Chapter 2
Extended Space-Times, Causal Structure and Penrose Diagrams

O radiant Dark! O darkly fostered ray
Thou hast a joy too deep for shallow Day!
George Eliot (The Spanish Gypsy)

2.1 Introduction and a Short History of Black Holes

It seems that the first to conceive the idea of what we call nowadays a black-hole was the English Natural Philosopher and Geologist John Michell (1724–1793). Member of the Royal Society, Michell already before 1783 invented a device to measure Newton's gravitational constant, namely the torsion balance that he built independently from its co-inventor Charles Augustin de Coulomb. He did not live long enough to put into use his apparatus which was inherited by Cavendish. In 1784 in a letter addressed precisely to Cavendish, John Michell advanced the hypothesis that there could exist heavenly bodies so massive that even light could not escape from their gravitational attraction. This letter surfaced back to the attention of contemporary scientists only in the later seventies of the XXth century [1]. Before that finding, credited to be the first inventor of black-holes was Pierre Simon Laplace (see Fig. 2.1). In the 1796 edition of his monumental book *Exposition du Système du Monde* [2] he presented exactly the same argument put forward in Michell's letter, developing it with his usual mathematical rigor. All historical data support the evidence that Michell and Laplace came to the same hypothesis independently. Indeed the idea was quite mature for the physics of that time, once the concept of escape velocity v_e had been fully understood.

Consider a spherical celestial body of mass M and radius R and let us pose the question what is the minimum initial vertical velocity that a point-like object located on its surface, for instance a rocket, should have in order to be able to escape to infinite distance from the center of gravitational attraction. Energy conservation provides the immediate answer to such a problem. At the initial moment $t = t_0$ the energy of the missile is:

$$E = \frac{1}{2} m_m v_e^2 - \frac{GMm_m}{R} \tag{2.1.1}$$

where G is Newton's constant. At a very late time, when the missile has reached $R = \infty$ with a final vanishing velocity its energy is just $0 + 0 = 0$. Hence E vanished

P.G. Frè, *Gravity, a Geometrical Course*, DOI 10.1007/978-94-007-5443-0_2,

Fig. 2.1 Pierre Simon Laplace (1749–1827) was born in Beaumont en Auge in Normandy in the family of a poor farmer. He could study thanks to the generous help of some neighbors. Later with a recommendation letter of d'Alembert he entered the military school of Paris where he became a teacher of mathematics. There he started his monumental and original research activity in Mathematics and Astronomy that made him one of the most prominent scientists of his time and qualified him to the rank of founder of modern differential calculus, his work being a pillar of XIXth century Mathematical Physics. A large part of his work on Astronomy was still done under the Ancien Regime and dates back to the period 1771–1787. He proved the stability of the Solar System and developed all the mathematical tools for the systematic calculus of orbits in Newtonian Physics. His results were summarized in the two fundamental books *Mecanique Cèleste* and *Exposition du Système du Monde*. Besides introducing the first idea of what we call nowadays a black-hole, Laplace was also the first to advance the hypothesis that the Solar System had formed through the cooling of a globular-shaped, rotating, cluster of very hot gas (a nebula). In later years of his career Laplace gave fundamental and framing contributions to the mathematical theory of probability. His name is attached to numberless corners of differential analysis and function theory. He received many honors both in France and abroad. He was member of all most distinguished Academies of Europe. He also attempted the political career serving as Minister of Interiors in one of the first Napoleonic Cabinets, yet he was soon dismissed by the First Consul as a person not qualified for that administrative job notwithstanding Napoleon's recognition that he was a great scientist. Politically Laplace was rather cynic and ready to change his opinions and allegiance in order to follow the blowing wind. Count of the First French Empire, after the fall of Napoleon he came on good terms with the Bourbon Restoration and was compensated by the King with the title of marquis

also at the beginning, which yields:

$$v_e = \sqrt{2\frac{GM}{R}} \tag{2.1.2}$$

If we assume that light travels at a finite velocity c, then there could exist heavenly bodies so dense that:

$$\sqrt{2\frac{GM}{R}} > c \tag{2.1.3}$$

In that case not even light could escape from the gravitational field of that body and no-one on the surface of the latter could send any luminous signal that distant observers could perceive. In other words by no means distant observers could see the surface of that super-massive object and even less what might be in its interior.

Obviously neither Michell nor Laplace had a clear perception that the speed of light c is always the same in every reference frame, since Special Relativity had to wait its own discovery for another century. Yet Laplace's argument was the following: let us assume that the velocity of light is some constant number a on the surface of the considered celestial body. Then he proceeded to an estimate of the speed of light on the surface of the Sun, which he could do using the annual light aberration in the Earth-Sun system. The implicit, although unjustified, assumption was that light velocity is unaffected, or weakly affected, by gravity. Analyzing such an assumption in full-depth it becomes clear that it was an anticipation of Relativity in disguise.

Actually condition (2.1.3) has an exact intrinsic meaning in General Relativity. Squaring this equation we can rewrite it as follows:

$$R > r_S \equiv 2\frac{GM}{c^2} \equiv 2m \tag{2.1.4}$$

where r_S is the Schwarzschild radius of a body of mass M, namely the unique parameter which appears in the Schwarzschild solution of Einstein Equations.

So massive bodies are visible and behave qualitatively according to human common sense as long as their dimensions are much larger then their Schwarzschild radius. Due to the smallness of Newton's constant and to the hugeness of the speed of light, this latter is typically extremely small. Just of the order of a kilometer for a star, and about 10^{-23} cm for a human body. Nevertheless, as we extensively discussed in Chap. 6 of Volume 1, sooner or later all stars collapse and regions of spacetime with outrageously large energy-densities do indeed form, whose typical linear size becomes comparable to r_S. The question of what happens if it is smaller than r_S is not empty, on the contrary it is a fundamental one, related with the appropriate interpretation of what lies behind the apparent singularity of the Schwarzschild metric at $r = r_S$.

As all physicists know, any singularity is just the signal of some kind of criticality. At the singular point a certain description of physical reality breaks down and it must be replaced by a different one: for instance there is a phase-transition and the degrees of freedom that capture most of the energy in an ordered phase become negligible with respect to other degrees of freedom that are dominating in a disordered phase. What is the criticality signaled by the singularity $r = r_S$ of the Schwarzschild metric? Is it a special feature of this particular solution of Einstein Equations or it is just an instance of a more general phenomenon intrinsic to the laws of gravity as stated by General Relativity? The answer to the first question is encoded in the wording *event horizon*. The answer to the second question is that event horizons are a generic feature of static solutions of Einstein equations.

An event-horizon $\mathfrak{H}$ is a hypersurface in a pseudo-Riemannian manifold $(\mathscr{M}, g)$ which separates two sub-manifolds, one $\mathfrak{E} \subset \mathscr{M}$, named the exterior, can communicate with infinity by sending signals to distant observers, the other $\mathrm{BH} \subset \mathscr{M}$, named the black-hole, is causally disconnected from infinity, since no signal produced in

BH can reach the outside region $\mathfrak{E}$. The black-hole is the region deemed by Michell and Laplace where the escape velocity is larger than the speed of light.

In order to give a precise mathematical sense to the above explanation of event-horizons a lot of things have to be defined and interpreted. First of all what is infinity and is it unique? Secondly which kind of hypersurface is an event-horizon? Thirdly can we eliminate the horizon singularity by means of a suitable analytic extension of the apparently singular manifold? Finally, how do we define causal relations in a curved Lorentzian space-time?

The present chapter addresses the above questions. The answers were found in the course of the XXth century and constitute the principal milestones in the history of black-holes.

Although Schwarzschild metric was discovered in 1916, less than six months after the publication of General Relativity, its analytic extension, that opened the way to a robust mathematical theory of black-holes, was found only forty-five years later, six after Einstein's death. In 1960, the American theorist Martin Kruskal (see Fig. 2.2) found a one-to-many coordinate transformation that allowed him to represent Schwarzschild space-time as a portion of a larger space-time where the locus $r = r_S$ is non-singular, rather it is a well-defined light-like hypersurface constituting precisely the event-horizon [6]. A similar coordinate change was independently proposed the same year also by the Australian-Hungarian mathematician Georges Szekeres [7].

These mathematical results provided a solid framework for the description of the final state in the gravitational collapse of those stars that are too massive to stop at the stage of white-dwarfs or neutron-stars. In Chap. 6 of Volume 1 we already mentioned the intuition of Robert Openheimer and H. Snyder who, in their 1939 paper, wrote: *When all thermonuclear sources of energy are exhausted, a sufficiently heavy star will collapse. Unless something can somehow reduce the star's mass to the order of that of the sun, this contraction will continue indefinitely...past white dwarfs, past neutron stars, to an object cut off from communication with the rest of the universe*. Such an object, could be identified with the interior of the event horizon in the newly found Kruskal space-time. Yet, since the Kruskal-Schwarzschild metric is spherical symmetric such identification made sense only in the case the parent star had vanishing angular momentum, namely was not rotating at all. This is quite rare since most stars rotate.

In 1963 the New Zealand physicist Roy Kerr, working at the University of Texas, found the long sought for generalization of the Schwarzschild metric that could describe the end-point equilibrium state in the gravitational collapse of a rotating star. Kerr metric, that constitutes the main topic of Chap. 3, introduced the third missing parameter characterizing a black-hole, namely the angular momentum J. The first is the mass M, known since Schwarzschild's pioneering work, the second, the charge Q (electric, magnetic or both) had been introduced already in the first two years of life of General Relativity. Indeed the Reissner-Nordström metric,[1] which

[1] Hans Jacob Reissner (1874–1967) was a German aeronautical engineer with a passion for mathematical physics. He was the first to solve Einstein's field equations with a charged electric source

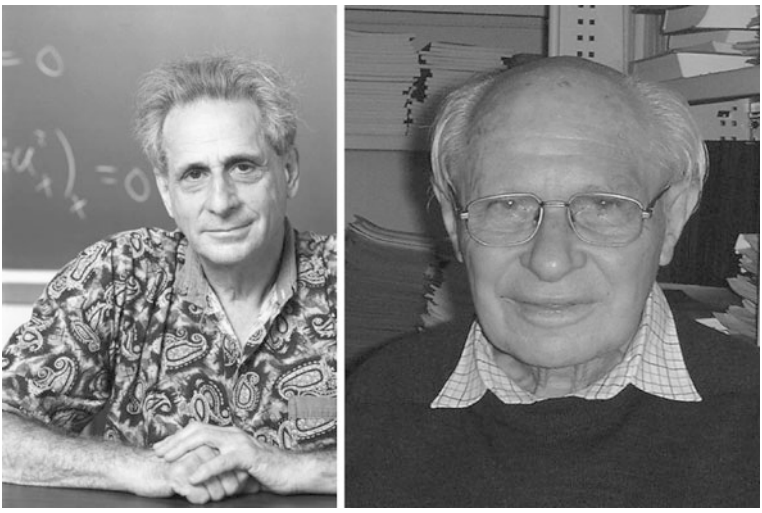

Fig. 2.2 Martin David Kruskal (1925–2006) on the *left* and George Szekeres (1911–2005) on the *right*. Student of the University of Chicago, Kruskal obtained his Ph.D from New York University and was for many years professor at Princeton University. In 1989 he joined Rutgers University were he remained the rest of his life. Mathematician and Physicist, Martin Kruskal gave very relevant contributions in theoretical plasma physics and in several areas of non-linear science. He discovered exact integrability of some non-linear differential equations and is reported to be the inventor of the concept of *solitons*. Kruskal 1960 discovery of the maximal analytic extension of Schwarzschild space-time came independently and in parallel with similar conclusions obtained by Georges Szekeres. Born in Budapest, Szekeres graduated from Budapest University in Chemistry. As a Jewish he had to escape from Nazi persecution and he fled with his family to China where he remained under Japanese occupation till the beginning of the Communist Revolution. In 1948 he was offered a position at the University of Adelaide in Australia. In this country he remained the rest of his life. Notwithstanding his degree in chemistry Szekeres was a Mathematician and he gave relevant contributions in various of its branches. He is among the founders of combinatorial geometry

solves coupled Einstein-Maxwell equations for a charged spherical body, dates back to 1916–1918.

The long time delay separating the early finding of the spherical symmetric solutions and the construction of the axial symmetric Kerr metric is explained by the high degree of algebraic complexity one immediately encounters when spherical

and he did that already in 1916 [3]. Emigrated to the United States in 1938 he taught at the Illinois Institute of Technology and later at the Polytechnic Institute of Brooklyn. Reissner's solution was retrieved and refined in 1918 by Gunnar Nordström (1881–1923) a Finnish theoretical physicist who was the first to propose an extension of space-time to higher dimensions. Independently from Kaluza and Klein and as early as 1914 he introduced a fifth dimension in order to construct a unified theory of gravitation and electromagnetism. His theory was, at the time, a competitor of Einstein's theory. Working at the University of Leiden in the Netherlands with Paul Ehrenfest, in 1918 he solved Einstein field equations for a spherically symmetric charged body [4] thus extending the Hans Reissner's results for a point charge.

symmetry is abandoned. Kerr's achievement would have been impossible without the previous monumental work of the young Russian theoretician A.Z. Petrov [5]. Educated in the same University of Kazan where, at the beginning of the XIXth century Lobachevskij had first invented non-Euclidian geometry, in his 1954 doctoral dissertation, Petrov conceived a classification of Lorentzian metrics based on the properties of the corresponding Weyl tensor. This leads to the concept of principal null-directions. According to Petrov there are exactly six types of Lorentzian metrics and, in current nomenclature, Schwarzschild and Reissner Nordström metrics are of Petrov type D. This means that they have two double principal null directions. Kerr made the hypothesis that the metric of a rotating black-hole should also be of Petrov type D and searching in that class he found it.

