}{\ell_{\text{P}}^2}\right) \geq 0, \tag{7.33}$$

with η a constant.

After Bekenstein, in a seminal paper Bardeen, Carter and Hawking (1973) [67] conjectured that the four laws of black hole mechanics had a direct link with the laws of thermodynamics. [3] Nonetheless, up to this point, there was no clear notion of a temperature for black holes, albeit the laws were strongly suggesting that the surface gravity κ had to be seen as the analogue of the black hole temperature T.

In 1974-75, Hawking finally found that by cleverly applying quantum field theory on black holes spacetimes [4] indeed they would radiate with a temperature determined by their surface gravity

$$T_H = \frac{\kappa}{2\pi}, \tag{7.34}$$

and from there everything was consistent. However, let us stress that the four laws are classical laws, while Hawking radiation is a quantum effect. It looks like, someway, somehow, general relativity knows about quantum field theory...

7.3.1 Null hypersurfaces, Surface integrals and other tools

Black hole thermodynamics is also called horizon thermodynamics. Horizons of stationary black holes are 3-d null hypersurfaces. As such, let us elaborate better about how to describe their structure more precisely.

First of all, any hypersurface can be defined or by requiring the constancy of some function of the coordinates over it, or equivalently by describing it by some parametric equations of the form $x^\alpha = x^\alpha(y^a)$ where y^a (with $a = 1, 2, 3 = v, \theta, \phi$ for the specific null case we are considering). [5] Noticeably, we can easily construct in this case a set of tangent vectors to the hypersurface. Indeed, for the vectors

$$e^\alpha_a = \frac{\partial x^\alpha}{\partial y^a} \tag{7.35}$$

one has $e^\alpha_a k_\alpha = 0$ where k_α is the orthogonal vector to the hypersurface.

Also, the metric intrinsic to a hypersurface Σ can be obtained by restricting the line element to displacements along the hypersurface.

$$ds^2_\Sigma = g_{\alpha\beta}\, \mathrm{d}x^\alpha\, \mathrm{d}x^\beta = g_{\alpha\beta}\left(\frac{\partial x^\alpha}{\partial y^a}\, \mathrm{d}y^\alpha\right)\left(\frac{\partial x^\beta}{\partial y^b}\, \mathrm{d}y^\beta\right) = h_{ab}\, \mathrm{d}y^a\, \mathrm{d}y^b \tag{7.36}$$

[3] Note, however, that the third law was rigorously proved by Werner Israel only in 1986 [68].

[4] Let us stress that, as we shall see soon, in order to recover Hawking radiation that there is no need for quantum gravity, only QFT in curved spacetime

[5] For example, a two-sphere in a 3d flat space can be described either by the condition $\Phi(x, y, z) : x^2 + y^2 + z^2 = R^2$ where R is the sphere radius, or we can define it by providing the relation between the x^α coordinates and those on the surface, i.e. $x = R\sin\theta\cos\phi$, $y = R\sin\theta\sin\phi$, $z = R\cos\theta$.

where $h_{ab} \equiv g_{\alpha\beta} e^\alpha_a e^\beta_b$ is of course the induced metric (first fundamental form) of Σ.

Let us now specialise to a null hypersurface. When dealing with null hypersurfaces it is often convenient to introduce a system of coordinates adapted to their generators: let us then introduce as one of the coordinates on the horizon a (not necessarily affine) parameter v along a generator. For each $v =$ const, we have a 2-d spatial section of the horizon $\mathcal{S}^2$. We shall then choose the other two coordinates as θ^A, with $A = 2, 3$, i.e. the two spatial coordinates on $\mathcal{S}^2$. These are constant on each generator of $\mathcal{H}^+$ and span the two dimensional surface orthogonal to it. Basically, moving along v corresponds to move along a generator of $\mathcal{H}^+$ while moving along a θ^A corresponds to move from one generator to another. Let us now consider the vectors tangent to $\mathcal{H}^+$

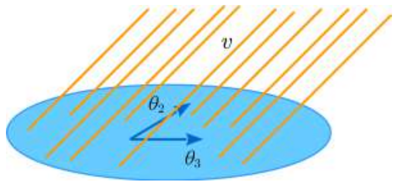


Figure 7.6: Pictorial representation of the null congruence generating $\mathcal{H}^+$ and a section.

$$\Xi^\alpha \equiv \left(\frac{\partial x^\alpha}{\partial v}\right)_{\theta^A}, \qquad \ell^\alpha_A \equiv \left(\frac{\partial x^\alpha}{\partial \theta^A}\right)_v, \tag{7.37}$$

and indeed $\Xi_\alpha \ell^\alpha_A = 0$ and furthermore $\mathcal{L}_\Xi \ell^\alpha_A = 0$ in stationary black hole solutions. In a Kerr black hole, we already know that $\Xi^\alpha = \xi^\alpha + \Omega_H \psi^\alpha$. Note that in this case $e^\alpha_1 = \Xi^\alpha$ and hence $h_{11} = g_{\alpha\beta}\, e^\alpha_1 e^\beta_1 = g_{\alpha\beta}\, \Xi^\alpha \Xi^\beta = 0$. So that in this case the induced metric reduces to

$$ds^2_\Sigma = \gamma_{AB}\, \mathrm{d}\theta^A\, \mathrm{d}\theta^B\,, \tag{7.38}$$

where $\gamma_{AB} \equiv g_{\alpha\beta} \ell^\alpha_A \ell^\beta_B$.

However, there are two bundles of null geodesics orthogonal to each horizon slice S^2. If Ξ^μ is the tangent vector to the first null congruence (the one associated with $\mathcal{H}^+$), we can define an auxiliary null vector N^ν and choose the normalization in such a way that $\Xi^\mu N_\mu = -1$. With this base of vectors we can in general decompose the inverse metric in 4 dimensions in a way such that [6].

$$g^{\alpha\beta} = -\left(\Xi^\alpha N^\beta + N^\alpha \Xi^\beta\right) + \gamma^{AB} \ell^\alpha_A \ell^\beta_B. \tag{7.39}$$

Remember, that the conditions $N^\nu N_\nu = 0$ and $\Xi^\mu N_\mu = -1$ are not sufficient to uniquely determine N^ν and hence the above decomposition. However, as in the case of the Raychaudhury equation physical quantities will come out to be independent of it.

7.3.2 Surface element

Normally, when dealing with the integral of a total divergence we apply the Stokes theorem

$$\int \mathrm{d}^4x \sqrt{g}\, \nabla^\alpha K_\alpha = \oint \mathrm{d}^3x \sqrt{h} n^\alpha K_\alpha = \oint \mathrm{d}\Sigma^\alpha K_\alpha. \tag{7.40}$$

where $\mathrm{d}\Sigma^\alpha \equiv \mathrm{d}^3x \sqrt{h}\, n^\alpha = n^\alpha\, \mathrm{d}\Sigma$ is called the **directional/vectorial surface element** that points in the direction of increasing values of the function whose constancy determines each hypersurface. However, in the case of a null hypersurface, this is problematic, because $h = 0$.

Hence, in order to generalize the above discussion to null surfaces we have to revert to a more fundamental description of the directional (hyper)-surface element: indeed, using the Levi–Civita

[6]For a non-null hypersurface the analogue decomposition would be $g^{\alpha\beta} = \sigma n^\alpha n^\beta + \gamma^{AB} \ell^\alpha_A \ell^\beta_B$

tensor $\epsilon_{\mu\alpha\beta\gamma}$ and the previously introduced basis of tangent vectors (7.35) one can write the surface element as

$$\mathrm{d}\Sigma_\mu = \epsilon_{\mu\alpha\beta\gamma}\mathbf{e}_1^\alpha \mathbf{e}_2^\beta \mathbf{e}_3^\gamma \,\mathrm{d}^3 y; \tag{7.41}$$

It can be shown (see e.g. [3] pages 64-65) that this is the correct volume element independently from the nature of the hypersurface. In particular, for timelike or spacelike hypersurfaces it is equivalent to $\mathrm{d}\Sigma_\mu = \sigma n_\mu \,\mathrm{d}\Sigma$, where again $\sigma = n^\mu n_\mu = \pm 1$ if the hypersurface is timelike or spacelike respectively.

Now, remember that in our, null, case, we have that $\mathbf{e}_1^\alpha = \Xi^\alpha$, $\mathrm{d}^3 y = \mathrm{d}v\,\mathrm{d}^2\theta$, and therefore the directional line element becomes

$$\mathrm{d}\Sigma_\mu = \Xi^\nu \,\mathrm{d}S_{\mu\nu}\,\mathrm{d}v, \tag{7.42}$$

where $\mathrm{d}S_{\mu\nu} = \epsilon_{\mu\nu\beta\gamma}\mathbf{e}_2^\beta \mathbf{e}_3^\gamma \,\mathrm{d}^2\theta$ is the bi-directional surface element for the two-surface.

Using the fact that $\epsilon_{\mu\nu\beta\gamma}\mathbf{e}_2^\beta \mathbf{e}_3^\gamma$ is orthogonal to any e_A^α and that it is antisymmetric in μ, ν, one can show [3] that $\mathrm{d}S_{\mu\nu}$ can be rewritten as the 2-d volume element $\mathrm{d}S \equiv \sqrt{\gamma}\,\mathrm{d}^2\theta$ times the antisymmetrized product of the two normals to this 2-surface, i.e. Ξ and N. Hence

$$\mathrm{d}S_{\mu\nu} = 2\,\Xi_{[\mu}N_{\nu]}\sqrt{\gamma}\,\mathrm{d}^2\theta\,. \tag{7.43}$$

Finally, combining (7.42) with (7.43), one then gets that the directional 3d surface element for a null hypersurface is (use $\Xi^\nu\Xi_\nu = 0$ and $\Xi^\nu N_\nu = -1$)

$$\mathrm{d}\Sigma_\mu = -\Xi_\mu \sqrt{\gamma}\,\mathrm{d}^2\theta\,\mathrm{d}v \tag{7.44}$$

Let us now apply these concepts to a proper definition of mass and angular momentum in stationary black hole spacetimes.

7.3.3 Komar Mass and the Smarr formula

In general, defining an energy in General Relativity is not simple; however, we have seen that the presence of Killing vectors in a spacetime is associated to conserved quantities.

Consider then the vector Q^a defined as $Q^a \equiv \xi_b R^{ab}$ where ξ is some Killing vector. We can see that

$$\nabla_a Q^a = \cancel{R^{ab}\nabla_a\xi_b} + \xi^b\nabla_a R^a{}_b = \xi^b\nabla_a\left(G^a{}_b + \frac{1}{2}g^a{}_b R\right) \overset{\text{Bianchi}}{=} \frac{1}{2}\xi^b\partial_b R = 0, \tag{7.45}$$

where the first term cancels due to the Killing equation, we made use of the contracted Bianchi identity $\nabla_a G^a{}_b = 0$, and the last equality is given by the fact that we are basically performing a Lie derivative along the isometry defined by ξ of a scalar built from the metric $g_{\mu\nu}$, which is invariant under ξ. Therefore, $\nabla_a Q^a = 0$, and Q^a is a conserved current.: i.e we have for non-null hypersurfaces

$$0 = \int_{\mathcal{M}} \nabla_a Q^a \sqrt{g}\,\mathrm{d}^4 x = \oint_{\partial\mathcal{M}} \sigma Q^a n_a \sqrt{h}\,\mathrm{d}^3 y = \int_{\Sigma_2} \sigma\,\mathrm{d}\Sigma_a\, Q^a - \int_{\Sigma_1} \sigma\,\mathrm{d}\Sigma_a\, Q^a, \tag{7.46}$$

as long as Q^a goes sufficiently fast to zero at ι^0. Therefore, we can define a conserved quantity I proportional to the above integral

$$I \propto \int_\Sigma \sigma\,\mathrm{d}\Sigma_a\, Q^a \tag{7.47}$$

Let us now specialise to the, for us, more relevant case in which we take a family of spacelike hypersurfaces Σ_i to foliate a stationary spacetime with a horizon. In this case we can write (remember the Killing vectors' property $R_{\mu\sigma}\xi^\sigma = \nabla_\nu\nabla_\mu\xi^\nu$ (see (2.86))

$$I \propto -\int_\Sigma \mathrm{d}\Sigma_\mu\, \xi^\nu R^\mu{}_\nu = -\int_\Sigma \mathrm{d}\Sigma_\mu\, \nabla_\nu\left(\nabla^\mu\xi^\nu\right) = -\oint \mathrm{d}^2\theta\sqrt{\gamma^{(2)}}\, n_{[\mu}r_{\nu]}\nabla^\mu\xi^\nu; \tag{7.48}$$

where n^μ is the timelike vector orthogonal to the three-dimensional spacelike hypersurface setting the directionality of $\mathrm{d}\Sigma_\mu$, while r^μ is a spacelike vector orthogonal to the two-sphere $\mathcal{S}^2$ at the boundary of Σ introduced by the application of the Stokes' theorem to remove the total derivative. The antisymmetrization is induced by the contraction with $\nabla^\mu \xi^\nu$ (Killing vector).

If we take such boundary $\mathcal{S}^2$ to be a section of $\mathcal{H}^+$ it is easy to see that the two null vectors Ξ^α and N^α we defined before can related to n^α and r^α as

$$\Xi^\alpha = \frac{1}{\sqrt{2}}(n^\alpha + r^\alpha), \qquad N^\alpha = \frac{1}{\sqrt{2}}(n^\alpha - r^\alpha). \tag{7.49}$$

With this prescription, we have that I is proportional to

$$I \propto -\oint_{\mathcal{S}} \mathrm{d}^2\theta \sqrt{\gamma^{(2)}}\, \Xi_{[\mu} N_{\nu]}\, \nabla^\mu \xi^\nu, \tag{7.50}$$

and by Equation (7.43), we can rewrite the integral in the much simpler form

$$I \propto -\frac{1}{2} \oint_{\mathcal{S}} \nabla^\mu \xi^\nu \,\mathrm{d}S_{\mu\nu}. \tag{7.51}$$

Now, consider the Kerr geometry; we have two Killing vectors, $\xi = \frac{\partial}{\partial t}$ and $\psi = \frac{\partial}{\partial \varphi}$ and therefore we have two associated conserved quantities. In particular, we can define the **Komar conserved mass**

$$M = -\frac{1}{8\pi G_N} \oint_{\mathcal{S}} \nabla^\mu \xi^\nu \,\mathrm{d}S_{\mu\nu}, \tag{7.52}$$

which is the mass conserved on a two-surface boundary of the spacetime, therefore the black hole mass associated to a 2-d slice of the event horizon; and the **Komar conserved angular momentum**

$$J = \frac{1}{16\pi G_N} \oint_{\mathcal{S}} \nabla^\mu \psi^\nu \,\mathrm{d}S_{\mu\nu}. \tag{7.53}$$

Note that in the above formulas the proportionality factors were fixed by the requirement that these quantities coincide with those obtained from Hamiltonian based definitions.

In general we can foliate a spacetime as $x^\alpha = (t, y^a)$ by providing a lapse function N and a shift function N^a so that the metric can be written as

$$ds^2 = -N^2\,\mathrm{d}t^2 + h_{ab}\,(\mathrm{d}y^a + N^a\,\mathrm{d}t)\left(\mathrm{d}y^b + N^b\,\mathrm{d}t\right) \tag{7.54}$$

Now one can choose initial values for the first and second fundamental forms on a given spacelike hypersurface Σ, i.e. for h_{ab} and the extrinsic curvature K_{ab}, and evolve them using the Einstein field equations (which also imply a set of constraints to be preserved during the evolution).

Remarkably, the constraint equations imply that when considering the gravitational Hamiltonian in the case in which h_{ab} and K_{ab} satisfy the Einstein vacuum field equations, one finds that it can be expressed as a boundary term i.e.

$$H_G^{sol} = -\frac{1}{8\pi} \oint_{S_t} \left[N(k - k_0) - N_a\left(K^{ab} - K h^{ab}\right) r_b\right] \sqrt{\gamma}\,\mathrm{d}^2\theta \tag{7.55}$$

where k is the trace of the extrinsic curvature of the 2-surface, k_0 is the trace of the extrinsic curvature of the 2-surface when embedded in flat space, and r_b is the normal to S_t. Note also that this Hamiltonian is is non-zero only for non-compact manifolds.

We can then define the gravitational mass of an asymptotically-flat spacetime as the limit value of H_G^{sol} when S_t is a 2-sphere at ι^0 evaluate by choosing $N = 1$ and $N^a = 0$ (so that this asymptotic quantity is associated to pure time translations). This is called the ADM mass.

$$M_{ADM} = -\frac{1}{8\pi} \lim_{S_t \to \infty} \oint_{S_t} (k - k_0)\sqrt{\gamma}\,\mathrm{d}^2\theta \tag{7.56}$$

Similarly one can define the angular momentum of an asymptotically-flat spacetime by taking the same limit and choosing $N = 0$ and $N^a = \phi^a$ where ϕ^a is the vector associated to asymptotic rotations in the cartesian plane.

Of course, we do know that we can consider alternative way to reach infinity and these in general will be associated to different notions of mass. In particular, we can consider to characterise the gravitational mass on a point of $\mathscr{I}^+$. This can be achieved by going to null coordinates u, v and consider the gravitational mass in the limit of $v \to \infty$. This defines the Bondi–Sachs mass as

$$M_{BS} = -\frac{1}{8\pi} \oint_{S(u,v\to\infty)} (k - k_0)\sqrt{\gamma}\, \mathrm{d}^2\theta \tag{7.57}$$

which is mostly relevant when one needs to describe the mass of a radiating body.

Actually, the above example, an isolated body emitting a constant flux of radiation, is useful in showing the difference between the ADM and Bondi–Sachs masses. In fact in this case, one has that on a $t = constant$ slice the ADM mass observed at infinity does not change as the sum of the mass of the body and that of the emitted radiation remains constant albeit both of them change. On the contrary the Bondi–Sachs mass describes the mass present on a given null slice at $u = constant$ as $v \to \infty$ hence it fails to intersect and take note of all the radiation emitted previously of the retarded time u. So in general, it will decrease at u increases.

For a more detail discussion about the $3+1$ decomposition, the gravitational Hamiltonian and the ADM and Bondi-Sachs masses see e.g. [[3] chapter 4]

Of course, in a black hole spacetime mass and angular momentum do not need to be associated only to the spacelike leaves of $\mathcal{H}^+$. In this case we can stick to the initial expression for I as

$$I \propto -\int_\Sigma \mathrm{d}\Sigma_\mu\, \xi^\nu R^\mu{}_\nu \tag{7.58}$$

where again $\mathrm{d}\Sigma^\mu \equiv \mathrm{d}^3x\sqrt{h}\, n^\mu$, and by making use of Einstein field equations

$$R_{\mu\nu} = 8\pi G_N \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T\right), \tag{7.59}$$

we can define a mass

$$M = 2\int_\Sigma \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T\right) n^\mu \xi^\nu \sqrt{h}\, \mathrm{d}^3y\,, \tag{7.60}$$

and an angular momentum

$$J = -\int_\Sigma \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T\right) n^\mu \psi^\nu \sqrt{h}\, \mathrm{d}^3y\,, \tag{7.61}$$

associated to matter fluxes across the hypersurface Σ.

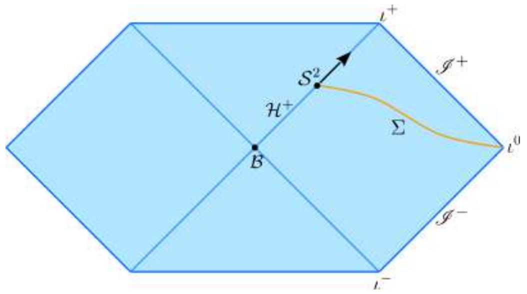


Figure 7.7: Maximal analytical extension of Schwarzschild with the chose hypersurface Σ.

Then, for example, in a black hole spacetime, across a spacelike Cauchy hypersurface foliating it , one can see that the total mass is given by

$$M = \underbrace{-\frac{1}{8\pi G_N}\oint_{\mathcal{H}^+(v)} \nabla^\mu \xi^\nu \, \mathrm{d}S_{\mu\nu}}_{\text{black hole mass } \equiv M_H} + \underbrace{\int_\Sigma \left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right)\eta^\mu \xi^\nu \sqrt{h}\, \mathrm{d}^3 y}_{\text{flux}}, \tag{7.62}$$

So the energy on a given hypersurface could be defined by the black hole mass plus the contribution of matter fluxes outside. A similar relation is valid for the total angular momentum J.

With the above conserved quantities associated to the horizon we can also derive a very useful formula. Consider the combination $M_H - 2\Omega_H J_H$

$$\begin{aligned}M_H - 2\Omega_H J_H &= -\frac{1}{8\pi G_N}\oint_{\mathcal{H}^+(v)} \nabla^\mu \underbrace{(\xi^\nu + \Omega_H \psi^\nu)}_{\Xi^\nu} \mathrm{d}S_{\mu\nu} = -\frac{1}{8\pi G_N}\int_{\mathcal{H}^+(v)} \nabla^\mu \Xi^\nu \cdot 2\,\Xi_{[\mu}N_{\nu]} \underbrace{\sqrt{\gamma}\,\mathrm{d}^2\theta}_{\equiv \mathrm{d}S} = \\ &= -\frac{1}{4\pi G}\oint \underbrace{\Xi_\mu \nabla^\mu \Xi^\nu}_{\kappa \Xi^\nu} N_\nu \,\mathrm{d}S = -\frac{1}{4\pi G_N}\oint \kappa\, \Xi^\nu N_\nu \,\mathrm{d}S = \frac{\kappa}{4\pi G_N}\oint_{\mathcal{H}^+(v)} \mathrm{d}S = \frac{\kappa}{4\pi G} A_{\mathrm{BH}}\,.\end{aligned} \tag{7.63}$$

So we finally get the **Smarr formula**

$$M_H = 2\Omega_H J_H + \frac{\kappa}{4\pi G_N} A_{\mathrm{BH}}. \tag{7.64}$$

This completes our minimal toolkit necessary for dealing with the four laws of black hole mechanics.

Zero-th Law

The zero-th law of thermodynamics states that the temperature of a body in thermal equilibrium is constant all over the body. In the case of black hole mechanics, this reduces to the fact that *the surface gravity is constant over the event horizon of a stationary black hole*: $\kappa = \text{const}$ over $\mathcal{H}^+$.

We want to prove that κ is constant both along each generator of the congruence and also from one generator to another. Consider the generator surface gravity

$$\kappa^2 = -\frac{1}{2}\nabla^\beta \xi^\alpha \nabla_\beta \xi_\alpha; \tag{7.65}$$

also remember that for a Killing vector the relation (2.85) holds $\nabla_\nu \nabla_\mu \xi_\alpha = R_{\alpha\mu\nu\beta}\xi^\beta$. Consider now the directional derivative of κ^2 along a generator of the horizon

$$\xi^\sigma \nabla_\sigma \kappa^2 = -\frac{1}{2}\xi^\sigma \nabla_\sigma \left(\nabla^\beta \xi^\alpha\right)(\nabla_\beta \xi_\alpha) = -\xi^\sigma \left(\nabla_\sigma \nabla_\beta \xi_\alpha\right)\left(\nabla^\beta \xi^\alpha\right) = -\xi^\sigma R_{\alpha\beta\sigma\gamma}\xi^\gamma \left(\nabla^\beta \xi^\alpha\right) = 0, \tag{7.66}$$

and this is zero since we have $\xi^\sigma \xi^\gamma$ which is symmetric in σ and γ, but the Riemann tensor is antisymmetric in the same indices. So the surface gravity is constant along each generator of the horizon.

What about the transverse direction on a leaf at constant v? I.e when we move from a generator to another one? The above derivation also implies that in general

$$\nabla_\sigma \kappa^2 = 2\kappa \partial_\sigma \kappa = -R_{\alpha\beta\sigma\gamma}\xi^\gamma \nabla^\beta \xi^\alpha, \tag{7.67}$$

so if we multiply both sides by ℓ^σ_A:

$$\ell^\sigma_A \nabla_\sigma \kappa^2 = 2\kappa \partial_\sigma \kappa \ell^\sigma_A = -R_{\alpha\beta\sigma\gamma}\ell^\sigma_A \xi^\gamma \nabla^\beta \xi^\alpha. \tag{7.68}$$

We would like to show that the right hand side of this equation has to be zero for a stationary black hole.

Let us specialise first to the case of a non-extremal (so $\kappa \neq 0$), eternal, black hole. The two-surface where the future event horizon meets the past event horizon is called the **bifurcation surface**, $\mathcal{B}$. This is a fixed point in the flow of the Killing vector field, in the sense that the Killing vector that generates the horizon vanishes there, $\xi^a|_{\mathcal{B}} = 0$. Then, via Eq. (7.68), we can conclude that $(\partial_\sigma \kappa)\, \ell^\sigma_A|_{\mathcal{B}} = 0$; now, since ℓ_A is Lie-dragged along v, and since we already proved that κ does not change along v, then we can deduce that $(\partial_\sigma \kappa)\, \ell^\sigma_A = 0$ on all of $\mathcal{H}^+$.

In the case where we are not dealing with an eternal black hole, but e.g. we have a black hole created by a collapse, we know that the late-time region in the Carter–Penrose diagram is equal to the eternal case; therefore, there exists a $\bar{v}$ such that, for any $v > \bar{v}$, the two situations are identical, and hence the surface gravity will be again constant for all the leaves at $v = const$ with $v > \bar{v}$.

A more rigorous proof, making use of the DEC, can be found in [67].

First Law

In thermodynamics, the first law basically states the conservation of energy; in the case of black hole mechanics, it can be show that, under an infinitesimal transformation, the following holds:

$$\mathrm{d}M = \frac{\kappa}{8\pi}\, \mathrm{d}A + \Omega_H\, \mathrm{d}J. \tag{7.69}$$

Let us prove it by comparing two close Kerr black holes.

Let us remember that for a Kerr black hole $A_{BH} = 16\pi M^2_{\mathrm{irr}} \equiv 8\pi \left(M^2 + \sqrt{M^4 - J^2}\right)$. Hence $A = A(M, J)$ and inverting $M = M(A, J)$; both A and J have the same dimension, which is $[A] = [J] = M^2$, therefore M must be an homogeneous function of degree $\frac{1}{2}$ of J and A:

$$M(\lambda A, \lambda J) = \lambda^{1/2} M(A, J); \tag{7.70}$$

according to Euler's theorem, for an homogeneous function f of degree n one can always write,

$$nf = x_i \frac{\partial f}{\partial x_i}, \tag{7.71}$$

then, in our case this becomes

$$\frac{1}{2} M = A \frac{\partial M}{\partial A} + J \frac{\partial M}{\partial J}. \tag{7.72}$$

Using the Smarr formula (7.64) we substitute for M so that

$$\frac{\kappa}{8\pi G} A + \Omega_H J = A \frac{\partial M}{\partial A} + J \frac{\partial M}{\partial J}, \tag{7.73}$$

which implies

$$A \left(\frac{\partial M}{\partial A} - \frac{\kappa}{8\pi G} \right) + J \left(\frac{\partial M}{\partial J} - \Omega_H \right) = 0\,. \tag{7.74}$$

For this to be true for any A and J one has to have

$$\frac{\partial M}{\partial A} = \frac{\kappa}{8\pi G} \qquad \frac{\partial M}{\partial J} = \Omega_H. \tag{7.75}$$

But now, the total variation of the mass can be written as

$$\mathrm{d}M = \frac{\partial M}{\partial A}\, \mathrm{d}A + \frac{\partial M}{\partial J}\, \mathrm{d}J, \tag{7.76}$$

and substituting the previous relations, we immediately find the first law of black hole mechanics

$$\boxed{\mathrm{d}M = \frac{\kappa}{8\pi G}\, \mathrm{d}A + \Omega_H\, \mathrm{d}J}. \tag{7.77}$$

Note that this derivation heavily relied on properties of the Kerr geometry, a vacuum solution. It is nonetheless possible to extend the law to situations including (infinitesimal) matter fluxes (see e.g. [3] page 212).

Second Law

In thermodynamics, the second law states that in a thermodynamical process, the entropy of an isolated system cannot decrease; the analogous in black hole mechanics is the area law, which we have already seen: if the NEC holds, then $\delta A \geq 0$. *The surface area of a black hole can never decrease* (Area Theorem, Hawking 1971).

The area theorem rests on two important results from Penrose. First, the focusing theorem: if at some time $\theta = \theta_0 < 0$ and if the NEC holds, than we shall reach a focusing point in a finite time, $\theta \to -\infty$ (we applied this in the singularity theorem). Second, the event horizon $\mathcal{H}^+$ of a black hole is generated by null geodesics which can have past but no future end-points: a null-ray from $\mathscr{I}^-$ can get into the horizon and become part of the generators of the horizon, but no generator can leave $\mathcal{H}^+$, which in turn implies that it cannot run into a caustic in the future once it has entered $\mathcal{H}^+$. (We already met this theorem explaining why black holes cannot bifurcate).

Then, with these two results, it is obvious that the area of the horizon cannot decrease: if it decreased at some point so that $\theta < 0$ for some generators, then the focussing theorem guarantees that these generator would run into a caustic at which $\theta = -\infty$ and then light-rays would leave $\mathcal{H}^+$ implying a violation of the "no end-points theorem".

Third Law

In standard thermodynamics, the third law takes two forms:

- **Nerst formulation** or **isoentropic formulation**: $S = \text{const}$ at $T = 0$, that means that the entropy does not depend on the macroscopic variables of the system; zero temperature states of a system are isoentropic. Whatever is the starting point, there is a universal value of the entropy at zero temperature which is furthermore equal to zero accordingly with the so called **Planck postulate**, $S(T = 0) = 0$.

- **Unattainability formulation**: it is impossible to reach the absolute zero temperature state with a finite number of thermodynamical steps; we can go to zero temperature asymptotically, but we cannot reach it.

It is interesting to notice that, for black holes, the third law is violated in the Nerst formulation. Consider an extremal black hole: the entropy is still proportional to the area, and the area still depends on the mass of the black hole. However, the third law in the unattainability formulation is realised: *the surface gravity of a black hole cannot reduced to zero within a finite advanced time.* This difference w.r.t. standard thermodynamics seems to hint towards the possibility that extremal black hole have to be considered a different class of objects w.r.t. non-extremal ones (see e.g. [69]).

This law was rigorously proved by Werner Israel only in 1986 [68]. Israel proved that the topological structure of the trapped surfaces of a non-extremal and an extremal black hole is different, and one cannot be deformed into the other if WEC holds. For a partial demonstration of this law and its connection with WEC see e.g. [3].

7.3.4 On the necessity of Hawking radiation

Now that we have shown the four laws of black hole mechanics, it is easy to see that we incur into paradoxes if we do not assume that black holes can emit Hawking radiation.

First of all, let us notice that black holes can apparently violate the second law of thermodynamics if they are not endowed with an entropy: indeed in this case it would be enough to drop some entropic system (e.g. a cup of tea) into a black hole to reduce the total entropy of the universe. In response to this Bekenstein conjectured that black hole have to had an entropy and the Area law suggested it had to be proportional to their area

$$\frac{S_{\mathrm{BH}}}{k_B} = \bar{\eta} \frac{A}{\ell_P^2}, \tag{7.78}$$

(where $\bar{\eta}$ was some unknown proportionality factor and the Area is normalised to the Planck area for dimensional reasons) and that this entropy had to satisfy a *generalised second law* by which $S_{tot} = S_{out} + S_{BH}$ never decreases.

Similarly, one is led to conjecture by the zeroth and first laws that there had also to be a temperature of associated to the black hole given by

$$k_B T_{\mathrm{BH}} = \frac{\hbar}{8\pi c}\frac{\kappa}{\bar{\eta}}. \tag{7.79}$$

However, classically the black hole cannot radiate. This leads to a paradox.

Let us consider a black hole of temperature T_{BH} in a box with a thermal bath of temperature T_{out}. Imagine that some radiation ΔE falls into the black hole; the box will have to decrease its entropy by an amount

$$\Delta S_{\mathrm{out}} = -\frac{\Delta E}{T_{\mathrm{out}}}, \tag{7.80}$$

while the black hole will increase its entropy by the amount

$$\Delta S_{\mathrm{BH}} = \frac{\Delta E}{T_{\mathrm{BH}}}. \tag{7.81}$$

If $T_{\mathrm{out}} < T_{\mathrm{BH}}$, then we are in trouble, because

$$\Delta S_{\mathrm{tot}} = \left(\frac{1}{T_{\mathrm{BH}}} - \frac{1}{T_{\mathrm{out}}}\right)\Delta E < 0, \tag{7.82}$$

and this would violate the generalized second law.

Now, let us instead conjecture that somehow the black hole radiates black body radiation instead

$$U = \sigma_{\mathrm{SB}} T_{\mathrm{rad}}^4 \qquad S = \frac{4}{3}\sigma_{\mathrm{SB}} T^3 = \frac{4}{3}\frac{U}{T_{\mathrm{rad}}}, \tag{7.83}$$

where U is the internal energy. Let us consider again the above ideal experiment, this time in its worst case scenario, i.e. $T_{\mathrm{out}} \ll T_{BH}$ so that we can take $T_{out} \to 0$; now, if the black hole emits energy $\mathrm{d}E > 0$ we have

$$\Delta S_{\mathrm{BH}} = -\frac{\mathrm{d}E}{T_{\mathrm{BH}}}, \tag{7.84}$$

and for the box

$$\Delta S_{\mathrm{out}} = \frac{4}{3}\frac{\mathrm{d}E}{T_{\mathrm{rad}}}; \tag{7.85}$$

but we assumed $T_{\mathrm{rad}} = T_{\mathrm{BH}}$ so the total variation of entropy is

$$\mathrm{d}S_{\mathrm{tot}} = \frac{1}{3}\frac{\mathrm{d}E}{T_{\mathrm{BH}}} > 0, \tag{7.86}$$

So the generalized second law this time holds if somehow black holes can radiate.

7.3.5 Hawking radiation and black hole entropy

As we all know, the resolution of this puzzle came in 1975 with the discovery by Hawking of quantum black hole radiation which takes his name. By a masterful calculation of quantum field theory on a black hole spacetime, Hawking found

$$k_B T_H = \frac{\hbar\kappa}{2\pi c} = \frac{\hbar c^3}{8\pi G M}, \tag{7.87}$$

and hence the previously introduced $\bar{\eta}$ turned out to be $\bar{\eta} = \frac{1}{4}$. This implies that the entropy of a black hole is indeed

$$\frac{S_{\mathrm{BH}}}{k_B} = \frac{A_{\mathrm{BH}}}{4\ell_P^2}. \tag{7.88}$$

The temperature scales with the inverse of the mass of the black hole

$$T \simeq 6 \times 10^{-8} \left(\frac{M_\odot}{M}\right) \mathrm{K}. \tag{7.89}$$

So black holes are very cold! (**Exercise** Compute the mass of a black hole that has a temperature equal to the temperature of the Cosmic Microwave Background, $T_{\mathrm{CMB}} = 2.725\,\mathrm{K}$.)

Black holes are unstable: the more they evaporate, the more their mass decreases, the more their temperature increases, the more they emit; this is a sign that black holes possess a negative heat capacity, and therefore are really thermodynamically unstable.

Also, the lifetime of a black hole scales like the cube of the mass, $\tau_{\mathrm{BH}} \propto M^3$, since

$$\frac{\mathrm{d}M}{\mathrm{d}t} = -\sigma T^4 A. \tag{7.90}$$

While a rigorous derivation of Hawking radiation will require us a full QFT in curved spacetime treatment, in what follows we shall provide an **heuristic derivation** based on the physical intuition that Hawking pair are produced from the vacuum by the tidal forces acting on virtual pairs in the vacuum in proximity of a black hole horizon.

The basic idea is that in a stationary spacetime, the creation of particles is due to the strong tidal forces nearby the event horizon of a black hole, breaking virtual pairs apart. Notice that such tidal forces are stronger the less massive the black hole is, since at the horizon

$$F_{\mathrm{tidal}} \sim \frac{GMm}{r^3}\Delta r \propto \frac{1}{M^2}\Delta r.$$

Now, consider pairs with a Compton length $\lambda_c = \frac{\hbar}{mc}$, and use the geodesic deviation equation:

$$A^\mu = \frac{D^2 S^\mu}{\mathrm{d}t^2} = R^\mu{}_{\nu\rho\sigma} T^\nu T^\rho S^\sigma,$$

where T^μ is the vector tangent to the set of geodesics and S^μ is the vector going from one geodesic to the other. Consider now the reference frame at rest with respect to the particle, $T^\mu = (1, 0)$; the condition for pair breaking is that the gravitational field must do a work $2m$ over a distance λ_c, therefore we can take $S^\mu = (0, \lambda_c)$. With this, the acceleration becomes

$$A = \frac{F}{m} = \frac{W}{md} = \frac{2m}{m\lambda_c} = \frac{2}{\lambda_c}.$$

Now, in the reference frame we defined, we have $A^i = R^i{}_{00j}\lambda_c$, so we need $|R^i{}_{00j}| \gtrsim \frac{2}{\lambda_c}$, meaning a curvature $\gtrsim \frac{1}{\lambda_c}$.

Now, once the virtual pair is broken apart, the two particles can be considered "on-shell", and if they are massive they will travel on timelike geodesics. If the black hole is stationary, than we know that energy is conserved; because of the vacuum outside of the black hole, the initial energy is zero, $E_{\mathrm{in}} = 0$; after the pair creation, the two particles will have energy $E_1 = -p_1 \cdot \xi$ and $E_2 = -p_2 \cdot \xi$. By energy conservation:

$$E_{\mathrm{in}} = 0 = E_1 + E_2 = -(p_1 \cdot \xi + p_2 \cdot \xi).$$

If we assume that, for example, particle 1 escapes to infinity, $E_1 > 0$, since ξ in this case is the usual Minkowski timelike Killing vector, and then $E_2 < 0$, meaning particle 2 must have a negative energy in some local inertial frame.

However, in quantum mechanics, a real particle must have a positive energy in any local inertial frame: the only way out of this apparent paradox is that particle 2 reaches a region where ξ becomes spacelike: the ergoregion beyond the horizon.

In the region where $|\xi|^2 > 0$ and ξ is spacelike, saying that $-p_2 \cdot \xi < 0$ is equivalent to say that the three-momentum of particle 2 is negative in some local inertial frame, and it is here that the existence of particle 2 is allowed. This is also the reason why the pair breaking must happen close to the horizon.

After particle 2 enters the black hole, the mass as measured at infinity must decrease:

$$M' = M + E_2 < M, \qquad E_2 < 0,$$

and therefore $\Delta M_{\text{BH}} = -|E_2| = E_1$, being this the energy radiated at infinity.

Part II

Quantum Field Theory in Curved Spacetimes

8 Preliminaries

8.1 Second quantization in curved spacetime

Motivated by the results obtained for black hole dynamics, we are interested in the quantization of field theories in non-trivial metric backgrounds. In this section of the notes we shall use the particle physics signature $(+,-,-,-)$ (metric determinants in square roots are alway assumed in absolute value even if not explicitly written).

In particular, let us focus on a real scalar field $\phi(x)$ minimally coupled to gravity. The minimal coupling procedure consists in replacing all non-generically covariant quantities in the action with the generically covariant counterparts

$$\begin{aligned} \eta^{\mu\nu} &\to g^{\mu\nu} \\ \partial_\mu &\to \nabla_\mu \\ \mathrm{d}^4x &\to \sqrt{-g}\,\mathrm{d}^4x \,, \end{aligned} \tag{8.1}$$

In this way one is assured that the equation of motion reduces to the usual Klein-Gordon equation in flat space. Therefore, the non-interacting Lagrangian density for $\phi(x)$ becomes

$$\mathscr{L} = \frac{\sqrt{g}}{2}\left[g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - m^2\phi^2\right] \,. \tag{8.2}$$

Notice that the term $\xi\phi^2 R$ goes to zero in flat space as well, so it is a possible interaction term. However, such term violates the equivalence principle, since it is peculiar of scalar fields (indeed fermions and vector fields would couple differently to gravity). Moreover, in d spacetime dimensions for $m = 0$ and $\xi = (d-2)/(4(d-1))$ the action for ϕ would be *conformally invariant.*

We now move to the canonical quantization of ϕ, that now becomes an operator (or better, an operator valued distribution). Of course, the equation of motion is

$$(\Box + m^2)\phi(x) = 0 \,, \tag{8.3}$$

where the box operator is the usual d'Alembertian operator $\Box = \nabla^\mu\partial_\mu$. First, we can consider a spatial foliation of $\mathcal{M}$ and write the metric $g_{\mu\nu}$ in ADM form. Then, in the canonical quantization we need to define a conjugate momentum π as

$$\pi(x) \equiv \frac{\partial\mathscr{L}}{\partial(\nabla_t\phi)} = \sqrt{g}\,g^{\mu t}\nabla_\mu\phi \,. \tag{8.4}$$

Then the commutation relations between the fields are

$$\begin{aligned} [\phi(t,\mathbf{x}),\pi(t,\mathbf{y})] &= \frac{i}{\sqrt{g}}\,\delta^{(3)}(\mathbf{x}-\mathbf{y})\,, \\ [\phi(t,\mathbf{x}),\phi(t,\mathbf{y})] &= [\pi(t,\mathbf{x}),\pi(t,\mathbf{y})] = 0 \,. \end{aligned} \tag{8.5}$$

Again, within the ADM formalism we can consider a foliation at time t that we call Σ_t then the **scalar product** between two solution of the Klein-Gordon (KG) equation ϕ_1, ϕ_2 is defined as

$$(\phi_1,\phi_2) \equiv -i\int_{\Sigma_t} \left(\phi_1\nabla_\mu\phi_2^* - \phi_2^*\nabla_\mu\phi_1\right) n^\mu \sqrt{h}\,\mathrm{d}^3x = -i\int_{\Sigma_t}\left(\phi_1 \overleftrightarrow{\nabla}_\mu \phi_2^*\right) n^\mu \sqrt{h}\,\mathrm{d}^3x \,. \tag{8.6}$$

This product is independent on the specific Σ_t we pick, indeed:

$$\begin{aligned}
(\phi_1,\phi_2)_{\Sigma_2} - (\phi_1,\phi_2)_{\Sigma_1} &= -i\left(\int_{\Sigma_2}\left(\phi_1 \overleftrightarrow{\nabla}_\mu \phi_2^*\right) n^\mu \sqrt{h}\,\mathrm{d}^3x - \int_{\Sigma_1}\left(\phi_1 \overleftrightarrow{\nabla}_\mu \phi_2^*\right) n^\mu \sqrt{h}\,\mathrm{d}^3x\right)\\
&= -i\int_{\mathcal{V}} \nabla^\mu\left(\phi_1 \overleftrightarrow{\nabla}_\mu \phi_2^*\right)\sqrt{g}\mathrm{d}^4x\\
&= -i\int_{\mathcal{V}}\left(\cancel{\nabla^\mu\phi_1\nabla_\mu\phi_2^*} + \phi_1\Box\phi_2^* - \phi_2^*\Box\phi_1 - \cancel{\nabla^\mu\phi_1\nabla_\mu\phi_2^*}\right)\sqrt{g}\mathrm{d}^4x\\
&= 0\,,
\end{aligned}$$

where in the second line we used Gauss's theorem (assuming that the fields go to zero fast enough at ι^0). In the fourth line we obtain zero since the fields satisfy the KG equation.

The scalar product $(.,.)$ defines the positive/negative-norm modes which form a complete orthonormal basis $\{f_i\}$ of solutions of the KG equation

$$(f_i,f_j) = \delta_{ij} \quad (f_i^*,f_j^*) = -\delta_{ij} \quad (f_i,f_j^*) = 0\,. \tag{8.7}$$

These modes also satisfy $\partial_t f = -i\omega f$ and $\partial_t f^* = +i\omega f^*$ (from this they are called sometime positive/negative energy modes, albeit $\omega > 0$ in both cases).

The field ϕ can then be expanded in such a basis

$$\phi(x) = \sum_i \left[\hat{a}_i\, f_i(x) + \hat{a}_i^\dagger f_i^*(x)\right]\,, \tag{8.8}$$

where the index i spans the whole basis and the operators $\hat{a}_i$, $\hat{a}_i^\dagger$ are the creation and annihilation operators which inherit the commutation relations

$$[\hat{a}_i,\hat{a}_j^\dagger] = \delta_{ij}, \qquad [\hat{a}_i,\hat{a}_j] = [\hat{a}_i^\dagger,\hat{a}_j^\dagger] = 0\,. \tag{8.9}$$

In this basis, these operators define a vacuum state $|0\rangle_f$ which is annihilated by all the $\hat{a}_i$'s:

$$\hat{a}_i|0\rangle_f = 0 \qquad \forall i\,. \tag{8.10}$$

Then the Fock space is built by n-particles states for the mode i as

$$|\mathrm{n_i}\rangle_f \equiv \frac{1}{\sqrt{n!}}\left(\hat{a}_i^\dagger\right)^n|0\rangle_f\,. \tag{8.11}$$

Finally, we can define the *number operator* $\hat{N}_{i|f}$ as

$$\hat{N}_{i|f} \equiv \hat{a}_i^\dagger\hat{a}_i\,. \tag{8.12}$$

Remark 1: the so defined vacuum state is invariant under Poincaré transformations. For rotations this is trivial; for boosts of velocity $v < 1$ we have that a frequency $\omega_{\mathbf{k}} > 0$ is transformed into $\omega'_{\mathbf{k}} = \gamma(v)(\omega_{\mathbf{k}} - v|\mathbf{k}|)$ but this is still a positive quantity.
Remark 2: also the *normal ordering* is a Poincaré invariant procedure.

However, when general coordinate transformations are allowed, all these prescriptions are not assured to hold. For instance, in this situation the choice of another basis *is not* necessarily equivalent.

Consider a new set of solutions $\{g_A\}$, (whose indexes are in principle different from $i, j, ..$ and we call $A, B, ...$). For this choice, the field now reads

$$\hat{\phi}(x) = \sum_A \left[\hat{b}_A g_A(x) + \hat{b}_A^\dagger g_A^*(x)\right] . \tag{8.13}$$

and the vacuum is now $|0\rangle_g \neq |0\rangle_f$ as it is defined as the state annihilated by the new set of operators $\hat{b}_A$ for any A. Although the two basis are different, the field $\phi(x)$ must be the same so we expect a relation between the bases. Since we are dealing with linear functional spaces there must be a linear relation between g_A and f_i. Then it must be the case that

$$g_A = \sum_i \left[\alpha_{Ai} f_i + \beta_{Ai} f_i^*\right] . \tag{8.14a}$$

$$f_i = \sum_A \left[\alpha_{Ai}^* g_A - \beta_{Ai} g_A^*\right] . \tag{8.14b}$$

These relations between the two sets of modes are called **Bogolyubov transformations** and the coefficients α_{Ai} and β_{Ai} are called **Bogolyubov coefficients**. [1]

These are explicitly given by the following scalar products:

$$\alpha_{Ai} = (g_A, f_i), \qquad \beta_{Ai} = -(g_A, f_i^*) . \tag{8.15}$$

From these result it is also easy to obtain the relation between the two operators $\hat{a}_i$ and $\hat{b}_A$.

$$\hat{a}_i = \sum_A \left[\hat{b}_A \alpha_{Ai} + \hat{b}_A^\dagger \beta_{Ai}^*\right] , \tag{8.16a}$$

$$\hat{b}_A = \sum_i \left[\hat{a}_i \alpha_{Ai}^* - \hat{a}_i^\dagger \beta_{Ai}^*\right] . \tag{8.16b}$$

The Bogolyubov coefficients satisfy the following normalisation conditions

$$\sum_k \left(\alpha_{Ak}\alpha_{ik}^* - \beta_{Ak}\beta_{ik}^*\right) = \delta_{Ai} , \tag{8.17a}$$

$$\sum_k \left(\alpha_{Ak}\beta_{ik} - \beta_{Ak}\alpha_{ik}\right) = 0 . \tag{8.17b}$$

A very natural question arises now. Given that the two bases are in principle inequivalent, what is the expectation value of the number operator $\hat{N}_{A|g}$ over the vacuum $|0\rangle_f$? In other words, how many particles does an observer in the basis $\{g_A\}$ sees in the vacuum of the observer in the basis $\{f_i\}$?

$$\begin{aligned}
{}_f\langle 0|\hat{N}_{A|g}|0\rangle_f &= {}_f\langle 0|\hat{b}_A^\dagger \hat{b}_A|0\rangle_f = \sum_{ij} {}_f\langle 0|(\alpha_{Ai}\hat{a}_i^\dagger - \beta_{Ai}\hat{a}_i)(\alpha_{Aj}^*\hat{a}_j - \beta_{Aj}^*\hat{a}_j^\dagger)|0\rangle_f \\
&= \sum_{ij} \beta_{Ai}\beta_{Aj}^* \; {}_f\langle 0|\hat{a}_i\hat{a}_j^\dagger|0\rangle_f = \sum_{ij} \beta_{Ai}\beta_{Aj}^* \; {}_f\langle 0|(\hat{a}_j^\dagger\hat{a}_i + \delta_{ij})|0\rangle_f = \sum_i |\beta_{Ai}|^2 .
\end{aligned} \tag{8.18}$$

The conclusion is that the mixing between the positive and negative energy modes appearing in the change of basis (8.14) implies that the observed associated to the quantization in the g_A basis percives the vacuum defined by on observer adopting the f_i basis as filled with particles. The concept of "particle" is Lorentz invariant, however it is not generally covariant.

The Bogolyubov transformations are relevant in several physical situations

[1] Note that in the extant literature one can also find the alternative spellings Bogoliubov and (more rarely) Bogolubov.

- They can be used to relate the observations of different observers in curved spacetimes. E.g. in a Schwarzschild spacetime they can relate the vacua of an observer static at infinity and of another one freely falling into the black hole.
- They can relate the vacua in the asymptotic past and future of a universe undergoing an expansion at some time, or in Minkowski space with a time varying potential.
- They are used in condensed matter physics in relating the inequivalent vacua of the atoms and of the quasi-particle in a Bose–Einstein condensate.

Let's see a simple example below.

8.2 Particle production by a time-varying potential

The above mentioned possible non-equivalence of particle vacua can be easily seen in the simple situation of a massive scalar $\phi(x)$ in flat spacetime obeying the usual KG equation but with an external time-dependent potential $U(t)$:

$$\left(\Box + m^2 + U(t)\right)\phi(x) = 0 \, , \tag{8.19}$$

Since the potential is only time dependent, the equation is linear in space, and a basis orthonormal modes will have in general the form $u_{\mathbf{k}} = N_{\mathbf{k}}\, g_{\mathbf{k}}(t) e^{i\mathbf{kx}}$ Then the classical field can be written as

$$\phi(x) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} N_{\mathbf{k}}\, g_{\mathbf{k}}(t) e^{i\mathbf{kx}} \, . \tag{8.20}$$

where the $N_{\mathbf{k}}$ is some normalization factor.

Inserting this ansatz in the KG equation (8.19) we can separate the time and space dependent equations. In particular for the time dependent modes we get:

$$\ddot{g}_{\mathbf{k}}(t) + \omega_{\mathbf{k}}(t) g_{\mathbf{k}}(t) = 0 \, , \tag{8.21}$$

with the time-dependent frequency $\omega_{\mathbf{k}}^2 \equiv |\mathbf{k}|^2 + m^2 + U(t)$.

Remark: If the function $U(t)$ is periodic, then (8.21) is in the form of a *Flouquet equation* (or called Mathieu equation for $U(t) = A\sin(\Omega t)$). Such a differential equation has a characteristic band-like structure for its solutions. For instance, for certain values of the parameters m and Ω (generally within bands) the solution is unstable and behaves as $g_{\mathbf{k}} \sim e^{\mu_{\mathbf{k}} t}$, where $\mu_{\mathbf{k}}$ is called Flouquet index. This parametric amplification is very important for example in the *pre-heating* stage after inflation (see e.g. [70] for a review), when the oscillations of the inflaton around the minimum of its potential lead to huge amplifications for the standard model fields.

We are instead interested in a situation for which the function $\omega_{\mathbf{k}}(t)$ varies over a finite time lapse and asymptotically takes constant values both in the far past and future

$$\begin{aligned} \lim_{t\to+\infty} \omega_{\mathbf{k}}(t) &= \omega_{out} \, , \\ \lim_{t\to-\infty} \omega_{\mathbf{k}}(t) &= \omega_{in} \, . \end{aligned} \tag{8.22}$$

Corresponding to the two different limits, we can expand the field ϕ in terms of plane-waves with frequency ω_{out} or ω_{in}. Let us indicate the two different basis as $\{u_{\mathbf{k}}^{(in)}\}$ and $\{u_{\mathbf{k}'}^{(out)}\}$. From the discussion of the previous section the relation between the "in" and "out" basis is

$$u_{\mathbf{k}'}^{(out)} = \sum_{\mathbf{k}} \left[\alpha_{\mathbf{kk}'} u_{\mathbf{k}}^{(in)} + \beta_{\mathbf{kk}'} u_{\mathbf{k}}^{*(in)}\right] \, . \tag{8.23}$$

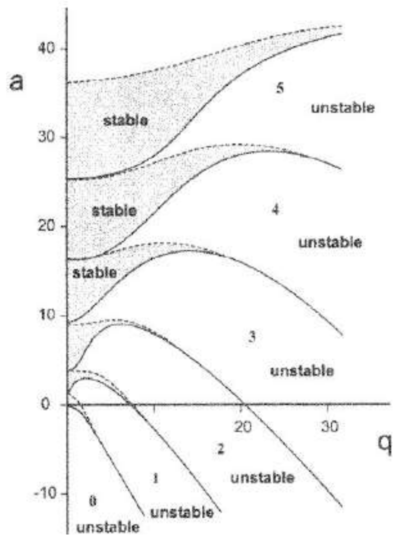


Figure 8.1: Instability bands associated to the Matthieu equation, i.e. an harmonic oscillator with a periodic frequency. In particular here the latter has been parametrized as $\omega(t) = [a(k) - 2q\cos(2z)]$ with $z = mt$. Figure from [71].

where the isotropy and homogeneity in space implies

$$\alpha_{\mathbf{k}\mathbf{k}'} = \left(u^{(out)}_{\mathbf{k}'}, u^{(in)}_{\mathbf{k}}\right) = \alpha_{\mathbf{k}}\,\delta^{(3)}(\mathbf{k}-\mathbf{k}') = \alpha_{\mathbf{k}}\,\delta_{\mathbf{k},\mathbf{k}'}\,, \quad \beta_{\mathbf{k}\mathbf{k}'} = \left(u^{(out)}_{\mathbf{k}'}, u^{*(in)}_{\mathbf{k}}\right) = \beta_{\mathbf{k}}\,\delta^{(3)}(\mathbf{k}+\mathbf{k}') = \beta_{\mathbf{k}}\delta_{\text{-}\mathbf{k},\mathbf{k}'}\,. \tag{8.24}$$

Formulas (8.24) are telling us is that particles can be created out of the vacuum in pairs and the overall momentum is conserved (obviously, given the invariance of the problem under space translations).

Indeed, if the initial quantum state is the vacuum for the "in" observer, then in the final state we are going to have production of particles out of the vacuum. From (8.18) we have:

$$\langle 0_{in}|\hat{N}^{(out)}_{\mathbf{k}'}|0_{in}\rangle = \sum_{\mathbf{k}} |\beta_{\mathbf{k}\mathbf{k}'}|^2\,. \tag{8.25}$$

Now, we would like to characterise the nature of the *in* vacuum state for a late time observer. For general $U(t)$ the above problem cannot be analytically solved but one can alternatively look for two opposite limits: indeed if $\dot{U}/U \sim 1/\tau$ is the typical time over which the transition between the two asymptotic values of U happens, and ω is the typical frequency associated to the excited modes, then we can have two regimes [72]

- *Sudden regime* $\omega\tau \ll 1$: modes for which this is realised have a small energy compared to the energy that U brings into the system. Hence, they can efficiently be excited and we expect a large particle production;
- *Adiabatic regime* $\omega\tau \gg 1$: modes in this regime corresponds to highly energetic particles whose production we then expect to be heavily (exponentially) suppressed.

Note that sometimes these two above regimes are characterised by the equivalent conditions $\dot{\omega}/\omega \gg \omega$ for the sudden regime, and $\dot{\omega}/\omega \ll \omega$ for the adiabatic one.

In particular one can see that in the sudden regime, assuming an instantaneous transition from ω_{in} to ω_{out} (e.g. at $t = 0$ due to a potential $U(t) \propto \delta(t)$), one gets

$$|\beta_{\mathbf{k}}|^2 \approx \frac{1}{4}\frac{(\omega_{in} + \omega_{out})^2}{\omega_{in}\omega_{out}}\,, \tag{8.26}$$

Note that in this case Bogolyubov coefficient is invariant under the exchange of ω_{in} and ω_{out} as only their relative value does matter. In the case of an adiabatic regime one has instead

$$|\beta_{\mathbf{k}}|^2 \approx \exp\left(-2\pi\ \omega_{out}\ \tau\right)\ , \tag{8.27}$$

Finally, note that in the above results we have used (8.24).

For the adiabatic case, one can say, in the Schrödinger picture, that during an adiabatic change of $\omega(t)$ the state continually adjusts to remain close to the instantaneous adiabatic ground state. This state at some time $t = t_0$ is the one annihilated by the lowering operator $a(f_{t_0})$ defined by the solution $f(t)$ satisfying the initial conditions at t_0 and corresponding to the "instantaneous positive frequency solution" at some time t. This instantaneous adiabatic ground state is sometimes called the "lowest order adiabatic ground state" at t_0.

We can write the total number of out-particles present at late times as (introducing the proper volume normalization)

$$N = \frac{V}{2\pi^2}\int_0^\infty |\beta_{kk'}|^2\omega_{out}^2\ \mathrm{d}\omega_{out}, \tag{8.28}$$

and hence the number spectrum will be

$$\frac{\mathrm{d}N}{V\ \mathrm{d}\omega_{out}} = \frac{1}{2\pi^2}|\beta_{kk'}|^2\omega_{out}^2 \tag{8.29}$$

The full spectrum will interpolate between these two regimes and will be characterised by a peak at frequencies $\omega_{out} \sim 1/\tau$. The resulting shape is shown below in Figure 8.2. Let us conclude that the

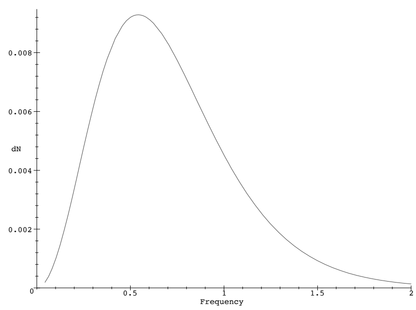


Figure 8.2: Number spectrum (photons per unit volume). The horizontal axis is ω_{out} in arbitrary units. The typical timescale τ is set equal to one. The vertical axis is in arbitrary units. The spectrum might resemble a black body but it is different. At low frequencies, where the sudden approximation holds, it scales like ω_{out}^2 instead of linearly. At high frequency instead, it does show an exponential Planck-like suppression.

case of particle creation in an expanding homogeneous and isotropic universe is a simple generalisation of the above case. Also in this setting we shall have everywhere in space back-to-back pair creation due to conservation of 3-momentum and overall picture remains basically unchanged.

8.3 Coherent and Squeezed states

We want now to describe more precisely the relation between the *in* and *out* vacua which we expect to be related via a unitary operator. A quick – although not rigorous – way to obtain such mapping is the following.

For concreteness, let us consider again the example of particle production associated to a time-varying potential. Recalling (8.16a) we can write

$$0 = \hat{a}^{(in)}_{\mathbf{k}}|0_{\rm in}\rangle = \sum_{\mathbf{k}'}\left[\alpha_{\mathbf{k}\mathbf{k}'}\hat{a}^{(out)}_{\mathbf{k}'} + \beta^*_{\mathbf{k}\mathbf{k}'}\hat{a}^{\dagger(out)}_{\mathbf{k}'}\right]|0_{\rm in}\rangle \overset{(8.24)}{=} \left[\alpha_{\mathbf{k}}\hat{a}^{(out)}_{\mathbf{k}} + \beta^*_{-\mathbf{k}}\hat{a}^{\dagger(out)}_{-\mathbf{k}}\right]|0_{\rm in}\rangle \,. \tag{8.30}$$

Hence

$$\hat{a}^{(out)}_{\mathbf{k}}|0_{\rm in}\rangle = -\frac{\beta^*_{-\mathbf{k}}}{\alpha_{\mathbf{k}}}\,\hat{a}^{\dagger(out)}_{-\mathbf{k}}|0_{\rm in}\rangle \,. \tag{8.31}$$

Now let's note that from

$$\left[\hat{a}^{(out)}_{\mathbf{k}'}, \hat{a}^{\dagger(out)}_{\mathbf{k}}\right] = \delta^{(3)}(\mathbf{k}-\mathbf{k}') \,, \tag{8.32}$$

the operator $\hat{a}$ can be seen formally as (think about the relation between x and $p = \partial/\partial x$) as $\hat{a} = \frac{\partial}{\partial \hat{a}^\dagger}$. Then the above relation can be written as

$$\frac{\partial|0_{\rm in}\rangle}{\partial a^\dagger_{\mathbf{k}(out)}} = -\frac{\beta^*_{-\mathbf{k}}}{\alpha_{\mathbf{k}}}\,\hat{a}^{\dagger(out)}_{-\mathbf{k}}|0_{\rm in}\rangle \tag{8.33}$$

Formally, one could "integrate" both sides to get

$$\log\left(\frac{|0_{\rm in}\rangle}{|0_{\rm out}\rangle}\right) = -\frac{\beta^*_{-\mathbf{k}}}{2\alpha_{\mathbf{k}}}\,\hat{a}^{\dagger(out)}_{-\mathbf{k}}\hat{a}^{\dagger(out)}_{\mathbf{k}} \,. \tag{8.34}$$

Then, by summing over all **k**'s this implies

$$\begin{aligned}|0_{\rm in}\rangle &= \prod_{\mathbf{k}}\mathcal{N}_{\mathbf{k}}\,\exp\left[-\frac{\beta^*_{-\mathbf{k}}}{2\alpha_{\mathbf{k}}}\hat{a}^{\dagger(out)}_{\mathbf{k}}\hat{a}^{\dagger(out)}_{-\mathbf{k}}\right]|0_{\rm out}\rangle \\ &= \prod_{\mathbf{k}}\mathcal{N}_{\mathbf{k}}\sum_{n=0}^{+\infty}(-1)^n\frac{\sqrt{2n!}}{n!}\left(\frac{\beta^*_{-\mathbf{k}}}{2\alpha_{\mathbf{k}}}\right)^n |{\rm n}_{\mathbf{k}}, {\rm n}_{-\mathbf{k}}\rangle \,,\end{aligned} \tag{8.35}$$

where $\mathcal{N}_{\mathbf{k}}$ are normalization coefficients and $|{\rm n}_{\mathbf{k}}, {\rm n}_{-\mathbf{k}}\rangle$ is a collection of $2n$ particles of which n have momentum $\mathbf{k}$ and the other n momentum $-\mathbf{k}$.

Therefore, the *in* vacuum appears as what is called a squeezed state for the *out* observer. Indeed, as we will outline in a moment, equation (8.35) is nothing but a particular form of a squeezed state.

Generally speaking, for single-mode states the most generic squeezed state is defined as [73]

$$|{\rm z},\zeta\rangle = \hat{D}(z)\hat{S}(\zeta)|0\rangle, \qquad z,\zeta \in \mathbb{C} \,, \tag{8.36}$$

where $\hat{D}(z)$ is the *displacement operator* whilst $\hat{S}(\zeta)$ is the *squeezing operator* and are defined as

$$\hat{D}(z) \equiv \exp\left[z\hat{a}^\dagger - z^*\hat{a}\right] \,, \tag{8.37}$$

$$\hat{S}(\zeta) \equiv \exp\left[-\frac{1}{2}\left(\zeta^*\left(\hat{a}\right)^2 - \zeta\left(\hat{a}^\dagger\right)^2\right)\right] \,. \tag{8.38}$$

Further, these additional relations hold (for the sake of clarity, we suppress the explicit dependence of the operators):

$$\hat{S}^\dagger\hat{a}^\dagger\hat{S} = (\cosh|\zeta|)\,\hat{a}^\dagger + \frac{\zeta^*}{|\zeta|}(\sinh|\zeta|)\,\hat{a} \,, \tag{8.39}$$

$$\hat{S}^\dagger\hat{a}\hat{S} = (\cosh|\zeta|)\,\hat{a} + \frac{\zeta}{|\zeta|}(\sinh|\zeta|)\,\hat{a}^\dagger \,, \tag{8.40}$$

$$\hat{S}^\dagger\hat{S} = \mathbb{1} \,, \tag{8.41}$$

$$\hat{D}^\dagger\hat{a}\hat{D} = \hat{a} + z \,, \tag{8.42}$$

$$\hat{D}^\dagger\hat{a}^\dagger\hat{D} = \hat{a}^\dagger + z^* \,. \tag{8.43}$$

A *coherent state* is an eigenstate of the annihilation operator, or equivalently can be seen as a squeezed state (8.36) with $\zeta = 0$. This type of state represents a "classical" solution, in the sense that its expectation value is

$$\langle \phi \rangle = zf + z^* f^* \,, \tag{8.44}$$

and its variance minimizes the quantum uncertainty:

$$\langle \phi^2 \rangle = \langle \phi \rangle^2 \,. \tag{8.45}$$

On the other hand, a *squeezed state* is obtained for $z = 0$ and is the "least classical" state. They are associated with the increase (w.r.t. the coherent case) of the uncertainty in one quantum variable w.r.t. its conjugate. Also this kind of states are associated with violations of energy conditions.

If we go back to (8.30) and confront it with (8.40) we can see that they have the same form if we identify

$$\alpha_{\mathbf{k}} = \cosh|\zeta| \,, \qquad \beta^*_{-\mathbf{k}} = \frac{\zeta}{|\zeta|}\sinh|\zeta| \,, \tag{8.46}$$

then, clearly, the relation between *in* and *out* annihilation operators becomes $\hat{S}^\dagger \hat{a}^{(out)}_{\mathbf{k}} \hat{S} = \hat{a}^{(in)}_{\mathbf{k}}$. Thence, if we write $|0_{\rm in}\rangle = \hat{\Xi}|0_{\rm out}\rangle$ we get:

$$0 = \hat{a}^{(in)}_{\mathbf{k}}|0_{\rm in}\rangle = \hat{S}^\dagger \hat{a}^{(out)}_{\mathbf{k}} \hat{S}\,\hat{\Xi}|0_{\rm out}\rangle \,. \tag{8.47}$$

This relation is satisfied only for $\hat{\Xi} = \hat{S}^\dagger$, so that

$$|0_{\rm in}\rangle = \hat{S}^\dagger |0_{\rm out}\rangle \tag{8.48}$$

I.e. particle production from the quantum vacuum via Bogolyubov coefficients is always realised by a squeezing operation. Of course, (8.35) has not exactly the same form as (8.38) but this is just because in the case of pair production from a time-varying potential we do not have a single-mode state but rather a two-mode squeezed state.

In general, a two-modes, k and l, squeezing operator takes the form

$$\hat{S}_{kl} = \exp\left[\frac{1}{2}\left(\zeta \hat{a}^\dagger_k \hat{a}^\dagger_l - \zeta^* \hat{a}_k \hat{a}_l\right)\right] \,. \tag{8.49}$$

and generates a squeezed state when applied on the two modes vacuum $|\zeta_{\rm kl}\rangle = \hat{S}_{kl}|0_{\rm k}, 0_{\rm l}\rangle$. In quantum optics the two modes are sometime called as the "idler "and the "signal".

It is interesting to consider the expectation value of the number operator of one of the two modes of the pair in such a squeezed state

$$\langle \zeta_{\rm kl}|\hat{N}_k|\zeta_{\rm kl}\rangle = \langle 0_k, 0_l|\hat{S}^\dagger \hat{a}^\dagger_k \hat{S}\hat{S}^\dagger \hat{a}_k \hat{S}|0_k, 0_l\rangle = \sinh^2|\zeta|\langle 0_k, 0_l|\hat{a}_k \hat{a}^\dagger_k|0_k, 0_l\rangle = \sinh^2(\zeta) = |\beta_k|^2 \,, \tag{8.50}$$

where we have used the relations (8.39) and (8.40).

A natural question arises at this point: can a squeezed state produce a thermal distribution? A brief manipulation implies

$$\begin{aligned} \sinh^2|\zeta| &= \frac{\sinh^2|\zeta|}{\cosh^2|\zeta| - \sinh^2|\zeta|} = \frac{1}{[\tanh|\zeta|]^{-2} - 1} && (8.51) \\ &= \text{Planck distribution if} \Longrightarrow \tanh|\zeta| \sim e^{-\hbar\omega_{\mathbf{k}}/(2k_B T)} \,, && (8.52) \end{aligned}$$

where indeed

$$\tanh^2|\zeta| = \frac{|\beta|^2}{|\alpha|^2} \,. \tag{8.53}$$

So, in this case we get a black body spectrum at an effective temperature

$$k_B T_{\rm squeezing} = \frac{\hbar\omega_k}{2\ln\coth|\zeta|} \,. \tag{8.54}$$

Note the dependence on ω_k, this implies that in general "signal" and "idler" do not need to have the same squeezing temperature.

Some comments are in order at this point. The above results about thermality and squeezing have a strong formal analogy with thermofield dynamics (TFD) [74], where a doubling of the physical Hilbert space of states is invoked in order to be able to rewrite the usual Gibbs (mixed state) thermal average of an observable as an expectation value with respect to a temperature dependent "vacuum" state (the thermofield vacuum, a pure state).

In the TFD approach, a trace over the unphysical (fictitious) states of the fictitious Hilbert space gives rise to thermal averages for physical observables completely analogous to the one in (8.50) if we make the identification (8.52).

The formal analogy with TFD allows us to conclude that, if we measure only one photon mode, the two-mode squeezed-state acts as a thermofield vacuum and the single-mode expectation values acquire a thermal character corresponding to a "temperature" $T_{\text{squeezing}}$ related with the squeezing parameter ζ

From this analysis, the conclusion is that squeezed states can be related to thermality if the special condition (8.52) holds. This implies an "exponential" behaviour of the Bogolyubov coefficients, needed for such a thermal spectrum can naturally arise in de Sitter space, where the scale factor goes exponentially $a(t) \sim e^{Ht}$, or in spacetimes with black holes. In the latter case the "peeling" of geodesics close to the horizon turns out to be related to the temperature of the black hole itself.

Additionally, it is important to realise that by going from one basis to another, unitarity is never lost because the squeezing operator is unitary: $\hat{S}^\dagger \hat{S} = \mathbb{1}$.

Finally, a Planck distribution *is not* equivalent to thermality. In fact, at thermal equilibrium all the n-point correlation functions of a field theory become trivially related to the 2-point correlation function. However, in a squeezed vacuum this is in general not the case [75].

9 Effects of Quantum Field Theory in Curved Spacetime

9.1 The Casimir effect

In 19th century P.C. Causseè described in his *"L' Album du Marin"* a mysterious phenomenon which was the cause of maritime disasters [76, 77]. Figure 9.1 shows the situation which he called "Calme avec grosse houle": no wind but still with a big swell running. In this situation he stressed that if two ships end up lying parallel at a close distance then often "une certaine force attractive" was appearing, pulling the two ships towards each other and possibly leading to a collision.

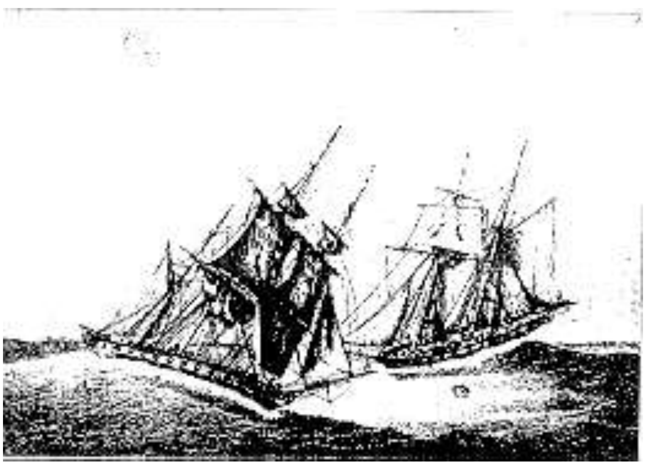


Figure 9.1: From Causseè "The Mariners Album" [76], two ships heavily rolling in a situation where there are still long waves but no wind.

This phenomenon was for many years considered as just another superstition of sailors because it was far from clear what force should be at work in these situations. It is only recently that this effect was given the name of the "Maritime Casimir Effect" [77].

The real Casimir effect, however, was discovered in 1948 by Casimir himself. His idea was to consider a system of two conducting plates facing each other, and to study the vacuum energy of the system.

First, let us consider a Cartesian system of coordinates (x_1, x_2, x_3) and a single thin conducting plate (a foil) positioned along $x_3 = 0$. What is the electromagnetic stress-energy tensor $T_{\mu\nu}$ for this vacuum configuration?

- Due to Lorentz invariance in the coordinates (t, x_1, x_2) we expect the stress energy tensor to be proportional to the Minkowski metric in these coordinates: $T_{\mu\nu} \propto \text{diag}(-1, 1, 1, ?)$
- Conformal invariance of the Electromagnetic Lagrangian, implies that our SET must also be traceless. This fix the last component to be $T_{\mu\nu} \propto \text{diag}(-1, 1, 1, -3)$
- Finally, the proportionality factor must be depdent only on the direction of the symmetry breaking, i.e. x_3, so: $T_{\mu\nu} = f(x_3)\,\text{diag}(-1, 1, 1, -3)$

However, being an energy density our SET must also have dimension 4 in energy. Unfortunately we have at our disposal only $\hbar$ (dimensions $E \times t$) and c (a velocity) and we miss a length scale (or mass or time). Hence we conclude that $f = 0$ and therefore $T_{\mu\nu} = 0$ everywhere.

Now, consider two plates positioned at a distance a between each other and along the plane $x_3 = 0$. Such a configuration is depicted in Fig 9.2.

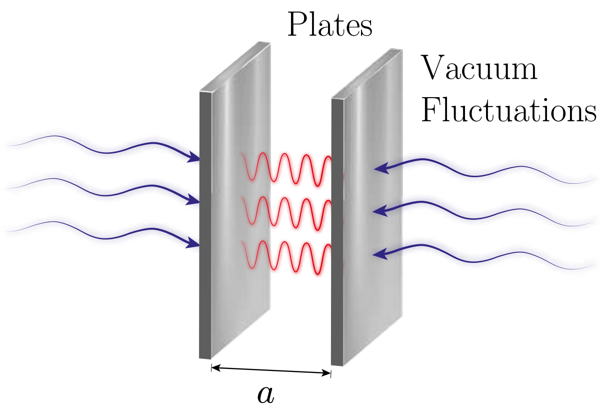


Figure 9.2: Casimir configuration for two conducting plates, facing each other at a distance a.

With the same argument as before one can show that the form of $\langle T_{\mu\nu} \rangle$ in this case we can give to $f(x_3)$ the dimension of an energy density (E/m^3) by requiring $f(x_3) = \alpha(\hbar c)/a^4$ so that finally

$$\langle T_{\mu\nu} \rangle = \alpha \frac{\hbar c}{a^4}\,\text{diag}(-1, 1, 1, -3), \tag{9.1}$$

where α is a real constant to be determined. From a QFT calculation, Casimir was able to show that $\alpha = \pi^2/720$ so the energy density in the vacuum state at zero temperature is

$$\rho_{Cas} = \langle T_{tt} \rangle = -\frac{\pi^2 \hbar c}{720 a^4}. \tag{9.2}$$

We can then calculate the energy between the plates as

$$E_{Cas} = a\rho = -\frac{\pi^2 \hbar c}{720 a^3}. \tag{9.3}$$

Which implies that it is energetically convenient to reduce the distance between the plates (minus sign). Consequently, there will be an attractive force F between the plates given by

$$F = -\frac{\partial E_{Cas}}{\partial a} = -\frac{\pi^2 \hbar c}{240 a^4}. \tag{9.4}$$

A direct measurement in 1958 at the Philips Labs found indeed that $F \approx 0.2 \cdot 10^{-5}\ N/cm^2$ for plates separated by a distance $a = 0.5\ \mu m$.