The decade from 1964 to 1974 witnessed a vigorous development of the mathematical theory of black-holes. Brandon Carter solved the geodesic equations for the Kerr-metric, discovering a fourth hidden first integral which reduces these differential equations to quadratures. In the same time through the work of Stephen Hawking, George Ellis, Roger Penrose and several others, general analytic methods were established to discuss, represent and classify the causal structure of space-times. Slowly a new picture emerged. Similarly to soliton solutions of other non-linear differential equations, black-holes have the characteristic features of a new kind of particles, mass, charge and angular momentum being their unique and defining attributes. Indeed it was proved that, irrespectively from all the details of its initial structure, a gravitational collapsing body sets down to a final equilibrium state parameterized only by (M, J, Q) and described by the so called Kerr-Newman metric, the generalization of the Kerr solution which includes also the Reissner Nordström charges (see Chap. 3, Sect. 3.2).

This introduced the theoretical puzzle of *information loss*. Through gravitational evolution, a supposedly coherent quantum state, containing a detailed fine structure, can evolve to a new state where all such information is unaccessible, being hidden behind the event horizon. The information loss paradox became even more severe when Hawking on one side demonstrated that black-holes can evaporate through a quantum generated thermic radiation and on the other side, in collaboration with Bekenstein, he established, that the horizon has the same properties of an entropy and obeys a theorem similar to the second principle of thermodynamics.

Hence from the theoretical view-point black-holes appear to be much more profound structures than just a particular type of classical solutions of Einstein's field equations. Indeed they provide a challenging clue into the mysterious realm of quantum gravity where causality is put to severe tests and needs to be profoundly revised. For this reason the study of black-holes and of their higher dimensional analogues within the framework of such candidates to a Unified Quantum Theory of all Interactions as Superstring Theory is currently a very active stream of research.

Ironically such a Revolution in Human Thought about the Laws of Causality, whose settlement is not yet firmly acquired, was initiated two century ago by the observations of Laplace, whose unshakable faith in determinism is well described by the following quotation from the *Essai philosophique sur les probabilités*. In

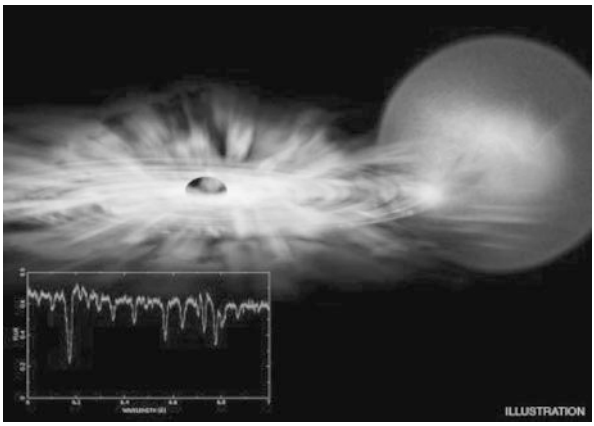

Fig. 2.3 J1655 is a binary system that harbors a black hole with a mass seven times that of the sun, which is pulling matter from a normal star about twice as massive as the sun. The Chandra observation revealed a bright X-ray source whose spectrum showed dips produced by absorption from a wide variety of atoms ranging from oxygen to nickel. A detailed study of these absorption features shows that the atoms are highly ionized and are moving away from the black hole in a high-speed wind. The system J1655 is a galactic object located at about 11,000 light years from the Sun

that book he wrote: *We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.* The vast intellect advocated by Pierre Simon and sometimes named the Laplace demon might find some problems in reconstructing the past structure of a star that had collapsed into a black hole even if that intellect had knowledge of all the conditions of the Universe at that very instant of time.

From the astronomical view-point the existence of black-holes of stellar mass has been established through many overwhelming evidences, the best being provided by binary systems where a visible normal star orbits around an invisible companion which drags matter from its mate. An example very close to us is the system J1655 shown in Fig. 2.3. Giant black-holes of millions of stellar masses have also been indirectly revealed in the core of active galactic nuclei and also at the center of our Milky Way a black hole is accredited.

In the present chapter, starting from the Kruskal extension of the Schwarzschild metric we establish the main framework for the analysis of the causal structure of space-times and we formulate the general definition of black-holes. In the next chapter we study the Kerr metric and the challenging connection between the laws of black-hole mechanics and those of thermodynamics.

2.2 The Kruskal Extension of Schwarzschild Space-Time

According to the outlined programme in this section we come back to the Schwarzschild metric (2.2.1) that we rewrite here for convenience

$$ds^2 = -\left(1 - 2\frac{m}{r}\right) dt^2 + \left(1 - 2\frac{m}{r}\right)^{-1} dr^2 + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \qquad (2.2.1)$$

and we study its causal properties. In particular we investigate the nature and the significance of the coordinate singularity at the Schwarzschild radius $r = r_S \equiv 2m$ which, as anticipated in the previous section, turns out to correspond to an *event horizon*. This explains the nomenclature *Schwarzschild emiradius* that in Chap. 4 of Volume 1 we used for the surface $r = m$.

2.2.1 Analysis of the Rindler Space-Time

Before analyzing the Kruskal extension of the Schwarzschild space-time, as a preparatory exercise we begin by considering the properties of a two-dimensional toy-model, the so called Rindler space-time. This is $\mathbb{R}^2$ equipped with the following Lorentzian metric:

$$ds^2_{Rindler} = -x^2\, dt^2 + dx^2 \qquad (2.2.2)$$

which, apparently, has a singularity on the line $H \subset \mathbb{R}^2$ singled out by the equation $x = 0$. A careful analysis reveals that such a singularity is just a coordinate artefact since the metric (2.2.2) is actually flat and can be brought to the standard form of the Minkowski metric via a suitable coordinate transformation:

$$\xi : \mathbb{R}^2 \to \mathbb{R}^2 \qquad (2.2.3)$$

The relevant point is that the diffeomorphism ξ is not surjective since it maps the whole of Rindler space-time, namely the entire $\mathbb{R}^2$ manifold into an open subset $I = \xi(\mathbb{R}^2) \subset \mathbb{R}^2 = \text{Mink}_2$ of Minkowski space. This means that Rindler space-time is incomplete and can be extended to the entire 2-dimensional Minkowski space Mink_2. The other key point is that the image $\xi(H) \subset \text{Mink}_2$ of the singularity in the extended space-time is a perfectly regular null-like hypersurface. These features are completely analogous to corresponding features of the Kruskal extension of Schwarzschild space-time. Also there we can find a suitable coordinate transformation $\xi_K : \mathbb{R}^4 \to \mathbb{R}^4$ which removes the singularity displayed by the Schwarzschild metric at the Schwarzschild radius $r = 2m$ and such a map is not surjective, rather it maps the entire Schwarzschild space-time into an open sub-manifold $\xi_K(\text{Schwarzschild}) \subset \text{Krusk}$ of a larger manifold named the Kruskal space-time. Also in full analogy with the case of the Rindler toy-model the image $\xi_K(H)$ of the coordinate singularity H defined by the equation $r = 2m$ is a regular null-like hypersurface of Kruskal space-time. In this case it has the interpretation of event-horizon delimiting a black-hole region.

The basic question therefore is: how do we find the appropriate diffeomorphism ξ or ξ_K? The answer is provided by a systematic algorithm which consists of the following steps:

1. derivation of the equations for geodesics,
2. construction of a complete system of incoming and outgoing null geodesics,
3. transition to a coordinate system where the new coordinates are the affine parameters along the incoming and outgoing null geodesics,
4. analytic continuation of the new coordinate patch beyond its original domain of definition.

We begin by showing how this procedure works in the case of the metric (2.2.2) and later we apply it to the physically significant case of the Schwarzschild metric.

The metric (2.2.2) has a coordinate singularity at $x = 0$ where the determinant $\det g_{\mu\nu} = -x^2$ has a zero. In order to understand the real meaning of such a singularity we follow the programme outlined above and we write the equation for null geodesics:

$$g_{\mu\nu}(x)\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0; \qquad -x^2\left(\dot{t}^2\right) + \left(\dot{x}^2\right) = 0 \tag{2.2.4}$$

from which we immediately obtain:

$$\left(\frac{dx}{dt}\right)^2 = \frac{1}{x^2} \quad \Rightarrow \quad t = \pm\int\frac{dx}{x} = \pm\ln x + \text{const} \tag{2.2.5}$$

Hence we can introduce the null coordinates by writing:

$$\begin{aligned} &t + \ln x = v; \quad v = \text{const} \quad \Leftrightarrow \quad \text{(incoming null geodesics)} \\ &t - \ln x = u; \quad u = \text{const} \quad \Leftrightarrow \quad \text{(outgoing null geodesics)} \end{aligned} \tag{2.2.6}$$

The shape of the corresponding null geodesics is displayed in Fig. 2.4. The first change of coordinates is performed by replacing x, t by u, v. Using:

$$x^2 = \exp[v-u]; \qquad \frac{dx}{x} = \frac{dv-du}{2}; \qquad dt = \frac{dv+du}{2} \tag{2.2.7}$$

the metric (2.2.2) becomes:

$$ds^2_{Rindler} = -\exp[v-u]\,du\,dv \tag{2.2.8}$$

Next step is the calculation of the affine parameter along the null geodesics. Here we use a general property encoded in the following lemma:

Lemma 2.2.1 *Let* $\mathbf{k}$ *be a Killing vector for a given metric* $g_{\mu\nu}(x)$ *and let* $\mathbf{t} = \frac{dx^\mu}{d\lambda}$ *be the tangent vector to a geodesic. Then the scalar product*:

$$E \equiv -(\mathbf{t}, \mathbf{k}) = -g_{\mu\nu}\frac{dx^\mu}{d\lambda}k^\nu \tag{2.2.9}$$

is constant along the geodesic.

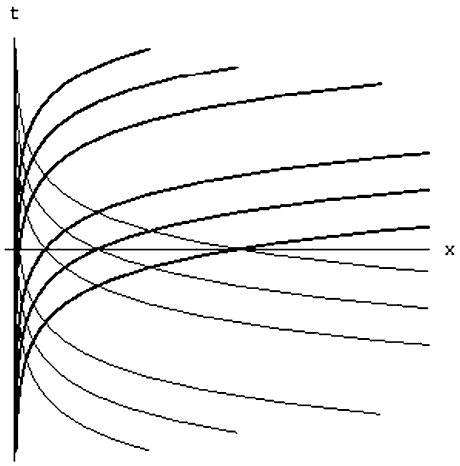

Fig. 2.4 Null geodesics of the Rindler metric. The *thin curves* are incoming ($v = \text{const}$), while the *thick* ones are outgoing ($u = \text{const}$)

Proof The proof is immediate by direct calculation. If we take the $d/d\lambda$ derivative of E we get:

$$\frac{dE}{d\lambda} = \underbrace{-\nabla_\rho g_{\mu\nu}\frac{dx^\rho}{d\lambda}\frac{dx^\mu}{d\lambda}k^\nu}_{\substack{=0 \text{ since metric} \\ \text{is cov. const.}}} \underbrace{- g_{\mu\nu}\left(\nabla_\rho \frac{dx^\mu}{d\lambda}\right)\frac{dx^\rho}{d\lambda}k^\nu}_{=0 \text{ for the geodesic eq.}}$$
$$\underbrace{- g_{\mu\nu}\nabla_\rho k^\nu, \frac{dx^\rho}{d\lambda}\frac{dx^\mu}{d\lambda}}_{=0 \text{ for the Killing vec. eq.}} \tag{2.2.10}$$

So we obtain the sum of three terms that are separately zero for three different reasons. □

Relying on Lemma 2.2.1 in Rindler space time we can conclude that $E = x^2\frac{dt}{d\lambda}$ is constant along geodesics. Indeed the vector field $\mathbf{k} \equiv \frac{d}{dt}$ is immediately seen to be a Killing vector for the metric (2.2.2). Then by means of straightforward manipulations we obtain:

$$d\lambda = \frac{1}{E}\exp[v-u]\frac{du+dv}{2} \quad \Rightarrow$$
$$\lambda = \begin{cases} \frac{\exp[-u]}{2E}\exp[v] & \text{on } u = \text{const outgoing null geodesics} \\ -\frac{\exp[v]}{2E}\exp[-u] & \text{on } v = \text{const incoming null geodesics} \end{cases} \tag{2.2.11}$$

The third step in the algorithm that leads to the extension map corresponds to a coordinate transformation where the new coordinates are proportional to the affine parameters along incoming and outgoing null geodesics. Hence in view of (2.2.11) we introduce the coordinate change:

$$U = -e^{-u} \quad \Rightarrow \quad dU = e^{-u}\,du; \qquad V = e^v \quad \Rightarrow \quad dV = e^v\,dv \tag{2.2.12}$$