9.1.1 1+1 dimensional case:

Let us know compute explicitly the Casimir energy density in the simpler case of a two dimensional spacetime. First, we note that in the Coulomb gauge $\vec{\nabla}\cdot\vec{A}=0$ the electromagnetic potential A_μ is equivalent to two independent real scalar fields. For convenience, we can quantise a single scalar and multiply our result by 2. The equation of motion for this scalar $\phi(t,x)$ is the Klein-Gordon equation $\Box\phi(t,x)=0$.

It is clear that the eventual difference in energy will not be due to the outside modes but to the different distribution of the field modes confined to between the plates (actually in this $1+1$ case limited to a segment of length a) w.r.t the one present in the absence of the plates. So, we shall need to find our solutions both with and without (free space) the two plates.

When the two plates are absent, we are effectively in Minkowski space, so the solution is

$$\phi(t,x)=\int\frac{\mathrm{d}p}{\sqrt{(2\pi)2\omega_p}}\left[\hat{a}_p e^{-ip\cdot x}+\hat{a}_p^\dagger e^{ip\cdot x}\right], \tag{9.5}$$

where $p\cdot x=t\omega_p-xk$ is the usual scalar product, $\omega_p=|k|$ is the frequency, and $\hat{a}_p$, $\hat{a}_p^\dagger$ are respectively the annihilation and creation operators satisfying the oscillator algebra $[\hat{a}_p,\hat{a}_k^\dagger]=\delta(p-k)$. The vacuum state of the outside is defined as the state annihilated by all the $\hat{a}_p$ so that $\hat{a}_p|0\rangle=0$. Therefore the vacuum energy in the free region is (from now on we take $\hbar=c=1$)

$$E_{\text{free}}=\langle 0|\hat{\mathscr{H}}|0\rangle\times\text{Volume}=\frac{1}{2}\int_{-\infty}^{+\infty}|k|\ \mathrm{d}k\frac{L}{2\pi}, \tag{9.6}$$

where $\hat{\mathscr{H}}$ is the Hamiltonian density for the single scalar field and we have introduced a normalisation length L for the one dimensional volume. Of course, this quantity diverges, so we decide to regularize it with an exponential function e^{-ks}. In the end we will eventually take the limit $s\to 0$. Also we want to compute the energy in a volume a so we shall multiply our expression by a/L

$$E_{\text{free}}{}^a=E_{\text{free}}\frac{a}{L}=\frac{1}{2}\int_{-\infty}^{+\infty}|k|e^{-|k|s}\ \mathrm{d}k\frac{a}{2\pi}=-\frac{2a}{2(2\pi)}\frac{\partial}{\partial s}\int_0^{+\infty}e^{-ks}\ \mathrm{d}k=\frac{a}{2\pi s^2} \tag{9.7}$$

In the inside region we must fulfil the Dirichlet boundary conditions on the plates:

$$\phi(t,0)=\phi(t,a)=0\qquad\forall t\in\mathbb{R} \tag{9.8}$$

hence, the wave-number assumes only discrete values $k_n=\pi n/a$ and the solution of the wave equation is

$$\phi(t,x)=\sum_{n=1}^{+\infty}\frac{1}{\sqrt{a\omega_n}}\left[\hat{a}_n e^{-it\omega_n}+\hat{a}_n^\dagger e^{it\omega_n}\right]\sin\left(k_n x\right), \tag{9.9}$$

and the vacuum energy is

$$E_{\text{closed}}=\langle 0|\hat{\mathscr{H}}|0\rangle\times\text{Volume}=\frac{1}{2}\sum_{n=1}^{+\infty}\frac{\pi n}{a}=\frac{\pi}{2a}\zeta(-1), \tag{9.10}$$

where $\zeta(-1)$ is the Riemann zeta function in -1, which formally diverges. [1] We regularize the infinity as before:

$$E_{\text{closed}|s}=\frac{1}{2}\sum_{n=1}^{+\infty}\frac{\pi n}{a}e^{-s\pi n/a}=\frac{\pi}{8a\sinh^2\left(\pi s/(2a)\right)}\simeq\frac{a}{2\pi s^2}-\frac{\pi}{24a}+\mathcal{O}(s). \tag{9.11}$$

[1] The Riemann's zeta function is defined as

$$\zeta(x)=\sum_{n=1}^{\infty}\frac{1}{n^x}$$

which converges for real $x>1$. It is possible to do an analytical continuation of this function to all $x\in\mathbb{C}$ with the exception of $x=1$ where $\zeta(x)$ has a pole. While the effectiveness of the Riemann zeta function to provide straightforwardly the correct result as a proper renormalisation albeit apparent miraculous it appears to be ubiquitous. See [5] for a more complete discussion.

Now, the Casimir energy inside the plates is the difference between Equation (9.11) and Equation (9.7):

$$\Delta E = -\frac{\pi}{24a}. \tag{9.12}$$

Note that the same result would have been obtained equivalently by utilizing the analytical continuation of the Riemann function in (9.10) that gives $\zeta(-1) = -\frac{1}{12}$.

Using the last observation it is now easy to obtain the real result in 4 dimensions. We just regularize the vacuum energy in the closed region and analytically continue the divergent result. From the previous calculation it is easy to get directly the form of the energy, that we multiply by a factor 2 for the electromagnetic case:

$$\begin{aligned}
E &= a^2 \int_{\mathbb{R}^2} \frac{\mathrm{d}^2 k}{(2\pi)^2} \sum_{n=1}^{+\infty} \sqrt{k^2 + \left(\frac{\pi n}{a}\right)^2} \\
&= \lim_{s\to -1/2} a^2 \int_{\mathbb{R}^2} \frac{\mathrm{d}^2 k}{(2\pi)^2} \sum_{n=1}^{+\infty} \left[k^2 + \left(\frac{\pi n}{a}\right)^2\right]^{-s} \\
&= \lim_{s\to -1/2} \sum_{n=1}^{+\infty} \frac{1}{4\pi} \frac{\Gamma(s-1)}{\Gamma(s)} \left(\frac{\pi n}{a}\right)^{2-2s} \\
&= -\frac{\pi^2}{6a}\zeta(-3) = -\frac{\pi^2}{720a},
\end{aligned}$$

which is the required result.

Also non-trivial topologies can have a Casimir energy, and in this case we talk about *topological Casimir effect*. For instance consider a spatially closed manifold (a circle) of length L. Then a scalar field defined over the circle has to obey periodic boundary conditions:

$$\begin{aligned}
\phi(t,0) &= \phi(t,L) \\
\partial_x \phi(t,0) &= \partial_x \phi(t,L)
\end{aligned}$$

With the same procedure as in the Casimir case, one can obtain the vacuum energy for the scalar. In this situation the energy is $E = -\pi/(6L)$. If we would have chosen anti-periodic boundary conditions we would have got $E = +\pi/(12L)$. In general the topology of spacetime is not fixed by the metric (neither by the Einstein equation). A natural but still open question arising is: why do we live in a simple-connected geometry?

The above results highlight an important point: one cannot see the Casimir effect as a produced by a reduction of available modes inside the plates: the allowed modes are still infinite. One has rather to think to the effect as something induced by a mode redistribution caused by the presence of the plates/boundary conditions. Also the required Casimir subtraction highlights the fact that this redistribution is different from the free case mainly at wavelengths set by the size of the confined region. This is intrinsically an infrared effect that leaves unmodified the structure of the UV divergences of the field which hence are the same between the free and the bounded case and can be removed via the Casimir subtraction.

Finally, let us point out an important feature common to a wide class of cases of vacuum polarization in external fields. We can return to our example of $1+1$ Casimir effect. Consider the energy density which in this case is simply given by Eq. (9.10) divided by the interval length a.

$$\varepsilon = \langle 0|T_{00}|0\rangle = \frac{1}{2a}\sum_{n=1}^{\infty} \omega_n\,, \qquad \omega_n = \frac{\pi n}{a} \tag{9.13}$$

In order to compute it we can adopt a slightly different procedure for making the Casimir subtraction. It is in fact sometimes very useful in two dimensional problems to use (especially for problems much more complicated that the one at hand) the so called Abel–Plana formula which is generically

$$\sum_{n=0}^{\infty} F(n) - \int_0^{\infty} F(\sigma)\mathrm{d}\sigma = \frac{F(0)}{2} + i\int_0^{\infty} \mathrm{d}\sigma \frac{F(i\sigma) - F(-i\sigma)}{\exp(2\pi\sigma) - 1} \equiv \text{reg}\sum_{0}^{\infty} F(n) \tag{9.14}$$

where $F(z)$ is an analytic function at integer points and σ is a dimensionless variable. This formula is a powerful tool in calculating spectra because the exponentially fast convergence of the integral in $F(\pm i\sigma)$ removes the need for explicitly inserting a cut-off.

In our specific case, $F(n) = \omega_n$ and $\sigma = \pi t/a$, and so Eq.(9.14) applied to the former energy density takes the form

$$\varepsilon = -\frac{1}{\pi}\int_0^{\infty} \frac{\omega \mathrm{d}\omega}{\exp(2a\omega) - 1} \tag{9.15}$$

Although it may appear surprising, we have found that the spectral density of the Casimir energy of a scalar field on a line interval does indeed coincide, apart from the sign, with a thermal spectral density at temperature $T = 1/a$.

9.2 The Unruh Effect

The Unruh effect takes place in a Minkowski spacetime when one considers the response of a detector which is uniformly accelerated. The upshot is that Minkowski vacuum appears as a thermal state to uniformly accelerated observers with proper acceleration a_p at a temperature equal to

$$K_B T_{Unruh} = \frac{\hbar a_p}{2\pi c}.$$

In order to see how this is realise let's understand first the causal structure perceived by these accelerated observers.

9.2.1 The Rindler Wedge

For our purposes it is sufficient to study the case of $1+1$ dimensions. Let us consider a flat spacetime with metric $g_{\mu\nu} = \text{diag}(+1, -1)$. The spacetime interval is simply $\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}x^2$. The velocity u^μ of an observer moving with constant proper acceleration a_p satisfies the following equation

$$u^\mu \nabla_\mu u^\nu = a_p^\nu. \tag{9.16}$$

One can easily check by integrating this equation that, in Cartesian coordinates (t, x), such an observer would lie along the hyperbola $x^2 - t^2 = a_p^{-2}$. Inspired by this, one can consider a new set of coordinates (η, ξ) under which this motion is just along the curve $\xi = \text{const}$. Of course, the required change of coordinates is an hyperbolic parametrization:

$$\begin{aligned} t &= \frac{e^{a\xi}}{a}\sinh(a\eta) \\ x &= \frac{e^{a\xi}}{a}\cosh(a\eta), \end{aligned} \tag{9.17}$$

where the parameter a is just a bookkeeping parameter, not a physical acceleration like a_p, which in principle we could put equal to one. In this system of coordinates (9.17) it easy to see that the ξ=constant lines are the hyperbola

$$x^2 - t^2 = [e^{a\xi}/a]^2,$$

which implies

$$a_p = ae^{-a\xi}.$$

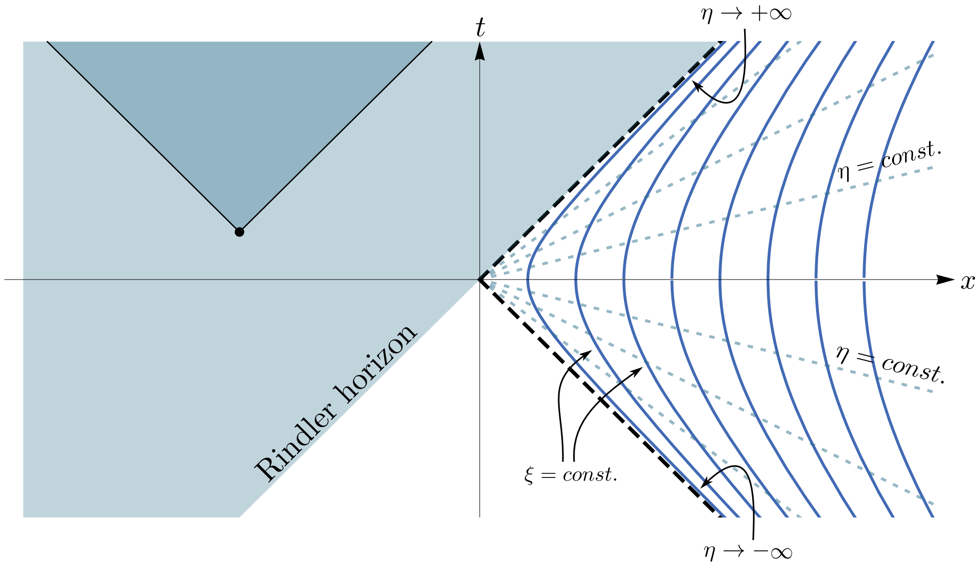


Figure 9.3: Eternally accelerating observers in Minkowski space. Their worldlines are shown in blue and labelled by ξ. Events in the shaded region such as the black dot are hidden to them. The Rindler horizon is the boundary between the shaded and unshaded regions. Rindler space corresponds to the right wedge outlined by the dashed black null lines. The straight dotted lines are lines of constant Rindler time η. Picture taken from "Black Holes notes", by Fay Dowker.

Also these hyperbola intersect the $t = 0$ axis at a distance $d = (\exp \xi a)/a = 1/a_p$. The η=constant surfaces are straight lines stemming from the origin at different angles as η varies. This is why η is also called the hyperbolic opening angle of the Rindler wedge $x > |t|$. See Fig. 9.3.

The change of coordinates (9.17) puts the metric in the conformal form

$$\mathrm{d}s^2 = e^{2\xi a} \left(\mathrm{d}\eta^2 - \mathrm{d}\xi^2 \right) . \tag{9.18}$$

Of course we can also introduce null coordinates related both to the Cartesian coordinates say $U = t - x$ and $V = t + x$ as well as to the Rindler ones, $u = \eta - \xi$ and $v = \eta + \xi$. In these coordinates the metric is respectively

$$ds^2 = \mathrm{d}U \, \mathrm{d}V = e^{2a\xi} \, \mathrm{d}u \, \mathrm{d}v.$$

The coordinates (9.17) cover only a patch of the full Minkowski space (the part on the right of Fig 9.3, $x > |t|$). Hence, an eternally accelerating observer sees an horizon (see Fig. 9.3): the Rindler horizon at $t = x$, or $\eta \to +\infty$. More specifically $\mathcal{H}^+ = \{(x, t) \in \mathcal{M} | x = t\}$. This horizon is generated by a Killing vector field, which is also the generator of boosts,

$$\chi_{\text{boost}} \equiv \frac{\partial}{\partial \eta} = a \left(x \frac{\partial}{\partial t} + t \frac{\partial}{\partial x} \right) . \tag{9.19}$$

and is elsewhere the tangent to the aforementioned hyperbola of constant ξ.

In Cartesian coordinates this vector has contravariant components $\chi^\mu = a(x, t)$ and hence $\chi_\mu = a(x, -t)$ covariant ones. Also $\chi^2 = a^2(x^2 - t^2)$. The surface gravity can be obtained easily from the "normal" definition,

$$\partial_\mu \chi^2 = -2\kappa \chi_{\mu|\mathcal{H}^+} . \tag{9.20}$$

Indeed by direct substitution and choosing the x component of the above equation

$$\partial_x a^2 (x^2 - t^2) \;=\; 2\kappa a t \, , \tag{9.21}$$

$$(2a^2 x \;=\; 2\kappa a t)_{|\mathcal{H}^+} . \tag{9.22}$$

Hence $\kappa = a$.

From the previous comment, one can realize that such a quantity is not the one observed by one of the Rindler observer. The physical, or proper, surface gravity is the one measured by an accelerating observer (at constant ξ). Along its curve, the proper time is $\mathrm{d}\tau = e^{a\xi}\mathrm{d}\eta$ so the vector of interest is $\partial_\tau = e^{-a\xi}\partial_\eta$. Therefore the acceleration per unit mass that we obtain is $\kappa_{\text{proper}} = \kappa e^{-a\xi} = a_p$.

At this point we are ready to perform the quantisation in the two frames, and see if accelerated observers measure a different number of particles.

9.2.2 Quantum field theory on the Rindler Wedge

In Minkowski coordinates, the quantization for a massless scalar ϕ is, as usual, done in terms of plane waves:

$$\phi(t,x) = \int_{\mathbb{R}} \left(\hat{a}_k \bar{u}_k + \hat{a}_k^\dagger \bar{u}_k^*\right) \mathrm{d}k, \tag{9.23}$$

where the plane waves are defined as

$$\bar{u}_k = \frac{e^{ikx - i\varpi t}}{\sqrt{(2\pi)2\varpi}}, \tag{9.24}$$

where $\varpi = |k|$ and $-\infty < k < +\infty$. These are positive norm modes with respect to the global timelike Killing vector ∂_t so that $\partial_t \bar{u}_k = -i\varpi u_k$. The Minkowski vacuum $|0_{\mathrm{m}}\rangle$ is defined as $\hat{a}_k|0_{\mathrm{m}}\rangle = 0$.

In the Rindler coordinates, however, we need to be careful. Indeed, the coordinates (9.17) cover only the right patch of spacetime (see Fig 9.3). The other side can be covered by modifying the definition of $x \to -x$ (9.17). For this reason, a complete basis for the scalar field ϕ in Rindler

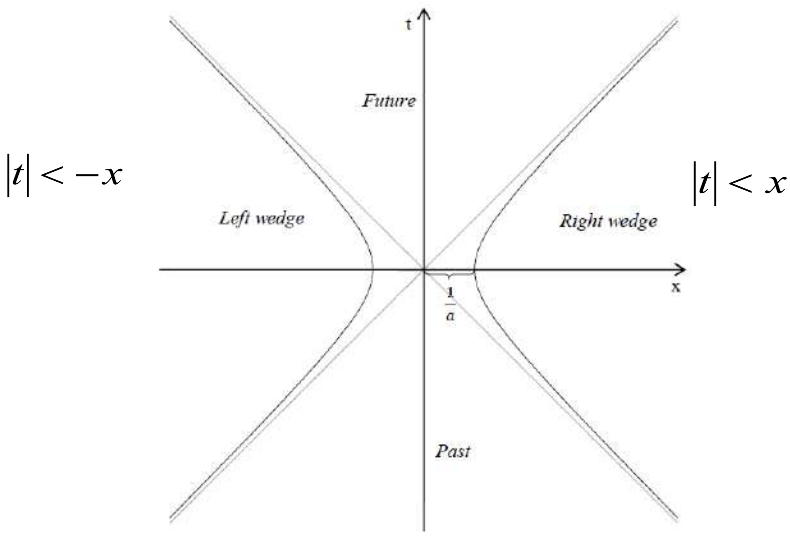


Figure 9.4: Minkowski space divided in Left and Right Rindler wedges

coordinates has to contain both right and left modes [2]. In particular, the field has to be expanded as

$$\phi(\eta,\xi) = \int_{\mathbb{R}} \left(\left[\hat{b}_k^L u_k^L + \hat{b}_k^{L\dagger} u_k^{*L}\right] + \left[\hat{b}_k^R u_k^R + \hat{b}_k^{R\dagger} u_k^{*R}\right]\right) \mathrm{d}k, \tag{9.25}$$

where in this case the modes are

$$u_k^R = \frac{e^{ikx - i\omega\eta}}{\sqrt{(2\pi)2\omega}}, \qquad u_k^L = \frac{e^{ikx + i\omega\eta}}{\sqrt{(2\pi)2\omega}}, \tag{9.26}$$

[2] We do not include modes in the future and past regions of Fig. 9.4 as these region are causally connected with the Rindler wedges and modes there can be hence decomposed in combinations of the wedges ones.

while they are zero when evaluated on the opposite wedge of the Rindler space ($u^R_{k\,|L} = u^L_{k\,|R} = 0$). Also in this case $\omega = |k|$ and $-\infty < k < +\infty$

Note also that u^R_k and u^L_k are positive-norm modes with respect to future pointing timelike Killing vectors which have opposite signs, i.e. $+\partial_\eta$ in the right wedge and $-\partial_\eta$ on the left wedge (think at the fact that lines of increasing η on the right side correspond to lower an lower lines on the left wedge). Hence the positive norm modes are

$$\partial_\eta u^R_k = -i\omega u^R_k, \qquad -\partial_\eta u^L_k = -i\omega u^L_k. \tag{9.27}$$

Finally, the Rindler vacuum $|0\rangle_{Rin}$ is defined as $\hat{b}^{R/L}_k|0\rangle_{Rin} = 0$.

Our goal is to compute the number of particles measured by an accelerating observer in the Minkowski vacuum. I.e. to compute $\beta_{\varpi,\omega} = -(u^R_\omega, \bar{u}^*_\varpi)$ That this Bogolyubov coefficient will be non zero is already evident from the fact that u^R_k does not go over smoothly to u^L_k at the origin, given that there the exponent has to jump between $\pm i\omega\eta$, while $\bar{u}_K$ is obviously analytic at the same point. I.e. Rindler positive norm modes cannot be made of a simple superposition of positive norm modes of the Minkowski base.

It was Unruh to first realise that a smart way to get around this problem at the origin was to introduce a suitable combination of the left and right Rindler modes which are indeed always analytical and bounded for all real (U, V) and everywhere in the lower half complex (U, V) planes (see e.g. [5] page 115)

$$\begin{aligned} f^{(1)}_k N_k &\equiv e^{\frac{\pi\omega}{2a}} u^R_k + e^{-\frac{\pi\omega}{2a}} u^{*L}_{-k} \\ f^{(2)}_k N_k &\equiv e^{-\frac{\pi\omega}{2a}} u^{*R}_{-k} + e^{\frac{\pi\omega}{2a}} u^L_k, \end{aligned} \tag{9.28}$$

where N_k are just normalization coefficients.

We can then use these combinations for expanding our scalar field so to find the following scalar products with the old basis

$$\begin{aligned} &(f^{(1)}_k, u^R_{k'}) = N_k^{-1} e^{\frac{\pi\omega}{2a}} \delta_{k,k'} \qquad &&(f^{(2)^*}_k, u^R_{k'}) = N_k^{-1} e^{-\frac{\pi\omega}{2a}} \delta_{-k,k'} \\ &(f^{(1)^*}_k, u_{k'}{}^R) = 0 &&(f^{(2)}_k, u_{k'}{}^R) = 0. \end{aligned} \tag{9.29}$$

In terms of these new functions the field ϕ can be written as

$$\phi(\eta,\xi) = \int_{\mathbb{R}} \left[2\sinh\left(\frac{\pi\omega}{a}\right)\right]^{-1/2} \left\{ \left(\hat{d}^{(1)}_k f^{(1)}_k + \hat{d}^{(1)\dagger}_k f^{*(1)}_k\right) + \left(\hat{d}^{(2)}_k f^{(2)}_k + \hat{d}^{(2)\dagger}_k f^{*(2)}_k\right) \right\} \, \mathrm{d}k \tag{9.30}$$

where the $\hat{d}^{(i)}_k$ are the operators associated to the new basis. Given that these modes share the same analyticity properties of the Minkowski modes, they must also share the same vacuum state. Hence the operators $\hat{d}_k$ also annihilate the Minkowski vacuum.

From the scalar products (9.29) we also get the Bogolyubov relations between the $\hat{b}_k$ and the $\hat{d}_k$

$$\begin{aligned} \hat{b}^L_k &= \left[2\sinh\left(\frac{\pi\omega}{a}\right)\right]^{-1/2} \left\{ e^{\frac{\pi\omega}{2a}} \hat{d}^{(2)}_k + e^{-\frac{\pi\omega}{2a}} \hat{d}^{(1)\dagger}_{-k} \right\} \\ \hat{b}^R_k &= \left[2\sinh\left(\frac{\pi\omega}{a}\right)\right]^{-1/2} \left\{ e^{\frac{\pi\omega}{2a}} \hat{d}^{(1)}_k + e^{-\frac{\pi\omega}{2a}} \hat{d}^{(2)\dagger}_{-k} \right\}. \end{aligned} \tag{9.31}$$

At this point, since we know how the $\hat{d}_k$ acts on the Minkowski vacuum $|0\rangle_M$ we are ready to evaluate the expectation value of the number operator on the right wedge of the Rindler space:

$${}_M\langle 0|\hat{b}^{\dagger R}_k \hat{b}^R_k|0\rangle_M = \frac{e^{-\frac{\pi\omega}{a}}}{2\sinh\left(\frac{\pi\omega}{a}\right)} \, {}_M\langle 0|\hat{d}^{(2)}_{-k}\hat{d}^{(2)\dagger}_{-k}|0\rangle_M = \frac{e^{-\frac{\pi\omega}{a}}}{2\sinh\left(\frac{\pi\omega}{a}\right)} = \frac{1}{e^{\frac{2\pi\omega}{a}} - 1} \tag{9.32}$$

This corresponds to a thermal distribution of particles with temperature $T = \frac{a}{2\pi}$. This is sometime called the Rindler wedge temperature, as it not associated to a special observer, and it can always be rescaled to be $T_{wedge} = 1/2\pi$. However, the physical temperature is experienced by an observer along $\xi = \text{const}$. Therefore, as we did for the proper surface gravity, we can obtain a proper temperature T_{proper} as

$$T_{\text{proper}} = \frac{T}{\sqrt{g_{00}}} = Te^{-a\xi} = \frac{a}{2\pi}e^{-a\xi} = \frac{\kappa_p}{2\pi} \tag{9.33}$$

This corresponds to the temperature measured by the accelerating observer.

Comments:

- A non zero Bogolyubov coefficient does not necessarily imply a non zero SET as there can be $\alpha\beta$ interference terms. Nonetheless, in the special case when one gets a thermal distribution, the quantum correlations are lost and interference terms vanish. So on a single Rindler wedge the SET will be the one corresponding to a thermal bath at the wedge temperature.

- The total SET is covariant under coordinate transformation and so if it is equal to zero in Minkowski (as it is set by definition), it should be zero also for accelerated observers. In fact it can be shown that the expectation value of the SET in the (Rindler) vacuum state of the accelerated observer is non zero but corresponds to a vacuum polarization which is equivalent to *subtracting* from the Minkowski vacuum a thermal bath at the Unruh temperature [75]. This contribution exactly cancels out the other one coming from the fact that a thermal bath of real particles is actually experienced from the Rindler observer. So in each wedge

$$\langle 0|T_{\mu\nu}|0\rangle_{\text{M}ink} = 0 = \langle 0|T_{\mu\nu}|0\rangle_{\text{R}ind} + \text{Thermal bath} \tag{9.34}$$

  Physically this tells us that we can regard the non-equivalence of the Minkowski and Rindler vacua as an example of vacuum polarization. Moreover this result can easily be understood in a thermofield dynamics framework. In this case the SET of Minkowski can be written as the sum of its corresponding values in the left and right Rindler wedges (which play the role of the two "twin spaces" of TFD [74])

$$\langle 0|T_{\mu\nu}|0\rangle_{\text{M}ink} = 0 = \langle 0|T_{\mu\nu}|0\rangle_L + \langle 0|T_{\mu\nu}|0\rangle_R \tag{9.35}$$

  Each of the terms on the right hand side contributes a thermal bath at the Unruh temperature but the sign of the timelike Killing vector is opposite on the opposite wedges. Hence $\langle 0|T_{\mu\nu}|0\rangle_L = -\langle 0|T_{\mu\nu}|0\rangle_R$. So we see how the TFD construction can help in rapidly arriving at otherwise non-intuitive results.

- The thermal distribution can be seen as an effect of the apparent horizon at $\bar{u} = \bar{v} = 0$ experienced by the Rindler observer. This apparent horizon is similar to the event horizon of a black hole although it has a less objective reality: it is dependent on the state of motion of some observers and is not observed, for example, by Minkowski observers.

  Moreover, this apparent horizon can be globally defined for observers who have undergone a uniform acceleration for an infinite time. In the case of uniform acceleration for a finite time, the thermal spectrum is replaced by a more complicated and general distribution.

- Remarkably, one can readily understand that the found result is very general, independent on the details of the field, the detector and of how the constant acceleration is reached. There is indeed a very general result due to Bisognano and Wichmann which states that when restricted to a Rindler wedge the vacuum state of a relativistic quantum field is a canonical thermal state with density matrix [78]

$$\hat{\rho} \propto \exp[-2\pi \text{H}_\eta/a],$$

where H_η is the so called **boost or Rindler Hamiltonian** given by

$$\mathrm{H}_\eta = \int_{\Sigma_t} T_{ab}\chi^a \,\mathrm{d}\Sigma^b = a \int x\, T_{tt}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

I.e. this is the Hamiltonian associated to the generator of boosts ∂_η. From ρ_R we can easily see then that the "temperature" of the Rindler wedge is always going to be $T_R = \frac{\hbar a}{2\pi}$

The above results are rooted in the fact that the Minkowski vacuum can be, not surprisingly, written as an entangled state [75]

$$|0\rangle_M \propto \prod_\omega \sum_n e^{-\pi\omega_n} |\mathrm{n_L}\rangle \otimes |\mathrm{n_R}\rangle \tag{9.36}$$

Where ω_n are the eigenvalues of H_η. Hence, if we trace over the states of the left (or right) side we obtain a thermal distribution.

$$\hat{\rho}_R \propto \mathrm{Tr}_L |0\rangle_M\langle 0| \propto \sum_n e^{-2\pi\omega_n} |\mathrm{n_R}\rangle\langle \mathrm{n_R}| \tag{9.37}$$

10

Evaporating Black Holes

10.1 Moving Mirrors

Before delving into black hole radiation it is interesting and instructive to analyse the behaviour of quantum radiation in presence of a moving boundaries in a flat gravitational background. More precisely, let us focus on the idealised situation in $1+1$ dimensions of a massless scalar field ϕ on a a flat metric with a moving boundary over which imposes on the field Dirichlet boundary conditions, so that this boundary can be considered a mirror in motion.

In light cone coordinates, $u = t - x$ and $v = t + x$, the trajectory of the mirror can be parametrised as $v = p(u)$ or $u = q(v)$, where $p(u)$, $q(v)$ are functions satisfying $u = q(v) = q(p(u))$ (i.e. $q = p^{-1}$).

10.1.1 Field theory with moving boundaries

The boundary condition for $\phi(u, v)$ takes the form

$$\phi(u, p(u)) = 0 \qquad \forall\, u \in \mathbb{R}. \tag{10.1}$$

Moreover, the wave equation has the very simple form

$$\Box\phi(u, v) = \frac{\partial^2}{\partial u \partial v}\phi(u, v) = 0. \tag{10.2}$$

A general solution for the above equation is of the form $\phi(u, v) = f(u) + g(v)$.

However, the boundary conditions (10.1) imply that

$$\begin{aligned}
\phi(u, p(u)) &= 0 = f(u) + g(p(u)) \quad \Longrightarrow \quad f(u) = -g(p(u)) && (10.3)\\
\phi(u, p(u)) &= 0 = f(q(v)) + g(v) \quad \Longrightarrow \quad g(v) = -f(q(v)) && (10.4)
\end{aligned}$$

So we can write the solutions in the presence of the moving mirror alternatively as

$$\phi(u, v) = f(u) - f(q(v)) = g(v) - g(p(u)). \tag{10.5}$$

We can interpret $f(u)$ and $g(v)$ as the right and left moving waves respectively, while $f(q(v))$, $g(p(u))$ represent waves reflected by the mirror.

Let us first notice, that the motion of the mirror naturally induces a Doppler shift for the incoming rays. To see this let us consider two rays, at advanced times v and $v + \delta v$, approaching the mirror. The corresponding bounced solution is characterized by a retarded time u and $u + \delta u$. From the parametric equation of the mirror we have

$$v + \delta v = p(u + \delta u) \simeq p(u) + p'(u)\delta u + \mathcal{O}(\delta u^2). \tag{10.6}$$

Thus $\delta v = p'(u)\delta u$. The frequency of the wave is inversely proportional to u, therefore

$$\delta\omega \propto \frac{1}{\delta u} \simeq \frac{p'(u)}{\delta v}. \tag{10.7}$$

In the case of a static mirror $p(u) = u$ there is no Doppler shift in the frequency, however e.g. for a moving mirror in uniform rectilinear motion for $p(u) = \kappa u$ (with κ =const) we get redshift or blueshift for $\kappa < 1$ or $\kappa > 1$ respectively (as the mirror moves indeed with velocity $V = (\kappa - 1)/(\kappa + 1)$).

10.1.2 Quantum field theory with moving boundaries

The quantization of the field $\phi(u, v)$ can be achieved as usual. First, we recall that, without the mirror, the left and right going modes are

$$\begin{aligned} g &\sim e^{-i\omega v}, \qquad \omega > 0 \;\; \text{(left going)} \\ f &\sim e^{-i\omega u}, \qquad \omega > 0 \;\; \text{(right going)} \end{aligned}$$

Similarly, in the presence of the mirror from (10.5) we have

$$u_\omega^{(in)} = \frac{1}{\sqrt{4\pi\omega}}\left[e^{-i\omega v} - e^{-i\omega p(u)}\right] \tag{10.8}$$

$$u_\omega^{(out)} = \frac{1}{\sqrt{4\pi\omega}}\left[e^{-i\omega u} - e^{-i\omega q(v)}\right]. \tag{10.9}$$

These are two possible bases for the mode expansion of $\phi(u, v)$ and in general we do expect that by bouncing off the mirror the reflected components can be a mixture of positive and negative norm modes of the free Minkowski base. This mode mixing, as we have seen, then leads to the conclusion that particle creation must generically occur.

In 1+1 dimensions it is possible to show [79] that the energy flux T_{uv} is given by the so-called Schwarzian derivative of $p(u)$

$$F(u) = \langle T_{uu}\rangle = \langle T_{xt}\rangle = -\frac{1}{24\pi}\left[\frac{p'''}{p'} - \frac{3}{2}\left(\frac{p''}{p'}\right)^2\right]. \tag{10.10}$$

In Cartesian coordinates the trajectory can be equivalently written as

$$\begin{cases} x = z(t) \\ u_{\text{mirror}} = t - z(t) \\ v_{\text{mirror}} = t + z(t) \end{cases} \tag{10.11}$$

thus $p(u) = t_{\text{mirror}}(u) + z(t_{\text{mirror}}(u))$. In terms of $z(t)$ the flux becomes

$$F(u) = \frac{\dddot{z}(\ddot{z}^2 - 1) - 3\dot{z}\ddot{z}^2}{12\pi(\dot{z} - 1)^4(\dot{z} + 1)^2}\Big|_{t_{\text{mirror}}} = -\frac{\dot{\alpha}}{12}\sqrt{\frac{1 + \dot{z}}{(1 + \dot{z})^3}}, \tag{10.12}$$

with $\alpha \equiv \frac{\ddot{z}}{(1-\dot{z})^{3/2}}$ being the proper acceleration of the mirror.

This result is also interesting because it also shows that that for trajectories of constant acceleration α the flux of particles vanishes, $F(u) = 0$ [79]. This is a case where the Bogolyubov coefficient β is in general non-zero but nonetheless the stress energy tensor vanishes [80].

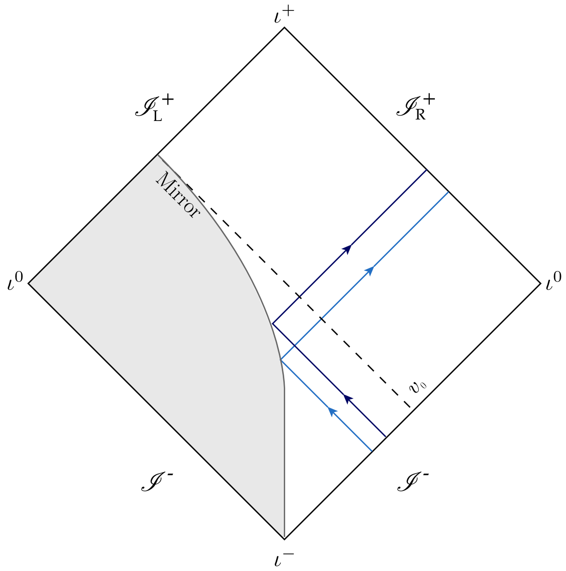


Figure 10.1: Penrose diagram representing the motion of a mirror in Minkowski space. The curve delimiting the shaded region represents the trajectory of the mirror. The two blue curves represent two light rays reflected by the mirror and ending up on $\mathscr{I}_{\mathrm{R}}^+$.

10.1.3 Special case: the exponentially receding mirror

We specialize now in the case depicted in Fig 10.1, where the mirror starts at ι^- and reaches asymptotically $\mathscr{I}_L^+$ with advanced time v_0.

An appropriate basis for an observer on $\mathscr{I}^-$ is given by the incoming modes of (10.8) that for the reader convenience we rewrite below

$$u_\omega^{(in)} = \frac{1}{\sqrt{4\pi\omega}}\left[e^{-i\omega v} - e^{-i\omega p(u)}\right] . \tag{10.13}$$

As depicted in Fig 10.1, the modes that can reach $\mathscr{I}_{\mathrm{R}}^+$ are those leaving $\mathscr{I}^-$ with $v < v_0$ and are given by purely right going modes plus reflected modes

$$u_{\omega,\mathrm{R}}^{(out)} = \frac{1}{\sqrt{4\pi\omega}}\left[e^{-i\omega u} - \theta(v_0 - v)e^{-i\omega q(v)}\right], \tag{10.14}$$

whilst on $\mathscr{I}_{\mathrm{L}}^+$ we have only those modes that left $\mathscr{I}^-$ at $v > v_0$ and hence were not reflected by the mirror

$$u_{\omega,\mathrm{L}}^{(out)} = \frac{1}{\sqrt{4\pi\omega}} e^{-i\omega v}\theta(v - v_0). \tag{10.15}$$

Therefore, the appropriate expansions for the field are:

$$\phi(u,v) = \int_0^{+\infty} \mathrm{d}\omega \left[\hat{a}_\omega^{(in)} u_\omega^{(in)} + \hat{a}_\omega^{\dagger(in)} u_\omega^{*(in)}\right] \tag{10.16}$$

$$\phi(u,v) = \int_0^{+\infty} \mathrm{d}\omega \left[\hat{a}_{\omega,\mathrm{L}}^{(out)} u_{\omega,\mathrm{L}}^{(out)} + \hat{a}_{\omega,\mathrm{L}}^{\dagger(out)} u_{\omega,\mathrm{L}}^{*(out)}\right] + [\mathrm{L} \to \mathrm{R}] . \tag{10.17}$$

The vacuum state $|0_{\mathrm{in}}\rangle$ on $\mathscr{I}^-$ is by definition the state annihilated by all the operators $\hat{a}_\omega^{(in)}$, while $|0_{\mathrm{out}}\rangle$ on $\mathscr{I}_{L,R}^+$ is annihilated by all the operators $\hat{a}_{\omega,L,R}^{(out)}$. In the Heisenberg picture, let us impose that the quantum state associated to the whole spacetime is $|0_{\mathrm{in}}\rangle$ at $\mathscr{I}^-$. We want then to describe the evolution of the destruction and creator operators via the Bogolyubov coefficients so to see how the same state appears on $\mathscr{I}^+$.

The last statement can be translated in the condition that there will be a particle creation by the moving mirror if any of the β Bogolyubov coefficients will be non zero. I.e. we need to check if

$$|\beta^{L,R}_{\omega,\omega'}|^2 = \left| \left(u^{*(out)}_{\omega'(L,R)}, u^{(in)}_{\omega} \right) \right|^2 \neq 0$$

Let us start by noticing that it is quite straightforward to see that $|\beta^{L}_{\omega,\omega'}|^2 = 0$. This is a direct consequence of the fact that all the rays reaching $\mathscr{I}^+_{\mathrm{L}}$ have not been scattered off the mirror, hence there is not reason for mode mixing to occur. On the contrary, on $\mathscr{I}^+_{\mathrm{R}}$ things are more interesting, since also reflected modes can arrive there. So it is definitely $|\beta^{R}_{\omega,\omega'}|^2$ we should try to calculate.

To do so, we need to consider a specific form for the trajectory. So let us choose

$$p(u) = \theta(-u)u + \theta(u)\left[v_0 - Ae^{-\kappa u}\right], \tag{10.18}$$

meaning that for $u < 0$ we have astatic mirror while for $u > 0$ we have an exponential recession. In this specific case the coefficients β in the Bogoliubov transformation can be worked out explicitly, and the final result turns out to be

$$|\beta^{\mathrm{R}}_{\omega\omega'}|^2 = \left| \left(u^{*(out)}_{\omega\mathrm{R}}, u^{(in)}_{\omega'} \right) \right|^2 = \frac{e^{-\pi\omega/\kappa}}{4\pi^2\omega\omega'} \left| \Gamma\left(1 + \frac{i\omega}{k}\right) \right|^2, \tag{10.19}$$

where $\Gamma(z)$ is the Gamma function which is defined for a complex number $z \in \mathbb{C}$ whose real part is positive $\Re(z) > 0$

$$\Gamma(z) = \int_0^\infty e^{-x} x^{z-1}\, \mathrm{d}x\,. \tag{10.20}$$

Let us recall the properties $\Gamma(1+z) = z\Gamma(z)$ and $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$ we obtain

$$\begin{aligned}
|\beta^{\mathrm{R}}_{\omega\omega'}|^2 &= \frac{e^{-\pi\omega/\kappa}}{4\pi^2\omega\omega'} \left| \Gamma\left(1 + \frac{i\omega}{\kappa}\right) \right|^2 \\
&= \frac{e^{-\pi\omega/\kappa}}{4\pi^2\omega\omega'} \Gamma\left(1 + \frac{i\omega}{\kappa}\right) \Gamma\left(1 - \frac{i\omega}{\kappa}\right) \\
&= \frac{e^{-\pi\omega/\kappa}}{4\pi^2\omega\omega'} \frac{i\omega}{\kappa} \Gamma\left(\frac{i\omega}{\kappa}\right) \Gamma\left(1 - \frac{i\omega}{\kappa}\right) \\
&= \frac{e^{-\pi\omega/\kappa}}{4\pi^2\omega\omega'} \frac{i\omega}{\kappa} \frac{\pi}{\sin\left(\frac{i\omega\pi}{\kappa}\right)} = \frac{e^{-\pi\omega/\kappa}}{4\pi^2\omega\omega'} \frac{\omega}{\kappa} \frac{\pi}{\sinh\left(\frac{\omega\pi}{\kappa}\right)} \\
&= \frac{1}{2\pi k\omega'} \frac{1}{e^{2\omega\pi/\kappa} - 1}.
\end{aligned} \tag{10.21}$$

where we have used $\sin(ix) = i\sinh(x)$ and the expression of sinh in terms of exponentials.

Therefore the number of particles measured by an observer on $\mathscr{I}^+_{\mathrm{R}}$ is

$$\langle 0_{\mathrm{in}} | \hat{N}^{(out)}_{\omega\mathrm{R}} | 0_{\mathrm{in}} \rangle = \frac{1}{2\pi\kappa} \frac{1}{e^{2\omega\pi/\kappa} - 1} \int_0^{+\infty} \frac{\mathrm{d}\omega'}{\omega'}. \tag{10.22}$$

The presence of a logarithmically divergent integral is due to the ill normalization property of Fourier modes (the divergence could be cured by using properly normalizable wave-packets).

We conclude that an exponentially receding mirror induces a thermal spectrum with a temperature $T = \frac{\kappa}{2\pi}$ on $\mathscr{I}^+_{\mathrm{R}}$. The parameter κ represents the acceleration of the mirror, and is the analogous of the proper acceleration of a Rindler observer.

10.2 Hawking Radiation

In the collapse of a star into a black hole one would expect the particle production to be restricted into a transitory time window associated with the time-varying geometry. As soon as the black hole has settled down, it would be reasonable for such radiation to disappear. Surprisingly enough, this consideration turned out to be wrong. Hawking proved [81] that a thermal spectrum continues to radiate for an indefinitely long time from any stationary black hole.

By recalling the results of the previous lectures about black holes thermodynamics, this fact should come as no surprise. Indeed, we already argued that black holes should radiate, at a quantum level, with a characteristic temperature related to their surface gravity. In this section we want to show explicitly this relation also by exploiting the analogy with the just seen case of an exponentially receding mirrors.

We will work in $3+1$ dimensions with a massless scalar field ϕ (representing radiation)[1] in a spherically symmetric collapsing star geometry. The most general form of the metric in this context is (note that we are back to the relativists' signature $(-,+,+,+)$)

$$\mathrm{d}s^2 = -C(r)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{C(r)} + r^2\mathrm{d}\Omega^2, \tag{10.23}$$

where $C(r)$ is a function of r only (and possibly of the conserved quantities M, Q, Λ) and the angular variables are θ and φ. In this geometry, the Klein-Gordon equation takes the form

$$\Box\phi = -\frac{1}{C}\frac{\partial^2\phi}{\partial t^2} + \frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2 C\frac{\partial\phi}{\partial r}\right] + \frac{1}{r^2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\phi}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2\phi}{\partial\varphi^2}\right] = 0. \tag{10.24}$$

A generic solution can be split into a (t,r) and (θ,φ) in a spherical harmonics decomposition $\phi(t,r,\Omega) = \sum_{ml}\Phi_l(t,r)\mathcal{Y}_l^m(\Omega)$, where $\mathcal{Y}_l^m(\theta,\phi)$ are eigenfunctions of the angular momentum operator (second squared bracket in (10.24))

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\mathcal{Y}_l^m}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2\mathcal{Y}_l^m}{\partial\varphi^2} = -l(l+1)\mathcal{Y}_l^m \tag{10.25}$$

Obviously, we are mostly interested in the equation for $\Phi_l(t,x)$

$$-\frac{1}{C}\frac{\partial^2\Phi_l}{\partial t^2} + \frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2 C\frac{\partial\Phi}{\partial r}\right] - \frac{l(l+1)}{r^2}\Phi_l = 0. \tag{10.26}$$

It is convenient to introduce, as previously done for the Schwarzschild solution, the Regge-Wheeler coordinate x from the differential relation

$$\mathrm{d}x = \frac{\mathrm{d}r}{C(r)}. \tag{10.27}$$

In this way we have $\frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} = C\frac{\partial}{\partial r}$ and, by defining $\psi_l = \Phi_l r$, we can simplify equation (10.26) further. Indeed we have for the term in square brackets

$$r^2 C\frac{\partial\Phi_l}{\partial r} = r^2\frac{\partial\Phi_l}{\partial x} = -C\psi_l + r\frac{\partial\psi_l}{\partial x}. \tag{10.28}$$

Hence equation (10.26) is written as

$$\begin{aligned}
&-\frac{1}{rC}\frac{\partial^2\psi_l}{\partial t^2} + \frac{1}{r^2C}\frac{\partial}{\partial x}\left[-C\psi_l + r\frac{\partial\psi_l}{\partial x}\right] - \frac{l(l+1)}{r^3}\psi_l = \\
&-\frac{1}{rC}\frac{\partial^2\psi_l}{\partial t^2} + \frac{1}{r^2C}\left[-\frac{\partial C}{\partial r}\frac{\partial r}{\partial x}\psi_l - \cancel{C\frac{\partial\psi_l}{\partial x}} + \cancel{C\frac{\partial\psi_l}{\partial x}} + r\frac{\partial^2\psi_l}{\partial x^2}\right] - \frac{l(l+1)}{r^3}\psi_l = \\
&= -\frac{1}{rC}\frac{\partial^2\psi_l}{\partial t^2} + \frac{1}{rC}\left[\frac{\partial^2\psi_l}{\partial x^2} - \frac{C}{r}C'\psi_l\right] - \frac{l(l+1)}{r^3}\psi_l = 0,
\end{aligned}$$

[1] The final result can be proved to hold even in presence of any type of field (vector and spinor fields for example).

where $C' = \frac{\partial C}{\partial r}$. Thus, the above equation in equivalent to a wave equation in a $1+1$ flat-space

$$-\frac{\partial^2 \psi_l}{\partial t^2} + \frac{\partial^2 \psi_l}{\partial x^2} + V_l(r)\psi_l = 0, \tag{10.29}$$

with potential $V_l(r)$ defined as

$$V_l(r) \equiv -C\left[\frac{C'}{r} + \frac{l(l+1)}{r^2}\right]. \tag{10.30}$$

This function approaches zero for $r \to +\infty$ and also close to the horizon $\mathcal{H}^+$ (recall the form of $C(r)$ in Schwarzschild). Hence, $V_l(r)$ represents the potential barrier a particle has to climb in order to escape from the gravitational attraction of the star.

The analogy with the moving mirror situation is now clearer: for $V_l(r) = 0$, equation (10.29) is equivalent to (10.2), furthermore the analogy holds also in terms of the global structure of the two spacetimes (e.g. see Fig 10.1 and Fig 10.2). In the case of the black hole geometry, the mirror is to be thought at the origin $r = 0$ and our problem is to find the analogue of the relation $v = p(u)$ relating coordinates on $\mathscr{I}^-$ and $\mathscr{I}^+$.

To do so, we need to match the coordinates inside and outside the collapsing star. Inside, the metric takes the generic form for a time dependent spherically symmetric metric

$$\mathrm{d}s^2 = \gamma(\tau,\chi)\left(-\mathrm{d}\tau^2 + \mathrm{d}\chi^2\right) + \rho^2(\tau,\chi)\,\mathrm{d}\Omega^2, \tag{10.31}$$

where γ and ρ can be chosen to be regular everywhere including $\mathcal{H}^+$.

Additionally, we can also introduce a set of null coordinates

$$\begin{cases} U = \tau - \chi \\ V = \tau + \chi \end{cases} \tag{10.32}$$

and fix the center at $\chi = 0$ (i.e. $U = V$).

On the outside we also have the usual set of null coordinates $u = t - x$ and $v = t + x$. However, we need to work with coordinates which are everywhere well behaved (includinhg $\mathcal{H}^+$) in order to perform the matching, and in this case these are the usual Kruskal coordinates:

$$\begin{cases} \mathscr{U} = -e^{-\kappa u} \\ \mathscr{V} = e^{\kappa v} \end{cases} \tag{10.33}$$

where κ is the surface gravity of the black hole. This set of coordinates is regular everywhere.

To perform the matching between (U,V) and $(\mathscr{U},\mathscr{V})$ we proceed as follows. We take two null rays γ_{H} and γ. The first, γ_{H}, is the generator of the horizon $\mathcal{H}^+$ and is extended back into the past where it hits $\mathscr{I}^-$ at some value of v that we fix being $v = v_0$. The second geodesic, γ, on the other hand is "infinitesimally close" to γ_{H} and starts from $\mathscr{I}^-$ at some $v < v_0$. See Fig 10.2. Equivalently, these two rays can be seen as propagating back from $\mathscr{I}^+$ (and thus being described by $u = +\infty$ and $u = u$ respectively). We assume they are free plane waves on $\mathscr{I}^+$ but, possibly, a collection of modes on $\mathscr{I}^-$.

The distance between γ_{H} and γ can be written, in terms of $\mathscr{U}$, as $\mathrm{d}\mathscr{U}$. This must correspond to a $\mathrm{d}U = \beta(\mathscr{U})\mathrm{d}\mathscr{U}$ in terms of the inside coordinates, with $\beta(\mathscr{U})$ is determined by the details of the collapse. Additionally, in terms of $\mathscr{V}$ and V we have $\mathrm{d}V = \zeta(\mathscr{V})\mathrm{d}\mathscr{V} = \zeta(v)\mathrm{d}v$ where, again, $\zeta(\mathscr{V})$ is determined by the details of the collapse. The second step in the last equation follows from the fact that the coordinate v is also well behaved on the horizon and thus $\mathscr{V}$ can be replaced with it.

Close to $\mathcal{H}^+$ we have $x \to -\infty$ thus $u = (t - x) \to +\infty$ and $\mathscr{U} \to 0$. At the same time, v approaches v_0. Hence:

$$\begin{cases} \mathrm{d}U \simeq \beta(0)\mathrm{d}\mathscr{U} \\ \mathrm{d}V \simeq \zeta(v_0)\mathrm{d}v \end{cases} \tag{10.34}$$

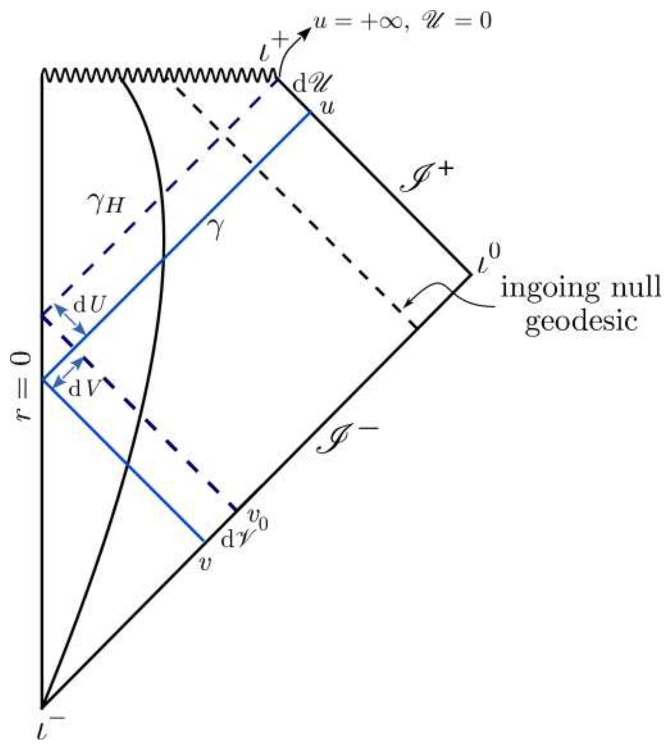


Figure 10.2: Penrose diagram representing a collapsing, spherically symmetric, star. The geodesic γ_{H} lies on the future event horizon $\mathcal{H}^+$ while the geodesic γ is infinitesimally close to γ_{H} and starts from $\mathscr{I}^-$ with advanced time v.

Finally, we recall that at the center of the star $\chi = 0$, so $U = V$ and thus $\mathrm{d}U = \mathrm{d}V$ (by the fact that the geometry is stationary). This makes us conclude that

$$\zeta(v_0)\mathrm{d}v \simeq \beta(0)\mathrm{d}\mathscr{U}\,. \tag{10.35}$$

This relation can be integrated:

$$\int_{v_0}^{v} \zeta(v_0)\mathrm{d}v' = \int_0^{\mathscr{U}} \beta(0)\mathrm{d}\mathscr{U}' \implies v = v_0 + \frac{\beta(0)}{\zeta(v_0)}\mathscr{U} \implies v = v_0 - \frac{\beta(0)}{\zeta(v_0)}e^{-\kappa u} \equiv p(u). \tag{10.36}$$

Hence, $v = p(u)$ takes the same form as in the case of the exponentially receding mirror (10.18) with $A = \beta(0)/\zeta(v_0)$!

Therefore, by following the previous steps we can conclude that for $V_l(r) = 0$ we must have

$$\langle 0_{\mathscr{I}^-}|\hat{N}_\omega^{\mathscr{I}^+}|0_{\mathscr{I}^-}\rangle = \frac{1}{2\pi\kappa}\frac{1}{e^{2\omega\pi/\kappa}-1}\int_0^{+\infty}\frac{\mathrm{d}\omega'}{\omega'}, \tag{10.37}$$

where $|0_{\mathscr{I}^-}\rangle$ is the vacuum on $\mathscr{I}^-$ and $\hat{N}_\omega^{\mathscr{I}^+}$ the number operator on $\mathscr{I}^+$. Let us note again that this this thermal behaviour of the black hole is rooted in κ which here controls the amount of "peeling" of geodesics close to $\mathcal{H}^+$, as shown in (10.36).

Up to now we have neglected the potential. The presence of a potential barrier inevitably points toward a reflection of some of the rays, while others can manage to tunnel through. In such a case the wave has the form

$$u_{\omega l}^{(out)}(u,v) \sim e^{-i\omega u} + u_{\omega l}^{(1)} + u_{\omega l}^{(2)} \tag{10.38}$$

Here the function $u_{\omega l}^{(1)} \sim \theta(v_0 - v)e^{-i\omega q(v)}K_l(\omega)$ describes transmitted modes that crossed the potential barrier and reach $\mathscr{I}^+$ after bouncing/passing through the center of the star, with $K_l(\omega)$ being an *attenuation coefficient*. The function $u^{(2)}_{\omega l}$ contains instead the reflected part of the wave, i.e. modes that bounced off the potential barrier and got to $\mathscr{I}^+$ without entering the collapsing star . Obviously, these are unprocessed modes which do not contribute to the Bogolyubov coefficients. So in the end $|\beta_{\omega\omega'}|^2 = |(u^{(in)}_{\omega' l}, u^{(out)*}_{\omega l})|^2 = |K_l|^2 |(u^{(in)}_{\omega' l}, u^{(1)*}_{\omega l})|^2$, only the transmitted modes contributes.

The upshot is that the correct number of particles observed on $\mathscr{I}^+$ gets corrected by a function of the attenuation coefficient

$$\langle 0_{\mathscr{I}^-} | \hat{N}^{\mathscr{I}^+}_\omega | 0_{\mathscr{I}^-} \rangle = \frac{1}{2\pi\kappa} \frac{\Gamma_l(\omega)}{e^{2\omega\pi/\kappa} - 1} \int_0^{+\infty} \frac{d\omega'}{\omega'}, \tag{10.39}$$

where $\Gamma_l(\omega)$ is the so-called *gray factor*, associated to the attenuation $K_l(\omega)$, that allows for the black hole to be at thermal equilibrium. Indeed, to have detailed balance we need the coefficient

$$R = \frac{\text{radiation emitted in } (\omega, l)}{\text{radiation absorbed in } (\omega, l)}, \tag{10.40}$$

to be independent on $\Gamma_l(\omega)$. Since the absorbed radiation does depend on this function, it must be the case also for the emitted radiation to have the same dependence, in such a way that $\Gamma_l(\omega)$ is factored out in R.

Before moving to a brief descriptions of the issues raised by the discovery of Hawking radiation it is perhaps worth discussing in some detail the nature of the quantum states that can be associated with a black hole spacetime.

- *Unruh state:* this is the state we just discussed above. It is defined as the that state which is vacuum on $\mathscr{I}^-$ and at $\mathcal{H}^+$ for a freely-falling. For an observer at infinity this is a thermal state due to Hawking radiation. It is the only non-singular state, vacuum on $\mathscr{I}^-$, on a spacetime where a black hole is formed by a collapse (like that shown in Fig 10.2). However, the case of an eternal black hole (Fig. 6.6) this state is singular on the past horizon $\mathcal{H}^-$ (we shall see why when we shall discuss explicitly the semiclassical collapse).

- *Hartle–Hawking state:* This is a state describing a thermal equilibrium between a black hole and a surrounding thermal bath at the Hawking temperature. As such it is defined as the state which is thermal at the Hawking temperature both on $\mathscr{I}^+$ and $\mathscr{I}^-$. Being a state of equilibrium it is the natural state for eternal black holes (Fig. 6.6) and indeed is the only one which is non-singular everywhere in these spacetimes.

- *Boulware state:* This is defined as the global state which is vacuum both on $\mathscr{I}^+$ and $\mathscr{I}^-$. It is a state appropriate for the exterior of stationary spherically symmetric stars. The annihilation operator is defined with respect to the timelike Killing vector ∂_t of this spacetime. Noticeably the expectation-value of $T_{\mu\nu}$ on this state diverges at the horizon [45]

$$_B\langle 0 | \, T_\mu{}^\nu | 0 \rangle_B \sim -\frac{1}{2\pi^2 \left(1 - \frac{2M}{r}\right)^2} \int \frac{\omega^2 d\omega}{e^{8\pi\omega} - 1} \text{diag}(-1, 1/3, 1/3, 1/3) \tag{10.41}$$

We see that such configuration due to the pre-factor $(1-2M/r)^{-2}$ develops a huge vacuum polarisation close to the horizon and that the expectation value of the stress energy tensor appears to correspond to the subtraction from the vacuum of a black body radiation at the Hawking temperature [75].

10.3 Open issues with Hawking radiation

Although apparently surprising, Hawking result follows from minimalistic assumptions about General Relativity and quantum theory for free fields in curved spacetime and the presence of interactions does not affect much the conclusion drawn in the previous section (actually, we shall see later that the thermality at the Hawking temperature can be deduced on purely geometrical grounds). However, as it is often the case in physics, the discovery of Hawking radiation led to new questions which we are going to pinpoint below.

10.3.1 Trans-planckian problem

Let us work in ingoing Eddington–Finkelstein coordinates. Let us evolve back in advanced time v a wave packet along a $u =$ const. ray, from $\mathscr{I}^+$ to $\mathscr{I}^-$. For a freely falling observer such wave packet peak frequency goes as the inverse of the advanced time in Kruskal coordinates, $\omega_{\text{peak}} = V^{-1} = e^{-\kappa v}$.[2]

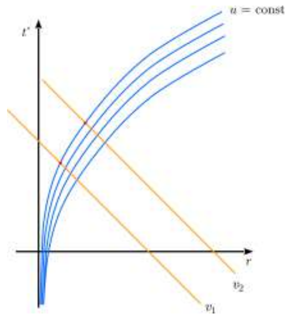


Figure 10.3: Ray-tracying backward in time a Hawking packet at $u = const$.

Therefore, in the interval $\Delta v = v_2 - v_1 > 0$ we have a blueshift

$$\frac{\omega_{\rm p}(v_1)}{\omega_{\rm p}(v_2)} \simeq e^{\kappa \Delta v}. \tag{10.42}$$

For a black hole with the mass of the Sun, $\kappa = (4M_{\odot})^{-1}$ and for a time interval $\Delta v \sim 2$sec we obtain $\omega_{\rm p}(v_1)/\omega_{\rm p}(v_2) \sim 10^5$. Hence while ray tracing the wave packet backward in time towards the horizon, the peak frequency undergoes an exponentially fast blueshift and, eventually, it reaches frequencies above the Planck mass $\omega \gg 10^{19}$ GeV, i.e. in principle beyond the regime of validity of out QFT in CS framework.Alternatively, we can say that the transplanckian problem consist in the fact that the IR Hawking quanta reaching $\mathscr{I}^+$ seems to originate from far UV modes in the vacuum at $\mathcal{H}^+$ or equivalently on $\mathscr{I}^-$.

Of course, frequencies are not per se covariant objects and one might think that boosting to a suitable Lorentz frame can remove the issue. However, we actually talking about UV s-waves emanating from the black hole and there is no way those can be boosted away.

While, in Hawking's original derivation this issue was not directly touched it was soon recognised by a non-negligible community, led by Bill Unruh and Ted Jacobson. It was partially the need to test the robustness of Hawking radiation that gave since the early nineties a strong impulse to an alternative and quite original framework, that of Analogue Gravity [43].

Analogue Gravity

And I cherish more than anything else the Analogies, my most trustworthy masters. They know all the secrets of Nature, and they ought least to be neglected in Geometry. Johannes Kepler.

[2]The Kruskal coordinates strictly speaking are not everywhere associated to freely falling observers. However, in proximity of the horizon of a stationary black hole and at infinity T and X do describe proper time and distance for geodesic observers.

In 1981, Unruh put forward a powerful analogy between curved spacetime and classical hydrodynamics, showing how different physics can be described mathematically in the same way. Consider for instance a perfect fluid (irrotational and non-viscous) satisfying the continuity and Euler equations

$$\begin{cases} \dot{\rho} + \nabla \cdot (\rho \mathbf{v}) = 0 \\ \rho \left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = -\nabla p \end{cases} \tag{10.43}$$

where ρ is the density, p the pressure, $\mathbf{v}$ the velocity, and $\nabla \times \mathbf{v} = 0 \implies \mathbf{v} = \nabla \phi$ being ϕ the so called velocity potential of the fluid. For a barotropic fluid, we also have the equation of state $\rho = \rho(p)$ with $c_s^2 \equiv \frac{\mathrm{d}p}{\mathrm{d}\rho}$. Albeit we considered the forced induced just by pressure, it is straightforward to add external forces, at least conservative body forces such as Newtonian gravity $\rho \nabla \Phi$ or dissipative effects induced by viscosity.

If we perturb this set of equations around a background $(\phi_0, \rho_0, p_0, \mathbf{v}_0)$, then in terms of the perturbations $(\phi_1, \rho_1, p_1, \mathbf{v}_1)$ the equations (10.43) can be recast into a single second order differential equation

$$\frac{\partial}{\partial t} \left[\frac{\rho_0}{c_s^2} \left(\frac{\partial \phi_1}{\partial t} + \mathbf{v}_0 \cdot \nabla \phi_1 \right) \right] = \nabla \cdot \left[\rho_0 \nabla \phi_1 - \frac{\mathbf{v}_0 \rho_0}{c_s^2} \left(\frac{\partial \phi_1}{\partial t} + \mathbf{v}_0 \cdot \nabla \phi_1 \right) \right]. \tag{10.44}$$

Equivalently, this can be recast into the simple form:

$$\Box \phi_1 = 0, \tag{10.45}$$

where $\Box$ is the D'Alembertian computed with the auxiliary metric

$$g_{\mu\nu} = \frac{\rho_0}{c_s^2} \begin{bmatrix} -(c_s^2 - \mathbf{v}_0^2) & -v_0^j \\ -v_0^i & \delta^{ij} \end{bmatrix}. \tag{10.46}$$

Furthermore, one can realize that the physics of the system can mimic the one of a black hole. For instance, we can reach $g_{00} = 0$ for $c_s^2 = v_0^2$ meaning that there is a trapped region for the waves.

To make this example more tangible, we can consider the hole in a bathtub. The background fluid has a larger velocity v_0^2 as we go closer to the hole. Therefore eventually we reach a point where indeed $g_{00} = 0$ and this equation delimits a region below which waves propagating on top of the fluid must flow downstream (see Fig 10.4). The analogy can be made even more precise by considering the Schwarzschild metric in *Painlevè-Gullstrand coordinates*, that are defined as

$$\begin{cases} t_{\mathrm{PG}} = t \pm \left[4M \operatorname{arctanh} \left(2\sqrt{\frac{2M}{r}} \right) - 2\sqrt{2Mr} \right] \quad r_{\mathrm{PG}} = r \\ \mathrm{d}t_{\mathrm{PG}} = \mathrm{d}t \pm \frac{\sqrt{2M/r}}{(1 - \frac{2M}{r})} \mathrm{d}r \quad +\text{=outgoing PG, -=ingoing PG} \end{cases} \tag{10.47}$$

The metric takes the form

$$\mathrm{d}s^2 = -\left(1 - \frac{2M}{r}\right) \mathrm{d}t_{\mathrm{PG}}^2 \pm \sqrt{\frac{2M}{r}} \mathrm{d}r\, t_{\mathrm{PG}} + \mathrm{d}r^2 + r^2 \mathrm{d}\Omega^2, \tag{10.48}$$

It should be noted that the + sign corresponds to a coordinate patch that covers the usual asymptotic region plus the region containing the future singularity of the maximally-extended Schwarzschild spacetime. Thus, it covers the future horizon and the black hole singularity. On the other hand the $-$ sign corresponds to a coordinate patch that covers the usual asymptotic region plus the region containing the past singularity. Thus it covers the past horizon and the white hole singularity.

We can now see that cast in this form the Schwarzschild metric is indeed of the acoustic form (modulo a conformal factor) and the analogy with fluid case can be made evident if we make the choice of the flow

$$|\mathbf{v}_0| \longleftrightarrow \sqrt{\frac{2M}{r}}, \qquad \rho \longleftrightarrow r^{-3/2}, \tag{10.49}$$

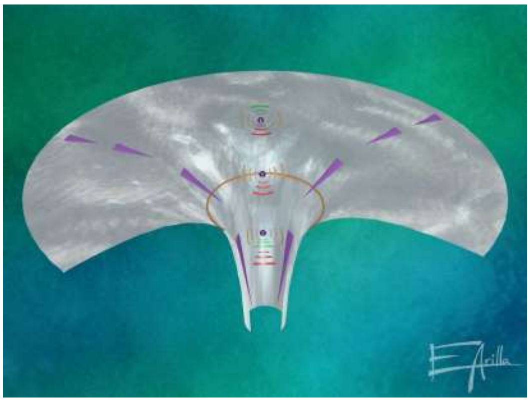


Figure 10.4: Representation of a hole with a trapped region (red circle). As the throat shrinks, wave packets need larger and larger group velocities to reach the flat region on top (from [43]).

where the condition on the density is imposed by consistency with the continuity equation. [3] Let us stress that many other black hole-like geometries can be designed by choosing suitable fluid flows. The Schwarzschild example above is just one among many (and not the most practical one to realise in a laboratory). All one needs is a fluid flow with an acoustic horizon as shown in Fig. 10.5

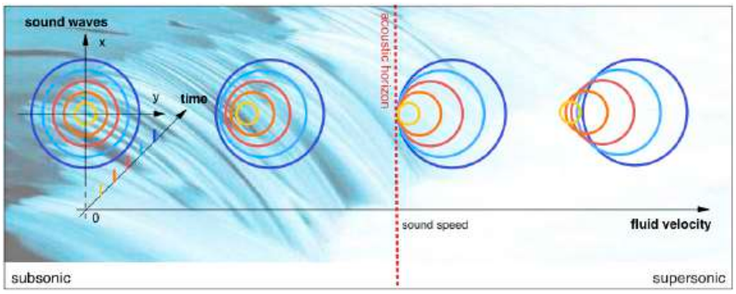


Figure 10.5: A moving fluid can form an "acoustic horizon" when supersonic flow prevents upstream motion of sound waves (from [43]).

To push the analogy even further, one could ask if there is an analogous of the Hawking radiation for the fluid. The answer is yes, and involves the *phonons*: the quantum fluctuations of the fluid. However, testing experimentally the quantum properties of a fluid is quite a hard task. Nonetheless, it is possible to simulate hHawking radiation is some systems like *superfluids* or *Bose-Einstein condensates* (BEC), where the coherence time is much longer than in a conventional fluid.

Within this analogy, the transplanckian problem can be seen in the following way. Below a characteristic energy scale K, the collection of atoms behaves as a fluid (or BEC), but as soon as we probe higher energies we start to resolve the atomic degrees of freedom of the system. This very fact is modelled by modifying the dispersion relation of the phonons to

$$\omega^2 = c_s^2 \left(k^2 + \eta \frac{k^n}{K^{n-2}} \right) . \tag{10.50}$$

where η s a dimensionless constant which can be negative or positive depending on the analogue system used. This modified dispersion relation (10.50) would introduce a Lorentz violating term in the equation of motion that have significant impact on the solutions and on their behaviour close to the horizon $\mathcal{H}^+$. Actually, if $\frac{\mathrm{d}\omega}{\mathrm{d}k} > 1$, then the group velocity of a wave close to $\mathcal{H}^+$ can exceed the speed of light/sound, and a particle back-traced starting from $\mathscr{I}^+$ could be seen to come from behind

[3] In order to satisfy the Euler equation one needs in this case also to add a suitable external potential, but this is not relevant for our discussion here.

the horizon. If, however, the modification (10.50) is such that $\frac{\mathrm{d}\omega}{\mathrm{d}k} < 1$ then the particle would change its group velocity sign close to the horizon and when back-traced it would be seen to come from $\mathscr{I}^-$ instead of piling up at $\mathcal{H}^+$. See Fig. 10.6.

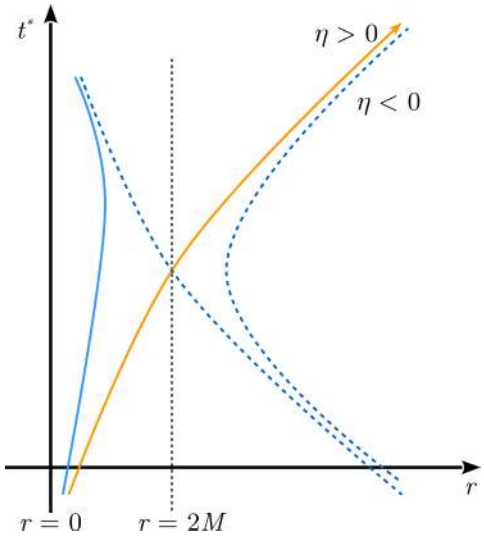


Figure 10.6: Modes associated to analogue Hawking radiation for η positive (solid lines) or negatives (not all the solutions for $n > 2$ dispersion relations are showed).

It has been shown that Hawking radiation is robust against different UV completions of General Relativity. More precisely, possible corrections to the Hawking temperature would be of order $\frac{\kappa}{\Lambda} \ll 1$ where Λ is parametrically related to K.

A worth-noting result in this context is the Lorentz violating theory by Petr Hořava formulated in 2009. He proposed a Lorentz violating UV completion of GR with the following dispersion relation for gravitons

$$\omega^2 = c^2 \left(k^2 + \frac{k^4}{M_*^2} + \frac{k^6}{M_*^4} \right), \tag{10.51}$$

which gives a power counting renormalizable theory of gravity. In this formulation, black holes solutions are endowed with a new impenetrable horizon called *universal horizon*. As the name suggest, not even particles with speeds approaching infinity are able to cross such horizon.

Finally, let us stress that the lesson from Analogue gravity is not that one needs UV Lorentz breaking in order to solve the transplanckian problem. On the contrary it shows, in a well defined system where the UV completion is under control, that the Hawking radiation is rather robust, or if you wish insensitive, about the details of the UV completion of your theory. This is good news for its reliability as a universal prediction.

10.3.2 The information loss problem

Hawking radiation predicts that a black hole will evaporate in a finite time. Therefore, a black hole cannot be eternal, and the Penrose diagram of a spacetime in which a black hole is formed at early times and later on completely evaporates will be of the form shown in Fig 10.7. In this spacetime, after a certain time, the singularity disappears. At the quantum level, on an early Cauchy hypersurface Σ_1, we can start from a state $|\Psi\rangle$ and after the black hole is completely gone we end up on a late Cauchy hypersurface Σ_2 with a thermal state $|\mathrm{T}\rangle$. This is a problem because quantum evolution is unitary, thence

$$|\Psi(\mathrm{t})\rangle = e^{i\hat{\mathrm{H}}t}|\Psi(0)\rangle, \tag{10.52}$$

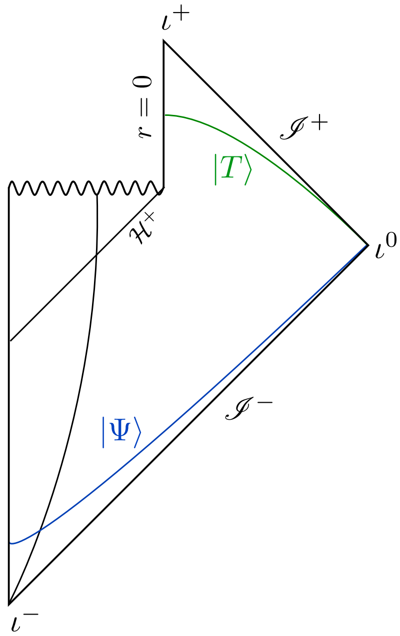


Figure 10.7: Spacetime of an evaporating black hole. After a while the singularity disappears. The quantum states $|\Psi\rangle$ and $|\mathrm{T}\rangle$ are defined over spacelike slices of the manifold.

where $\hat{\mathrm{H}}$ is the Hamiltonian operator of the system but the above argument seems to imply that a pure state $|\Psi\rangle$ can in principle evolve in a thermal, mixed, state $|\mathrm{T}\rangle$. Let's try to formalize this more precisely.

The density matrix of a system is an operator

$$\hat{\rho} = \sum_n P_n |\psi_\mathrm{n}\rangle\langle\psi_\mathrm{n}| \tag{10.53}$$

satisfying $\mathrm{Tr}\,\hat{\rho} = \sum_n P_n = 1$. The above general state is what we call a mixed state.

In general for a given observable $\hat{A}$ one has that the expectation value is $<\hat{A}>= \mathrm{Tr}[\hat{\rho}\hat{A}] = \sum_n P_n\langle\psi_\mathrm{n}|A|\psi_\mathrm{n}\rangle$. I.e. a quantum system is in general a superposition of several quantum states and a measurement will find it a particular state $|\psi_\mathrm{n}\rangle$ with probability P_n. The outcomes of measurements on mixed states are different pure quantum states eigenvalues with probability distribution P_n.