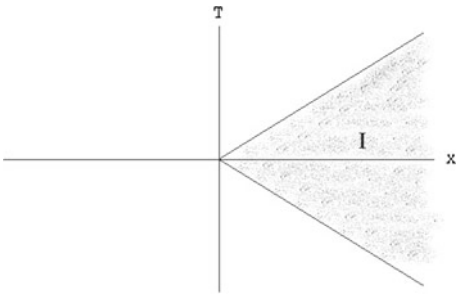

Fig. 2.5 The image of Rindler space-time in two-dimensional Minkowski space-time is the shaded region I bounded by the two null surfaces $X = T$ $(X > 0)$ and $X = -T$ $(X > 0)$. These latter are the image of the coordinate singularity $x = 0$ of the original metric

by means of which the Rindler metric (2.2.8) becomes:

$$ds^2_{Rindler} = -dU \otimes dV \tag{2.2.13}$$

Finally, with a further obvious transformation:

$$T = \frac{V + U}{2}; \qquad X = \frac{V - U}{2} \tag{2.2.14}$$

the Rindler metric (2.2.13) is reduced to the standard two-dimensional Minkowski metric in the plane $\{X, T\}$:

$$ds^2_{Rindler} = -dT^2 + dX^2 \tag{2.2.15}$$

Putting together all the steps, the coordinate transformation that reduces the Rindler metric to the standard form (2.2.15) is the following:

$$x = \sqrt{X^2 - T^2}; \qquad t = \text{arctanh}\left[\frac{T}{X}\right] \tag{2.2.16}$$

In this way we have succeeded in eliminating the apparent singularity $x = 0$ since the metric (2.2.15) is perfectly regular in the whole $\{X, T\}$ plane. The subtle point of this procedure is that by means of the transformation (2.2.12) we have not only eliminated the singularity, but also *extended* the space-time. Indeed the definition (2.2.12) of the U and V coordinates is such that V is always positive and U always negative. This means that in the $\{U, V\}$ plane the image of Rindler space-time is the quadrant $U < 0$; $V > 0$. In terms of the final X, T variables the image of the original Rindler space-time is the angular sector I depicted in Fig. 2.5. Considering the coordinate transformation (2.2.16) we see that the image in the extended space-time of the apparent singularity $x = 0$ is the locus $X^2 = T^2$ which is perfectly regular but has the distinctive feature of being *a null-like surface*. This surface is also the boundary of the image I of Rindler space-time in its maximal extension. Furthermore setting $X = \pm T$ we obtain $t = \pm\infty$. This means that in the original Rindler space any test particle takes an infinite amount of *coordinate time* to reach the boundary locus $x = 0$: this is also evident from the plot of null geodesics in Fig. 2.4. On the other hand the *proper time* taken by a test particle to reach such a locus from any other point is just finite.

All these features of our toy model apply also to the case of Schwarzschild space-time once it is extended with the same procedure. The image of the coordinate singularity $r = 2m$ will be a null-like surface, interpreted as *event horizon*, which can be reached in a *finite proper-time* but only after an *infinite interval of coordinate time*. What will be new and of utmost physical interest is precisely the interpretation of the locus $r = 2m$ as an *event horizon* $\mathfrak{H}$ which leads to the concept of Black-Hole. Yet this interpretation can be discovered only through the Kruskal extension of Schwarzschild space-time and this latter can be systematically derived via the same algorithm we have applied to the Rindler toy model.

2.2.2 Applying the Same Procedure to the Schwarzschild Metric

We are now ready to analyze the Schwarzschild metric (2.2.1) by means of the tokens illustrated above. The first step consists of reducing it to two-dimensions by fixing the angular coordinates to constant values $\theta = \theta_0$, $\phi = \phi_0$. In this way the metric (2.2.1) reduces to:

$$ds^2_{Schwarz.} = -\left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 \tag{2.2.17}$$

Next, in the reduced space spanned by the coordinates r and t we look for the null-geodesics. From the equation:

$$-\left(1 - \frac{2m}{r}\right)\dot{t}^2 + \left(1 - \frac{2m}{r}\right)^{-1} \dot{r}^2 = 0 \tag{2.2.18}$$

we obtain:

$$\frac{dt}{dr} = \pm \frac{r}{r - 2m} \quad \Rightarrow \quad t = \pm r^*(r) \tag{2.2.19}$$

where we have introduced the so called Regge-Wheeler *tortoise coordinate* defined by the following indefinite integral:

$$r^*(r) \equiv \int \frac{r}{r - 2m}\, dr = r + 2m \log\left(\frac{r}{2m} - 1\right) \tag{2.2.20}$$

Hence, in full analogy with (2.2.6), we can introduce the null coordinates

$$\begin{aligned} &t + r^*(r) = v; \quad v = \text{const} \quad \Leftrightarrow \quad \text{(incoming null geodesics)} \\ &t - r^*(r) = u; \quad u = \text{const} \quad \Leftrightarrow \quad \text{(outgoing null geodesics)} \end{aligned} \tag{2.2.21}$$

and the analogue of Fig. 2.4 is now given by Fig. 2.6. Inspection of this picture reveals the same properties we had already observed in the case of the Rindler toy model. What is important to stress in the present model is that each point of the

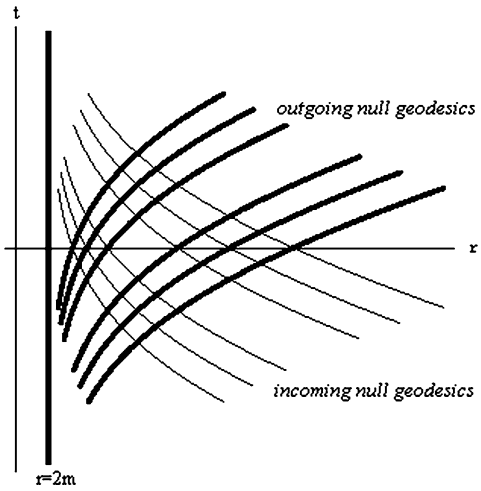

Fig. 2.6 Null geodesics of the Schwarzschild metric in the r, t plane. The *thin curves* are incoming ($v = \text{const}$), while the *thick* ones are outgoing ($u = \text{const}$). Each point in this picture represents a 2-sphere, parameterized by the angles θ_0 and ϕ_0. The *thick vertical line* is the surface $r = r_S = 2m$ corresponding to the coordinate singularity. As in the case of the Rindler toy model the null–geodesics incoming from infinity reach the coordinate singularity only at asymptotically late times $t \to > +\infty$. Similarly outgoing null-geodesics were on this surface only at asymptotically early times $t \to -\infty$

diagram actually represents a 2-sphere parameterized by the two angles θ and ϕ that we have freezed at the constant values θ_0 and ϕ_0. Since we cannot make four-dimensional drawings some pictorial idea of what is going on can be obtained by replacing the 2-sphere with a circle $\mathbb{S}^1$ parameterized by the azimuthal angle ϕ. In this way we obtain a three-dimensional space-time spanned by coordinates t, $x = r\cos\phi$, $y = r\sin\phi$. In this space the null-geodesics of Fig. 2.6 become two-dimensional surfaces. Indeed these null-surfaces are nothing else but the projections $\theta = \theta_0 = \pi/2$ of the true null surfaces of the Schwarzschild metric. In Fig. 2.7 we present two examples of such projected null surfaces, one incoming and one outgoing.

Having found the system of incoming and outgoing null-geodesics we go over to point (iii) of our programme and we make a coordinate change from t, r to u, v. By straightforward differentiation of (2.2.20), (2.2.21) we obtain:

$$dr = -\left(1 - \frac{r_S}{r}\right)\frac{du - dv}{2}; \qquad dt = \frac{du + dv}{2} \tag{2.2.22}$$

so that the reduced Schwarzschild metric (2.2.17) becomes:

$$ds^2_{Schwarz.} = -\left(1 - \frac{r_S}{r}\right) du \otimes dv \tag{2.2.23}$$

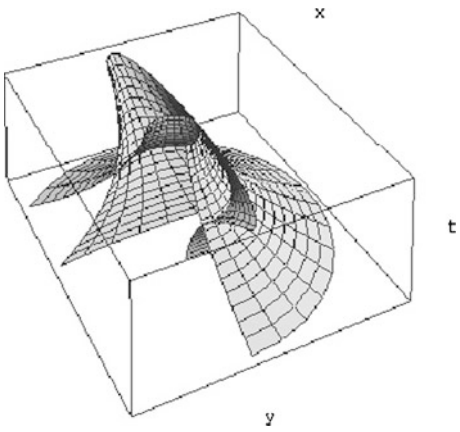

Fig. 2.7 An example of two null surfaces generated by null geodesics of the Schwarzschild metric in the r, t plane

Using the definition (2.2.20) of the tortoise coordinate we can also write:

$$\left(1-\frac{r_S}{r}\right) = -\exp\left[\frac{v-u}{2r_S}\right]\exp\left[-\frac{r}{r_S}\right] \tag{2.2.24}$$

which combined with (2.2.22) yields:

$$ds^2_{Schwarz.} = \exp\left[-\frac{r}{r_S}\right]\exp\left[\frac{v-u}{2r_S}\right]\frac{r_S}{r}\,du\otimes dv \tag{2.2.25}$$

In complete analogy with (2.2.12) we can now introduce the new coordinates:

$$U = -\exp\left[-\frac{u}{2r_S}\right]; \qquad V = \exp\left[-\frac{u}{2r_S}\right] \tag{2.2.26}$$

that play the role of affine parameters along the incoming and outgoing null geodesics.

Then by straightforward differentiation of (2.2.26) the reduced Schwarzschild metric (2.2.25) becomes:

$$ds^2_{Schwarz.} = -4\frac{r_S^3}{r}\exp\left[-\frac{r}{r_S}\right]dU\otimes dV \tag{2.2.27}$$

where the variable $r = r(U, V)$ is the function of the independent coordinates U, V implicitly determined by the transcendental equation:

$$r + r_S\log\left(\frac{r}{r_S}-1\right) = r_S\log(-UV) \tag{2.2.28}$$

In analogy with our treatment of the Rindler toy model we can make a final coordinate change to new variables X, T related to U, V as in (2.2.14). These, together

with the angular variables θ, ϕ make up the Kruskal coordinate patch which, putting together all the intermediate steps, is related to the original coordinate patch t, r, θ, ϕ by the following transition function:

$$\begin{array}{l}\text{polar}\\ \text{versus}\\ \text{Kruskal}\\ \text{coord.}\end{array}\quad \begin{cases}\theta = \theta\\ \phi = \phi\\ (\frac{r}{r_S} - 1)\exp[\frac{r}{r_S}] = T^2 - X^2\\ \frac{t}{r_S} = \log(\frac{T+X}{T-X}) \equiv 2\,\text{arctanh}\,\frac{X}{T}\end{cases} \tag{2.2.29}$$

In Kruskal coordinates the Schwarzschild metric (2.2.1) takes the final form:

$$ds^2_{\text{Krusk}} = 4\frac{r_S^3}{r}\exp\left[\frac{r}{r_S}\right]\left(-dT^2 + dX^2\right) + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \tag{2.2.30}$$

where the $r = r(X, T)$ is implicitly determined in terms of X, T by the transcendental equations (2.2.29).

2.2.3 A First Analysis of Kruskal Space-Time

Let us now consider the general properties of the space-time $(\mathscr{M}_{\text{Krusk}}, g_{\text{Krusk}})$ identified by the metric (2.2.30) and by the implicit definition of the variable r contained in (2.2.29). This analysis is best done by inspection of the two-dimensional diagram displayed in Fig. 2.8. This diagram lies in the plane $\{X, T\}$, each of whose points represents a two sphere spanned by the angle-coordinates θ and ϕ. The first thing to remark concerns the physical range of the coordinates X, T. The Kruskal manifold $\mathscr{M}_{\text{Krusk}}$ does not coincide with the entire plane, rather it is the infinite portion of the latter comprised between the two branches of the hyperbolic locus:

$$T^2 - X^2 = -1 \tag{2.2.31}$$

This is the image in the X, T-plane of the $r = 0$ locus which is a genuine singularity of both the original Schwarzschild metric and of its Kruskal extension. Indeed from (5.9.6)–(5.9.11) of Volume 1 we know that the intrinsic components of the curvature tensor depend only on r and are singular at $r = 0$, while they are perfectly regular at $r = 2m$. Therefore no geodesic can be extended in the X, T plane beyond (2.2.31) which constitutes a boundary of the manifold.