On the other hand a pure state is always of the form $\hat{\rho} = |\psi_{\bar{\mathrm{n}}}\rangle\langle\psi_{\bar{\mathrm{n}}}|$ and hence for these states the outcome of a given measurement is pre-determined: a complete measurement of the system provides always the same answer. Moreover, for pure states we also have $\hat{\rho}^2 = |\psi_{\bar{\mathrm{n}}}\rangle\langle\psi_{\bar{\mathrm{n}}}||\psi_{\bar{\mathrm{n}}}\rangle\langle\psi_{\bar{\mathrm{n}}}| = \hat{\rho}$ so that $\mathrm{Tr}\,\hat{\rho}^2 = \mathrm{Tr}\,\hat{\rho} = 1$ while for a mixed state $\mathrm{Tr}\,\hat{\rho}^2 = \sum_n P_n^2 < 1$.

The concept of "information" in the system is captured by the von Neumann entropy $S \equiv -\mathrm{Tr}\,(\hat{\rho}\log\hat{\rho})$. For pure states $S=0$, whilst for mixed states we could have a situation of the sort:

$$\hat{\rho} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \tag{10.54}$$

with states $|0\rangle$ and $|1\rangle$. In this case $S = -\log\frac{1}{2} = \log 2 > 0$.

In quantum mechanics, starting from a pure state there is no way, through unitary evolution, to end up with a mixed state. Indeed, for $\hat{U}\hat{U}^\dagger = \mathbb{1}$ and $|\psi\rangle' = U|\psi\rangle$ then $\hat{\rho}' = U|\psi\rangle\langle\psi|U^\dagger$, meaning that

$$\left(\hat{\rho}'\right)^2 = U|\psi\rangle\langle\psi|U^\dagger U|\psi\rangle\langle\psi|U^\dagger = U|\psi\rangle\langle\psi|U^\dagger = \hat{\rho}'. \tag{10.55}$$

So if $|\psi\rangle$ is a pure state, then $|\psi\rangle' = U|\psi\rangle$ will also be a pure state. *If* we are not making measurements on the system (no wave-function collapse), nor we are tracing over part of the quantum system, then there is no physical process in standard quantum mechanics that could bring a pure state $|\Psi\rangle$ to a thermal (mixed) state $|\mathrm{T}\rangle$ without violating unitarity. Nonetheless, as we already stressed, it seems that black holes can exactly do this.

Here we would like to highlight some notable attempts to solve this problem.

- Hawking tried to solve this problem by introducing a non-unitary *super-scattering matrix* [82], $\hat{\$}$, an operator such that $\hat{\rho}' = \hat{\$}\hat{\rho}$ with $\hat{\$} \neq S\bar{S}$ where S is the usual scattering matrix and $\bar{S}$ is complex conjugate. This idea, however, was soon abandoned due to the difficulty to built a consistent quantum mechanics consistent with it [83].

- Another possibility is that information is released in the quantum gravity regime $M \sim M_{Pl}$ at the end of the evaporation. However, if the black hole is this small there seems to be not enough energy left to encode all the missing information. Indeed, it was noticed by Don Page [84] that any decrease in the entanglement entropy of the Hawking radiation (a process often called "purification") has to start happening before half of the original black hole entropy is left (i.e. the black hole area is halved) otherwise there would be no more enough internal degrees for purifying the total Hawking flux by the end of the evaporation process at $\tau \sim M^3$. The time at which this is exactly realised is $t_{\mathrm{Page}} \sim R_h^3/l_{Pl}^2$ and is called the *Page-time*. In this approach,

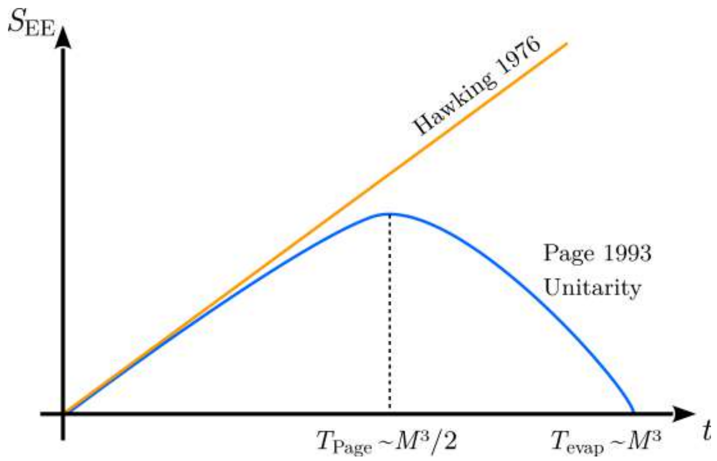


Figure 10.8: Behaviour of the Hawking radiation entanglement entropy in the standard (Hawking) scenario and in Page's scenario.

the entropy of the black hole S is interpreted as Boltzmann entropy, and then there is a time (the Page-time) t_{page} at which

$$S(t_{\mathrm{Page}}) = \frac{S(t_0)}{2}, \quad A(t_{\mathrm{Page}}) = \frac{A(t_0)}{2}, \tag{10.56}$$

where A is the area and t_0 is the time of formation of the black hole. For $t < t_{\mathrm{page}}$, a perturbative analysis should not be able to find any information leaking out; then, for $t > t_{\mathrm{page}}$, as the black hole starts to shrink and entropy decreases, we have lots of information coming out.

- A third, albeit no more very popular idea, is that of a *remnant*: the evaporation could stop at a certain point leaving a residual object with a huge entropy. A common objection is that since there are infinite numbers of ways of forming black holes and letting them evaporate, this remnant must have an infinite number of quantum states in order to encode the information in the initial state. Since the remnants carry mass, it must be possible to pair-produce them in a gravitational field and the total pair-production rate should be proportional to the number of remnant species, and therefore infinite. Thus, it would seem that remnants can be experimentally ruled out by the observed absence of copious pair-production.

However, it is clear that the standard argument that infinite pair-production is inevitable for all types of remnants is too naive, and arguments/calculations (e.g. in Euclidean quantum gravity) have been given that in some theories the pair production rate is finite.

Nonetheless, it seems hard to devise a conservation law forbidding remnant decay since that would also forbid remnant formation while, in the absence of a conservation law, it is hard to understand why matrix elements connecting a massive remnant to the vacuum plus outgoing radiation should be exactly zero. Alternatively one could say that a process leading to the formation of an eternal remnant would violate CPT symmetry.

- Another option is the so-called *firewall*: one may try to solve the information loss by allowing a slow leaking of information through the Hawking radiation itself. Here the argument is based on the fact that the following four, seemingly solid, assumptions

  1. Black holes form and evaporate, as seen by observers close to $\mathscr{I}^+$, in a way fully compatible with standard QFT (no unitarity violation)
  2. Outside the (stretched) horizon of a massive black hole, physics can be described by a set of semiclassical equations
  3. Bekenstein–Hawking entropy does determine the black hole density of states for observers outside the stretched horizon
  4. A freely falling observer experiences nothing out of the ordinary (no new physics) at horizon crossing and until the singularity is met.

  Albeit each of these assumptions is *per se* quite conservative, one can show that they are not reciprocally compatible [85]. The basic argument is that for the outside evolution to be unitary one has to require that the Hawking radiation starts to be purified no later than the Page time. However, such a purification (decrease in the Hawking radiation entanglement entropy) requires an entanglement between late and early Hawking quanta (more precisely each late Hawking quanta will have to be entangled with the whole system of early Hawking quanta). This is required if we want that the entablement entropy of the sum of Early and Late Hawking quanta say S_{EL} is smaller that S_E (which is what it should happen if one is on the descending branch of the Page curve in Fig. 10.8). However, we have seen before when we analysed the Unruh effect that in order to have a regular state everywhere, and in particular at the horizon, the vacuum state should be of the form (9.36).

$$|0\rangle_M \propto \prod_\omega \sum_n e^{-\pi\omega_n} |\mathrm{n_L}\rangle \otimes |\mathrm{n_R}\rangle$$

  The same consideration hold for the Unruh vacuum associated to a black hole formed by a collapse. Hence the late Hawking quanta have to be entangled with the early ones (for any hope to preserve unitarity) and at the same time to their ingoing partners (request of "no drama" at the horizon). But here it comes the problem: this violates the so called monogamy of quantum entanglement for the late Hawking quanta.

  We can bring this to focus by using the property of strong sub-additivity: if one divides a system in three subsystems A, B and C (so that the Hilbert space factorises) then sub-additivity of entropy requires that $S_{AB} + S_{BC} \geq S_A + S_C$. In our case let us take A as the early Hawking quanta, B as the late Hawking quanta and C as the late Hawking partners. In this case we know that $S_{BC} = 0$ as the Unruh state is just a squeezed state of the initial Minkowski vacuum on $\mathscr{I}^-$ (and hence it is pure if one takes both B and C). But we also know that $S_B \sim O(1)$ given that Hawking quanta are thermally populated. Also $S_C = S_B$ (they form a pure state). However, strong sub-additivity would then imply $S_{AB} \geq S_A + S_C > S_A$, i.e. exactly the contrary of what we need for decreasing the entanglement entropy! (and it can be shown that small corrections

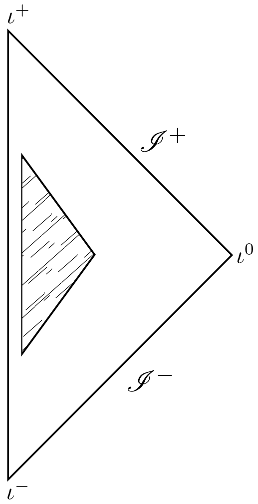


Figure 10.9: Regular spacetime: the black hole horizon is replaced by a trapped region (shaded area) and the singularity is removed.

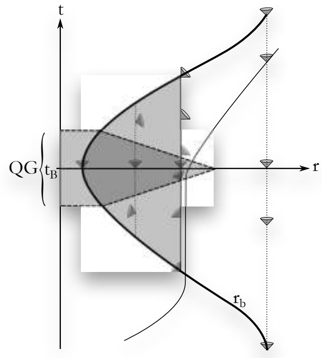


Figure 10.10: Bouncing black hole spacetime by quantum gravity: the horizon is replaced by a trapped region plus a region where classical gravity does not hold and quantum effects dominate. Such region has to extend outside the trapping horizon. (From [87])

cannot fix this problem). It is then clear that if we want to insist that the black hole evaporation has o be a unitary process at least one of the above assumptions must be ditched.

One popular conjecture is that it is the fourth assumption the one to be wrong — alternative one can give up one of the other assumptions (e.g. introducing non-locality, or UV Lorentz violation or violation of standard EFT at large distances). The violation of the fourth assumption implies that B and C do not combine in a pure state (the Unruh vacuum). However, this is the only non-singular state that can be compatible with the black hole casual structure, deviation from it will lead to arbitrary large excitations on the horizon which would take the form of a "firewall" (a highly energetic "wall") at the horizon of any black hole (even very large ones after a sufficient long time).

- Another possible way out is conjecturing that the problem lies in the scenario depicted by Figure 10.7 . Indeed it is easy to see that our problem are stemming from the causal structure of the full evaporation scenario. In particular by the presence of a full fledged event horizon. Can we coincive scenarios where only a trapping horizons form? One possibility is that quantum gravity regularise the inner singularity in black hole solutions so leading to scenarios like *regular black holes* (see e.g [86] and references therein) as depicted in Fig 10.9 or a bounce (see e.g. [87] and references therein) as in Fig 10.10. It might be the case that, somehow, quantum gravity removes the singularity from spacetime. Additionally, since $J^-(\mathscr{I}^+)$ covers now the whole manifold, we don't have an event horizon anymore, but rather a transient trapped region. In this case of regular black holes the main issue seems to reside in the stability of these objects given that the presence of a inner horizon seems to be associated to the so called mass inflation phenomenon by which the horizon could destabilise on very short timescales (at least without fine tuning the solution) [86]. In the case of black hole bounces one is facing instead the necessity to be compatible with the astrophysical observation of long living black hole like objects. This can be achieved by requiring a large difference between the timescale of the bounce for freely falling observers w.r.t that of static observers at infinity but still shorter of the evaporating time ($\tau \sim O(M^2)$ rather then $O(M^3)$). This is a model dependent feature which is still unclear if it is preferred by computable quantum gravity models.

- Finally, another interesting possibility is the presence of additional symmetries on the horizon that are not obviously realised. For instance, Bondi, Metzner and Sachs (BMS) have shown that *asymptotically flat* spacetimes have a very large, infinite dimensional symmetry group (BMS group) which is a generalisation of the Poincarè group and acts non-trivially on $\mathscr{I}^+$. Associated to this symmetries (called super-translations and super-rotations) are a series of conserved charges. Remarkably, it was recently shown that the same symmetries arise also close to the event horizon of a black hole and that they are associated there to"soft hairs". It was then conjectured that this realisation might help resolving the information paradox by encoding in such conserved "soft" gravitons the information associated to infalling matter and Hawking partners. Upon evaporation these soft hairs conservation would then imply a necessary entanglement between the black hole and the radiation.

While attractive also this resolution seems not fully consistent yet. To start with it can be shown that the soft gravitons/hairs decouple from finite energy states and can be factored out. It is then unclear how these IR feature could end up encoding the UV correlations which seems needed to reduce the entanglement entropy of the radiation. This scenarios heavily relies on the symmetries of four dimensional asymptotically flat spacetimes, but we already know that the information problem can be as well formulated in higher dimensions and in more general scenarios like for example AdS black holes.

10.4 Black Hole Entropy Interpretations

As already shown, the Bekenstein–Hawking entropy for a black holes is $S_{\mathrm{BH}} = \frac{A}{4L_P^2}$. What degrees of freedom are responsible for it? Do we have a microscopic interpretation for this entropy?

Let us have a panoramic view of proposals people consider in order to address these questions.

- **Boltzmann entropy:**

  One hypothesis is that S_{BH} corresponds to a "Boltzmann" entropy, meaning that it can be interpreted as arising from the counting of the miscrostates responsible for the black hole configuration. In this proposal $S_{\mathrm{BH}} = k_{\mathrm{B}} \log \mathcal{W}$, where $\mathcal{W}$ is the number of microscopic configurations giving the same M, J and Q. However, where are these degrees of freedom? How can this proposal address the loss of information problem? These are still open issues in this approach.

- **Entanglement entropy:**

  A second approach is to relate the Bekenstein-Hawking entropy to the *entropy of entanglement*, or Von-Neumann entropy $S_{\mathrm{ent}} = -k_{\mathrm{B}} \operatorname{Tr} \rho \log \rho$. Indeed, due to the Hawking radiation emitted by the black hole, the density matrix is mixed and it could yield the right value for the entropy. With more advanced methods that are beyond the scope of the course, one can compute this entanglement entropy, and finds $S_{\mathrm{ent}} = \frac{A}{\Lambda^2}$, where A is indeed the area of the black hole and Λ is a UV cutoff needed in order to make the result finite. The reason for this divergence is that null-rays starting from $\mathscr{I}^-$ accumulate on the horizon, in such a way that the overall result diverges. The aforementioned cutoff can be interpreted as the *thickness* of the horizon: due to quantum gravity fluctuations one cannot localize the horizon with a precision below the Planck length. Hence, this motivated to take $\Lambda \sim M_{\mathrm{Pl}}$ so to recover the right value for the entropy. Unfortunately this is a very rough estimate and more rigorous results are needed to claim an equivalence with S_{BH}. Indeed, this is the result for a single field living around the black hole, but what about other SM fields? The idea is that these fields provide additional divergent terms to S_{ent}. Requiring the result to be finite provides a renormalization condition for the Newton constant and the possible other constants for extended theories of gravity [88].

- **AdS/CFT duality:**

  The AdS/CFT duality relates a gravity theory living on the bulk AdS-space to a lower dimensional CFT on the boundary. In this context there is a different notion of entanglement entropy

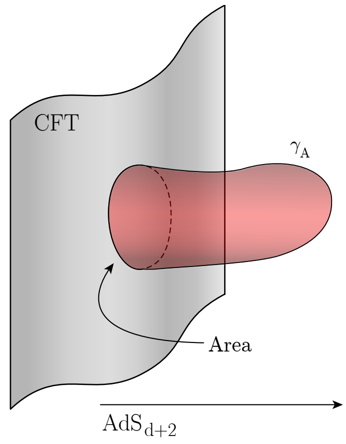


Figure 10.11: Schematic representation of the surface γ_A in AdS.

for a region of area A on the boundary. Given the region A, one takes the minimal surface γ_A on the bulk subtending A. The entanglement entropy is then found to be related to the bulk area:

$$S_{\rm ent} = \frac{\text{Area}(\gamma_A)}{4G_N^{d+2}} \, .$$

This idea was recenty proposed by S. Ryu and T. Takayanagi, but it is still unclear how it could be extended beyond the AdS?CFT scenario.

- **Black hole entropy as Noether charge:**

  An alternative interpretation of the black hole entropy is due to Wald and collaborators. Albeit it does not per se explain the microscopic origin of the entropy, it has the advantage to provide a computational tool for generalising the Bekenstein–Hawking entropy to more general gravitational theories than GR.

  Let us start by considering a collection of dynamical fields in D spacetime dimensions, collectively denoted by ϕ, including a metric tensor g_{ab} plus other possible matter fields, whose dynamics is determined by a diffeomorphism-invariant Lagrangian D-form $\mathbb{L} = \mathcal{L}\epsilon$, with ϵ the spacetime volume element.

  Under a generic variation $\delta\phi$ of the fields, we saw that the variation of $\mathbb{L}$ can be expressed as a sum of a bulk term plus a boundary one:

$$\delta\mathbb{L} = \mathbb{E}_\phi \delta\phi + d\Theta(\phi, \delta\phi) \tag{10.57}$$

  where the $(D-1)$-form Θ is locally constructed out of ϕ and $\delta\phi$. From (10.57) we read that the e.o.m. are $\mathbb{E}_\phi \doteq 0$ for each ϕ[4].

  In particular one can consider infinitesimal variations along a vector field ξ, $\delta\phi = \mathcal{L}_\xi\phi$. By diffeomorphism invariance, to any vector field ξ corresponds a Nother current $(D-1)$-form

$$\mathbb{J}[\xi] = \Theta(\phi, \mathcal{L}_\xi\phi) - \xi \cdot \mathbb{L} \tag{10.58}$$

  which is conserved on shell as indeed

$$d\mathbb{J}[\xi] = d\Theta(\phi, \mathcal{L}_\xi\phi) - d\xi \cdot \mathbb{L} = -\mathbb{E}_\phi \mathcal{L}_\xi\phi \doteq 0. \tag{10.59}$$

  Here we have used that $\mathbb{L}$ is exact by construction, and hence also closed $d\mathbb{L} = 0$ and we have made use of (10.57) and that $\delta\mathbb{L} = \xi \cdot d\mathbb{L} + d(\xi \cdot \mathbb{L}) = d(\xi \cdot \mathbb{L}) = d\xi \cdot \mathbb{L}$.

[4]From now on the dot indicates equalities holding on shell.

The conservation of $\mathbb{J}$ for any ξ implies the existence of a $(D-2)$-form $\mathbb{Q}[\xi]$ [89]

$$\mathbb{J}[\xi] \doteq d\mathbb{Q}[\xi]\,. \tag{10.60}$$

This is the so called "Noether potential" associated to ξ.

$\mathbb{Q}$ enters in the definition of the conserved charges: indeed the Hamiltonian variation, associated with the flow of ξ, over an initial value surface Σ with boundary $\partial\Sigma$, is given by [90]

$$\delta H[\xi] = \int_{\partial\Sigma} \big[\delta\mathbb{Q}[\xi] - \xi\cdot\Theta(\phi,\delta\phi)\big]; \tag{10.61}$$

it is then natural to identify the variations of the energy E and the angular momentum $\vec{J}$ at infinity as[5]

$$\delta E = \int_{S_\infty} \big[\delta\mathbb{Q}[t] - t\cdot\Theta(\phi,\delta\phi)\big], \tag{10.62}$$

$$\delta\vec{J} = -\int_{S_\infty} \big[\delta\mathbb{Q}[\vec{\psi}] - \vec{\psi}\cdot\Theta(\phi,\delta\phi)\big] = -\int_{S_\infty} \delta\mathbb{Q}[\vec{\psi}], \tag{10.63}$$

where S_∞ is the outer boundary of $\partial\Sigma$, and the last equality of (10.63) follows from the fact that $\vec{\psi}$ is tangential to S_∞. (Notice that, as usual, the angular charges are defined up to a conventional minus sign.).

Now, our theory admits a $D-1$ form $\mathbb{B}(\phi)$ such that $\int \xi\cdot\Theta(\phi,\delta\phi) = \delta\int \xi\cdot\mathbb{B}(\phi)$ (like e.g. in GR), then one can define the conserved Hamiltonian charge as

$$H[\xi] = \int_{S_\infty} \mathbb{Q}[\xi] - \xi\cdot\mathbb{B}; \tag{10.64}$$

in particular the angular momentum is exactly the Noether charge at infinity, modulo a sign:

$$\vec{J} = -\int_{S_\infty} \mathbb{Q}[\vec{\psi}]. \tag{10.65}$$

Let us now specialise to a generic stationary black hole spacetime. If the field ξ is taken to be the general Killing field generating the horizon (and null on it) $\xi = t + \vec{\Omega}_H\cdot\vec{\psi}$, then equation (10.61) implies the first law of black hole mechanics: let (i) ξ be a dynamical symmetry, meaning that $\mathcal{L}_\xi\phi \doteq 0$ for all the ϕ's, and (ii) $\delta\phi$ be a variation of the dynamical fields around the BH solution, such that $\delta\phi$ solves the linearized e.o.m.; then $\delta H[\xi] \doteq 0$, iii) we take an initial value surface with boundary $\partial\Sigma = S_\infty \cup B$, with B the bifurcation surface of the black hole. Then given that

$$\delta H[\xi] = \delta E[t] - \vec{\Omega}_H\cdot\delta\vec{J} + \int_{\mathbb{B}} \delta\mathbb{Q}[\xi] - \xi\Theta(\phi,\delta\phi) \tag{10.66}$$

It follows [90, 91] that, being $\delta H[\xi] \doteq 0$ and $\xi\Theta(\phi,\delta\phi) = 0$ on $\mathbb{B}$, one gets

$$\delta E \doteq \frac{\kappa}{2\pi}\delta S + \vec{\Omega}_H\cdot\delta\vec{J} \tag{10.67}$$

where S is $2\pi/\kappa$ times the integral of $\mathbb{Q}$ over the bifurcation surface:

$$S = \frac{2\pi}{\kappa}\int_B \mathbb{Q}[\xi], \tag{10.68}$$

[5] δE contains also work term contributions from long range fields, such as gauge fields.

Since $\kappa/2\pi$ is the Hawking temperature, one interprets S as the thermodynamical entropy of the BH[6].

Finally, for a general gravitational Lagrangian eq.(10.68) can be egeneralized as [91]:

$$S = -2\pi \int_B E_R^{abcd} \hat{\epsilon}_{ab} \hat{\epsilon}_{cd} \bar{\epsilon} \quad , \quad E_R^{abcd} = \frac{\delta \mathcal{L}}{\delta R_{abcd}} \tag{10.69}$$

where $\bar{\epsilon}$ is the area element of B and $\hat{\epsilon}_{ab}$ is the binormal to B.

[6]Note however that this identification fails if the dynamical fields have divergent components at the bifurcation surface. This circumstance occours, for example, in the case of gauge fields, but one can see that in this case the divergences at the horizon can be gauged out by an appropriate gauge fixing, thus recovering the correct expression for the entropy.

11 Semiclassical Gravity

11.1 Semiclassical Einstein Equations

Phenomena like Hawking radiation are computed assuming quantum matter as test field on the classical gravitational background. Of course, this is just an approximation and for example the description of the black hole evaporation requires to take in to account the back-reaction of quantum fields on the spacetime. Unfortunately, missing a full fledged quantum theory of matter and gravitation one has to resort to different kind of approximations to treat this problem.

For instance, if matter is described by some field $\Phi(x)$ with stress-energy tensor $T_{\mu\nu}[\Phi, g]$ in a quantum state $|\Psi\rangle$, and the metric is considered classical, then there are two interesting objects to consider:

$$G_{\mu\nu}[g]\,, \qquad \langle\Psi|\hat{T}_{\mu\nu}[\Phi, g]|\Psi\rangle. \tag{11.1}$$

Provided that it makes any sense, we can postulate the so called semiclassical Einstein equations

$$G_{\mu\nu}[g] = 8\pi G_N \,\langle\Psi|\hat{T}_{\mu\nu}[\Phi, g]|\Psi\rangle\,. \tag{11.2}$$

Whenever this is a good approximation for the dynamics of the metric $g_{\mu\nu}$, we say that we are in a *semi-classical* approximation. Note that classical solutions are obtained from (11.2) in the limit $\hbar \to 0$, where the classical stress-energy tensor is given by

$$T^{cl}_{\mu\nu} \equiv \lim_{\hbar\to 0} \langle\Psi|\hat{T}_{\mu\nu}[\Phi, g]|\Psi\rangle. \tag{11.3}$$

It makes sense to define a quantum stress-energy tensor as

$$T^{Q}_{\mu\nu} \equiv \langle\Psi|\hat{T}_{\mu\nu}[\Phi, g]|\Psi\rangle - T^{cl}_{\mu\nu}; \tag{11.4}$$

this quantity is clearly already of order $\sim \mathcal{O}(\hbar)$.

But is this a justified ansatz? For example, is it consistent to quantise matter while neglecting quantum fluctuations in the gravitational sector? If we do not do so, we have to consider the metric as an operator, $\hat{g}_{\mu\nu}$, and the wave-function (or better wave-functional) as a tensor product $|\Psi, \mathrm{g}\rangle \equiv |\Psi\rangle \otimes |\mathrm{g}\rangle$. The operatorial version of the Einstein equations would then have the form

$$\langle\Psi, g|G_{\mu\nu}[\hat{g}]|\Psi, g\rangle = 8\pi G_N \,\langle\Psi, g|\hat{T}_{\mu\nu}[\Phi, g]|\Psi, g\rangle. \tag{11.5}$$

Complications arise since $G_{\mu\nu}[\hat{g}]$ is a non-linear function of the metric operator, so in general $\langle G_{\mu\nu}[\hat{g}]\rangle \equiv \langle\Psi, g|G_{\mu\nu}[\hat{g}]|\Psi, g\rangle \neq G_{\mu\nu}[\langle\hat{g}\rangle]$. Nonetheless, we can always expand the metric around a classical solution as

$$\hat{g}_{\mu\nu} = g^{cl}_{\mu\nu}\,\mathbb{1} + \hat{\gamma}_{\mu\nu}. \tag{11.6}$$

Formally, one can then obtain (keeping only terms up to quadratic in $\gamma_{\mu\nu}$)

$$G_{\mu\nu}[\langle\hat{g}\rangle] - \langle G_{\mu\nu}[\hat{g}]\rangle = 8\pi G_N \langle \hat{t}_{\mu\nu}[\hat{\gamma}]\rangle \,, \tag{11.7}$$

where $\hat{t}_{\mu\nu}$ can be seen as the contribution to the stress-energy tensor due to quantum fluctuations of the metric.

Hence, we can write the operatorial version of the Einstein equations (11.5) as

$$G_{\mu\nu}[\langle\hat{g}\rangle] = 8\pi G_N \left(\langle \hat{T}_{\mu\nu}\rangle + \langle \hat{t}_{\mu\nu}[\hat{\gamma}]\rangle\right) . \tag{11.8}$$

This procedure suggests that, in the limit where the matter/gravitational state is highly populated (e.g. macroscopic values of the fields that can be approximated by coherent states), the term $\langle t_{\mu\nu}[\hat{\gamma}]\rangle$ becomes negligible and $\langle \hat{g}_{\mu\nu}\rangle$ behave as a classical metric. So formally, in this limit, equation (11.8) does reduce to (11.2).

There are, however, few remarkable problems with this approach. First of all, we are not even sure that a full quantum theory admits an operatorial form like (11.8). A second natural objection appears when de-localized quantum states are considered. For instance, in a semi-classical approach, a matter state of the form

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left(|\mathrm{M}_1\rangle + |\mathrm{M}_2\rangle\right) , \tag{11.9}$$

where $|\mathrm{M_i}\rangle$ is a generic mass state, would not make sense at all in the "averaged" equation (11.2).

Even when these issues are left aside, expression (11.2) remains an extremely hard equation to solve. Indeed, only trivial exact solutions are known at the present time and one usually needs to use approximation techniques in order to handle such a complicate problem.

11.1.1 Approximation Techniques

In dealing with Einstein semiclassical equation several approximation techniques have been devised.

- **Linearization:** The metric can be expanded around flat space as $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, with $h_{\mu\nu}$ considered as a small perturbation. By keeping only linear terms in the field equation, and by taking a vacuum state for matter such that $T^{cl}_{\mu\nu} = 0$, one finds that the perturbation satisfies

$$\Box \left(h_{\mu\nu} - \frac{1}{2} h\, \eta_{\mu\nu}\right) = -\frac{16\pi G_N}{c^2}\, T^{Q}_{\mu\nu} + \mathcal{O}(h^2). \tag{11.10}$$

  In this way, the graviton $h_{\mu\nu}$ appears just as a usual quantum field in flat space. For instance, graviton-matter scattering can be computed in this method (very time consuming though). The only drawback is that General Relativity is non-renormalizable, hence not every Green function is well defined.

- **Test field limit:** this is tantamount to quantum field theory in curved spacetime. One takes a fixed background geometry, a field in its vacuum state (so that the classical SET is zero) and then perform a quantization of this field of the background so to calculate $T^{Q}_{\mu\nu}$ (which being of order $O(\hbar)$ is naturally suppressed) neglecting its back-reaction on the background metric.

- **Full back-reaction:** First, one starts with a classical metric $g^{cl}_{\mu\nu}$ and computes $T^{Q}_{\mu\nu}$ in a vacuum. At this point, the metric is allowed to fluctuate $g_{\mu\nu} = g^{cl}_{\mu\nu} + h_{\mu\nu}$, and fluctuations $h_{\mu\nu}$ are to be computed form the linearised Einstein equation with $T^{Q}_{\mu\nu}$ as a source on the right-hand side. When this is done, one can re-compute $T^{Q}_{\mu\nu}$ in presence of the computed $g_{\mu\nu}$ and re-iterate the previous calculations until convergence is (hopefully) obtained. In practical situations, back-reactions are considered non-relevant, so they are usually neglected (at most the first iteration is performed).

This is all nice and well, but so far we have no prescription on how to compute $T^Q_{\mu\nu}$. The most prominent issue in its calculation is the fact that generally it diverges. This is not a surprise, indeed this is the case also in flat space (recall the divergent Casimir energy). These divergences, even for non-interacting field theories, can all be traced back to coincident-point problems: the stress-energy tensor is made of products of field evaluated at the same space-time point. For example, $T_{\mu\nu}$ for a scalar is

$$T_{\mu\nu}(x) = \frac{1}{2}\nabla_\mu\phi(x)\nabla_\nu\phi(x) - \frac{1}{2}g_{\mu\nu}V(\phi(x)). \tag{11.11}$$

In particular, products of the sort $\sim \phi(x)\phi(x)$ produce UV divergences. In a mathematical language, quantum fields are operator-valued distributions, but a product between two distributions is not a well-defined mathematical concept (for instance $\sim \delta(x)\delta(x)$ has no meaning mathematically speaking). This UV problem in usual Quantum Field Theory is dealt with by the use of either *normal ordering* or *Casimir subtraction*. It is easy to realize that these procedures cannot work in curved spacetime.

- First of all, normal ordering is the procedure of moving all the creation operator to the left and annihilation operators to the right of a given expression. However, we saw that the operators $\hat{a}$ and $\hat{a}^\dagger$ can be mixed by a Bogolyubov transformation in curved space, so normal order is not "conserved" in a generic change of basis.

- For what regards the Casimir subtraction scheme it is also easy to see that problems arise. Consider the FRW metric. As it is well known, this metric is conformally equivalent to Minkowski, so it can be brought to the form

  $$\mathrm{d}s^2_{\mathrm{FRW}} = C(\eta)\left[\mathrm{d}\eta^2 - \mathrm{d}\chi^2\right], \tag{11.12}$$

  where $\mathrm{d}\eta = a(t)^{-1}\,\mathrm{d}t$. A particular solution can be given by $C(\eta) = \cos^2(A\eta)$ so that the scale factor is $a(t) = \left[1-(At)^2\right]^{1/2}$.

  Also in this case, the stress-energy tensor for a scalar field in vacuum formally diverges (as expected) $\langle T_{\mu\nu}\rangle_{\mathrm{FRW}} \to \infty$. However, like in the case of the Casimir energy, we can regularize its components with the introduction of a regulator $\sim \exp[-\alpha(\omega^2+A^2)^{1/2}]$, and by then letting $\alpha \to 0$ once we have removed the divergences. An option is to remove them by subtracting the divergent expectation value in Minkowski $\langle T_{\mu\nu}\rangle_{\mathrm{M}}$. For example, the regularized energy density would look like

  $$\langle T^0_0\rangle_{\mathrm{FRW}} = [32\pi^2 C^2(\eta)]^{-1}\left\{\frac{48}{\alpha^2} + \left(D^2(\eta) - 8A^2\right)\frac{1}{\alpha^2} + A^2\left(\frac{D^2(\eta)}{2} - A^2\right)\log\alpha\right\} + \mathcal{O}(\alpha^0), \tag{11.13}$$

  with $D(\eta) \equiv \frac{\partial}{\partial\eta}\log C$. The Minkowski limit would correspond to assigning $C = 1$, and hence $D = A = 0$. Indeed, in Minkowski one finds

  $$\langle T^0_0\rangle_{\mathrm{M}} = \frac{48}{32\pi^2\alpha^2} + \mathcal{O}(\alpha^0). \tag{11.14}$$

  So we immediately realise that by subtracting (11.14) from (11.13) does not save the day. What we learnt is that in curved spaces divergences are also curvature dependent so the analogue of te Casimir subtraction can easily fail.

11.2 Renormalization of the Stress-Energy Tensor

As we have seen, UV divergences in curved spacetime appear in subtle ways and naive renormalization procedures usually fail. Thus, we need more solid procedures in order to make sense of $\langle T_{\mu\nu}\rangle$.

There are three main ways to regularize and renormalize the stress-energy tensor, all of which can also be used in flat space:

- Dimensional regularization;
- Riemann-zeta function renormalization;
- Point-splitting renormalization.

Given that he first and the second renormalization procedures are rigorous only in Riemannian manifolds [1]; for this reason, we focus on the point-splitting procedure.

11.2.1 Point-splitting Renormalization

To solve the problem of terms like $\phi(x)\phi(x)$, we introduce an auxiliary point y and a geodesics γ_0 connecting y with x. Then the stress-energy tensor in a vacuum state can be written as

$$\langle 0|\hat{T}_{\mu\nu}(x)|0\rangle \equiv \lim_{y\to x}\langle 0|\hat{T}_{\mu\nu}(x,y;\gamma_0)|0\rangle, \tag{11.15}$$

where it is understood that, since $T_{\mu\nu}$ is quadratic in the fields, one of them is evaluated at x while the other at y. It can be shown that for small distances (between y and x) this procedure does not depend on the particular choice of γ_0. Another option would be to take an average over all possible paths connecting the two points, but we are not going to dwell into this.

The result of this procedure has the following structure [45]

$$\begin{aligned}\langle 0|\hat{T}_{\mu\nu}(x)|0\rangle_{\text{ren}} &= \mathscr{D}_{\mu\nu}(x,y;\gamma_0)\, G^{(1)}(x,y;\gamma_0)_{\text{ren}}\\ G^{(1)}(x,y;\gamma_0)_{\text{ren}} &\equiv \langle 0|\,\{\phi(x),\phi(y)\}\,|0\rangle_{\text{ren}}.\end{aligned} \tag{11.16}$$

The object $\mathscr{D}_{\mu\nu}(x,y;\gamma_0)$ is a second-order differential operator that depends solely on the background geometry (and is therefore regular), whilst $G^{(1)}(x,y;\gamma_0)_{\text{ren}}$ is the renormalized *Hadamard Green function* for the field $\phi(x)$. The form (11.16) is written on the assumption that ultimately $T_{\mu\nu}(x)$ contains second-order derivatives and that it is constructed with products of two fields.

Green functions

In general, there is a plethora of interesting Green functions.

- **Hadamard Green function:**

$$G^{(1)}(x,y;\gamma_0) \equiv \langle 0|\,\{\phi(x),\phi(y)\}\,|0\rangle \tag{11.17}$$

- **Pauli-Jordan or Schwinger Green function:**

$$iG(x,y) \equiv \langle 0|[\phi(x),\phi(y)]|0\rangle \tag{11.18}$$

- **Wightman Green functions:**

$$G^{+}(x,y) \equiv \langle 0|\phi(x)\phi(y)|0\rangle \tag{11.19a}$$

$$G^{-}(x,y) \equiv \langle 0|\phi(y)\phi(x)|0\rangle \tag{11.19b}$$

The above Green functions are related as

$$\begin{aligned} iG(x,y) &= G^{+}(x,y) - G^{-}(x,y)\\ G^{(1)}(x,y) &= G^{+}(x,y) + G^{-}(x,y)\end{aligned}$$

[1] It has also been argued that Zeta regularization and dimensional regularization are equivalent since they use the same principle of using analytic continuation in order for a series or integral to converge.