Let us now consider the image of the constant r surfaces. Here we have to distinguish two cases: $r > r_S$ or $r < r_S$. We obtain:

$$\begin{aligned}&\{X, T\} = \{h\cosh p, h\sinh p\}; \qquad h = e^{\frac{r}{r_S}}\sqrt{\frac{r}{r_S} - 1} \quad \text{for } r > r_S\\ &\{X, T\} = \{h\sinh p, h\cosh p\}; \qquad h = e^{\frac{r}{r_S}}\sqrt{1 - \frac{r}{r_S}} \quad \text{for } r < r_S\end{aligned} \tag{2.2.32}$$

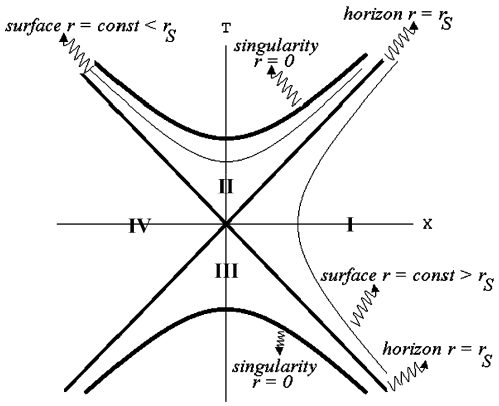

Fig. 2.8 A two-dimensional diagram of Kruskal space-time

These are the hyperbolae drawn in Fig. 2.8. Calculating the normal vector $N^\mu = \{\partial_p T, \partial_p X, 0, 0\}$ to these surfaces, we find that it is time-like $N^\mu N^\nu g_{\mu\nu} < 0$ for $r > r_S$ and space-like $N^\mu N^\nu g_{\mu\nu} > 0$ for $r < r_S$. Correspondingly, according to a discussion developed in the next section, the constant r surfaces are space-like outside the sphere of radius r_S and time-like inside it. The dividing locus is the pair of straight lines $X = \pm T$ which correspond to $r = r_S$ and constitute a null-surface, namely one whose normal vector is light-like. This null-surface is the *event horizon*, a concept whose precise definition needs, in order to be formulated, a careful reconsideration of the notions of Future, Past and Causality in the context of General Relativity. The next two sections pursue such a goal and by their end we will be able to define Black-Holes and their Horizons. Here we note the following. If we solve the geodesic equation for time-like or null-like geodesics with arbitrary initial data inside region II of Fig. 2.8 then the end point of that geodesic is always located on the singular locus $T^2 - X^2 = -1$ and the whole development of the curve occurs inside region II. The formal proof of this statement is involved and it will be overcome by the methods of Sects. 2.3 and 2.4. Yet there is an intuitive argument which provides the correct answer and suffices to clarify the situation. Disregarding the angular variables θ and ϕ the Kruskal metric (2.2.30) reduces to:

$$ds^2_{\text{Krusk}} = F(X,T)\big(-dT^2 + dX^2\big); \quad F(X,T) = 4\frac{r_S^3}{r}\exp\left[\frac{r}{r_S}\right] \tag{2.2.33}$$

so that it is proportional to two-dimensional Minkowski metric $ds^2_{\text{Mink}} = -dT^2 + dX^2$ through the positive definite function $F(X,T)$. In the language of Sect. 2.4 this fact means that, reduced to two-dimensions, Kruskal and Minkowski metrics are conformally equivalent. According to Lemma 2.4.1 proved later on, conformally equivalent metrics share the same light-like geodesics, although the time-like and space-like ones may be different. This means that in two-dimensional Kruskal space-time light travels along straight lines of the form $X = \pm T + k$ where k is some constant. This is the same statement as saying that at any point p of the $\{X, T\}$ plane the tangent vector to any curve is time-like or light-like and oriented to the future if

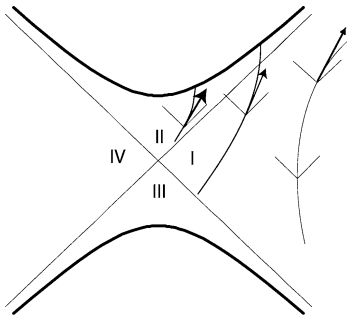

Fig. 2.9 The light-cone orientations in Kruskal space-time and the difference between physical geodesics in regions I and II

its inclination α with respect to the X axis is in the following range $3\pi/4 \geq \alpha \geq \pi/4$. This applies to the whole plane, yet it implies a fundamental difference in the destiny of physical particles that start their journey in region I (or IV) of the Kruskal plane, with respect to the destiny of those ones that happen to be in region II at some point of their life. As it is visually evident from Fig. 2.9, in region I we can have curves (and in particular geodesics) whose tangent vector is time-like and future oriented at any of their points which nonetheless avoid the singular locus and escape to infinity. In the same region there are also future oriented time-like curves which cross the horizon $X = \pm T$ and end up on the singular locus, yet these are not the only ones, as already remarked. On the contrary all curves that at some point happen to be inside region II can no longer escape to infinity since, in order to be able to do so, their tangent vector should be space-like, at least at some of their points. Hence the horizon can be crossed from region I to region II, never in the opposite direction. This leads to the existence of a Black-Hole, namely a space-time region, (II in our case) where gravity is so strong that not even light can escape from it. No signal from region II can reach a distant observer located in region I who therefore perceives only the presence of the gravitational field of the black hole swapping infalling matter.

To encode the ideas intuitively described in this section into a rigorous mathematical framework we proceed next to implement our already announced programme. This is the critical review of the concepts of Future, Past and Causality within General Relativity, namely when we assume that all physical events are points p in a pseudo-Riemannian manifold $(\mathscr{M}, g)$ with a Lorentzian signature.

2.3 Basic Concepts about Future, Past and Causality

Our discussion starts by reviewing the basic properties of the light-cone (see Fig. 2.10). In Special Relativity, where space-time is Minkowski-space, namely a pseudo-Riemannian manifold which is also affine, the light cone has a global meaning, while in General Relativity light-cones can be defined only locally, namely at each point $p \in \mathscr{M}$. In any case the Lorentzian signature of the metric implies that $\forall p \in \mathscr{M}$, the tangent space $T_p\mathscr{M}$ is isomorphic to Minkowski space and it admits the same decomposition in time-like, null-like and space-like sub-manifolds. Hence

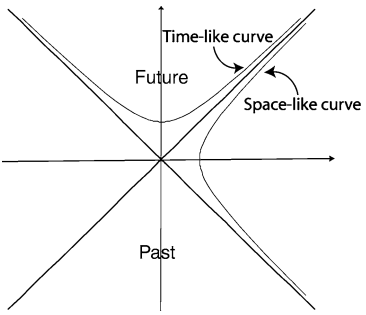

Fig. 2.10 The structure of the light-cone

the analysis of the light-cone properties has a general meaning also in General Relativity, although such analysis needs to be repeated at each point. All the complexities inherent with the notion of global causality arise from the need of gluing together the locally defined light-cones. We will develop appropriate conceptual tools to manage such a gluing after our review of the local light-cone properties.

2.3.1 The Light-Cone

When a metric has a Lorentzian signature, vectors t can be of three-types:

1. Time-like, if $(t,t) < 0$ in mostly plus convention for $g_{\mu\nu}$.
2. Space-like, if $(t,t) > 0$ in mostly plus convention for $g_{\mu\nu}$.
3. Null-like, if $(t,t) = 0$ both in mostly plus and mostly minus convention for $g_{\mu\nu}$.

At any point $p \in \mathscr{M}$ the light-cone $\mathscr{C}_p$ is composed by the set of vectors $t \in T_p\mathscr{M}$ which are either time-like or null-like. In order to study the properties of the light-cones it is convenient to review a few elementary but basic properties of vectors in Minkowski space.

Theorem 2.3.1 *All vectors orthogonal to a time-like vector are space-like.*

Proof Using a mostly plus signature, we can go to a diagonal basis such that:

$$g(X,Y) = g_{00}X^0Y^0 + (\mathbf{X},\mathbf{Y}) \tag{2.3.1}$$

where $g_{00} < 0$ and $(\,,\,)$ denotes a non-degenerate, positive-definite, Euclidian bilinear form on $\mathbb{R}^{n-1}$. In this basis, if $X \perp T$ and T is time-like we have:

$$\begin{aligned} -g_{00}T^0T^0 &> (\mathbf{T},\mathbf{T}) \\ -g_{00}T^0X^0 &= (\mathbf{T},\mathbf{X}) \leq \sqrt{(\mathbf{T},\mathbf{T})(\mathbf{X},\mathbf{X})} \end{aligned} \tag{2.3.2}$$

Then we get:

$$\frac{-g_{00}T^0X^0}{\sqrt{-g_{00}T^0T^0}} < \frac{(\mathbf{T},\mathbf{X})}{\sqrt{(\mathbf{T},\mathbf{T})}} \leq \sqrt{(\mathbf{X},\mathbf{X})} \tag{2.3.3}$$

Squaring all terms in (2.3.3) we obtain

$$-g_{00}X^0X^0 < (\mathbf{X},\mathbf{X}) \quad \Rightarrow \quad g(X,X) > 0 \tag{2.3.4}$$

namely the four-vector X is space-like as asserted by the theorem. □

Another useful property is given by the following

Lemma 2.3.1 *The sum of two future-directed time-like vectors is a future-directed time-like vector.*

Proof Let t and T be the two vectors under considerations. By hypothesis we have

$$\begin{aligned} g(t,t) &< 0; \qquad t^0 > 0 \\ g(T,T) &< 0; \qquad T^0 > 0 \end{aligned} \tag{2.3.5}$$

Since:

$$\begin{aligned} \sqrt{-g_{00}}\,t^0 &> (\mathbf{t},\mathbf{t}) \\ \sqrt{-g_{00}}\,T^0 &> (\mathbf{T},\mathbf{T}) \\ \sqrt{-g_{00}}\,t^0T^0 &> \sqrt{(\mathbf{t},\mathbf{t})(\mathbf{T},\mathbf{T})} > (\mathbf{t},\mathbf{T}) \end{aligned} \tag{2.3.6}$$

we have:

$$\begin{gathered} g(t+T,t+T) = g(t,t) + g(T,T) + 2g(t,T) \\ \Downarrow \\ -g_{00}\big((t^0)^2 + (T^0)^2 + 2t^0T^0\big) > (\mathbf{t},\mathbf{t}) + (\mathbf{T},\mathbf{T}) + 2(\mathbf{t},\mathbf{T}) \end{gathered} \tag{2.3.7}$$

which proves that $t+T$ is time-like. Moreover $t^0 + T^0 > 0$ and so the sum vector is also future-directed as advocated by the lemma. □

On the other hand with obvious changes in the proof of Theorem 2.3.1 the following lemma is established

Lemma 2.3.2 *All vectors X, orthogonal to a light-like vector L are either light-like or space-like.*

Let us now consider in the manifold $(\mathscr{M}, g)$ surfaces Σ defined by the vanishing of some smooth function of the local coordinates:

$$p \in \Sigma \quad \Leftrightarrow \quad f(p) = 0 \quad \text{where } f \in \mathbb{C}^\infty(\mathscr{M}) \tag{2.3.8}$$

By definition the normal vector to the surface Σ is the gradient of the function f:

$$n_\mu^{(\Sigma)} = \nabla_\mu f = \partial_\mu f \tag{2.3.9}$$

Indeed any tangent vector to the surface is by construction orthogonal to $n^{(\Sigma)}$:

$$g\left(t^{(\Sigma)}, n^{(\Sigma)}\right) = 0 \tag{2.3.10}$$

Definition 2.3.1 A surface Σ is said to be *space-like* if its normal vector $n^{(\Sigma)}$ is everywhere *time-like* on the surface. Conversely Σ is *time-like* if $n^{(\Sigma)}$ is **space-like**. We name *null surfaces* those Σ whose normal vector $n^{(\Sigma)}$ is *null-like*.

Null surfaces have very intriguing properties. First of all their normal vector is also tangent to the surface. This follows from the fact that the normal vector is orthogonal to itself. Furthermore we can prove that any null-surface is generated by null-geodesics. Indeed we can easily prove that the normal vector $n^{(\Sigma)}$ to a null surface is the tangent vector to a null-geodesics. Indeed we have:

$$\begin{aligned} 0 &= \nabla_\mu\left(\nabla_\nu f \nabla^\nu f\right) = 2\nabla^\nu f \nabla_\nu \nabla_\mu f \\ &= n^\nu \nabla_\nu n_\mu \end{aligned} \tag{2.3.11}$$

and the last equality is precisely the geodesic equation satisfied by the integral curve to the normal vector $n^{(\Sigma)}$.

A typical null-surface is the event-horizon of a black-hole.

2.3.2 Future and Past of Events and Regions

Let us now consider the pseudo-Riemannian space-time manifold $(\mathscr{M}, g)$ and at each point $p \in \mathscr{M}$ introduce the local light-cone $\mathscr{C}_p \subset T_p\mathscr{M}$. In this section we find it convenient to change convention and use a mostly minus signature where $g_{00} > 0$.

Definition 2.3.2 The local light-cone $\mathscr{C}_p$ (see Fig. 2.11) is the set of all tangent vectors $t \in T_p\mathscr{M}$, such that:

$$g_{\mu\nu} t^\mu t^\nu \geq 0 \tag{2.3.12}$$

and it is the union of the future light-cone with the past light-cone:

$$\mathscr{C}_p = \mathscr{C}_p^+ \bigcup \mathscr{C}_p^- \tag{2.3.13}$$

where

$$\begin{aligned} t \in \mathscr{C}_p^+ \quad &\Leftrightarrow \quad g(t,t) \geq 0 \quad \text{and} \quad t^0 > 0 \\ t \in \mathscr{C}_p^- \quad &\Leftrightarrow \quad g(t,t) \geq 0 \quad \text{and} \quad t^0 < 0 \end{aligned} \tag{2.3.14}$$

The vectors in $\mathscr{C}_p^+$ are named *future-directed*, while those in $\mathscr{C}_p^-$ are named *past-directed*.

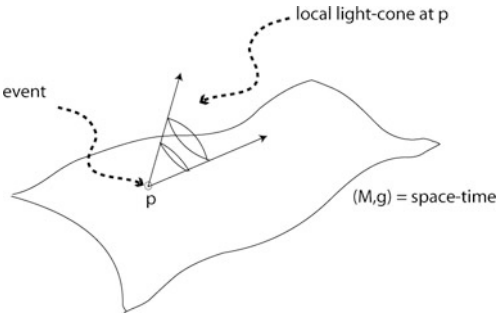

Fig. 2.11 At each point of the space-time manifold, the tangent space $T_p\mathscr{M}$ contains the sub-manifold $\mathscr{C}_p$ of time-like and null-vectors which constitutes the local light-cone

We can now transfer the notions of time orientation from vectors to curves by means of the following definitions:

Definition 2.3.3 A differentiable curve $\lambda(s)$ on the space-time manifold $\mathscr{M}$ is named a *future-directed time-like curve* if at each point $p \in \lambda$, the tangent vector to the curve t^μ is future-directed and time-like. Conversely $\lambda(s)$ is *past-directed time-like* if such is t^μ.