- **Feynman Green function:**

$$
\begin{aligned}
iG_{\mathrm{F}}(x,y) &\equiv \langle 0|T\phi(x),\phi(y)|0\rangle \\
&= \theta(y_0-x_0)G^{-}(x,y)+\theta(x_0-y_0)G^{+}(x,y)
\end{aligned}
\tag{11.20}
$$

- **Retarded Green function:**

$$
G_{\mathrm{R}}(x,y) \equiv -\theta(x_0-y_0)G(x,y) \tag{11.21}
$$

- **Advanced Green function:**

$$
G_{\mathrm{A}}(x,y) \equiv \theta(y_0-x_0)G(x,y) \tag{11.22}
$$

Green functions for thermal states

In curved spacetimes, not only vacuum states are relevant, but also thermal states are important. A thermal state $|\mathrm{T}\rangle$ is a statistical distribution of pure states. To deal with them we have to specify the probability distribution for the thermal bath. In the *Gran-canonical ensemble* the density matrix is, as usual, given by

$$
\hat{\rho}_T = e^{\beta(\Omega+\mu\hat{N}-\hat{\mathrm{H}})}, \tag{11.23}
$$

where $\beta=(k_B T)^{-1}$, $\hat{\mathrm{H}}$ is the Hamiltonian, μ is the chemical potential, $\hat{N}$ is the number operator, and the *gran-potential* Ω fixes the normalization for $\hat{\rho}_T$. For any operator $\hat{A}$, the expectation value in a thermal state is

$$
\langle\hat{A}\rangle_{\mathrm{T}} = \sum_i \rho_i^T\langle\psi_i|\hat{A}|\psi_i\rangle = \mathrm{Tr}\left[\hat{\rho}\hat{A}\right]. \tag{11.24}
$$

where $\rho_i^T = \langle\psi_{\mathrm{i}}|\hat{\rho}_T|\psi_{\mathrm{i}}\rangle$.

The Wightman functions evaluated over such a thermal state (with $\mu=0$) have the following striking property:

$$
G_{\mathrm{T}}^{\pm}(x_0,x;y_0,y) = G_{\mathrm{T}}^{\pm}(x_0+i\beta,x;y_0,y), \tag{11.25}
$$

so a thermal environment induces a periodicity along the imaginary time-axis. Moreover the Hadamard Green functions satisfy

$$
G_{\mathrm{T}}^{(1)}(x_0,x;y_0,y) = \sum_{k\in\mathbb{Z}} G^{(1)}(x_0+ik\beta,x;y_0,y), \tag{11.26}
$$

i.e. the thermal Green functions can be written as an infinite imaginary sum of the corresponding $T=0$ Green functions. This last result can be used to show [5] the useful fact that periodicity in Euclidean time implies thermality in Lorentzian time. From this very fact one could be able to derive thermally in stat ionary back hole spacetimes in a geometric way.

Semiclassical Euclidean quantum gravity techniques [92] play a key role in the investigation of the thermodynamics of black holes [93]. Given the classical Einstein–Hilbert action plus Gibbons–Hawking term for gravity and the action of classical matter fields, one can formulats the Euclidean path integral by means of a Wick rotation $t\to -i\tau$ [a].

In particular, one finds that the dominant contribution to the Euclidean path integral is given by gravitational instantons (i.e. non-singular solutions of the Euclidean Einstein equations).

In spacetimes with event horizons, metrics extremizing the Euclidean action are gravitational instantons only after removal of a conical singularity at the horizon [92]. A period must therefore be fixed in the imaginary time, $\tau\to\tau+\beta$, which becomes a sort of angular coordinate.

To be concrete we can take the standard Schwarzschild metric in metric signature:

$$
\mathrm{d}s_{\mathrm{E}}^2 = \left(1-\frac{r_{\mathrm{h}}}{r}\right)\mathrm{d}\tau^2 + \left(1-\frac{r_{\mathrm{h}}}{r}\right)^{-1}\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2 \tag{11.27}
$$

This solution still shows singularities at $r=r_{\mathrm{h}}=2M$ and $r=0$. If we now define a new radial variable, $\rho = 2\sqrt{r - r_{\rm h}}$, the metric acquires the form

$$
ds_{\rm E}^2 = \left(\frac{\rho^2}{4} + r_{\rm h}\right)^{-1} \frac{\rho^2}{4} d\tau^2 + \left(\frac{\rho^2}{4} + r_{\rm h}\right) d\rho^2 + \left(\frac{\rho^2}{4} + r_{\rm h}\right) d\Omega^2 \tag{11.28}
$$

For $\rho \to 0$ one then gets:

$$
ds_{\rm E}^2 = \frac{1}{r_{\rm h}} \frac{\rho^2}{4} d\tau^2 + r_{\rm h}\, d\rho^2 + r_{\rm h}^2\, d\Omega^2 \tag{11.29}
$$

If one now defines $\bar{\rho} \equiv \rho\sqrt{r_{\rm h}}$ and $\bar{\tau} \equiv \tau/(2r_{\rm h})$ the metric takes the so called conical form:

$$
ds_{\rm E}^2 = \bar{\rho}^2\, d\bar{\tau}^2 + d\bar{\rho}^2 + r_{\rm h}^2\, d\Omega^2 \tag{11.30}
$$

Smoothness at $\bar{\rho} = 0$ (no conical singularity) requires that the rescaled imaginary time behaves like an angle. I.e. $\bar{\tau} = \bar{\tau} + 2\pi$ which in turns imply $\tau = \tau + 2\pi/kappa$. So, the required period β corresponds exactly to the inverse Hawking temperature $\beta_{\rm H} = 2\pi/\kappa$ and because of the relation between the periodicity of the Euclidean Green's functions and the thermal character of the corresponding Green's functions in Lorentzian signature, this implies that the state of fields on this spacetime will have to be thermal at the Hawking temperature.

In such a way thermodynamics appears as a requirement of consistency of quantum field theory on spacetimes with Killing horizons, and in this sense we shall define such a thermodynamics as "intrinsic" to the spacetime rather than the property of a specific field. (Note however, that the situation changes dramatically in the case of extremal black holes where one can see that there is no way to fix the period of the imaginary time, and hence the intrinsic temperature. [b])

Hence, in these cases, the effective action is truly a free energy function βF. Indeed, we can consider the partition function for solutions where a temperature can be fixed in the way described above.

The partition function in a canonical ensemble Z can be written as

$$
Z = {\rm Tr}\ \exp -(\beta H) = \int \mathcal{D}\left[\phi, g_{\mu\nu}\right] \exp\left[-I\left(\phi, g_{\mu\nu}\right)\right], \tag{11.34}
$$

Using the fact that a black hole solution is an extremum of the (Euclidean) gravitational action, at the tree level of the semiclassical expansion one then obtains

$$
Z \sim \exp\left[-I_{\rm E}\right], \qquad I_{\rm E} = \frac{1}{16\pi}\int_M \left[(-R + 2\Lambda) + L_{matter})\right] + \frac{1}{8\pi}\int_{\partial M} [K], \tag{11.35}
$$

where $I_{\rm E}$ is the on-shell Euclidean action and $[K] = K - K_0$ is the difference between the extrinsic curvature of the manifold and that of a reference background. The Euclidean action is then related to the free energy F as $I_{\rm E} = \beta_{\rm H} F$ and the corresponding entropy can be computed as

$$
S \equiv \beta^2 \frac{\partial F}{\partial \beta} \tag{11.36}
$$

and similarly the internal energy and the heat capacity are respectively given by $U = \partial_\beta(\beta F)$ and $C = -\beta^2 \partial_\beta^2(\beta F)$. This formulas provide the black hole thermodynamical quantities at the tree level while in the case I_E is the one loop effective action of the fields on the black hole background the same formula can be used to derive the analogue quantities of the radiation of the background (and in this case the entropy coincides with the entanglement entropy of the radiation).

[a] Working in the Euclidean signature is a standard technique in QFT in flat backgrounds. It has the advantages of making the path integral "weighted" by a real factor and in general transforms the equation of motion of free fields from hyperbolic to elliptical, a feature that often leads to just one Green function satisfying the boundary conditions.

[b]Indeed for an extremal black hole general form of the metric is

$$\mathrm{d}s_{\mathrm{E}}^2 = \left(1 - \frac{r_{\mathrm{h}}}{r}\right)^2 \mathrm{d}\tau^2 + \left(1 - \frac{r_{\mathrm{h}}}{r}\right)^{-2} \mathrm{d}r^2 + r^2 \mathrm{d}\Omega^2 \tag{11.31}$$

Performing the same transformations as before this can be cast in the form:

$$\mathrm{d}s_{\mathrm{E}}^2 = \left(\frac{\rho^2}{4} + r_{\mathrm{h}}\right)^{-2} \left(\frac{\rho^2}{4}\right)^2 \mathrm{d}\tau^2 + \left(\frac{\rho^2}{4} + r_{\mathrm{h}}\right)^2 \frac{4}{\rho^2} \mathrm{d}\rho^2 + \left(\frac{\rho^2}{4} + r_{\mathrm{h}}\right) \mathrm{d}\Omega^2 \tag{11.32}$$

which in the $\rho \to 0$ limit gives

$$\mathrm{d}s_{\mathrm{E}}^2 = \left(\frac{1}{r_{\mathrm{h}}}\right)^2 \frac{\rho^4}{16} \mathrm{d}\tau^2 + \frac{4r_h^2}{\rho^2} \mathrm{d}\rho^2 + r_{\mathrm{h}}^2 \mathrm{d}\Omega^2 \tag{11.33}$$

It is now easy to see that in this case passing to a system of coordinates equivalent to the previous one given by $\bar{\tau}$ and $\bar{\rho}$ does not give rise to any "conical-like" metric. This implies that, for extremal black holes, there is no way to fix the period of the imaginary time and hence the intrinsic temperature.

Let us go back to (11.16). Although the problem of renormalizing $T_{\mu\nu}$ is reformulated in terms of Green functions, we still need to remove the infinities from them.

In general, $\mathscr{D}_{\mu\nu}$ is UV free and contains only geometrical quantities. One has instead to deal with the Green function infinities in the coincidence limit. The basic idea behind the renormalization of these infinities is that they are universal and related to the UV structure of the vacuum. As such, this divergence structure of the Green functions should show a universal behaviour similar to the one of Minkowski vacuum if the state is well defined all over our spacetime. State that satisfy this assumption are called *Hadamard states*.

In Minkowski, the divergent structure of the Hadamard function is

$$\lim_{y \to x} G^{(1)}(x, y; \gamma_0) = \frac{1}{4\pi^2 \sigma_0(x, y)} + \alpha \log \sigma_0(x, y) + \varpi(x, y), \tag{11.37}$$

where $\sigma_0(x, y)$ is half of the squared geodesics distance

$$\sigma_0(x, y) = \frac{1}{2}(x - y)^\mu (x - y)_\mu + \epsilon^2 + 2i\epsilon(x_0 - y_0)\,, \qquad \epsilon \to 0^+$$

Note that the terms dependent on ϵ have to be introduced to deal with null distances. Note that the terms depending on $\sigma_0(x, y)$ contain the divergent parts, while the function $\varpi(x, y)$ is in general a finite quantity.

In curved space, the divergent structure of the Hadamard function, evaluated on a Hadamard state, does reproduce the same UV behaviour, and indeed one finds

$$\lim_{y \to x} G^{(1)}(x, y; \gamma_0) = \frac{\sqrt{\Delta(x, y)_{\gamma_0}}}{4\pi^2} \left[\frac{1}{\sigma_{\gamma_0}(x, y)} + v_{\gamma_0}(x, y) \log |\sigma_{\gamma_0}(x, y)| + \varpi_{\gamma_0}(x, y)\right]; \tag{11.38}$$

where it is self-evident that the divergent structure of (11.38) is the same as the one in flat spacetime (11.37). The only difference is the appearance of the (regular) spacetime dependent functions $v_{\gamma_0}(x, y)$ and $\varpi_{\gamma_0}(x, y)$ and most of all the overall factor $\sqrt{\Delta(x, y)_{\gamma_0}}$, where

$$\Delta(x, y)_{\gamma_0} \equiv \det \left[\frac{\nabla^x_\mu \nabla^y_\nu \sigma_{\gamma_0}(x, y)}{\sqrt{g(x) g(y)}}\right] \tag{11.39}$$

is the so-called *van Vleck determinant* (these result change in dimensions different from 4). This is an exquisitely geometrical object which takes into account the focussing effects due to tidal forces (see e.g. [45]).

Finally, the renormalized Green function can be obtained by subtracting the divergent part of (11.38):

$$G^{(1)}_{\mathrm{ren}}(x, y) = \lim_{y \to x} G^{(1)}(x, y; \gamma_0) - \frac{\sqrt{\Delta(x, y)_{\gamma_0}}}{4\pi^2} \left[\frac{1}{\sigma_{\gamma_0}(x, y)} + v_{\gamma_0}(x, y) \log |\sigma_{\gamma_0}(x, y)|\right]. \tag{11.40}$$

The form of $\mathscr{D}_{\mu\nu}$ is in general a complicated expression. For the particular case of a conformally-coupled massless scalar field can be found for example in [45, p. 285] (too long and clumsy to be worth reporting here). However, once it is determined and applied to (11.40) the RSET is finally derived.

Note that there is nothing "new" in this renormalization. Indeed the first term in (11.38) is related to a wave-function renormalization whereas the logarithmic term represents a mass renormalization. One can see this is the case for example by considering the Green function in Minkowski for a massive scalar field: in order to have a regular $x \to y$ limit one needs to subtract a term $\sim 1/x^2$ and a term $\sim \log(x)$. Even in curved spacetimes, because of local Lorentz invariance, the structure of the divergences is the same as in Minkowski and Equation (11.38) is only a covariant generalization of the flat-space case (if the state we consider is well-behaved i.e. is Hadamard).

However, in curve spacetime this renormalization procedure is not completely unambiguous. Indeed, the Hadamard subtraction scheme is always ambiguous up to further finite renoprmalization:: in principle it is always possible to subtract the divergent part plus some arbitrary finite function so changing the resulting finite part $\varpi(x, y)$.

In order to address this ambiguity, there are some useful physical criteria that can be laid down so to fix $\langle T^{\mu\nu}(x)\rangle_{\text{ren}}$ These are often named the **Wald's criteria** for a well-defined renormalized stress-energy tensor:

1. Conservation equation: it must be the case that the renormalized stress-energy tensor remains conserved $\nabla_\mu \langle T^{\mu\nu}(x)\rangle_{\text{ren}} = 0$ (compatibility with Contracted Bainchi identity in GR);
2. Causality: $\langle T^{\mu\nu}(x)\rangle_{\text{ren}}$ for a fixed "in" state $\langle T^{\mu\nu}(x)\rangle_{\text{ren}}$ at a point p must depend only on $\bar{J}^-(p)$ (causal past of p); (the time reversal of this statement must be true for any fixed "out" state).
3. Standard results for off-diagonal elements: for positive-norm states the elements $\langle \Phi|T_{\mu\nu}|\Psi\rangle$ must behave as classical objects; This is simply the observation that as $\langle \Phi|T_{\mu\nu}|\Psi\rangle$ is anyway finite for orthonormal states $\langle \Phi|\Psi\rangle = 0$ [2], then its value must be formally the same one as in the classical case.
4. Minkowski limit: for vanishing curvature, the stress-energy tensor must reduce consistently to its normal ordered version in Minkowski.

Given these requirements it is possible to proof that albeit they cannot pinpoint the corrected RSET to use, any two SET will differ just by a local, conserved, geometrical tensor

$$H_{\mu\nu} \equiv T_{\mu\nu} - \bar{T}_{\mu\nu}. \tag{11.41}$$

Let us stress that this tensor depends only on the spacetime geometry and not on the state, which indeed it can be always decomposed as

$$H_{\mu\nu} = c_1 g_{\mu\nu} + c_2 G_{\mu\nu} + c_3 \left(\frac{1}{\sqrt{g}} \frac{\delta}{\delta g^{\mu\nu}} \sqrt{g} R^2 \right) + \dots . \tag{11.42}$$

Clearly, the coefficients c_i appearing in this expression can be interpreted as providing a renormalization for the bare quantities appearing on the left-hand-side of the Einstein equations. For instance, c_1 renormalizes the Cosmological constant Λ, c_2 renormalizes the Newton constant, and so on. Therefore, the renormalization of $T_{\mu\nu}$ immediately provides a renormalization for matter and gravitation couplings (and can induce higher order curvature terms even if they are not present in the bare action).

[2] Think for example (formally) $\langle \Phi|H|\Psi\rangle = \sum_k \omega_k \left(\langle \Phi|a_k^\dagger a_k|\Psi\rangle + 1/2 \langle \Phi|\mathbb{1}|\Psi\rangle \right) = \sum_k \omega \left(\langle \Phi|a_k^\dagger a_k|\Psi\rangle \right)$.

11.3 Semiclassical Collapse

In most situations, gravitational quantum effects are negligible: gravitational interactions are technically irrelevant at low energies and become strong only around Planckian energies. This means that in a scattering experiments we can neglect gravity when we deal with energies below $E \sim G^{-1/2} \sim M_{\rm Pl}$.

On the other hand classical non-linearities of General Relativity arise in many situations, for instance close to the formation of an horizon. In these scenarios one could wonder whether quantum effects can become important e.g. due to large blueshift effects close to the horizon.

Indeed, one could see that the transplanckian problem discussed in section 10.3.1 originates already during the collapse albeit being usually formulated in static spacetimes.

We can, as usual, encode the dynamics of the geometry in the relation $U = p(u)$ between the affine null coordinates U and u, regular on $\mathscr{I}^-$ and $\mathscr{I}^+$, respectively. Neglecting back-scattering, a mode of the form

$$\psi_\omega(r,t) \approx \frac{1}{(2\pi)^{3/2}(2\omega)^{1/2}\, r}\, \mathrm{e}^{-i\omega u}\ , \tag{11.43}$$

near $\mathscr{I}^-$ takes, near $\mathscr{I}^+$, the form

$$\varphi_\Omega(r,t) \approx \frac{1}{(2\pi)^{3/2}(2\Omega)^{1/2}\, r}\, \mathrm{e}^{-i\Omega p(u)}\ . \tag{11.44}$$

This can be regarded, approximately, as a mode of the type presented in equation (11.43), but now with u-dependent frequency $\omega(u,\Omega) = \dot{p}(u)\, \Omega$, where the dot denotes differentiation with respect to u. (Of course, this formula just expresses the redshift undergone by a signal in travelling from $\mathscr{I}^-$ to $\mathscr{I}^+$.)

In general we can expect a mode to be excited if the standard adiabatic condition

$$|\dot{\omega}(u,\Omega)|/\omega^2 \ll 1 \tag{11.45}$$

does *not* hold. It is not difficult to see that this happens for frequencies smaller than

$$\Omega_0(u) \sim |\ddot{p}(u)|/\dot{p}(u)^2\ . \tag{11.46}$$

One can then think of $\Omega_0(u)$ as a frequency marking, at each instant of retarded time u, the separation between the modes that have been excited ($\Omega \ll \Omega_0$) and those that are still unexcited ($\Omega \gg \Omega_0$).

Moreover, Planck-scale modes (as defined on $\mathscr{I}^-$) are excited in a finite amount of time, even *before* the actual formation of any trapped region. Indeed, they start to be excited when the surface of the star is above the classical location of the horizon by a proper distance D of about one Planck length, as measured by Schwarzschild static observers. We can see this by observing that the red-shift factor satisfies close to the horizon (where $p(u) \simeq U_{\mathcal{H}} - A_1 e^{-\kappa u}$)

$$(1 - 2M/r)^{1/2} \sim \omega/\Omega = \dot{p}(u) \sim \kappa/\Omega_0\ , \tag{11.47}$$

where $\kappa = (4M)^{-1}$ is the surface gravity. This then implies $(r - 2M) \sim \kappa/\Omega_0^2$, where we have used $\kappa M \sim 1$. Hence

$$D \sim (r - 2M)\,(1 - 2M/r)^{-1/2} \sim 1/\Omega_0\ . \tag{11.48}$$

Hence, the trans–Planckian problem has its roots at the very onset of the formation of the trapping horizon.

Should we have to worry about this? Indeed, the resurfacing of the transplanckain issue also in a semiclassical description of the gravitational collapse rise the question of wether quantum mechanical

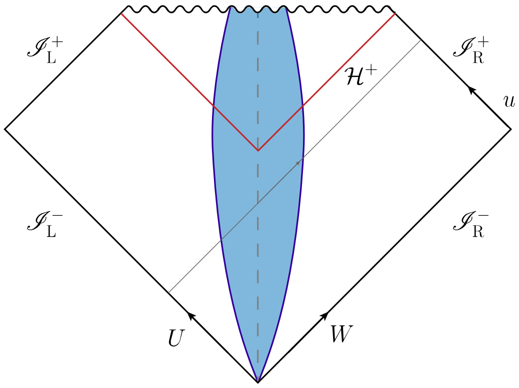


Figure 11.1: Penrose diagram for the collapse of a star in $1 + 1$-dimensions.

effects can affect the formation of an horizon and appears to be of great relevance if we want to assess the robustness of Hawking radiation. After all, the gravitational collapse of a star is expected to happen in a regime in which we can apply semiclassical gravity: we have quantum fields living in a classical curved space and quantum effects are expected to be relatively small. Let us then consider this problem in more detail (this discussion is based on [94]).

Consider a collapsing star. Outside of the star the vacuum state is Boulware and it is regular outside the horizon. We want to assess whether backreaction becomes important during the formation of the horizon (when classical non-linearities are large).

Since we are dealing with quantum fields it is better to work in Heisenberg representation, where the quantum state is fixed during the evolution and is globally defined over our spacetime. We call this state $|\mathrm{C}\rangle$ (as for Collapse). At early times this is chosen to be the vacuum state for an observer at $\mathscr{I}^-$ (i.e. the state is annihilated by the annihilation operator associated to free waves at $\mathscr{I}^-$). Having an Unruh state at late times would imply that a free-falling observer at $\mathcal{H}^+$ would not detect particles and that $\langle T_{\mu\nu}\rangle = 0$. Moreover, for a far-away observer at $\mathscr{I}^+$ such state would be at late times a thermal state at the Hawking temperature.

Let us work in a manageable $1+1$ geometry for a spherical collapse, and consider a scalar field on such background. Albeit quite simple this set-up is sufficiently general for drawing lessons about the possible relevance of quantum backreaction.

Before starting the calculation we can make some guesses on what we expect to happen based on what we learned in the previous lectures. Since we are considering a spectator free-field in the background geometry of the collapse, its renormalized stress-energy tensor (RSET) can only depend on the geometrical quantities (invariants) and physical parameters (such as the mass of the field). At the horizon we only have coordinate singularities, and invariant quantities such as the Riemann tensor squared are finite on $\mathcal{H}^+$. Therefore we expect the RSET to be finite as well. It is not trivial however to see how the Boulware state (appropriate for a spacetime with a static start) could be turned into the Unruh one by the onset of horizon formation.

More formally there is a theorem due to Fulling, Sweeny and Wald [95] that guarantees that in a globally hyperbolic spacetime (i.e. no Cauchy horizons) if a state is Hadamard on $\mathscr{I}^-$ (or any given Cauchy hypersurface) it will remain so everywhere. In a semiclassical description of a collapse under consideration, this can be used to predict that the state $|\mathrm{C}\rangle$, which coincides with vacuum state for observers on $\mathscr{I}^-$, *cannot* also be a vacuum state for observers on $\mathscr{I}^+$ as this would imply that $|\mathrm{C}\rangle$ is the Boulware state, which however it is not regular at $\mathcal{H}^+$.

We can choose coordinates U and W defined as the coordinates respectively on on $\mathscr{I}_L^-$ and $\mathscr{I}_R^-$ as in Fig. 11.1. The metric in these null-coordinates reads

$$g = -\mathcal{C}(U, W)\, \mathrm{d}U\ \mathrm{d}W\ . \tag{11.49}$$

Note that outside of the star, because of Birkhoff's theorem, we have a static geometry. We are also going to assume there is a relation between the coordinate U and the null coordinate on $\mathscr{I}_{\mathrm{R}}^+$, that we call u, of the form $U = p(u)$. This function $p(u)$ tells us about the dynamics of the collapse: rays going from $\mathscr{I}_\mathrm{L}^-$ to $\mathscr{I}_\mathrm{R}^+$ are blue/red shifted and this information is contained in $p(u)$. Also, we can write the metric as a function of the coordinates (u, W) as

$$g = -\bar{\mathcal{C}}(u, W)\, \mathrm{d}u\, \mathrm{d}W = -\mathcal{C}\, \dot{p}(u)\, \mathrm{d}u\, \mathrm{d}W\ . \tag{11.50}$$

We therefore have the relations

$$\mathcal{C} = \bar{\mathcal{C}}/\dot{p}\ , \qquad \partial_u = \frac{\partial U}{\partial u}\partial_U = \dot{p}\partial_U\ . \tag{11.51}$$

The RSET for any massless quantum field (corresponding to a vacuum on $\mathscr{I}^-$) is

$$T_{UU} \propto \mathcal{C}^{1/2}\partial_U^2\mathcal{C}^{-1/2}\ , \tag{11.52}$$

$$T_{WW} \propto \mathcal{C}^{1/2}\partial_W^2\mathcal{C}^{-1/2}\ , \tag{11.53}$$

$$T_{UW} \propto R\ . \tag{11.54}$$

This relations are given explicitly in [5] and are related to the so called trace anomaly.

Trace anomaly. For a conformal invariant field the trace of the classical SET is zero. Upon quantization one in general has that the total one loop effective action W will also be in general conformal invariant so that if $g \to \Omega(x)^2 g$ then

$$\langle T^\mu_\mu \rangle = -\frac{\Omega(x)}{\sqrt{g}}\frac{\delta W}{\delta\Omega}\bigg|_{\Omega=1} = 0 \tag{11.55}$$

In order to renomalize our action will shall still need to separate the finite and divergent parts of the effective action $W = W_{ren} + W_{div}$, however the resulting two terms are not separately conformal invariant and the both have a non-zero trace. The origin of this is that the action is conformal invariant in $d = 4$ but not exactly so e.g. if one performs a dimensional regularisation to handle the divergent part $d = 4 + \epsilon$. Even in the limit $\epsilon \to 0$ this leaves an imprint in the form of the trace anomaly. For a conformal invariant field on a conformally flat spacetime one can show that all of the renormalised SET will be accounted for by the trace anomaly coming from W_{ren}.

For us the prefactors are not important, what is relevant is the dependence of the geometrical quantities. The components T_{UW} and T_{WW} are manifestly well-behaved everywhere: they are the same as in an always static geometry and do not contain the dynamical quantity $\dot{p}(u)$. On the other hand T_{UU} can be problematic, so we are going to check whether it can become singular.

First, using (11.50) and (11.51) we can write

$$T_{UU} \propto \mathcal{C}^{1/2}\partial_U^2\mathcal{C}^{-1/2} = \frac{1}{\dot{p}}\left[\bar{\mathcal{C}}^{1/2}\partial_u^2\bar{\mathcal{C}}^{-1/2} - \dot{p}^{1/2}\partial_u\dot{p}^{-1/2}\right]\ . \tag{11.56}$$

The first term on the right-hand side is a term that would be present outside of a static star, and contains the divergent term of the Boulware vacuum $\sim (1 - 2M/r)^2$. The second term (containing $\dot{p}$) is instead related to the dynamics of the collapse and describes the related particle flux.

To track accurately quantum effects at the onset of the horizon it is obviously best to use system of horizon-crossing coordinates. Therefore we introduce a horizon penetrating system: the *Painlevè-Gullstrand coordinates* (x, t) defined in (10.47).[3] The metric takes the form (omitting the *PG* subscripts for simplicty)

$$g = -c^2(x,t)\, \mathrm{d}t^2 + \left[\mathrm{d}x - v(x,t)\, \mathrm{d}t\right]^2\ . \tag{11.57}$$

[3]For the reader convenience we recall them here

$$\begin{cases} t_{\mathrm{PG}} = t_{sch} \pm \left[4M\,\mathrm{arctanh}\left(2\sqrt{\frac{2M}{r_{sch}}}\right) - 2\sqrt{2Mr_{sch}}\right] \\ x_{\mathrm{PG}} = r_{sch} \end{cases}$$

The horizon is located at $c+v=0$. The null-rays trajectories are obtained from the relation $\mathrm{d}x = (\pm c + v)\,\mathrm{d}t$ respectively for left- and right-moving rays (the dependence on x, t will be always understood).

Outside the star the metric is of course static, so c and v become time-independent. When we reach infinity the metric has to be asymptotically the flat metric, so we have $\lim_{x\to\infty} c(x) = 1$, $\lim_{x\to\infty} v(x) = 0$.

We need to connect the two systems of coordinates. To do so we integrate the equation for the right-going null-rays starting from an event (t, x) just outside the star to another one (t_f, x_f) at $\mathscr{I}^+_{\mathrm{R}}$

$$\mathrm{d}t = \frac{\mathrm{d}x}{c(x)+v(x)} \Longrightarrow t_f - t = \int_x^{x_f} \frac{\mathrm{d}x'}{c(x')+v(x')} \,. \tag{11.58}$$

Assuming asymptotic flatness, $c(+\infty) = 1$ and $v(+\infty) = 0$, the null coordinate u is then given by [4].

$$u \equiv \lim_{t_f\to\infty}(t_f - x_f) = t - x_f + \int_x^{x_f} \frac{\mathrm{d}x'}{c(x')+v(x')} = t - \int^x \frac{\mathrm{d}x'}{c(x')+v(x')} = \,. \tag{11.59}$$

In contrast, along an incoming ray leaving asymptotic past infinity $\mathscr{I}^-_{\mathrm{R}}$ at an event (t_i, x_i) and remaining outside the star, we have instead

$$\mathrm{d}t = \frac{\mathrm{d}x}{-c(x)+v(x)} \Longrightarrow t - t_i = -\int_{x_i}^{x} \frac{\mathrm{d}x'}{c(x')-v(x')} \,. \tag{11.60}$$

Then the null coordinate W is defined as

$$W \equiv \lim_{t_i\to-\infty}(t_i + x_i) = t + x_i + \int_{x_i}^{x} \frac{\mathrm{d}x'}{c(x')-v(x')} = t + \int^x \frac{\mathrm{d}x'}{c(x')-v(x')} \,. \tag{11.61}$$

The expressions (11.59) and (11.61) for u and W give us useful expressions outside the star:

$$\begin{aligned}
\partial_x u &= -\frac{1}{c+v} \,, \\
\partial_x U &= \partial_x p(u) = \dot{p}(u)\partial_x u = -\frac{\dot{p}(u)}{c+v} \,, \\
\partial_t U &= \dot{p}(u)\partial_t u = \dot{p}(u) \,, \\
\partial_x W &= \frac{1}{c-v} \,, \\
\partial_t W &= 1 \,.
\end{aligned} \tag{11.62}$$

We see indeed that the dynamics enters only in U and not in W. We will also need the change of variables for the partial derivative

$$\partial_u = \frac{\partial t}{\partial u}\partial_t + \frac{\partial x}{\partial u}\partial_x = \frac{c+v}{2c}\partial_t - \frac{c^2-v^2}{2c}\partial_x \,. \tag{11.63}$$

Other useful relations, which can be derived by substituting and comparing the line elements (11.49) and (11.57), are

$$\partial_t U = -(c+v)\partial_x U \,, \quad \partial_t W = (c-v)\partial_x W \,, \quad \mathcal{C}(t,x) = -\frac{1}{\partial_x U \partial_x W} \,. \tag{11.64}$$

[4] Note that $x_f = \int^{x_f} \frac{\mathrm{d}x'}{c(x')+v(x')}$

At this point we can change coordinates in the RSET (using eq. (11.62) and (11.64))

$$
\begin{aligned}
T_{tt} &= (\partial_t U)^2 T_{UU} + 2(\partial_t U)(\partial_t W) T_{UW} + (\partial_t W)^2 T_{WW} \\
&= \dot{p}^2 T_{UU} - 2\dot{p} T_{UW} + T_{WW} \ , \qquad (11.65)\\
T_{xx} &= (\partial_x U)^2 T_{UU} + 2(\partial_x U)(\partial_x W) T_{UW} + (\partial_x W)^2 T_{WW} \\
&= \frac{\dot{p}^2}{(c+v)^2} T_{UU} - \frac{2\dot{p}}{c^2 - v^2} T_{UW} + \frac{1}{(c-v)^2} T_{WW} \ , \qquad (11.66)\\
T_{tx} &= (\partial_t U)(\partial_x U) T_{UU} + (\partial_t U \partial_x W + \partial_x U \partial_t W) T_{UW} + (\partial_t W)(\partial_x W) T_{WW} \\
&= -\frac{\dot{p}^2}{c+v} T_{UU} + \frac{2\dot{p} v}{c^2 - v^2} T_{UW} + \frac{1}{c-v} T_{WW} \ . \qquad (11.67)
\end{aligned}
$$

The problematic terms for the backreaction are those involving $c+v$, that goes to zero at $\mathcal{H}^+$. We can see that T_{tt} is well behaved at the horizon while T_{tx} is always less divergent than T_{xx} there. Also no divergences are coming from the parts proportional to T_{WW}.

So, let us focus on the region near the horizon, and also assume $c(x) = 1$ so that v encodes all the x dependence of the problem (this choice is equivalent in a specific choice for the collapse). The horizon can be taken at $x = 0$, and therefore at $\mathcal{H}^+$ we have $v = -1$. Let us expand v around this point

$$
v(x) \simeq -1 + \kappa x + \gamma x^2 + \dots \ . \qquad (11.68)
$$

The coefficient $\kappa = \frac{\partial v}{\partial x}|_{x=0}$ represent the *surface gravity* of the black hole (this can be checked for instance in Schwarzschild).

At this point we can focus on T_{UU}:

$$
T_{UU} \propto \frac{1}{\dot{p}} \left[\bar{\mathcal{C}}^{1/2} \partial_u^2 \bar{\mathcal{C}}^{-1/2} - \dot{p}^{1/2} \partial_u \dot{p}^{1/2} \right] \ . \qquad (11.69)
$$

By expanding around the horizon we get

$$
\bar{\mathcal{C}} = \dot{p}\,\mathcal{C} = -\frac{\dot{p}}{\partial_x U \partial_x W} = -\frac{1}{\partial_x u \partial_x W} = 1 - v^2(x) \approx 2\kappa x \, . \qquad (11.70)
$$

So that

$$
\bar{\mathcal{C}}^{1/2} \partial_u^2 \bar{\mathcal{C}}^{-1/2} \simeq (2\kappa x)^{1/2} \kappa x \partial_x \left[\kappa x \partial_x (2\kappa x)^{-1/2} \right] = \frac{\kappa^2}{4} + \mathcal{O}(x^2) \ . \qquad (11.71)
$$

Note that if this was the full result, we would have T_{tx}, T_{xx} diverging as $1/x$ and $1/x^2$ for $x \to 0$ (implying a large backreaction).

Let us however consider now also the contribution of the dynamical term. At late times ($u \to \infty$), we already know that

$$
p(u) \simeq U_{\mathcal{H}} - A_1 e^{-\kappa u} \qquad (11.72)
$$

but what about earlier times? One can show that in this case one generically gets extra sub-leading corrections to the previous formula [94].

$$
p(u) \simeq U_{\mathcal{H}} - A_1 e^{-\kappa u} + \frac{A_2}{2} e^{-2\kappa u} - \frac{A_3}{3!} e^{-3\kappa u} + \dots \ . \qquad (11.73)
$$

Therefore we can see that

$$
-\dot{p}^{1/2} \partial_u^2 \dot{p}^{-1/2} = -\frac{1}{2} \left[\frac{\dddot{p}}{\dot{p}} - \frac{3}{2} \left(\frac{\ddot{p}}{\dot{p}} \right)^2 \right] = -\frac{\kappa^2}{4} + \left[-\frac{A_3}{2A_1} + \frac{3}{4} \left(\frac{A_2}{A_1} \right)^2 \right] \kappa^2 e^{-2\kappa u} + \mathcal{O}(e^{-3\kappa u}) \ . \qquad (11.74)
$$

After this calculation we can make some comments.

- not surprisingly the flux term in this 1+1 problem is ruled by the Schwarzian derivative.

- In the end, by summing the two contributions (11.71) and (11.74) in (11.69) we obtain a cancellation of the leading terms (those in $\kappa^2/4$) so that the components T_{tx}, T_{xx} are indeed regular. This is the "Unruh state miracle" and it is what makes this the only regular state compatible with the spacetime associated to a gravitational collapse leading to the formation of a future event horizon $\mathcal{H}^+$. The vacuum polarization static term divergence at $\mathcal{H}^+$ (the same of the Boulware state) is exactly balanced by an equal and opposite divergence coming from the Hawking flux term.

- For what regards the subleading contributions, it is easy to see that they do not introduce other divergences. Indeed note that for small x (so that $1 + v \sim \kappa x$) we have

$$u = t - \int \frac{\mathrm{d}x}{c+v} \simeq t - \frac{1}{\kappa} \log x + const\ , \tag{11.75}$$

  so that $e^{-\kappa u} \propto x e^{-\kappa t}$ and so $e^{-2\kappa u} \sim (x^2)$. Therefore the subleading terms in (11.74) are able to regulate the singularities $\sim 1/x^2$.

- The subleading terms that survive are exponentially small in time. For example, we get $T_{xx} \sim O(1)e^{-2\kappa t}$. So after the formation of the horizon these terms rapidly decay and no information about the collapse remains encoded (via a residual REST) on the horizon. This is an another manifestation of the *no-hair theorem* for black holes as any collapse related feature is rapidly shed away.

- Moreover, if at some t we write $x = (r(t) - r_h) \simeq -\lambda(t - t_{\mathcal{H}}) + \dots$ (where $t_{\mathcal{H}}$ is the time of formation of the horizon) then one can show that

$$\frac{A_2}{A_1} \propto \lambda^{-1}\ , \quad \frac{A_3}{A_1} \propto \lambda^{-2}\ . \tag{11.76}$$

  The constant λ represent the velocity of matter during the collapse. For a freely falling collapse this is very close to the speed of light. However, if some unknown new physics allows the collapse proceeds slowly, then one could start witnessing large quantum contributions. This suggests the guess made at the beginning is right: collapsing stars with no relevant support at horizon-formation experience negligible quantum back-reaction (since as we have just seen quantum effects are negligible if not allowed to grow in first place by something else slowing down the collapse). This could have been foreseeable. If we start with Minkowski vacuum on $\mathscr{I}^-$ and we have to remain in a Hadamard state during a Cauchy evolution then we can keep renormalising the SET. As long as the star collapse in near free falling, we can always go in the corresponding local inertial frame where the renormalisation of the SET should not show large deviation from the Minkowski vacuum. Once a horizon is formed the RSET at the horizon should rapidly become zero as needed for the onset the Unruh vacuum being the latter the only regular vacuum compatible with this geometry.

- Our deductions do not change if instead of an event horizon one consider simply a trapping horizon existing for a finite time. Simply in this case they would be relevant for observations on a specific region of spacetime (e.g. the finite portion of $\mathscr{I}^+$ associated to light rays which arrive there after having experienced an exponential redshift.