Similarly we have:

Definition 2.3.4 A differentiable curve $\lambda(s)$ on the space-time manifold $\mathscr{M}$ is named a *future-directed causal curve* if at each point $p \in \lambda$, the tangent vector to the curve t^μ is either a future-directed *time-like* or a future-directed *null-like* vector. Conversely $\lambda(s)$ is a *past-directed causal* curve when the tangent t^μ, time-like or null-like, is past directed.

Relying on these concepts we can introduce the notions of Chronological Future and Past of a point $p \in \mathscr{M}$.

Definition 2.3.5 The Chronological Future (Past) of a point p, denoted $I^{\pm}(p)$ is the subset of points of $\mathscr{M}$, defined by the following condition:

$$I^{\pm}(p) = \left\{ q \in \mathscr{M} \;\middle|\; \begin{array}{l} \exists \text{ future- (past-)directed time-like} \\ \quad \text{curve } \lambda(s) \text{ such that} \\ \quad \lambda(0) = p; \quad \lambda(1) = q \end{array} \right\} \tag{2.3.15}$$

In other words the Chronological Future or Past of an event are all those events that can be connected to it by a future-directed or past-directed time-like curve.

Let now $S \subset \mathscr{M}$ be a region of space-time, namely a continuous sub-manifold of the space-time manifold.

Definition 2.3.6 The Chronological Future (Past) of the region S, denoted $I^{\pm}(S)$ is defined as follows:

$$I^{\pm}(S) = \bigcup_{p \in S} I^{\pm}(p) \tag{2.3.16}$$

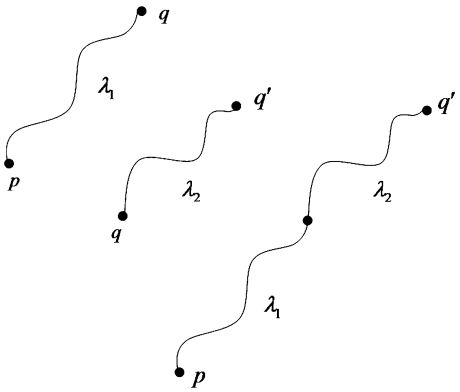

Fig. 2.12 The union of two time-like future-directed curves is still a time-like future directed curve

An elementary property of the Chronological Future is the following:

$$I^{\pm}\big(I^{\pm}(S)\big) = I^{\pm}(S) \tag{2.3.17}$$

The proof is illustrated in Fig. 2.12.

If $q' \in I^{\pm}(I^{\pm}(S))$ then, by definition, there exists at least one point $q \in I^{\pm}(S)$ to which q' is connected by a time-like future directed curve $\lambda_2(s)$. On the other hand, once again by definition, q is connected by a future-directed time-like curve $\lambda_1(s)$ to at least one point $p \in S$. Joining λ_1 with λ_2 we obtain a future-directed time-like curve that connects q' to p, which implies that $q \in I^{+}(S)$.

In a similar way, if $\overline{S}$ denotes the closure, in the topological sense, of the region S, we prove that:

$$I^{+}(\overline{S}) = I^{+}(S) \tag{2.3.18}$$

In perfect analogy with Definition 2.3.5 we have:

Definition 2.3.7 The Causal Future (Past) of a point p, denoted $J^{\pm}(p)$ is the subset of points of $\mathscr{M}$, defined by the following condition:

$$J^{\pm}(p) = \left\{ q \in \mathscr{M} \;\middle|\; \begin{array}{l} \exists \text{ future- (past-)directed causal} \\ \quad \text{curve } \lambda(s) \text{ such that} \\ \quad \lambda(0) = p; \quad \lambda(1) = q \end{array} \right\} \tag{2.3.19}$$

and the Causal Future(Past) of a region S, denoted $J^{\pm}(S)$ is:

$$J^{\pm}(S) = \bigcup_{p \in S} J^{\pm}(p) \tag{2.3.20}$$

An important point which we mention without proof is the following. In flat Minkowski space $J^{\pm}(p)$ is always a closed set in the topological sense, namely it contains its own boundary. In general curved space-times $J^{\pm}(p)$ can fail to be closed.

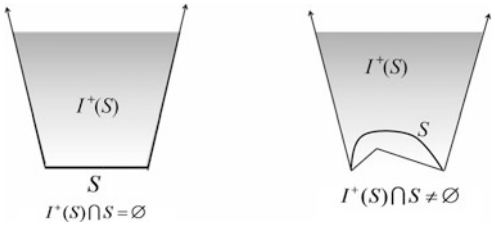

Fig. 2.13 In two-dimensional Minkowski space we show an example of achronal set. In the picture on *left* the segment S parallel to the space axis is achronal because it does not intersect its chronological future. On the other hand, in the picture on the *right*, the line S, although one dimensional is not achronal because it intersects its own chronological future

Achronal Sets

Definition 2.3.8 Let $S \subset \mathscr{M}$ be a region of space-time. S is said to be *achronal* if and only if

$$I^+(S) \bigcap S = \emptyset \tag{2.3.21}$$

The relevance of achronal sets resides in the following. When considering classical or quantum fields $\phi(x)$, conditions on these latter specified on an achronal set S are consistent, since all the events in S do not bear causal relations to each other. On the other hand one cannot freely specify initial conditions for fields on regions that are not achronal because their points are causally related to each other. In Fig. 2.13 we illustrate an example and a counterexample of achronal sets in two-dimensional Minkowski space.

Time-Orientability We mentioned above the splitting of the local light-cones in the future $\mathscr{C}_p^+$ and past $\mathscr{C}_p^-$ cones. Clearly, just as all the tangent spaces are glued together to make a fibre-bundle, the same is true of the local light-cones. The subtle point concerns the nature of the transition functions. Those of the tangent bundle $T\mathscr{M} \to \mathscr{M}$ to an n-dimensional manifold take values in $\mathrm{GL}(n, \mathbb{R})$. The light-cone, on the other hand, is left-invariant only by the subgroup $\mathrm{O}(1, n-1) \subset \mathrm{GL}(n, \mathbb{R})$. Furthermore the past and future cones are left invariant only by the subgroup of the former connected with the identity, namely $\mathrm{SO}(1, n-1) \subset \mathrm{O}(1, n-1)$. Hence the tipping of the light-cones from one point to the other of the space-time manifold are described by those transition functions of the tangent bundle that take values in the cosets $\mathrm{GL}(n, \mathbb{R})/\mathrm{O}(1, n-1)$ or $\mathrm{GL}(n, \mathbb{R})/\mathrm{SO}(1, n-1)$. The difference is subtle. Let $\mathrm{H_p} \subset \mathrm{GL}(n, \mathbb{R})$ be the subgroup isomorphic to $\mathrm{SO}(1, n-1)$, which leaves invariant the future and past light-cones at $p \in \mathscr{M}$ and let $\mathrm{H_q} \subset \mathrm{GL}(n, \mathbb{R})$ be the subgroup, also isomorphic to $\mathrm{SO}(1, n-1)$, which leaves invariant the future and past light cones at the point $q \in \mathscr{M}$. The question is the following. Are $\mathrm{H_p}$ and $\mathrm{H_q}$ conjugate to each other under the transition function $g(p, q) \in \mathrm{GL}(n, \mathbb{R})$ of the tangent bundle, that connects the tangent plane at p with the tangent plane at q, namely is it true that $\mathrm{H_q} = g(p, q)\mathrm{H_p}g^{-1}(p, q)$? If the answer is yes for all pair of points

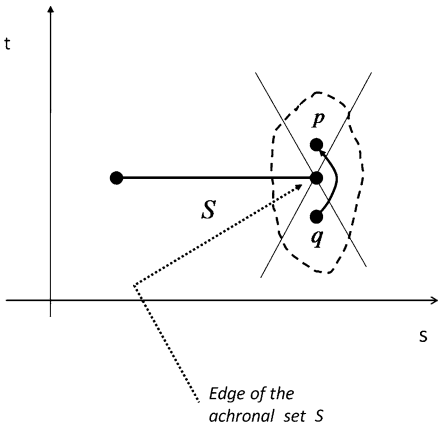

Fig. 2.14 The edge of an achronal set in two-dimensional Minkowski space. Notwithstanding how small can be the neighborhood $\mathcal{O}$ of the end point of the segment S, which we singled out with the dashed line, it contains a pair of points q and p, the former in the past of the end-point, the latter in its future, which can be connected by a time-like curve getting around the segment S and not intersecting it. Clearly this property does not hold for any of the interior points of the segment

p, q in $\mathcal{M}$, then the manifold $(\mathcal{M}, g)$ is said to be **time-orientable**. In this case the definition of future and past orientations varies continuously from one point to the other of the manifold without singular jumps. Yet there exist cases where the answer is no. When this happens the corresponding manifold is not time-orientable and all global notions of causality loose their meaning. In all the sequel we assume time-orientability.

For time orientable space-times we have the following theorem that we mention without proof

Theorem 2.3.2 *Let $(\mathcal{M}, g)$ be time-orientable and let $S \subset \mathcal{M}$ be a continuous connected region. The boundary of the chronological future of S, denoted $\partial I^+(S)$ is an achronal $(n-1)$-dimensional sub-manifold.*

Domains of Dependence The future domains of dependence are those sub-manifolds of space-time which are completely causally determined by what happens on a certain achronal set S. Alternatively the past domains of dependence are those that completely causally determine what happens on S. To discuss them we begin by introducing one more concept, that of edge.

Definition 2.3.9 Let S be an achronal and closed set. We define edge of S the set of points $a \in S$ such that for all open neighborhoods $\mathcal{O}_a$ of a, there exists two points $q \in I^-(a)$ and $p \in I^+(a)$ both contained in $\mathcal{O}_a$ which are connected by at least one time-like curve that does not intersect S.

The definition of edge is illustrated in Fig. 2.14. A very important theorem that once again we mention without proof is the following:

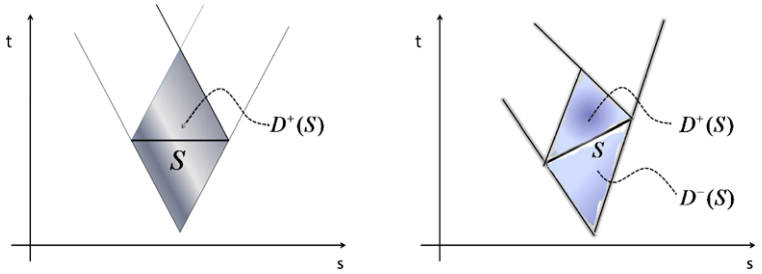

Fig. 2.15 Two examples of Future and Past domains of dependence for an achronal region S of two-dimensional Minkowski space

Theorem 2.3.3 *Let $S \subset \mathscr{M}$ be an achronal closed region of a time-orientable n-dimensional space-time $(\mathscr{M}, g)$ with Lorentz signature. Let us assume that* $\text{edge}(S) = \emptyset$. *Then S is an $(n-1)$-dimensional sub-manifold of $\mathscr{M}$.*

The relevance of this theorem resides in that it establishes the appropriate notion of places in space-time, where one can formulate initial conditions for the time development. These are achronal sets without an edge and, as intuitively expected, they correspond to the notion of space ($(n-1)$-dimensional sub-manifolds) as opposed to time.

These ideas are made more precise introducing the appropriate mathematical definitions of domains of dependence.

Definition 2.3.10 Let S be a closed achronal set. We define the Future (Past) Domain of Dependence of S, denoted $D^{\pm}(S)$ as follows:

$$D^{\pm}(S) = \left\{ p \in \mathscr{M} \;\middle|\; \begin{array}{l} \text{every past- (future-)directed time-like} \\ \text{curve through } p \text{ intersects } S \end{array} \right\} \tag{2.3.22}$$

The above definition is illustrated in Fig. 2.15. The meaning of $D^{\pm}(S)$ was already outlined above. What happens in the points $p \in D^{+}(S)$ is completely determined by the knowledge of what happened in S. Conversely what happened in S is completely determined by the knowledge of what happened in all points of $p \in D^{-}(S)$.

The *Complete Domain of Dependence* of the achronal set S is defined below:

$$D(S) \equiv D^{+}(S) \bigcup D^{-}(S) \tag{2.3.23}$$

All the introduced definitions were preparatory for the appropriate formulation of the main concept, that of Cauchy surface.

Cauchy surfaces

Definition 2.3.11 A closed achronal set $\Sigma \subset \mathscr{M}$ of a Lorentzian space-time manifold $(\mathscr{M}, g)$ is named a Cauchy surface if and only if its domain of dependence

coincides with the entire space-time, as follows:

$$D(\Sigma) = \mathscr{M} \tag{2.3.24}$$

A Cauchy surface is without edge by definition. Hence it is an $(n-1)$-dimensional hypersurface. If a Cauchy surface Σ exists, data on Σ completely determine their future development in time. This is true for all fields lying on $\mathscr{M}$ but also for the metric. Knowing for instance the perturbations of the metric on a Cauchy surface we can calculate (analytically or numerically) their future evolution without ambiguity.

Definition 2.3.12 A Lorentzian space-time $(\mathscr{M}, g)$ is named *Globally Hyperbolic* if and only if it admits at least one Cauchy surface.

Globally Hyperbolic space-times are the good, non-patological solutions of Einstein equations which allow a consistent and global formulation of causality. A major problem of General Relativity is to pose appropriate conditions on matter fields such that Global Hyperbolicity of the metric is selected. Unified theories should possess such a property.

2.4 Conformal Mappings and the Causal Boundary of Space-Time

Given the appropriate definitions of Future and Past discussed in the previous section, in order to study the causal structure of a given space-time $(\mathscr{M}, g)$, one has to cope with a classical problem met in the theory of analytic functions, namely that of bringing the point at infinity to a finite distance. Only in this way the behavior at infinity can be mastered and understood. Behavior of what? This is the obvious question. In complex function theory the behavior under investigation is that of functions, in our case is that of geodesics or, more generally, of causal curves. These latter are those that can be traveled by physical particles and the issue of causality is precisely the question of who can be reached by what. Infinity plays a distinguished role in this game because of an intuitively simple feature that characterizes those systems which the space-times $(\mathscr{M}, g)$ under consideration here are supposed to describe. The feature alluded above corresponds to the concept of an isolated dynamical system. A massive star, planetary system or galaxy is, in any case, a finite amount of energy concentrated in a finite region which is separated from other similar regions by extremely large spatial distances. The basic idea of General Relativity foresees that space-time is curved by the presence of energy or matter so that, far away from concentrations of the latter, the metric should become the flat one of empty Minkowski space. This was the boundary condition utilized in the solution of Einstein equations which lead to the Schwarzschild metric and it is the generic one assumed whenever we use Einstein equations to describe any type of star or of other localized energy lumps. Mathematically, the property of $(\mathscr{M}, g)$

which encodes such a physical idea is named *asymptotic flatness*. The point at infinity corresponds to the regions of the considered space-time $(\mathscr{M}, g)$ where the metric g becomes indistinguishable from the Minkowski metric g_{Mink} and, by hypothesis, these are at very large distances from the center of gravitation. We would like to study the structure of such an asymptotic boundary and its causal relations with the finite distance space-time regions. Before proceeding in this direction it is mandatory to stress that asymptotic flatness is neither present nor required in other physical contexts, notably that of cosmology. When we apply General Relativity to the description of the Universe and of its Evolution, energy is not localized rather it is overall distributed. There is no asymptotically far empty region and most of what we discuss here has to be revised.

This being clarified let us come back to the posed problem. Assuming that a flat boundary at infinity exists how can we bring it to a finite distance and study its structure? The answer is suggested by the analogy with the theory of analytic functions we already anticipated and it is provided by the notion of *conformal transformations*. In the complex plane, conformal transformations change distances but preserve angles. In the same way the conformal transformations we want to consider here are allowed to change the metric, that is the instrument to calculate distances, yet they should preserve the causal structure. In plain words this means that time-like, space-like and null-like vector fields should be mapped into vector fields with the same properties. Under these conditions causal curves are mapped into causal curves, although geodesics are not necessarily mapped into geodesics. Shortening the distances, infinity can come close enough to be inspected.

We begin by presenting an explicit instance of such conformal transformations corresponding to a specifically relevant case, namely that of Minkowski space. From the analysis of this example we will extract the general rules of the game to be applied also to the other cases.

2.4.1 Conformal Mapping of Minkowski Space into the Einstein Static Universe

Let us consider flat Minkowski metric in polar coordinates:

$$ds^2_{\text{Mink}} = -dt^2 + dr^2 + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \tag{2.4.1}$$

and let us perform the following change of coordinates:

$$t + r = \tan\left[\frac{T+R}{2}\right] \tag{2.4.2}$$

$$t - r = \tan\left[\frac{T-R}{2}\right] \tag{2.4.3}$$

$$\theta = \theta \tag{2.4.4}$$

$$\phi = \phi \tag{2.4.5}$$

where T, R are the new coordinates replacing t, r. By means of straightforward calculations we find that in the new variables the flat metric becomes:

$$ds^2_{\text{Mink}} = \Omega^{-2}(T, R)\, ds^2_{\text{ESU}} \tag{2.4.6}$$

$$ds^2_{\text{ESU}} = -dT^2 + dR^2 + \sin^2 R\big(d\theta^2 + \sin^2\theta\, d\phi^2\big) \tag{2.4.7}$$

$$\Omega(T, R) = \frac{1}{2}\cos\left[\frac{T+R}{2}\right]\cos\left[\frac{T+R}{2}\right] \tag{2.4.8}$$

This apparently trivial calculation leads to many important conclusions.

First of all let us observe that, considered in its own right, the metric ds^2_{ESU}, named the *Einstein Static Universe*, is the natural metric on a manifold $\mathbb{R}\times\mathbb{S}^3$. To see this it suffices to note that because of its appearance as argument of a sine, the variable R is an angle, furthermore, parameterizing the points of a three-sphere:

$$1 = X_1^2 + X_2^2 + X_3^2 + X_4^2 \tag{2.4.9}$$

as follows:

$$\begin{aligned} X_1 &= \cos R \\ X_2 &= \sin R\cos\theta \\ X_3 &= \sin R\sin\theta\cos\phi \\ X_4 &= \sin R\sin\theta\sin\phi \end{aligned} \tag{2.4.10}$$

another straightforward calculation reveals that:

$$\sum_{i=1}^{4} dX_i^2 = dR^2 + \sin^2 R\big(d\theta^2 + \sin^2\theta\, d\phi^2\big) \tag{2.4.11}$$

This demonstrates that $ds^2_{\text{ESU}} = -dT^2 + ds^2_{\mathbb{S}^3}$. The metric ds^2_{ESU} receives the name of Einstein Static Universe for the following reason. It is just an instance of a family of metrics, which we will consider in later chapters while studying cosmology, that are of the following type:

$$ds^2 = -dt^2 + a^2(t)\, ds^2_{\text{3D}} \tag{2.4.12}$$

where ds^2_{3D} is the Euclidian metric of a homogeneous isotropic three-manifold, in the present case the three-sphere, and $a(t)$ is a function of the cosmic time, named the scale-factor. In the case of ds^2_{ESU} the scale factor is just one and for this reason the corresponding universe is static. Einstein, who was opposed to the idea of an evolving world discovered that by the addition of the cosmological constant his own equations admitted static cosmological solutions, in particular ds^2_{ESU}. Yet it was soon proved that Einstein's static universe is unstable and the great man later considered the cosmological constant the biggest mistake of his life. He was, in

this respect, twice wrong, since the cosmological constant does indeed exist, yet the universe evolves nonetheless. All these questions we shall address in later chapters; at present what is important for us is the following. By means of the coordinate transformation (2.4.2)–(2.4.5), we have realized a mapping:

$$\psi : \mathscr{M}_{\text{Mink}} \to \mathscr{M}_{\text{ESU}} \simeq \mathbb{R} \otimes \mathbb{S}^3 \tag{2.4.13}$$

that injects the whole of Minkowski space into a finite volume region of the Einstein Static Universe, whose corresponding differentiable manifold is isomorphic to $\mathbb{R} \otimes \mathbb{S}^3$. In order to verify the statement we just made it suffices to compare the ranges of the coordinates T, R, θ, ϕ respectively corresponding to the whole $\mathbb{R} \otimes \mathbb{S}^3$ and to the image of Minkowski-space through the ψ-mapping:

$$\psi(\mathscr{M}_{\text{Mink}}) \subset \mathbb{R} \otimes \mathbb{S}^3 \tag{2.4.14}$$

This comparison is presented below:

$\mathbb{R} \otimes \mathbb{S}^3$	Minkowski
$-\infty < T < +\infty$	$-\pi < T + R < \pi$
$0 \leq R \leq \pi$	$-\pi < T - R < \pi$
$0 \leq \theta \leq \pi$	$0 \leq \theta \leq \pi$
$0 \leq \phi \leq 2\pi$	$0 \leq \phi \leq 2\pi$

(2.4.15)

The specified ranges of the $T \pm R$ variables in Minkowski case are elementary properties of the function $\arctan(x)$ which maps the infinite interval $\{-\infty, \infty\}$ into the finite one $\{-\pi, \pi\}$. To each point T, R is attached a two-sphere $\mathbb{S}^2$ parameterized by the angles θ, ϕ. It is difficult to visualize four-dimensional spaces, yet, if we replace the three-sphere by a circle, we can visualize $\mathbb{R} \otimes \mathbb{S}^3$ as an infinite cylinder and the sub-manifold $\psi(\mathscr{M}_{\text{Mink}})$ corresponds to the finite shaded region of the cylinder displayed in Fig. 2.16. The reader will notice that we have decomposed the boundary of $\psi(\mathscr{M}_{\text{Mink}})$ into various components i^0, $i^{\pm}$, $J^{\pm}$. To understand the meaning of such a decomposition we need to stress another fundamental property of the mapping ψ defined by (2.4.2)–(2.4.5). As it is evident from (2.4.6) Minkowski metric and the metric of ESU are not identical, yet they differ only by the square of an overall function of the coordinates. This property is precisely what defines the concept of a *conformal mapping*.

Definition 2.4.1 Let $(\mathscr{M}, g)$ be a (pseudo-)Riemannian manifold of dimension m and $(\widetilde{\mathscr{M}}, \tilde{g})$ another (pseudo-)Riemannian manifold with the same dimension. A differentiable map:

$$\psi : \mathscr{M} \to \widetilde{\mathscr{M}} \tag{2.4.16}$$

is named *conformal* if and only if on the image $\text{Im}\,\psi \equiv \psi(\mathscr{M})$ the following condition holds true:

$$\exists \Omega \in \mathbb{C}^{\infty}(\text{Im}\,\psi) \setminus \tilde{g}|_{\text{Im}\,\psi} = \Omega^2 \psi^* g \tag{2.4.17}$$

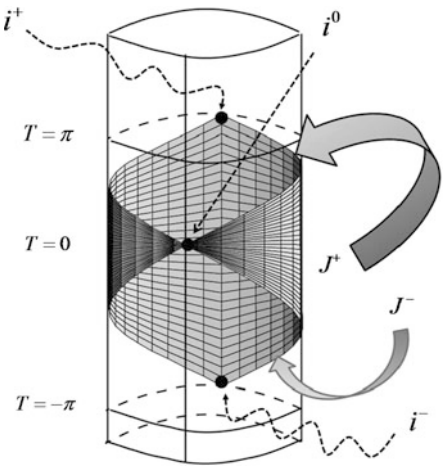

Fig. 2.16 The shaded region corresponds to the image, inside the Static Einstein Universe, of Minkowski space by means of the conformal mapping ψ. This picture visualizes the causal boundary of Minkowski space composed of a spatial infinity i^0 a future and a past time-like infinity $i^{\pm}$ and a future and past light-like infinity $J^{\pm}$

where $\psi^* g$ denotes the pull-back of the metric g. The function Ω is named the *conformal factor*.

As anticipated above, the basic property of conformal mappings is that they preserve the causal structure. On $\psi(\mathscr{M}) \subset \widetilde{\mathscr{M}}$ we have two metrics, namely $\tilde{g}|_{\mathrm{Im}\,\psi}$ and $\psi^* g$. Generically curves that are geodesics with respect to the former are not geodesics with respect to the latter; yet curves that are causal in one metric are causal also in the other and vice-versa. Furthermore light-like geodesics are common to $\tilde{g}|_{\mathrm{Im}\,\psi}$ and $\psi^* g$. Indeed we have the following:

Lemma 2.4.1 *Consider two metrics G and g on the same manifold $\mathscr{M}$ related by a positive definite conformal factor $\Omega^2(x)$, namely*:

$$G_{\mu\nu}\, dx^\mu \otimes dx^\nu = \Omega^2(x) g_{\mu\nu}\, dx^\mu \otimes dx^\nu \tag{2.4.18}$$

The light-like geodesics with respect to the metric G are light-like geodesics also with respect to the metric g and vice-versa.

Proof The proof is performed in two steps. First of all let us note that the differential equation for geodesics takes the form

$$\frac{d^2 x^\rho}{d\lambda^2} + \Gamma^\rho_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} = 0 \tag{2.4.19}$$

when we use an affine parameter λ. It can be rewritten with respect to an arbitrary parameter $\sigma = \sigma(\lambda)$. By means of direct substitution equation (2.4.19) transforms into:

$$\frac{d^2 x^\rho}{d\sigma^2} + \Gamma^\rho_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} = -\left(\frac{d\sigma}{d\lambda}\right)^{-2} \frac{d^2\sigma}{d\lambda^2} \frac{dx^\rho}{d\sigma} \tag{2.4.20}$$

Secondly let us compare the Christoffel symbols of the metric G, named $\Gamma^{\rho}_{\mu\nu}$ with those of the metric g, named $\gamma^{\rho}_{\mu\nu}$. Once again by direct evaluation we find:

$$\Gamma^{\rho}_{\mu\nu} = \gamma^{\rho}_{\mu\nu} + 2\partial_{(\mu} \ln \Omega \delta^{\rho}_{\nu)} - g_{\mu\nu}\partial^{\rho} \ln \Omega \tag{2.4.21}$$

Hence we obtain:

$$\frac{d^2x^{\rho}}{d\sigma^2} + \Gamma^{\rho}_{\mu\nu}\frac{dx^{\mu}}{d\sigma}\frac{dx^{\nu}}{d\sigma} = \frac{d^2x^{\rho}}{d\sigma^2} + \gamma^{\rho}_{\mu\nu}\frac{dx^{\mu}}{d\sigma}\frac{dx^{\nu}}{d\sigma} - \left(g_{\mu\nu}\frac{dx^{\mu}}{d\sigma}\frac{dx^{\nu}}{d\sigma}\right)\partial^{\rho} \ln \Omega + \left(\frac{d}{d\sigma} \ln \Omega\right)\frac{dx^{\rho}}{d\sigma} \tag{2.4.22}$$