So in the end it seems very unlikely then that quantum physics can prevent the formation of an horizon in standard gravitational collapse. Nonetheless, this does not guarantee that the regularisation of spacetime singularities imposed by quantum gravity could not dramatically change the black hole structure and fate at later time (see e.g. the scenarios listed in [61]). For example, one could explore bouncing scenarios where a series of multiple bounces could lead conditions where the above mentioned corrections might not be negligible.

Anyway, modulo the above mentioned scenarios resorting to transient quantum gravity regimes, we can safely conclude that the above analysis corroborates the conclusion that gravitational collapse will generically lead to the formation of a trapping horizon even when taking into account the semiclassical back-reaction, and that the onset of a thermal at the Hawking temperature on $\mathscr{I}^+$ at late times will be unavoidable, just by regularity of the quantum state, once such a horizon formes.

12 Time Machines and Chronology Protection

Sometimes, this is seen as something related to science-fiction, but in reality this is a serious problem of General Relativity, since the theory allows time travels. The fact that this is a problem in a certain sense is good, because represents a point to investigate in order to extend the theory, as time travel solutions can be seen of further signs that General Relativity is an incomplete theory.

12.1 Time travel paradoxes

Another good reason to explore time machine solutions in General relativity is the fact that they are intrinsically paradoxical (at least without entailing radical departures from our vision of how physics works). Note that by paradoxical we mean here true logical inconsistencies rather that the kind of apparent paradoxes that sometime we encounter in discussing the implications of physical theories.

To be clear and example of the latter type is the so called **twin paradox** in Special Relativity, where one of two twins is sent in an interstellar travel at relativistic speed while the other stay on Earth. According to the one that stayed on Earth, the twin who left, once back to Earth, should be younger. However, because of the relativity of reference frames, according to the travelling twin, the reverse should apply as well. However, this is not a real paradox, since there is a way to solve the problem: in order to return, the travelling twin has to decelerate and accelerate back, thus breaking the symmetry of the problem, and therefore the travelling twin ends up being the only one remaining younger.

The real paradoxes associated to time machines are in situations involving travelling backward in time; travelling forward in time is not a problem: this is the situation in the aforementioned twin paradox or in science-fiction movies, like *Interstellar* where one hour for people sitting on a close orbit to a black hole could corresponds to several years for someone farther away. More precisely, there are two types of real paradoxes associated to backward time-travel: the grandfather paradox and the bootstrap paradox.

In the **grandfather paradox**, a man travels back in time and kills his own grandfather before he meets his grandmother. In this way, the time traveller can never be born, and in this way he will not be able to travel back in time. Therefore the grandfather lives, the time traveller is born and he has the chance to get back in time, and so on. Thus, it is not clear from which present the time traveller can come from, and this causes a paradox.

In the **bootstrap paradox**, for example, you can go back in time to give your past self the winning numbers of a lottery. At this point, you have generated information from nowhere. This type of paradoxes can then be also transformed in a sort of grandfather paradox: imagine for example that, once you won the lottery, later in your future you conclude that money do not buy happiness and you decide to not travel back in time to give yourself the winning numbers...

Actually, there are numerous examples one can formulate for the bootstrap paradox; the science-fiction British TV Show *Doctor Who* provides one such example in one episode:

So, there's this man, he has a time machine. Up and down history he goes — zip, zip, zip, zip, zip — getting into scrapes. Another thing he has is a passion for the works of Ludwig van Beethoven. Then, one day, he thinks, "What's the point in having a time machine if you don't get to meet your heroes?" So, off he goes to 18th century Germany, but he can't find Beethoven anywhere. No one's heard of him. Not even his family have any idea who the time traveller is talking about. Beethoven literally doesn't exist. [...] The time traveller panics. He can't bear the thought of a world without the music of Beethoven. Luckily, he'd brought all of his Beethoven sheet music for Ludwig to sign. So, he copies out all the concertos and the symphonies, and he gets them published. He becomes Beethoven. And history continues with barely a feather ruffled. My question is this: who put those notes and phrases together? Who really composed Beethoven's Fifth?

(Doctor Who, *Before the Flood*, written by Toby Whithouse, 2015)

12.2 Time Machines

After this preliminary discussion about the paradoxes of time travel let us now try to formulate more precisely what we meant by a time machine in physics.

A time machine is a region of spacetime which allows to travel back in time. This region can be an integral part of a given solution of the Einstein field equations or it can be dynamically generated in an otherwise perfectly sane spacetime by using some sort of device.

For a technical definition of what we mean by travel back in time we have to consider world-lines associated to light or material observers. From the concepts of causal curve and chronological curves, we can in fact introduce the concepts of **closed causal curve** (CCC) and **closed timelike curve** (CTC). So we have time machines in a spacetime if in at least so some region of it one finds CCC or CTC. Closed causal curves produce causality violations, while closed timelike curves produce chronology violations.

Notice that chronology violation imply causality violation, but the converse is not true. Indeed, if one has only CCC the above mentioned paradoxes cannot be encountered: nonetheless it is possible to show, and rather intuitive, that such spacetimes if perturbed can easily lead to CTC. This is why time machines are associated to the existence of a chronology violation region (i.e. a region endowed with CTC).

In order to define a latter we can start by noticing that:
A point p is in a ***chronology violating region*** *if the intersection of the chronological past and the chronological future of p is not zero*:

$$I^0(p) \equiv I^+(p) \cap I^-(p) \neq \emptyset; \tag{12.1}$$

the **total chronology violating region** in a manifold $\mathcal{M}$ is then defined as

$$I^0(\mathcal{M}) = \bigcup_{p\in\mathcal{M}} I^0(p); \tag{12.2}$$

Obviously, the same concepts can be defined for causality violation, where it is sufficient to replace I with J in the definitions above.

Then we can define a **chronological horizon** as

$$\mathcal{H}^+(I) = \partial\left[I^+(I^0(\mathcal{M}))\right]; \tag{12.3}$$

chronological horizons are, as we have already said, a type of Cauchy horizons; they mark the onset of a region in which we can have chronology violation, therefore they mark the boundary between a "sane" region of the manifold and a region where closed timelike curves are allowed.

But is time travel really allowed in General Relativity? Is there any spacetime which is solution of the Einstein field equations, in which closed timelike curves are present? The answer is yes.

There are three categories of solutions of the field equations that allow time travels: globally rotating solutions, wormholes solutions, and faster-than-light travel solutions (warp drives and Krasnikov tubes).

12.2.1 Time machines from globally rotating solutions

In this category are included the van Stockum spacetimes, the Gödel Universe and Kerr Black Holes. The basic mechanism generating time machine regions in spinning spacetimes can be summarised with the fact that for a sufficiently high velocity, there are generically regions where the tilting of the lightcones can lead to the formation of closed timelike curves.

The van Stockum spacetime

In this solution of General Relativity, we have an infinite cylinder of dust of density $\rho(r)$ that satisfies the condition

$$\frac{\mathrm{d}\rho}{\mathrm{d}r} > 0; \tag{12.4}$$

the solution is given by the following metric:

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 - 2ar^2\,\mathrm{d}t\,\mathrm{d}\varphi + (1 - a^2r^2)r^2\,\mathrm{d}\varphi^2 + \frac{\mathrm{d}z^2 + \mathrm{d}r^2}{\exp(a^2r^2/2)}, \tag{12.5}$$

where $a > 0$ is a constant such that a^2 is proportional to the mass density of the dust at $r = 0$. The coordinates range in the intervals

$$t \in (-\infty, +\infty), \quad \varphi \in [-\pi, \pi], \quad r \in (0, +\infty), \quad z \in (-\infty, +\infty). \tag{12.6}$$

We can see that there are closed timelike curves if we consider an azimuthal curve at fixed (r, z, t),

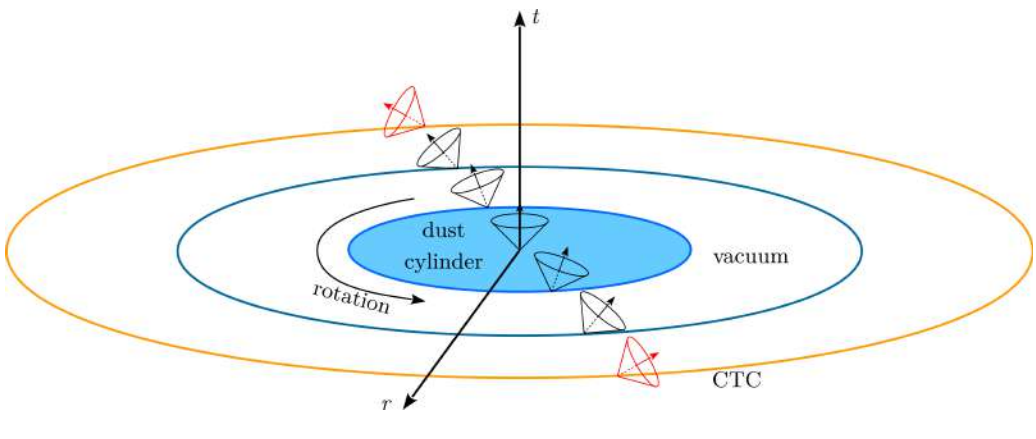


Figure 12.1: Graphical representation of the tilting of lightcones in a van Stockum spacetime. At a particular value of the radius, the tilting is such that closed null curves are possible and, beyond that radius, also closed timelike curves are allowed.

which is an integral curve of the angle ϕ and hence is always closed. If we compute the invariant length

$$s_\phi = \int \mathrm{d}s \Longrightarrow s_\phi^2 = (1 - a^2r^2)r^2(2\pi)^2, \tag{12.7}$$

we can see that s^2_ϕ becomes negative (timelike) for $r > \frac{1}{a}$. I.e. for sufficiently high values of the radius, we have closed timelike curves.

Finally, let us note that also Tipler cylinders are included in this family (with a massive cylinder replacing the cylindrical distribution of dust) as well as Gott's spacetimes (these are spacetimes with two infinitely long, parallel, cosmic strings longitudinally spinning around each other).

The Gödel Universe

The Gödel universe [96], is very similar to Einstein static Universe, in the sense that it contains a homogeneous distribution of dust and a cosmological constant, but the dust is swirling so it describes a universe endowed with a global rotation. Therefore, this solution is not isotropic but only homogeneous, in the sense that the axis of rotation is the same at any point and appears the same at any location.

The stress-energy tensor of this universe is

$$T^{\mu\nu}_{\rm tot} = \rho v^\mu v^\nu - \frac{\Lambda}{8\pi G} g^{\mu\nu} = (\bar\rho + \bar p) v^\mu v^\nu + \bar p g^{\mu\nu}, \tag{12.8}$$

with

$$\bar p = -\frac{\Lambda}{8\pi G} \qquad \bar\rho = \rho + \frac{\Lambda}{4\pi G} \quad \text{and} \quad \rho = \frac{\Omega^2}{4\pi G_N} \quad \text{with} \quad \Omega^2 = -\Lambda; \tag{12.9}$$

we can pick the velocity of the fluid in a way that it is always timelike, $v^\mu = (1, 0, 0, 0)$, and in addition we can define a vorticity $\omega^\mu = (0, 0, 0, \Omega)$, where Ω is the angular velocity of the universe [45, page 271]. Then, the solution is given by

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 - 2\mathrm{e}^{\sqrt{2}\Omega y}\,\mathrm{d}t\,\mathrm{d}x - \frac{1}{2}\mathrm{e}^{2\sqrt{2}\Omega y}\,\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{12.10}$$

Gödel universe has many interesting features that would deserve a much more in depth discussion (see e.g. [97]), but for us its main characteristic is that it also entails light-cones tilting and CTC. In order to see this, it is more convenient to go to cylindrical coordinates, where the metric becomes [45]

$$\mathrm{d}s^2 = -\,\mathrm{d}\bar t^2 + 4\Omega^{-1}\sinh^2(\sqrt{2}\Omega r)\,\mathrm{d}\varphi\,\mathrm{d}\bar t + 2\Omega^{-2}\sinh^2(\sqrt{2}\Omega r)\left[1 - \sinh^2(\sqrt{2}\Omega r)\right]\,\mathrm{d}\varphi^2 + \mathrm{d}r^2 + \mathrm{d}z^2; \tag{12.11}$$

as we did before, we can compute the invariant length along φ and see where this can become negative,

$$s^2_\varphi \propto \left[1 - \sinh^2(\sqrt{2}\Omega r)\right] < 0\,, \qquad \text{if} \quad \sinh(\sqrt{2}\Omega r) > 1 \quad \text{i.e.} \quad \Omega r > \frac{\ln(1+\sqrt{2})}{\sqrt{2}}\,, \tag{12.12}$$

So, around each point of Gödel spacetime (being this universe homogeneous) one can find sufficiently far away orbits there are closed timelike curves. See Figure 12.2.

This solution is considered unphysical, since the rotation of the whole Universe seems physically unmotivated, and also because a fine tuning of the cosmological constant is needed. Indeed, the van Stockum and Gödel solutions are sometime dismissively called "GIGO solutions", which stands for "Garbage In - Garbage Out", in the sense that with a very strange/unphysical input in the Einstein equations, it is obvious to have a very strange/unphysical output. Finally, it is also worth noticing that this universe also a non-Machian aspect of GR: it has the same content as Einstein static Universe, still it has a completely different behaviour, since there is a rotation not "visible" in the distribution of the matter content.

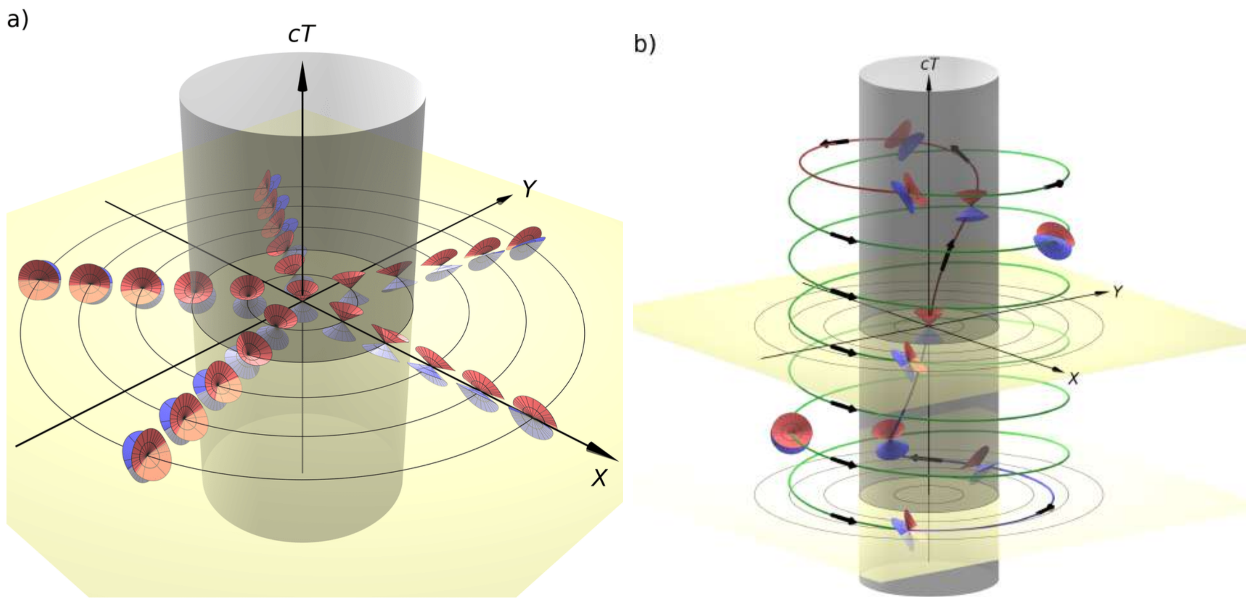


Figure 12.2: Left: graphical representation of the tilting of the lightcones in a Gödel Universe. The grey cylinder represents the Gödel horizon. Right: example of a Closed Timelike Curve in a Gödel spacetime. From [98].

Kerr Black Holes

We have already seen Kerr black holes, described by a metric given by

$$\begin{aligned} \mathrm{d}s^2 = -\left(\frac{\Delta - a^2\sin^2\theta}{\Sigma}\right)\mathrm{d}t^2 - 2a\sin^2\theta\left(\frac{r^2+a^2-\Delta}{\Sigma}\right)\mathrm{d}t\,\mathrm{d}\varphi + \\ + \left[\frac{(r^2+a^2)^2 - \Delta a^2\sin^2\theta}{\Sigma}\right]\sin^2\theta\,\mathrm{d}\varphi^2 + \frac{\Sigma}{\Delta}\mathrm{d}r^2 + \Sigma\,\mathrm{d}\theta^2, \\ \Delta = r^2 - 2Mr + a^2, \qquad \Sigma = r^2 + a^2\cos^2\theta \end{aligned}$$

We have seen that in $r = 0$ there is a ring singularity; however, this singularity is only at $\theta = \frac{\pi}{2}$. Therefore, approaching $r = 0$ at an angle different than $\pi/2$ will not result in encountering the singularity. It can be shown that traversing the plane on which there is the singularity through the ring is analogous to extend the Kerr metric to negative values of r (e.g. [58]).

Now, consider the Killing vector of Kerr related to the conservation of angular momentum along φ, $\psi = \frac{\partial}{\partial\varphi}$. Its norm is given by

$$|\psi|^2 = g_{\varphi\varphi} = \frac{(r^2+a^2)^2 - \Delta a^2\sin^2\theta}{\Sigma}\sin^2\theta\,, \tag{12.13}$$

Now it is easy to see that such norm becomes negative, whenever

$$\begin{aligned} (r^2+a^2)^2 - \Delta a^2\sin^2\theta < 0 \\ \Rightarrow r^4 + (2-\sin^2\theta)a^2r^2 + 2Ma^2r\sin^2\theta + a^4\cos^2\theta < 0. \end{aligned} \tag{12.14}$$

This is realised for small negative values of r – which means close to the ring singularity after one has travelled through the ring – and for θ close to $\frac{\pi}{2}$ (not equal to $\frac{\pi}{2}$ in order to avoid the ring singularity); since ψ describes closed curves, when the previous condition is realized we are in presence of closed timelike curves and hence of chronology violation conditions. Also in this case, people is often unconcerned about this unpleasant feature because it happens well within the Cauchy horizon, which, as said, is conjectured to be unstable anyway.

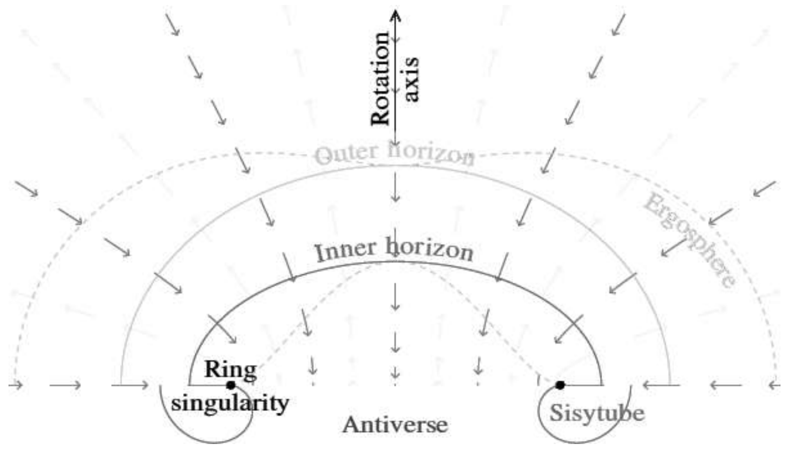


Figure 12.3: Schematic representation of the basic structure of a Kerr black hole. The region denominated “Sisytube” is the toroidal region close to the singularity where closed timelike curves are allowed. (Credits: Andrew Hamilton, http://jila.colorado.edu/~ajsh/insidebh/index.html)

12.2.2 Wormholes

Wormholes are shortcuts through spacetime, connecting or points of teh same universe (intra-universe wormhole) or points of different, disconnected universes (inter-universe wormholes). From a topological point of view, an intra-universe wormhole (a wormhole between two points in the same universe) is a compact region of spacetime whose boundary is topologically trivial, but whose interior is not simply connected. Formalising this idea leads to definitions such as the following [45]: *If a Minkowski spacetime contains a compact region W, and if the topology of W is of the form $W \sim R \times \Sigma^3$ where Σ^3 is a three-manifold of nontrivial topology, whose boundary has topology of the form $\partial\Sigma \sim S^2$ and if, furthermore, the hypersurfaces Σ^3 are all spacelike, then the region W contains a quasi-permanent intra-universe wormhole.*

Wormholes can come in different shape and sizes. Let us hence do here a little bit of taxonomy.

Wheeler Wormholes

These are structures at the Planck scale, where you imagine to have a foam, where geometry is not uniform and can have topologically complex structure. Since we do not really know what happens at the Planck scale, it is generally believed that they are not a problem as far as chronology violations are concerned.

Schwarzschild wormholes — Einstein–Rosen Bridges

Schwarzschild wormholes, also known as Einstein–Rosen bridges, are connections between the two asymptotically flat regions the maximally extended version of the Schwarzschild metric (describing an eternal black hole with no charge and no rotation). In this spacetime, it is possible to come up with coordinate systems such that if a hypersurface of constant time is chosen (e.g. $T = 0$ and $\theta = \pi/2$) then the “embedding diagram”, drawn depicting the curvature of space at that time deducible from the Schwarzschild line element, will look like a tube connecting the two exterior eulclidean regions, known as an “Einstein–Rosen bridge”. See Fig. 12.4. These form of wormholes is not traversable (as it is referring to a space-like surface) and even worse not even stable: in 1962 it was shown by Wheeler and Fuller [99] that if the Einstein–Rosen bridge connects two parts of the same universe, then it will pinch off too quickly for light (or any particle moving slower than light) that falls in from one exterior region to make it to the other exterior region.

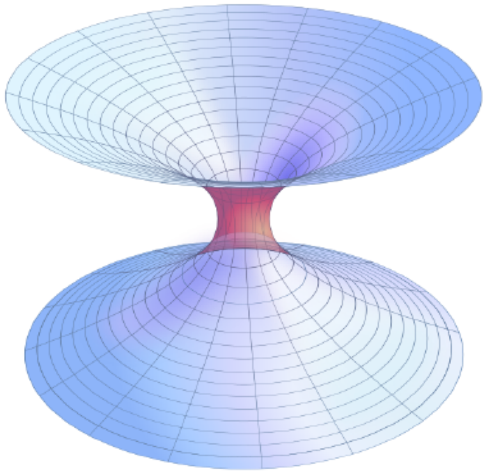


Figure 12.4: Embedding of a Schwarzschild wormhole (From Wikipedia).

Morris-Thorne Traversable Wormholes

In an attempt to improve over the the Einstein–Rosen bridge and construct a traversable wormhole Morris and Thorne proposed in 1998 a new set of solution of Einstein equations [100, 101]. [1] In principle, these traversable wormholes are not forbidden or unstable, however they need to violate energy conditions and would be very difficult to produce. Let us add that the above comment does not apply only to the Morris–Thorne solutions: these are just a special case of traversable wormholes for which the same point would apply. [2].

The Morris–Thorne solution in General Relativity is given by a time-independent, non-rotating, spherically symmetric metric:

$$\mathrm{d}s^2 = -\mathrm{e}^{-2\Phi(\ell)}\,\mathrm{d}t^2 + \mathrm{d}\ell^2 + r^2(\ell)\left[\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right]. \tag{12.15}$$

This metric needs some requirements in order to describe a wormhole:

1. the range for the coordinate ℓ should be $\ell \in (-\infty, +\infty)$;

2. the structure of $r(\ell)$ must be such that there are no event horizons, since we want a traversable solution;

3. this must be an asymptotically flat and regular solution at infinity, which implies that

$$\lim_{\ell\to\pm\infty} \frac{r(\ell)}{|\ell|} = 1, \tag{12.16}$$

   and therefore $r(\ell) \sim |\ell| + \mathcal{O}(1)$ at large ℓ. Moreover, since we want to recover Minkowski at infinity, we also want $\lim_{\ell\to\pm\infty}\Phi(\ell) = \Phi_\pm = $ const which can be then set to 1 by a proper redefinition of the time coordinate.

4. The metric component have to be at least C^2 in ℓ

Solving the whole solution, one finds the radius of the wormhole to be

$$r_{\text{wormhole}} = r_0 \equiv \min\{r(\ell)\}; \tag{12.17}$$

[1]Noticeably, these are the structures appearing in *Interstellar* or in *Contact* by Carl Sagan, and apparently Kip Thorne was stimulated in his research by a question of Carl Sagan which was at the time writing the novel *Contact*.

[2]For example a wormhole solution can be built by cutting the Schwarzschild solution at some r very close to r_H and gluing it with an identical metric on the other side: inserting the so obtained metric the Einstein equations would still lead to a SET associated to large violations of the energy conditions)

basically, ℓ is a coordinate going through the wormhole, being $-\infty$ at one side and $+\infty$ at the other side; anyway, ℓ can be set in a way that $r_0 = r(\ell = 0)$, and therefore $\ell = 0$ at the center of the wormhole, at the "throat".

Given the above requirements the spacetime described by this solution is then a sort of shortcut between two asymptotically flat regions of the same universe or of disconnected universes. See Fig. 12.5.

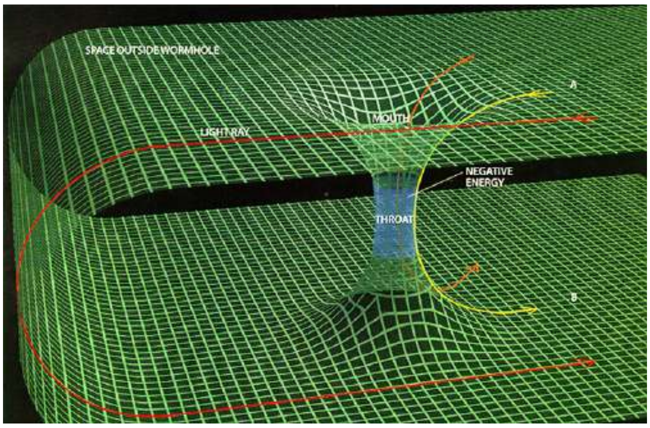


Figure 12.5: A traversable Lorentzian wormhole. From Web.

In order to see that this metric is associated, as anticipated, to large violation of the energy conditions, all one needs to do is to inserte the solution into the Einstein equations and find the stress-energy tensor. The latter takes the form

$$T_{\mu\nu} = \mathrm{diag}\left(\rho(r), \tau(r), p_\theta(r), p_\varphi(r)\right), \tag{12.18}$$

where ρ is the density, $\tau(r) = -p_r(r)$ is the tension, p_θ is the pressure along the θ component and p_φ the pressure along the φ component. One can show that there is always a radius r_* such that, for any r between r_0 and r_*, $\rho - \tau < 0$, which means a violation of the NEC, which in turn implies a violation of WEC, SEC and DEC.

Notice that all these aforementioned energy condition violations do not forbid *a priori* the existence of such a solution. As a matter of fact, we already know that our universe violates at least the SEC and not only that: quantum phenomenal like the Hawking radiation or even the Casimir effect can violate most if not all of energy conditions, therefore it might be that quantum effects could be engineered by an advanced civilisations so to realise this type of structures. Nonetheless, let us stress that this will be very difficult as generically large amount of exotic matter will be required.

Indeed let us estimate the value of the energy density and of the tension at the throat. These are [45]

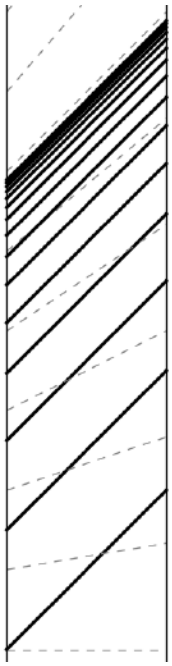
$$\rho(r_0) \leq \frac{1}{8\pi G r_0^2} \qquad \tau(r_0) = \frac{1}{8\pi G r_0^2} = 5 \times 10^{36} \left(\frac{10\ \mathrm{m}}{r_0}\right)^2 \ \mathrm{N\ cm}^2, \tag{12.19}$$

which implies that, for example, to have a traversable wormhole with $r_0 \sim 1$ m, one has to have

$$M_{\mathrm{throat}} \simeq -\frac{r_0 c^2}{G} \sim M_{♃}, \tag{12.20}$$

which means that we need at the wormhole throat something of the order of a Jupiter's mass of exotic matter, in order to keep open a wormhole about one meter wide.

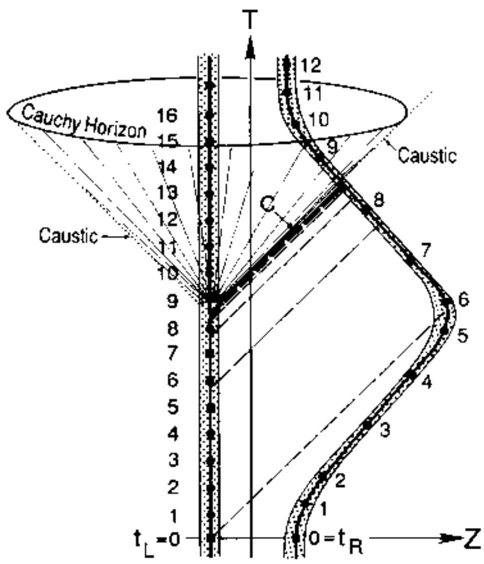
But how could a wormhole become a time machine? For example, we could set the two ends of the wormhole in a way that, by entering at a time t, you exit in the same place at a time $t - \Delta t$, and in this way you would have basically travelled back in time.



In order to achieve a situation like this, one possibility, consist in placing one of the ends of the wormhole close to a black hole (or some other very compact object) and the other basically in Minkowski space, in order to have a time dilation between the two clocks (shown in the picture on the left by the dashed grey lines). If we now send a light ray (solid black lines) through the wormhole, every time it passes through the wormhole it will feel a greater effect of time dilation, as illustrated here.

After some time, the light ray will be on the left end (the exit) of the wormhole at the same time as it will be on the right end (the entering) of the wormhole. From that point on, it will be possible to exit the wormhole *before* entering it on the other side, and we would then have a chronology violation.

Figure 12.6: Wormhole time machine constructed by delaying time at one mouth.



An alternative, could be to Lorentz move away at relativistic speeds one of the two ends for some time and the boost it back close to the initial mouth. The tow mouths in this case are like the twin in the famous twin "paradox" of special relativity. The non-inertial path of the boosted mouth motion generates a time shift between the two mouths (the right mouth to "age" less than the left as seen from the exterior) and then getting them closer and closer can generate first a CCC (the chronological horizon) and then CTCs.

Figure 12.7: Spacetime diagram for the conversion of a spherical, traversable wormhole into a time machine by one mouth boost. Figure from [100, 101]

To be more precise let us see how a CTC can come into effects with these sort of expedients. Consider the worldlines of the two wormhole mouths, l_1 and l_2 at distance l; suppose that $l_1^\mu(t) = (t,0,0,0)$ and $l_2^\mu(t) = (t+T,0,0,l)$. The invariant length of the geodesic connecting the mouths at the same t can be calculated as before:

$$s^2(t) = \|l_2^\mu(t) - l_1^\mu(t)\|^2 = \|(T,0,0,l)\|^2 = -T^2 + l^2, \qquad (12.21)$$

and it can be seen that if $l^2 \leq T^2$, then $s^2(t) \leq 0$ and therefore we would have a closed timelike curve.

When $s^2(t) > 0$ it is alway possible to find, via some Lorentz boost, a frame where the two wormhole mouths are at equal times. The required rapidity of the boost would be $\beta = T/l$ However, as $\ell^2 \to T^2$ one can easily see that such rapidity would approach one, and for $l < T$, one would need to go faster than the speed of light in order to undo the time machine.

12.2.3 Alcubierre Warp Drives

These Alcubierre warp drive [102] describes a metric allowing for faster-than-light (FTL) travel. Note that classically the speed of the bubble can exceed the speed of light albeit locally this is still the limit speed for the propagation of signals (we saw an example of this in cosmology when discussion the meaning of the Hubble "horizon").

In particular, it is characterised by are bubble-like structure where there is an expansion of the spacetime in the back of the bubble and a contraction at its front. The inside of the bubble is basically flat Minkowski spacetime, and the bubble can travel at an arbitrary speed. [3] Also in this case large violations of the energy condition are required. Finally, the Alcubierre warp drive is just a special case of a family of solutions with the same characteristics (i.e. there are several other types of warp drives).

The structure of such a pacetime is schematically shown Figure 12.8: one has a "bubble" of flat spacetime, such that behind it there is an expansion of spacetime and in front of it a contraction, that makes the bubble to move forward. In order for this to be stable, one needs areas with large violation of energy conditions localised on the sides at the transition between expansion and contraction.

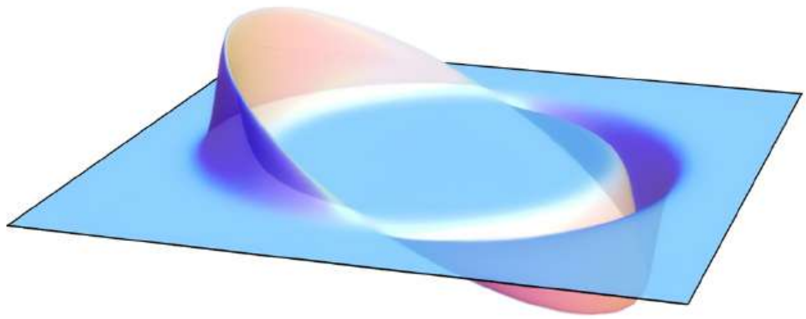


Figure 12.8: Spacetime structure in Alcubierre warp drive (from Wikipedia user AllenMcC.)).

The metric describing this solution is

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + (\,\mathrm{d}x - v_c f(r)\,\mathrm{d}t)^2 + \,\mathrm{d}y^2 + \,\mathrm{d}z^2, \tag{12.22}$$

where v_c is the speed of the center of the bubble, and $f(r)$ is a function describing the shape of the bubble. Here, r is the distance from the center of the bubble,

$$r = \sqrt{[x - x_c(t)]^2 + y^2 + z^2}; \tag{12.23}$$

the function $f(r)$ is usually taken to be $f(0) = 1$ and $f(\infty) = 0$. We can also see that the worldline $\{x_c(t), 0, 0, 0\}$ is a geodesic, therefore at the centre of the bubble one does not feel any acceleration.

Now, while having a superluminal warp drive does not violate any physical law *per se* (as long as one admits EC violating matter) it is easy to say that it can be easily turned into a time machine.

Consider Figure 12.9. If you can travel faster than light in *any* reference frame – for example, along the red spacelike trajectory from the origin, and than following the blue trajectory – you can always pick a Lorentz frame where the arrival point belongs to the past of the starting point. In this

[3]Also in this case there is a connection with science fiuction as apparently the investigation on teh possibility to built such a structure was inspired by the famous warp drives used in the TV series *Star Trek*.

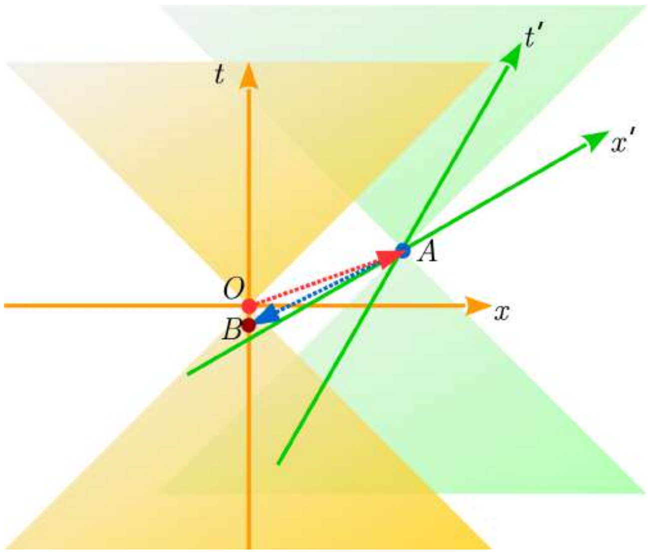


Figure 12.9: Light-cone representation of a situation where superluminal travel could generate a chronology violation. The green reference frame travels at subluminal speed with respect to the orange one. The orange frame sends a superluminal signal in its future, from O to A. The green frame sends a superluminal signal in its future from A to B. This causes B to arrive in the orange frame before OA is sent.

way, you can create a closed timelike curve. However this is not true if we have a particle that in *a given* reference frame is superluminal: in this way we would have a preferred system of reference and, in this case, Lorentz invariance would not hold anymore. If Lorentz invariance is removed, this paradox is removed from the onset. So keep it mind it is superluminal travel *plus* Lorentz invariance that make a time machine possible.

Once you have built a warp drive, this cannot communicate with the outside world, if it is travelling faster than light. Then, in order to travel to, say, Alpha Centauri, one should build first a kind of railway, similar to a magnetic monorail, with synchronised devices to create EC violations at the right time in the right place (e.g. by generating localized, large, vacuum polarization). Then you can come back using this "railway" at an arbitrary speed. There have been alternative ideas that allow the spaceship itself to generate its own faster than light "railway". Most noticeably, there is the so called **Krasnikov tube** [103], where the light cones are opened and tilted during the trip so to allow FTL in one direction. Also in this case the tube has to be created with a first trip on the destination but even worse the amount of required NEC/DEC violation is huge as it increases with the length of the tube that needs to be kept permanently opened.

12.3 Possible solutions

Now, how can we solve this problems?

As far as warp drives are concerned, it seems that quantum effects in curved spacetime forbid this possibility. What can we say about wormholes, for example? Are there possible way out to avoid time machines? It is thought that the solution could be one of these:

1. **Splitting Timelines**: once a paradox happens, the timeline bifurcates; in this case, one has to do physics on a non-Hausdorff manifold, that have a disconnected structure with branches. More precisely, *a topology is said to be Hausdorff if and only if for any two different points* x_1 *and* x_2*, there exist opens sets* O_1 *and* O_2 *such that* $x_1 \in O_1$*,* $x_2 \in O_2$*, and* $O_1 \cap O_2 = \emptyset$. For example, a non-Hausdorff set is what one obtains by removing the zero from the real numbers axis and replacing it with two values 0_1 and 0_2. In this case, consider a small interval $(-\epsilon, +\epsilon)$

with $\epsilon > 0$. Then the two sets $(-\epsilon, 0) \cup \{0_1\} \cup (0, \epsilon)$ and $(-\epsilon, 0) \cup \{0_2\} \cup (0, \epsilon)$ are both open sets containing respectively just 0_1 and 0_2 but their intersection is not empty.