Let us now apply the identity (2.4.22) to the case where the curve $x^{\mu}(\sigma)$ is a light-like geodesics for the metric $g_{\mu\nu}$ and σ is an affine parameter for it. Then all terms on the right hand side of equation (2.4.22) written in the first line vanish. Indeed:

$$\frac{d^2x^{\rho}}{d\sigma^2} + \gamma^{\rho}_{\mu\nu}\frac{dx^{\mu}}{d\sigma}\frac{dx^{\nu}}{d\sigma} = 0 \tag{2.4.23}$$

is the geodesic equation in the affine parameterization and

$$g_{\mu\nu}\frac{dx^{\mu}}{d\sigma}\frac{dx^{\nu}}{d\sigma} = 0 \tag{2.4.24}$$

is the light-like condition on the tangent vector to the considered curve. It follows that the same curve $x^{\mu}(\sigma)$ satisfies the geodesic equation also with respect to the metric $G_{\mu\nu}$ provided we are able to solve the following differential equation:

$$-\left(\frac{d\sigma}{d\lambda}\right)^{-2}\frac{d^2\sigma}{d\lambda^2} = \frac{d}{d\sigma} \ln \Omega \tag{2.4.25}$$

for a function $\lambda(\sigma)$ which will play the role of affine parameter with respect to the new metric. Such an integration is easily performed. Indeed by means of straightforward steps we first reduce (2.4.25) to:

$$\ln\left(\frac{d\sigma}{d\lambda}\right) = -\ln \Omega + \text{const} \tag{2.4.26}$$

and then with a further integration we obtain:

$$\lambda = k_1 \int \Omega(\sigma)\, d\sigma + k_2 \tag{2.4.27}$$

where $k_{1,2}$ are the two integration constants. So a light-like geodesics with respect to the metric $g_{\mu\nu}$ satisfies the geodesic equation also with respect to any metric G conformal to g with an affine parameter λ given by (2.4.27). Moreover the tangent vector to the curve is obviously light-like in the metric G if it is light-like in the metric g. This concludes the proof of the lemma. □

Let us summarize. We have found a conformal mapping of Minkowski space into a finite region of another pseudo-Riemannian manifold so that the boundary at infinity has been brought to finite distance and can be inspected. This boundary is decomposed into the following pieces:

$$\partial\psi(\mathscr{M}_{\text{Mink}}) = i^0 \bigcup i^+ \bigcup i^- \bigcup J^+ \bigcup J^- \tag{2.4.28}$$

that have been appropriately marked in Fig. 2.16. What is their meaning? It is listed below:

(1) i^0, named *Spatial Infinity* is the endpoint of the ψ image of all space-like curves in $(\mathscr{M}, g)$.
(2) i^+, named *Future Time Infinity* is the endpoint of the ψ image of all future-directed time-like curves in $(\mathscr{M}, g)$.
(3) i^-, named *Past Time Infinity* is the endpoint of the ψ image of all past-directed time-like curves in $(\mathscr{M}, g)$.
(4) J^+, named *Future Null Infinity* is the endpoint of the ψ image of all future-directed light-like curves in $(\mathscr{M}, g)$.
(5) J^-, named *Past Null Infinity* is the endpoint of the ψ image of all past-directed light-like curves in $(\mathscr{M}, g)$.

In the above listing we have denoted by $(\mathscr{M}, g)$ Minkowski space with its flat metric. The reason to use such a notation is that the same structure of the boundary applies to all *asymptotically flat space-times* in the definition we shall shortly provide.

In order to verify the above interpretation of the boundary it is convenient to disregard the two-sphere coordinates θ, ϕ restricting one's attention to radial geodesics or curves only. In this way Minkowski space becomes effectively two-dimensional with the metric $ds^2 = -dt^2 + dr^2$. The conformal transformation (2.4.2), (2.4.3) maps the half plane $(\infty \geq t \geq -\infty,\ \infty \geq r \geq 0)$ into a finite region of the half-plane $(\infty \geq T \geq -\infty,\ \infty \geq R \geq 0)$. This finite region is the triangle displayed in Fig. 2.17. Radial geodesics in Minkowski space are straight lines in the (t, r) half-plane. They are time-like if the angular coefficient is bigger than 45 degrees, space-like if it is less than 45 degrees and they are light-like when it is exactly $\pi/2$. In Fig. 2.18 we display the conformal transformation of these geodesics from which it is evident that the time-like ones end up in the time-infinities while the space-like ones end up in spatial infinity. The image of the light-like geodesics are still segments of straight-lines at 45 degrees which end on the null-infinities defined above. Analytically the above statements can be verified by calculating some elementary limits. The image of a straight line $t = \alpha r$ is given by:

$$\begin{aligned} T(\alpha, r) &= \arctan\big[(\alpha+1)r\big] + \arctan\big[(\alpha-1)r\big] \\ R(\alpha, r) &= \arctan\big[(\alpha+1)r\big] - \arctan\big[(\alpha-1)r\big] \end{aligned} \tag{2.4.29}$$

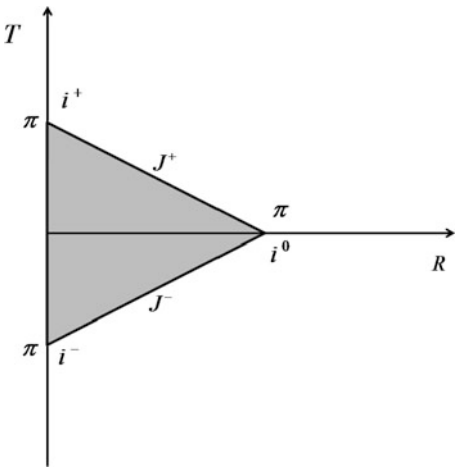

Fig. 2.17 The Penrose diagram of Minkowski space

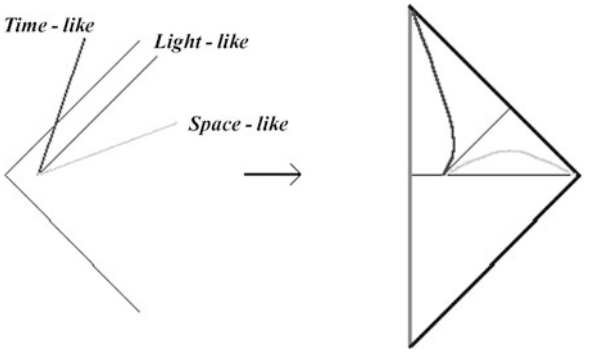

Fig. 2.18 The conformal mapping of Minkowski geodesics into the Penrose triangle

and we find:

$$\lim_{r\to\infty} T(\alpha, r) = \begin{cases} \pi & \text{if } \alpha > 1 \\ 0 & \text{if } 1 > \alpha > -1 \\ -\pi & \text{if } \alpha < -1 \end{cases} \tag{2.4.30}$$

$$\lim_{r\to\infty} R(\alpha, r) = \begin{cases} 0 & \text{if } \alpha > 1 \\ \pi & \text{if } 1 > \alpha > -1 \\ 0 & \text{if } \alpha < -1 \end{cases} \tag{2.4.31}$$

More generally we can consider curves $t = f(r)$. The same limits as above hold true if we replace α with $f'(r)$.

This concludes our discussion of the causal boundary of Minkowski space which was possible thanks to the conformal mapping of the latter into a finite region of the Einstein Static Universe. From this discussion we learnt a lesson that enables us to extract some general definition of conformal flatness allowing the inspection of the causal boundary of more complicated space-times such as, for instance, the Kruskal extension of the Schwarzschild solution.

2.4.2 Asymptotic Flatness

In this section we describe the definition of asymptotic flatness according to Ashtekar [8].

Definition 2.4.2 A space-time $(\mathscr{M}, g)$ is asymptotically flat if there exists another larger space-time $(\widetilde{\mathscr{M}}, \tilde{g})$ and a *conformal mapping*:

$$\psi : \mathscr{M} \to \psi(\mathscr{M}) \subset \widetilde{\mathscr{M}} \tag{2.4.32}$$

with conformal factor Ω:

$$\tilde{g} = \Omega^2 \psi^* g \quad \text{on } \psi(\mathscr{M}) \tag{2.4.33}$$

such that the following conditions are verified:

(1) Naming i^0 *spatial infinity*, namely the locus in $\psi(\mathscr{M})$ where terminate all space-like curves, which is required to be a single point, we have:

$$\widetilde{\mathscr{M}} - \psi(\mathscr{M}) = \overline{J^+(i^0)} \bigcup \overline{J^-(i^0)}$$

(2) The boundary of $\mathscr{M}$, named $\partial\mathscr{M}$ is decomposed as follows:

$$\partial\mathscr{M} = i^0 \bigcup \mathscr{J}^+ \bigcup \mathscr{J}^-$$

where by definition we have set:

$$\mathscr{J}^\pm = \partial J^\pm(i^0) - i^0$$

(3) There exists a neighborhood $V \subset \partial\psi(\mathscr{M})$ such that for every $p \in V$ and every neighborhood $\mathscr{O}_p$ of that point we can find a sub-neighborhood $\mathscr{U}_p \subset \mathscr{O}_p$ with the property that no causal curve intersects $\mathscr{U}_p$ more than once.
(4) The conformal factor Ω can be extended to an overall function on the whole $\widetilde{\mathscr{M}}$
(5) The conformal factor Ω vanishes on $\mathscr{J}^+$ and $\mathscr{J}^-$ but its derivative $\nabla_\mu \Omega$ does not on the same locus.

In order to appreciate all the points of the above definition it is convenient to look at Fig. 2.19 and compare with the case of Minkowski space. The starting point of the analysis is the obvious observation that any causal curve which departs from spatial infinity $i^0 \equiv (\pi, 0)$ cannot penetrate in the triangle representing Minkowski space and therefore lies in $\widetilde{\mathscr{M}} - \psi(\mathscr{M})$. If the causal curve is future-directed it goes up, while if it is past directed it goes down so that point (1) of Definition 2.4.2 is indeed verified. Let us next consider the boundary of the causal future and causal past of spatial infinity. They are given by the upper and lower side, respectively, of the triangle in Fig. 2.19, which intersect in i^0. Hence point (2) of Definition 2.4.2 is also verified. Let us note that according to this definition $\mathscr{J}^\pm$ are just the Causal

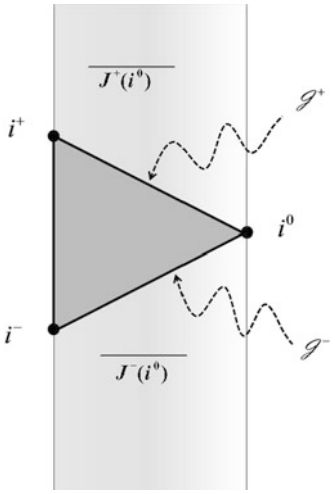

Fig. 2.19 The causal boundary of Minkowski space following Ashtekar definition

Future and Causal Past of the considered space-time, namely the locus where terminate future-directed and past-directed causal curves, respectively. In the case of Minkowski space we were able to make a finer distinction by decomposing:

$$\mathscr{I}^{\pm} = i^{\pm} \bigcup J^{\pm} \tag{2.4.34}$$

where $i^{\pm}$ correspond to Future and Past Time-Infinity, while $J^{\pm}$ are Future and Past Null-Infinities.

Point (3) of the definition is also visually evident in the case of Minkowski space and aims at excluding pathological space-times where causal curves might have chaotic behavior.

Points (4) and (5) are also extracted from the example of Minkowski space mapped into the Einstein Static Universe. There the conformal factor is

$$\Omega = \frac{1}{2}\cos\left[\frac{T+R}{2}\right]\cos\left[\frac{T-R}{2}\right] \equiv \cos T + \cos R$$

which vanishes on the two straight-lines:

$$\{\xi, \xi + \pi\}; \qquad \{\xi, -\xi + \pi\} \tag{2.4.35}$$

so, in particular on the two loci $\mathscr{I}^{\pm}$.

2.5 The Causal Boundary of Kruskal Space-Time

Let us now consider the Kruskal extension of the Schwarzschild metric given in (2.2.30) where the variable r is implicitly defined by its relation with T and X,

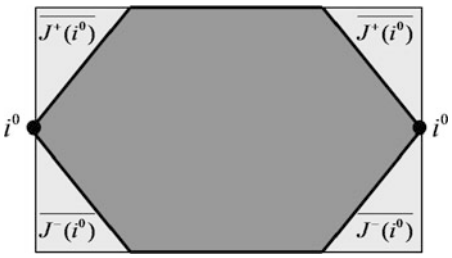

Fig. 2.20 The Spatial Infinity of Kruskal space-time and its Future and Past

namely:

$$T^2 - X^2 = \left(\frac{r}{r_S} - 1\right) \exp\left[\frac{r}{r_S}\right] \tag{2.5.1}$$

Let us introduce the further change of variables defined below:

$$\begin{aligned} T &= \frac{1}{2}\tan\left(\frac{\tau+\rho}{2}\right) + \frac{1}{2}\tan\left(\frac{\tau-\rho}{2}\right) \\ X &= \frac{1}{2}\tan\left(\frac{\tau+\rho}{2}\right) - \frac{1}{2}\tan\left(\frac{\tau-\rho}{2}\right) \end{aligned} \tag{2.5.2}$$

By means of straightforward substitutions we find that:

$$ds^2_{\text{Krusk}} = \Omega^{-2}\, \widetilde{ds}^2_{\text{Krusk}} \tag{2.5.3}$$

$$\begin{aligned} \widetilde{ds}^2_{\text{Krusk}} &= 4\frac{r^3}{r_S}\exp\left[-\frac{r}{r_S}\right]\left(-d\tau^2 + d\rho^2\right) \\ &\quad + r^2(\cos\tau + \cos\rho)^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \end{aligned} \tag{2.5.4}$$

$$0 = \tan\left(\frac{\tau+\rho}{2}\right)\tan\left(\frac{\tau-\rho}{2}\right) + \left(\frac{r}{r_S} - 1\right)\exp\left[\frac{r}{r_S}\right] \tag{2.5.5}$$

$$\Omega = (\cos\tau + \cos\rho) \tag{2.5.6}$$

This calculation shows that the map ψ defined by the coordinate substitution (2.5.2) is indeed a conformal map, the new metric being $\widetilde{ds}^2_{\text{Krusk}}$ defined by (2.5.4) and the conformal factor being Ω defined in (2.5.6). Let us then verify that Kruskal space-time is asymptotically flat and study the causal structure of its boundary. To this effect let us consider Fig. 2.20. Just as in the case of Minkowski space we represent the four-dimensional space-time by means of a two-dimensional picture where each point actually stands for a two-sphere spanned by the coordinates $\{\theta, \phi\}$. The points are located in the $\{\tau, \rho\}$ plane and such kind of visualization receives the name of *Penrose diagram* (Fig. 2.21).