2. **Novikov Conjecture**: in this case the idea is that time paradoxes are just apparent ones because histories are indeed consistent and periodic. This is equivalent to say if you are about to create a time travel paradox you will not not succeed, but on the contrary, your time travel and your actions in the past had to happen for history to unfold the way it has [104, 105]. This solution is very popular among philosophers of science, less so among physicists because of in some toy models (e.g. Minkowski spacetime with periodic boundary conditions in time and space plus a massless scalar field) it was shown that not always the required periodic solution do exists. Moreover, such conjecture seems also difficult to accomodate with the probabilistic nature of quantum mechanics.

3. **Multiverse Solution**, where there is an infinite number of Universes, and in any of them anything can happen, and there could be a sort of path integral that is connected to the probability of something happening or not happening with respect to other Universes. This is also related to the Many World Interpretations of Quantum Mechanics by Everett.

4. **Hawking's Chronology protection conjecture**, that states that quantum physics will forbid time machines to form; in the case of the warp drive, we have already seen that quantum effects forbid superluminal warp drives; in the case of the wormhole, we can see that we have an increasing concentration of photons near the critical point, an accumulation, and in proximity of the formation of the chronological horizon, we would see a diverging stress-energy tensor. However, there is a theorem by Kay, Radzikowski and Wald [106] according to which a quantum state fails to be Hadamard on chronological horizons. Since we need an Hadamard state to renormalize the stress-energy tensor, this implies that close to a chronological horizon the stress-energy tensor cannot be renormalized, and the final word cannot be set. The final answer should be Quantum Gravity, that should forbid the formation of time machines.

12.3.1 Preemptive chronology protection?

While we have only circumstantial evidence for the validity of the chronology protection conjecture, there is at least in the case of the Alcubierre warp drive some evidence this can be indeed true. In fact, in the case of the warp drive it can be shown that even before attempting to use it to form CTC an QFT driven instability sets in as long as the propagation is superluminal [107, 108]).

Consider the Carter–Penrose diagram of a warp drive as shown in Figure 12.10. Once the superluminal warp drive is established, there are two apparent horizons, where the back wall acts like a black hole and the front wall acts like a white hole: inside a superluminal bubble, nothing coming from the back can come in, since it is moving faster than light to every observer outside, and therefore it acts like a black horizon; however, if someone tries to send a signal outside of the bubble from the front, the bubble from the outside is receding faster than light, and therefore the front wall acts like a white horizon.

Now one can understand that an instability appears: the back wall, being a black horizon, emits Hawking radiation, that flows to the white horizon, which then becomes an accumulator of this radiation, and this gives a growing instability. Let us see this in more detail.

12.3.2 Superluminal warp drive instability

In the actual computation we shall restrict our attention to the $1 + 1$ dimensions case (since in this case one can carry out a complete analytic treatment) and follow [107]. Changing coordinates to those associated with an observer at the center of the bubble, the warp-drive metric (12.22) becomes

$$ds^2 = -c^2 dt^2 + [dr - \bar{v}(r)dt]^2\,, \qquad \bar{v} = v - v_c\,, \tag{12.24}$$

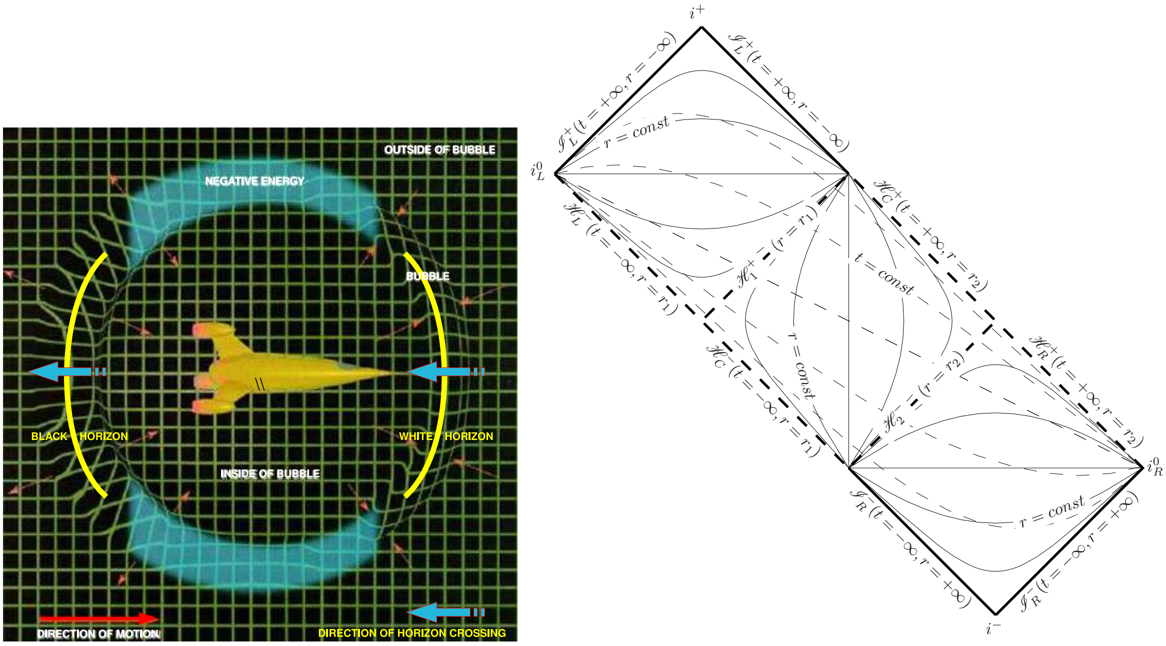


Figure 12.10: Left panel, structure of Alcubierre spacetime (from [109]), right panel corresponding Carter-Penrose diagram of an eternal Alcubierre warp drive (from [107]).

where $r \equiv x - x_c(t)$ is now the signed distance from the center. Let us consider a dynamical situation in which the warp-drive geometry interpolates between an initial Minkowski spacetime $[\hat{v}(t,r) \to 0$, for $t \to -\infty]$ and a final stationary superluminal $(v_c > c)$ bubble $[\hat{v}(t,r) \to \bar{v}(r)$, for $t \to +\infty]$. The causal structure will be the one shown in Fig. 12.11 To an observer living inside the bubble this geometry has two horizons, a *black horizon* $\mathcal{H}^+$ located at some $r = r_1$ and a *white horizon* $\mathcal{H}^-$ located at $r = r_2$. Here let us just add that from the point of view of the Cauchy development of $\mathscr{I}^-$ these spacetimes posses Cauchy horizons (see $\mathcal{H}_C^+$). Note that these are an artefact of the idealised case in which the warp drive keeps on going at superluminal speed forever. As such we shall not deem them physically relevant in what follow.

Light-ray propagation

Let us now consider light-ray propagation in the above described geometry. Only the behaviour of right-going rays determines the universal features of the RSET, just like outgoing modes do in the case of a black hole collapse. Therefore, we need essentially the relation between the past and future null coordinates U and u respectively on $\mathscr{I}^-$ and $\mathscr{I}^+$, labelling right-going light rays.

There are two special right-going rays defining, respectively, the asymptotic location of the black and white horizons. In terms of the right-going past null coordinate U let us denote these two rays by $U_{\rm BH}$ and $U_{\rm WH}$, respectively. The finite interval $U \in (U_{\rm WH}, U_{\rm BH})$ is mapped to the infinite interval $u \in (-\infty, +\infty)$ covering all the rays traveling inside the bubble. For rays which are close to the black horizon, the relation between U and u can be approximated as a series of the form [107]

$$U(u \to +\infty) \simeq U_{\rm BH} + A_1 e^{-\kappa_1 u} + \frac{A_2}{2} e^{-2\kappa_1 u} + \dots \, . \tag{12.25}$$

Here A_n are constants (with $A_1 < 0$) and $\kappa_1 > 0$ represents the surface gravity of the black horizon. This relation is the standard result for the formation of a black hole through gravitational collapse. As a consequence, the quantum state which is vacuum on $\mathscr{I}^-$ will show, for an observer inside the warp-drive bubble, Hawking radiation with temperature $T_H = \kappa_1/2\pi$.

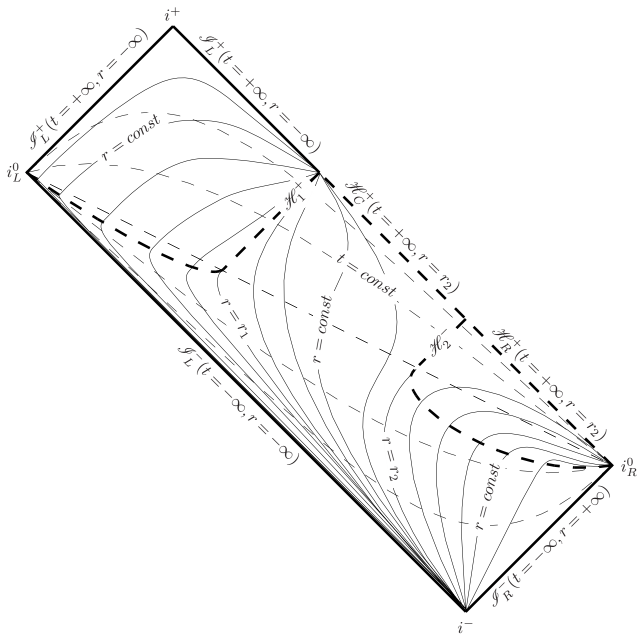


Figure 12.11: Carter-Penrose diagram of a dynamically formed Alcubierre warp drive (from [107]).

Equivalently, we find that the corresponding expansion in proximity of the white horizon is [107]

$$U(u \to -\infty) \simeq U_{\mathrm{WH}} + D_1 e^{\kappa_2 u} + \frac{D_2}{2} e^{2\kappa_2 u} + \dots \ , \tag{12.26}$$

where $D_2 > 0$ and κ_2 is the white hole surface gravity and is also defined to be positive. The interpretation of this relation in terms of particle production is not as clear as in the black horizon case and a full study of the renormalised stress energy tensor (RSET) is required.

Renormalized stress-energy tensor

In past null coordinates U and W the metric can be written as

$$ds^2 = -C(U, W) dU dW \ . \tag{12.27}$$

In the stationary region at late times, we can use the previous future null coordinate u and a new coordinate $\tilde{w}$, defined as

$$\tilde{w}(t, r) = t + \int_0^r \frac{dr}{c - \bar{v}(r)} \ . \tag{12.28}$$

In these coordinates the metric is expressed as

$$ds^2 = -\bar{C}(u, \tilde{w}) du d\tilde{w} \ , \quad \text{with} \quad C(U, W) = \frac{\bar{C}(u, \tilde{w})}{\dot{p}(u)\dot{q}(\tilde{w})} \ , \tag{12.29}$$

where $U = p(u)$ and $W = q(\tilde{w})$. In this way, $\bar{C}$ depends only on r through $u, \tilde{w}$.

For concreteness, we refer to the RSET associated with a quantum massless scalar field living on the spacetime. The RSET components for this case are again given in Ref. [5] and using the relationships $U = p(u)$, $W = q(\tilde{w})$ and the time-independence of u and $\tilde{w}$, one can calculate the RSET components in the stationary (late times) region.[107]

Let us now focus on the energy density inside the bubble, in particular at the energy ρ as measured by a set of free-falling observers, whose four velocity is $u_c^\mu = (1, \bar{v})$ in (t, r) components. For these observers neglecting transient terms one obtains [107] $\rho = T_{\mu\nu} u_c^\mu u_c^\nu = \rho_{\rm st} + \rho_{\rm dyn}$, where we define a static term $\rho_{\rm st}$, depending only on the r coordinate through $\bar{v}(r)$,

$$\rho_{\rm st} \equiv -\frac{1}{24\pi} \left[\frac{\left(\bar{v}^4 - \bar{v}^2 + 2\right)}{\left(1 - \bar{v}^2\right)^2} \bar{v}'^{\,2} + \frac{2\bar{v}}{1 - \bar{v}^2} \bar{v}'' \right] , \tag{12.30}$$

and a, time-dependent, dynamic term

$$\rho_{\rm dyn} \equiv \frac{1}{48\pi} \frac{\mathcal{F}(u)}{(1 + \bar{v})^2} , \quad \text{where} \quad \mathcal{F}(u) \equiv \frac{3\ddot{p}^2(u) - 2\dot{p}(u)\,\dddot{p}(u)}{\dot{p}^2(u)} . \tag{12.31}$$

Physical interpretation

Let us start by looking at behavior of the RSET in the center of the bubble at late times. Here $\rho_{\rm st} = 0$, because $\bar{v}(r = 0) = \bar{v}'(r = 0) = 0$. One can evaluate $\rho_{\rm dyn}$ from Eq. (12.31) by using a late-time expansion for $\mathcal{F}(u)$, which gives $\mathcal{F}(u) \approx \kappa_1^2$, so that $\rho(r = 0) \approx \kappa_1^2/(48\pi) = \pi T_H^2/12$, where $T_H \equiv \kappa_1/(2\pi)$ is the usual Hawking temperature. This result confirms that an observer inside the bubble measures a thermal flux of radiation at temperature T_H.

Let us now study ρ on the horizons $\mathcal{H}^+$ and $\mathcal{H}^-$. Here, both $\rho_{\rm st}$ and $\rho_{\rm dyn}$ are divergent because of the $(1 + \bar{v})$ factors in the denominators. Using the late time expansion of $\mathcal{F}(u)$ in the proximity of the black horizon one gets

$$\lim_{r \to r_1} \mathcal{F}(u) = \kappa_1^2 \left\{ 1 + \left[3\left(\frac{A_2}{A_1}\right)^2 - 2\frac{A_3}{A_1} \right] e^{-2\kappa_1 t} \left(r - r_1\right)^2 + \mathcal{O}\left(\left(r - r_1\right)^3 \right) \right\} , \tag{12.32}$$

and expanding both the static and the dynamic terms up to order $\mathcal{O}(r - r_1)$, one obtains that the diverging terms ($\propto (r - r_1)^{-2}$ and $\propto (r - r_1)^{-1}$) in $\rho_{\rm st}$ and $\rho_{\rm dyn}$ exactly cancel each other exactly as it happens in the semiclassical collapse [107]. We hence get an Unruh-like state setting up in correspondence to the formation of the warp drive. It is now clear that the total ρ is $\mathcal{O}(1)$ on the horizon and does not diverge at any finite time. By looking at the subleading terms,

$$\rho = \frac{e^{-2\kappa_1 t}}{48\pi} \left[3\left(\frac{A_2}{A_1}\right)^2 - 2\frac{A_3}{A_1} \right] + A + \mathcal{O}\left(r - r_1\right) , \tag{12.33}$$

where A is a constant, we see that on the black horizon the contribution of the transient radiation (different from Hawking radiation) dies off exponentially with time, on a time scale $\sim 1/\kappa_1$.

Close to the white horizon, the divergences in the static and dynamical contributions cancel each other, as in the black horizon case. However, something distinctive occurs with the subleading contributions. In fact, they now becomes

$$\rho = \frac{e^{2\kappa_2 t}}{48\pi} \left[3\left(\frac{D_2}{D_1}\right)^2 - 2\frac{D_3}{D_1} \right] + D + \mathcal{O}\left(r - r_1\right) . \tag{12.34}$$

This expression shows an exponential increase of the energy density with time. This means that ρ grows exponentially and eventually diverges along $\mathcal{H}^-$. This is just the statement that a Unruh-like state cannot be regular also on a white horizon.

In a completely analogous way, one can study ρ close to the Cauchy horizon. Performing an expansion at late times ($t \to +\infty$) one finds that the RSET diverges also there [107]. Note that the above mentioned divergences are very different in nature. The divergence at late times on $\mathcal{H}^+$ stems from the untamed growth of the transient disturbances produced by the white horizon formation. The RSET divergence on the Cauchy horizon is due instead to the well known infinite blue-shift suffered by light rays while approaching this kind of horizon. While the second can be deemed inconclusive because of the Kay–Radikowski–Wald theorem, the first one is inescapable. Summarising: the backreaction of the RSET will doom the warp drive making it semiclassically unstable. This seems to suggest a preemptive implementation of Hawking's chronology protection conjecture.

Bibliography

[1] Robert M Wald. *General relativity*. Chicago, IL: Chicago Univ. Press, 1984.

[2] Sean M. Carroll. *Spacetime and geometry. An introduction to general relativity*. 2004.

[3] Eric Poisson. *A Relativist's Toolkit: The Mathematics of Black-Hole Mechanics*. Cambridge University Press, 2004. DOI: 10.1017/CBO9780511606601.

[4] T. Padmanabhan. *Gravitation: Foundations and Frontiers*. Cambridge University Press, 2010. DOI: 10.1017/CBO9780511807787.

[5] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge, UK: Cambridge Univ. Press, 1984. DOI: 10.1017/CBO9780511622632.

[6] A A Grib, S G Mamayev, and V M Mostepanenko. *Vacuum quantum effects in strong fields*. St. Petersburg: Fridmann Lab., 1994.

[7] Viatcheslav Mukhanov and Sergei Winitzki. *Introduction to quantum effects in gravity*. Cambridge University Press, 2007.

[8] Vittorio Berzi and Vittorio Gorini. "Reciprocity Principle and the Lorentz Transformations". In: *Journal of Mathematical Physics* (1969).

[9] R. Torretti. *Relativity and geometry*. 1983.

[10] Sebastiano Sonego and Massimo Pin. "Foundations of anisotropic relativistic mechanics". In: *Journal of Mathematical Physics* (2009).

[11] George Bogoslovsky. "Some physical displays of the space anisotropy relevant to the feasibility of its being detected at a laboratory". In: *arXiv e-prints* (2007).

[12] Harvey R. Brown. *Physical relativity. Space-time structure from a dynamical perspective*. 2005.

[13] S. Liberati. "Tests of Lorentz invariance: a 2013 update". In: *Classical and Quantum Gravity* (2013).

[14] Andrea Pelissetto and Massimo Testa. "Getting the Lorentz transformations without requiring an invariant speed". In: *American Journal of Physics* (2015).

[15] Eolo Di Casola, Stefano Liberati, and Sebastiano Sonego. "Nonequivalence of equivalence principles". In: *American Journal of Physics* (2015).

[16] Eolo Di Casola, Stefano Liberati, and Sebastiano Sonego. "Weak equivalence principle for self-gravitating bodies: A sieve for purely metric theories of gravity". In: *Phys. Rev. D* (2014).

[17] In: (2005). URL: http://www.slimy.com/~steuard/teaching/tutorials/GRtorsion.pdf.

[18] P. K. Townsend. *Black Holes*. 1997. arXiv: gr-qc/9707012 [gr-qc].

[19] Erwin Schrödinger. *Space-Time Structure*. Cambridge Science Classics. Cambridge University Press, 1985. DOI: 10.1017/CBO9780511586446.

[20] Kris Pardo et al. "Limits on the number of spacetime dimensions from GW170817". In: *Journal of Cosmology and Astro-Particle Physics* (2018).

[21] R. Bufalo, M. Oksanen, and A. Tureanu. "How unimodular gravity theories differ from general relativity at quantum level". In: *Eur. Phys. J.* C75.10 (2015), p. 477. DOI: 10.1140/epjc/s10052-015-3683-3. arXiv: 1505.04978 [hep-th].

[22] Stanley Deser. "Selfinteraction and gauge invariance". In: *Gen. Rel. Grav.* 1 (1970), pp. 9–18. DOI: 10.1007/BF00759198. arXiv: gr-qc/0411023 [gr-qc].

[23] T. Padmanabhan. "From gravitons to gravity: Myths and reality". In: *Int. J. Mod. Phys.* D17 (2008), pp. 367–398. DOI: 10.1142/S0218271808012085. arXiv: gr-qc/0409089 [gr-qc].

[24] Carlos Barceló, Raúl Carballo-Rubio, and Luis J. Garay. "Unimodular gravity and general relativity from graviton self-interactions". In: *Phys. Rev.* D89.12 (2014), p. 124019. DOI: 10.1103/PhysRevD.89.124019. arXiv: 1401.2941 [gr-qc].

[25] Domenico Giulini. "What is (not) wrong with scalar gravity?" In: *Stud. Hist. Phil. Sci.* B39.1 (2008), pp. 154–180. DOI: 10.1016/j.shpsb.2007.09.001. arXiv: gr-qc/0611100 [gr-qc].

[26] Yvonne Choquet-Bruhat. *General Relativity and the Einstein Equations*. United Kingdom: Oxford University Press, 2009.

[27] Steven Willison. "Lovelock gravity and Weyl's tube formula". In: *Phys. Rev.* D80 (2009), p. 064018. DOI: 10.1103/PhysRevD.80.064018. arXiv: 0904.3224 [gr-qc].

[28] Sveta Golod and Tsvi Piran. "Choptuik's Critical Phenomenon in Einstein-Gauss-Bonnet Gravity". In: *Phys. Rev.* D85 (2012), p. 104015. DOI: 10.1103/PhysRevD.85.104015. arXiv: 1201.6384 [gr-qc].

[29] C. Brans and R. H. Dicke. "Mach's principle and a relativistic theory of gravitation". In: *Phys. Rev.* 124 (1961). [,142(1961)], pp. 925–935. DOI: 10.1103/PhysRev.124.925.

[30] Thomas P. Sotiriou and Valerio Faraoni. "f(R) theories of gravity". In: *Reviews of Modern Physics* (2010).

[31] Antonio De Felice and Shinji Tsujikawa. "f(R) Theories". In: *Living Reviews in Relativity* (2010).

[32] Thomas P. Sotiriou. "f(R) gravity and scalar tensor theory". In: *Classical and Quantum Gravity* (2006).

[33] Gregory Walter Horndeski. "Second-Order Scalar-Tensor Field Equations in a Four-Dimensional Space". In: *International Journal of Theoretical Physics* (1974).

[34] Neil Barnaby and Niky Kamran. "Dynamics with infinitely many derivatives: The Initial value problem". In: *JHEP* 02 (2008), p. 008. DOI: 10.1088/1126-6708/2008/02/008. arXiv: 0709.3968 [hep-th].

[35] David Langlois and Karim Noui. "Degenerate higher derivative theories beyond Horndeski: evading the Ostrogradski instability". In: *JCAP* 1602.02 (2016), p. 034. DOI: 10.1088/1475-7516/2016/02/034. arXiv: 1510.06930 [gr-qc].

[36] Jibril Ben Achour et al. "Degenerate higher order scalar-tensor theories beyond Horndeski up to cubic order". In: *JHEP* 12 (2016), p. 100. DOI: 10.1007/JHEP12(2016)100. arXiv: 1608.08135 [hep-th].

[37] I. L. Shapiro. "Physical aspects of the space-time torsion". In: *Phys. Rept.* 357 (2002), p. 113. DOI: 10.1016/S0370-1573(01)00030-8. arXiv: hep-th/0103093 [hep-th].

[38] Thomas P. Sotiriou and Stefano Liberati. “Metric-affine f(R) theories of gravity”. In: *Annals Phys.* 322 (2007), pp. 935–966. DOI: 10.1016/j.aop.2006.06.002. arXiv: gr-qc/0604006 [gr-qc].

[39] Ted Jacobson. “Einstein-aether gravity: A Status report”. In: *PoS* QG-PH (2007), p. 020. DOI: 10.22323/1.043.0020. arXiv: 0801.1547 [gr-qc].

[40] Thomas P. Sotiriou. “Horava-Lifshitz gravity: a status report”. In: *J. Phys. Conf. Ser.* 283 (2011), p. 012034. DOI: 10.1088/1742-6596/283/1/012034. arXiv: 1010.3218 [hep-th].

[41] Charles W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. San Francisco: W. H. Freeman, 1973. ISBN: 9780716703440, 9780691177793.

[42] G. Jannes et al. “Experimental demonstration of the supersonic-subsonic bifurcation in the circular jump: A hydrodynamic white hole”. In: *Phys. Rev.* E83 (2011), p. 056312. DOI: 10.1103/PhysRevE.83.056312. arXiv: 1010.1701 [physics.flu-dyn].

[43] Carlos Barceló, Stefano Liberati, and Matt Visser. “Analogue Gravity”. In: *Living Reviews in Relativity* (2005).

[44] I. Racz and R. M. Wald. “Extensions of spacetimes with Killing horizons”. In: *Classical and Quantum Gravity* (1992).

[45] Matt Visser. *Lorentzian wormholes. From Einstein to Hawking*. 1995.

[46] Bethan Cropp, Stefano Liberati, and Matt Visser. “Surface gravities for non-Killing horizons”. In: *Class. Quant. Grav.* 30 (2013), p. 125001. DOI: 10.1088/0264-9381/30/12/125001. arXiv: 1302.2383 [gr-qc].

[47] Carlos Barcelo et al. “Hawking-like radiation from evolving black holes and compact horizonless objects”. In: *JHEP* 02 (2011), p. 003. DOI: 10.1007/JHEP02(2011)003. arXiv: 1011.5911 [gr-qc].

[48] Vitor Cardoso et al. “Quasinormal modes and Strong Cosmic Censorship”. In: *Phys. Rev. Lett.* 120.3 (2018), p. 031103. DOI: 10.1103/PhysRevLett.120.031103. arXiv: 1711.10502 [gr-qc].

[49] G. F. R. Ellis and T. Rothman. “Lost horizons”. In: *American Journal of Physics* 61.10 (1993), pp. 883–893. DOI: 10.1119/1.17400. eprint: https://doi.org/10.1119/1.17400. URL: https://doi.org/10.1119/1.17400.

[50] Tamara M. Davis and Charles H. Lineweaver. “Expanding confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe”. In: *Proc. Astron. Soc. Austral.* (2003). [Publ. Astron. Soc. Austral.21,97(2004)]. DOI: 10.1071/AS03040. arXiv: astro-ph/0310808 [astro-ph].

[51] Matt Visser. “Physical observability of horizons”. In: *Phys. Rev.* D90.12 (2014), p. 127502. DOI: 10.1103/PhysRevD.90.127502. arXiv: 1407.7295 [gr-qc].

[52] Koushik Balasubramanian and Sujan P. Dabholkar. “Time-Dependent Warping and Non-Singular Bouncing Cosmologies”. In: (2014). arXiv: 1401.7015 [hep-th].

[53] Abhay Ashtekar and Badri Krishnan. “Dynamical horizons and their properties”. In: *Phys. Rev.* D68 (2003), p. 104030. DOI: 10.1103/PhysRevD.68.104030. arXiv: gr-qc/0308033 [gr-qc].

[54] Abhay Ashtekar and Badri Krishnan. “Isolated and Dynamical Horizons and Their Applications”. In: *Living Reviews in Relativity* 7.1 (2004), p. 10. DOI: 10.12942/lrr-2004-10. URL: https://doi.org/10.12942/lrr-2004-10.

[55] Roger Penrose. “Gravitational Collapse and Space-Time Singularities”. In: *Phys. Rev. Lett.* 14 (3 1965), pp. 57–59.

[56] I. D. Novikov and Valeri P. Frolov. *PHYSICS OF BLACK HOLES*. Vol. 27. Dordrecht, Netherlands: Kluwer Academic, 1989. DOI: 10.1007/978-94-017-2651-1.

[57] S. W. Hawking. "The Occurrence of Singularities in Cosmology. III. Causality and Singularities". In: *Proceedings of the Royal Society of London Series A* (1967).

[58] S. W. Hawking and G. F. R. Ellis. "The Cosmic Black-Body Radiation and the Existence of Singularities in Our Universe". In: *Astrophys. J.* (1968).

[59] B. P. Abbott et al. "GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs". In: *Phys. Rev. X* 9 (3 2019), p. 031040.

[60] The EHT Collaboration et al. "First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole". In: *ApJL* 875 (2019), p. 1. URL: https://iopscience.iop.org/article/10.3847/2041-8213/ab0ec7.

[61] Raúl Carballo-Rubio et al. "Geodesically complete black holes". In: (2019). arXiv: 1911.11200 [gr-qc].

[62] Kip S. Thorne, Richard H. Price, and Douglas A. MacDonald. *Black holes: The membrane paradigm.* 1986.

[63] Matt Visser. "The Kerr spacetime: A Brief introduction". In: *Kerr Fest: Black Holes in Astrophysics, General Relativity and Quantum Gravity Christchurch, New Zealand, August 26-28, 2004*. 2007. arXiv: 0706.0622 [gr-qc].

[64] Daniel Kapec and Alexandru Lupsasca. "Particle motion near high-spin black holes". In: *Class. Quant. Grav.* 37.1 (2020), p. 015006. DOI: 10.1088/1361-6382/ab519e. arXiv: 1905.11406 [hep-th].

[65] Richard Brito, Vitor Cardoso, and Paolo Pani. "Superradiance". In: *Lect. Notes Phys.* 906 (2015), pp.1–237. DOI: 10.1007/978-3-319-19000-6. arXiv: 1501.06570 [gr-qc].

[66] W. Unruh. "Separability of the Neutrino Equations in a Kerr Background". In: *Phys. Rev. Lett.* 31 (20 1973), pp. 1265–1267.

[67] J. M. Bardeen, B. Carter, and S. W. Hawking. "The four laws of black hole mechanics". In: *Communications in Mathematical Physics* (1973).

[68] W. Israel. "Third Law of Black-Hole Dynamics: A Formulation and Proof". In: *Phys. Rev. Lett.* 57.4 (1986), p. 397. DOI: 10.1103/PhysRevLett.57.397.

[69] F. Belgiorno and M. Martellini. "Black holes and the third law of thermodynamics". In: *Int. J. Mod. Phys.* D13 (2004), pp. 739–770. DOI: 10.1142/S0218271804004876. arXiv: gr-qc/0210026 [gr-qc].

[70] Rouzbeh Allahverdi et al. "Reheating in Inflationary Cosmology: Theory and Applications". In: *Annual Review of Nuclear and Particle Science* 60.1 (2010), pp. 27–51. DOI: 10.1146/annurev.nucl.012809.104511. eprint: https://doi.org/10.1146/annurev.nucl.012809.104511. URL: https://doi.org/10.1146/annurev.nucl.012809.104511.

[71] Peter Roper, Paul Bressloff, and Andre Longtin. In: (2000).

[72] Ted Jacobson. "Introduction to quantum fields in curved space-time and the Hawking effect". In: *Lectures on quantum gravity. Proceedings, School of Quantum Gravity, Valdivia, Chile, January 4-14, 2002*. 2003, pp. 39–89. DOI: 10.1007/0-387-24992-3_2. arXiv: gr-qc/0308048 [gr-qc].

[73] L. H. Ford. "Quantum field theory in curved space-time". In: *Particles and fields. Proceedings, 9th Jorge Andre Swieca Summer School, Campos do Jordao, Brazil, February 16-28, 1997*. 1997, pp. 345–388. arXiv: gr-qc/9707062 [gr-qc].

[74] YASUSHI TAKAHASHI and HIROOMI UMEZAWA. "THERMO FIELD DYNAMICS". In: *International Journal of Modern Physics B* 10.13n14 (1996), pp. 1755–1805. DOI: 10.1142/S0217979296000817. eprint: https://doi.org/10.1142/S0217979296000817. URL: https://doi.org/10.1142/S0217979296000817.

[75] D. W. Sciama, P. Candelas, and D. Deutsch. "Quantum Field Theory, Horizons and Thermodynamics". In: *Adv. Phys.* 30 (1981), pp. 327–366. DOI: 10.1080/00018738100101457.

[76] P.C. Causseè. l'Album du Marin. Charpentier, Nantes, 1836.

[77] Sipko L. Boersma. "A maritime analogy of the Casimir effect". In: *American Journal of Physics* 64.5 (1996), pp. 539–541. DOI: 10.1119/1.18150. eprint: https://doi.org/10.1119/1.18150. URL: https://doi.org/10.1119/1.18150.

[78] Ted Jacobson. "Black holes and Hawking radiation in spacetime and its analogues". In: *Lect. Notes Phys.* 870 (2013), pp. 1–29. DOI: 10.1007/978-3-319-00266-8_1. arXiv: 1212.6821 [gr-qc].

[79] S. A. Fulling and P. C. W. Davies. "Radiation from a Moving Mirror in Two Dimensional Space-Time: Conformal Anomaly". In: *Proceedings of the Royal Society of London Series A* (1976).

[80] Stefano Liberati, Tony Rothman, and Sebastiano Sonego. "Nonthermal nature of incipient extremal black holes". In: *Phys. Rev. D* 62 (2 2000), p. 024005.

[81] S. W. Hawking. "Particle creation by black holes". In: *Communications in Mathematical Physics* (1975).

[82] S. W. Hawking. "Superscattering matrix for two-dimensional black holes". In: *Phys. Rev. D* 50 (6 1994), pp. 3982–3986.

[83] Andrew Strominger. "Les Houches lectures on black holes". In: *NATO Advanced Study Institute: Les Houches Summer School, Session 62: Fluctuating Geometries in Statistical Mechanics and Field Theory Les Houches, France, August 2-September 9, 1994*. 1994. arXiv: hep-th/9501071 [hep-th].

[84] Don N. Page. "Information in black hole radiation". In: *Phys. Rev. Lett.* 71 (23 1993), pp. 3743–3746.

[85] Ahmed Almheiri et al. "Black Holes: Complementarity or Firewalls?" In: *JHEP* 02 (2013), p. 062. DOI: 10.1007/JHEP02(2013)062. arXiv: 1207.3123 [hep-th].

[86] Raúl Carballo-Rubio et al. "On the viability of regular black holes". In: *JHEP* 07 (2018), p. 023. DOI: 10.1007/JHEP07(2018)023. arXiv: 1805.02675 [gr-qc].

[87] Daniele Malafarina. "Classical collapse to black holes and quantum bounces: A review". In: *Universe* 3.2 (2017), p. 48. DOI: 10.3390/universe3020048. arXiv: 1703.04138 [gr-qc].

[88] Joshua H. Cooperman and Markus A. Luty. "Renormalization of Entanglement Entropy and the Gravitational Effective Action". In: *JHEP* 12 (2014), p. 045. DOI: 10.1007/JHEP12(2014)045. arXiv: 1302.1878 [hep-th].

[89] Robert M. Wald. "On identically closed forms locally constructed from a field". In: *Journal of Mathematical Physics* 31.10 (1990), pp. 2378–2384. DOI: 10.1063/1.528839. URL: https://doi.org/10.1063/1.528839.

[90] Robert M. Wald. "Black hole entropy is the Noether charge". In: *Phys. Rev. D* 48 (8 1993), R3427–R3431.

[91] Vivek Iyer and Robert M. Wald. "Some properties of the Noether charge and a proposal for dynamical black hole entropy". In: *Phys. Rev. D* 50 (2 1994), pp. 846–864.

[92] G W Gibbons and S W Hawking. *Euclidean Quantum Gravity*. WORLD SCIENTIFIC, 1993. DOI: 10.1142/1301. eprint: https://www.worldscientific.com/doi/pdf/10.1142/1301. URL: https://www.worldscientific.com/doi/abs/10.1142/1301.

[93] G. W. Gibbons and S. W. Hawking. "Action Integrals and Partition Functions in Quantum Gravity". In: *Phys. Rev.* D15 (1977), pp. 2752–2756. DOI: 10.1103/PhysRevD.15.2752.

[94] Carlos Barceló et al. “Fate of gravitational collapse in semiclassical gravity”. In: *Phys. Rev. D* (2008).

[95] Stephen A. Fulling, Mark Sweeny, and Robert M. Wald. “Singularity structure of the two-point function quantum field theory in curved spacetime”. In: *Comm. Math. Phys.* 63.3 (1978), pp. 257–264. URL: https://projecteuclid.org:443/euclid.cmp/1103904566.

[96] Kurt Gödel. “An Example of a New Type of Cosmological Solutions of Einstein’s Field Equations of Gravitation”. In: *Reviews of Modern Physics* (1949).

[97] In: (2013). DOI: 10.1088/1367-2630/15/1/013063. URL: https://doi.org/10.1088%2F1367-2630%2F15%2F1%2F013063.

[98] Frank Grave. PhD thesis. 2010.

[99] Robert W. Fuller and John A. Wheeler. “Causality and Multiply Connected Space-Time”. In: *Phys. Rev.* 128 (2 1962), pp. 919–929.

[100] Michael S. Morris, Kip S. Thorne, and Ulvi Yurtsever. “Wormholes, Time Machines, and the Weak Energy Condition”. In: *Phys. Rev. Lett.* 61 (13 1988), pp. 1446–1449.

[101] Michael S. Morris and Kip S. Thorne. “Wormholes in spacetime and their use for interstellar travel: A tool for teaching general relativity”. In: *American Journal of Physics* 56.5 (1988), pp. 395–412. DOI: 10.1119/1.15620. eprint: https://doi.org/10.1119/1.15620. URL: https://doi.org/10.1119/1.15620.

[102] In: (1994). DOI: 10.1088/0264-9381/11/5/001. URL: https://doi.org/10.1088%2F0264-9381%2F11%2F5%2F001.

[103] S. V. Krasnikov. “Hyperfast travel in general relativity”. In: *Phys. Rev. D* (1998).

[104] I. D. Novikov. *Evolution of the universe*. 1983.

[105] John Friedman et al. “Cauchy problem in spacetimes with closed timelike curves”. In: *Phys. Rev. D* (1990).

[106] Bernard S. Kay, Marek J. Radzikowski, and Robert M. Wald. “Quantum Field Theory on Spacetimes with a Compactly Generated Cauchy Horizon”. In: *Communications in Mathematical Physics* (1997).

[107] Stefano Finazzi, Stefano Liberati, and Carlos Barceló. “Semiclassical instability of dynamical warp drives”. In: *Phys. Rev. D* (2009).

[108] Carlos Barceló, Stefano Finazzi, and Stefano Liberati. “On the impossibility of superluminal travel: the warp drive lesson”. In: *arXiv e-prints* (2010).

[109] Eric Davis. In: (2004).