As in the case of Minkowski space we first look for *Spatial Infinity* and we find that in the plane $\{\tau, \rho\}$ it is given by the following pair of points:

$$i^0 \equiv \{\pi, 0\} \bigcup \{-\pi, 0\} \tag{2.5.7}$$

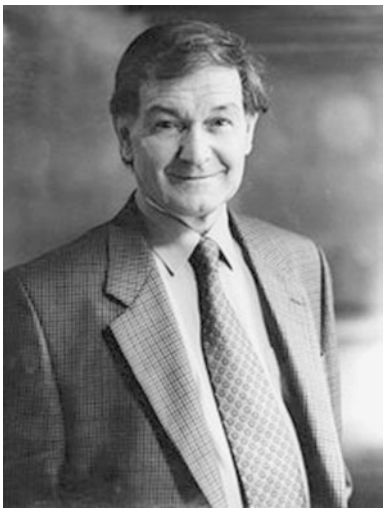

Fig. 2.21 Sir Roger Penrose, was born in 1931 in Colchester (England) and he is Emeritus Rouse Ball Professor of Mathematics at the University of Oxford. His main contributions have been to Mathematical Physics in the fields of Relativity and Quantum Field Theory. He invented the *twistor* approach to Lorentzian field theories which maps geometrical metric data of a real manifold into holomorphic data in a complex manifold with signature $(2, 2)$. He was the first to propose the cosmic censorship hypothesis according to which space-time singularities are always hidden behind event-horizons and he conceived the idealized *Penrose mechanism* which shows how energy can be extracted from rotating black-holes. Probably the most famous of his results is the quasi-periodic *Penrose tiling* of the plane with five-fold rotational symmetry. Roger Penrose is also an amateur philosopher whose views on consciousness and its relation with quantum physics are quite original and source of intense debate

Indeed this is the locus where terminate the images of all space-like curves. The duplication of i^0 is due to the periodicity of the trigonometric functions and it occurs also in Minkowski case. There it was disregarded because all copies of i^0, namely $\{(2n+1)\pi, 0\}$, $(n \in \mathbb{Z})$ could be identified without ambiguity. In the Kruskal case, instead, as we are going to see, $i^0_{\rm I} = \{\pi, 0\}$ and $i^0_{\rm IV} = \{-\pi, 0\}$ must be considered as distinct physical points since they are separated by the black-hole region which we are now going to discuss.

Following the scheme outlined in previous section, we search for the causal future and causal past of i^0 inside the extended manifold $(\widetilde{\mathscr{M}}_{\rm Krusk}, \tilde{g}_{\rm Krusk})$. At this level a fundamental new feature appears with respect to Minkowski case where, reduced to the plane $\{T, R\}$, the manifold $\widetilde{\mathscr{M}}_{\rm Mink}$ was identified with the infinite vertical strip depicted in Fig. 2.19. In the Kruskal case, on the contrary, also the embedding manifold $\widetilde{\mathscr{M}}_{\rm Krusk}$ corresponds to a finite region of the $\{\tau, \rho\}$ plane, namely the following rectangular region:

$$\{\tau, \rho\} \in \widetilde{\mathscr{M}}_{\rm Krusk} \quad \Leftrightarrow \quad -\frac{\pi}{2} \leq \tau \leq \frac{\pi}{2} \quad \text{and} \quad -\pi \leq \rho \leq \pi \tag{2.5.8}$$

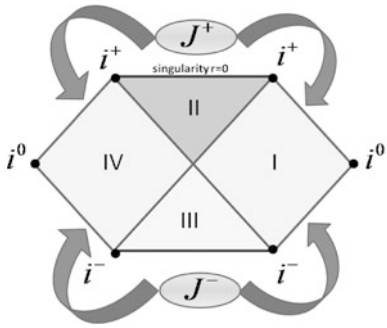

Fig. 2.22 The Penrose diagram of Kruskal space-time

The upper and lower limits on the variable τ are consequences of the form of the metric $\tilde{g}_{\text{Krusk}}$ defined in (2.5.4). This latter becomes singular when $r = 0$ and from (2.5.5) we realize that this singularity is mapped into $\tau = \pm\frac{\pi}{2}$. Hence no causal curve can trespass such limits which become a boundary for the manifold $(\widetilde{\mathscr{M}}_{\text{Krusk}}, \tilde{g}_{\text{Krusk}})$. The range of the variable ρ is fixed instead by modding out the periodicity $\rho \to \rho + 2n\pi$.

Once (2.5.8) is established, it is fairly easy to conclude that the Causal Future and Causal Past of Spatial Infinity are indeed the lighter regions of the rectangle depicted in Fig. 2.20. The corresponding boundaries are:

$$\begin{aligned}
\partial\overline{J^+(i^0)} &= \left\{\frac{\pi}{2}\xi, -\frac{\pi}{2}\xi + \pi\right\} \bigcup \left\{\frac{\pi}{2}\xi, \frac{\pi}{2}\xi - \pi\right\}; \quad \xi \in [0,1] \\
\partial\overline{J^-(i^0)} &= \left\{-\frac{\pi}{2}\xi, -\frac{\pi}{2}\xi + \pi\right\} \bigcup \left\{-\frac{\pi}{2}\xi, \frac{\pi}{2}\xi - \pi\right\}; \quad \xi \in [0,1]
\end{aligned} \tag{2.5.9}$$

and on these boundaries the conformal factor (2.5.6) vanishes. Hence all necessary conditions are satisfied and the Kruskal extension of Schwarzschild space-time is indeed asymptotically flat.

We can now inspect the causal structure of conformal infinity and we are led to consider the more detailed diagram of Fig. 2.22, which is the conformal image in the $\{\tau, \rho\}$ plane of the diagram 2.8 drawn in the $\{T, X\}$ plane. We easily identify in Fig. 2.22 the points $i^{\mp}$ that correspond to time-like Past and Future Infinity, respectively. Just as it was the case for Spatial Infinity also these Infinities have a double representation in the diagram. Similarly Past and Future Null Infinities are twice represented and correspond to the segments with ± 45 degrees orientation shown in Fig. 2.22. The conformal image of the singularity $r = 0$ is also double and it is provided by the two segments, upper and lower, parallel to the ordinate axis depicted in Fig. 2.22. The conformal image of the event horizon $X^2 - T^2 = 0$ is provided by the two internal lines splitting the hexagon of Fig. 2.22 into four separate regions.

Let us know consider, using the language developed in previous sections, the Causal Past of Future-Null Infinity namely $J^-(\mathscr{J}^+)$. By definition this is the set of all space-time events p such that there exists at least one causal curve starting at p

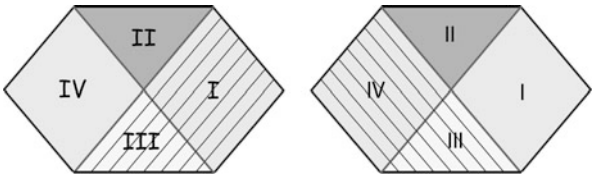

Fig. 2.23 The Causal Past of Future Null Infinity is composed of two-sheets. The Causal Past of $\mathscr{J}_{\mathrm{I}}^{+}$ and the Causal Past of $\mathscr{J}_{\mathrm{IV}}^{+}$. The first corresponds to the region shaded by lines in the picture on the *left*, the second to the region shaded by lines in the picture on the *right*

and ending on $\mathscr{J}^{+}$. Since $\mathscr{J}^{+}$ is the union of two disconnected loci:

$$\mathscr{J}^{+} = \mathscr{J}_{\mathrm{I}}^{+} \bigcup \mathscr{J}_{\mathrm{IV}}^{+} \tag{2.5.10}$$

we actually have:

$$J^{-}\left(\mathscr{J}^{+}\right) = J^{-}\left(\mathscr{J}_{\mathrm{I}}^{+}\right) \bigcup J^{-}\left(\mathscr{J}_{\mathrm{IV}}^{+}\right) \tag{2.5.11}$$

A simple inspection of the Penrose diagram shows that the Causal Past of Future Null Infinity is given by the regions shown in Fig. 2.23, namely we have:

$$\begin{aligned} J^{-}\left(\mathscr{J}^{+}\right) &= \mathrm{I} \bigcup \mathrm{III} \bigcup \mathrm{IV} \\ J^{-}\left(\mathscr{J}_{\mathrm{I}}^{+}\right) &= \mathrm{I} \bigcup \mathrm{III} \\ J^{-}\left(\mathscr{J}_{\mathrm{IV}}^{+}\right) &= \mathrm{III} \bigcup \mathrm{IV} \end{aligned} \tag{2.5.12}$$

This conclusion is simply reached with the following argument. The image of light-like geodesics in the Penrose diagram is given by the straight lines with ± 45 degrees orientation; hence it suffices to trace all lines that have such an inclination and which intersect Future Null Infinity. The result is precisely that of (2.5.12), depicted in Fig. 2.23.

In this way we discover a very important feature of region II, namely we find that it has empty intersection with the Causal Past of Future Null Infinity: $\mathrm{II} \bigcap J^{-}(\mathscr{J}^{+}) = \emptyset$. This property provides a rigorous mathematical formulation of that *object cut off from communication with the rest of the universe* which was firstly conceived by Openheimer and Snyder as end-point result of the gravitational collapse of super massive stars.

Inspired by the case of Kruskal space-time we can now present the general definition of black-holes:

Definition 2.5.1 Let $(\mathscr{M}, g)$ be an asymptotically flat space time and let $\mathscr{J}^{+}$ denote the Future Null Infinity component of its causal boundary. A black-hole region $\mathrm{BH} \subset \mathscr{M}$ is a sub-manifold of this space-time with the following defining property:

$$\mathrm{BH} \bigcap J^{-}\left(\mathscr{J}^{+}\right) = \emptyset \tag{2.5.13}$$

The event horizon is the boundary of the black-hole region separating it from the Causal Past of Future Null Infinity, namely:

$$\mathfrak{H} = \partial \mathrm{BH} \bigcap \partial J^{-}\left(\mathscr{I}^{+}\right) \tag{2.5.14}$$

Let us now comment on the properties of region III of Kruskal space-time. Differently from the black-hole region II, where all future-directed causal curves end up on the singularity, in region III this is the inevitable property of past-directed causal curves. Namely every one who happens to be in region III at some instant of time had origin in the singularity and came out from there. Furthermore all future-directed causal curves starting in III necessarily cross the horizon and end up either in the flat asymptotic region I or in its copy IV. Hence III is just the time reversal of a black-hole, named a *white hole*. A white hole emits matter rather than swallowing it and therefore evaporates as soon as it is formed.

Although white holes are a part of the classical Kruskal vacuum solution, it is doubtful that they might exist in Nature. When one considers the gravitational collapse of realistic stars, the presence of matter fields in the Einstein equations removes the presence of the white hole sector from their solutions. Furthermore, as it will become clearer in next chapter where we study the fascinating relation between the Laws of Thermodynamics and those of Black-Hole Mechanics, white holes violate the second principle of thermodynamics, reducing rather than increasing the entropy and this is one more reason for their absence from the physical universe.

References

1. Gibbons, G.: The man who invented black holes [his work emerges out of the dark after two centuries]. New Scientist **28**, 1101 (1979)
2. Laplace, P.S.: Exposition du Système du Monde, 1st edn. Paris (1796)
3. Reissner, H.: Über die Eigengravitation des elektrischen Feldes nach der Einstein'schen Theorie. Ann. Phys. **50**, 106–120 (1916). doi:10.1002/andp.19163550905
4. Nordström, G.: On the energy of the gravitational field in Einstein's theory. Verhandl. Koninkl. Ned. Akad. Wetenschap., Afdel. Natuurk., Amsterdam **26**, 1201–1208 (1918)
5. Petrov, A.Z.: Classification of spaces defined by gravitational fields. Uch. Zapiski Kazan Gos. Univ. **144**, 55 (1954). English translation Petrov, A.Z, Gen. Relativ. Gravit. **22**, 1665 (2000). doi:10.1023/A:1001910908054
6. Kruskal, M.D.: Maximal extension of Schwarzschild metric. Phys. Rev. **119**, 1743 (1960)
7. Szekeres, G.: On the singularities of a Riemannian manifold. Publ. Math. (Debr.) **7**, 285 (1960)
8. Ashtekar, A.: Asymptotic structure of the gravitational field at spatial infinity. Gen. Relativ. Gravit. **2** (1980)

http://www.springer.com/978-94-007-5442-3

Gravity, a Geometrical Course
Volume 2: Black Holes, Cosmology and Introduction to Supergravity
Frè, P.G.
2013, XX, 452 p., Hardcover
ISBN: 978-94-007-5442-3