CHAPTER TWO

Structure of Spacetime Singularities

R. Penrose

In the first lecture by Stephen Hawking, singularity theorems were discussed. The essential content of these theorems is that under reasonable (global) physical conditions, singularities must be expected. They do not say anything about the nature of the singularities, or where the singularities are to be found. On the other hand, the theorems are very general. A natural question to ask is, therefore, what the geometric nature of a spacetime singularity is. It is usually assumed that the characteristic of a singularity is that the curvature diverges. However, this is not exactly what the singularity theorems by themselves imply.

Singularities occur in the big bang, in black holes, and in the big crunch (which might be regarded as a union of black holes). They also might appear as naked singularities. Related to this question is what is called *cosmic censorship*, namely the hypothesis that these naked singularities do not occur.

To explain the idea of cosmic censorship, let me recall a bit the history of the subject. The first explicit example of a solution of Einstein's equations describing a black hole was the collapsing dust cloud of Oppenheimer and Snyder (1939). There is a singularity inside, but it is not visible from outside, as it is surrounded by the event horizon. This horizon is the surface inside of which events cannot send signals out to infinity. It was tempting to believe that this picture is generic, i.e., that it represents the general gravitational collapse. However, the OS model has a special symmetry (namely, spherical symmetry), and it is not obvious that it is really representative.

As the Einstein equations are generally hard to solve, one looks instead for global properties that imply the existence of singularities. For example, the OS model has a trapped surface, which is a surface whose area will decrease along light rays that are initially orthogonal to it (fig. 2.1).

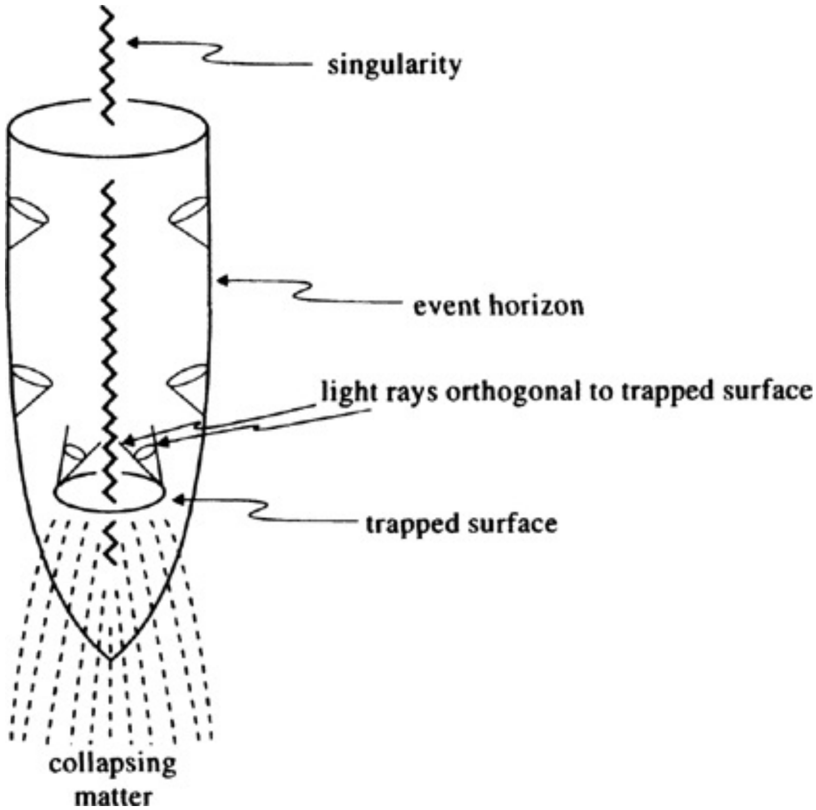

Figure 2.1 The Oppenheimer-Snyder collapsing dust cloud, illustrating a trapped surface.

One might try to show that the existence of a trapped surface implies that there is a singularity. (This was the first singularity theorem I was able to establish, on the basis of reasonable causality assumptions but no spherical symmetry being assumed; see Penrose 1965.) One can also derive similar results by assuming the existence of a converging light cone (Hawking and Penrose 1970; this occurs when all the light rays emitted in different directions from a point start to converge toward each other at a later time).

Stephen Hawking (1965) observed, very early on, that one can also turn my original argument upside down on a cosmological scale, i.e., apply it to the time-reversed situation. A reversed trapped surface then implies that there had been a singularity in the past (making appropriate causality assumptions). Now, the (time-reversed) trapped surface is very large, being on a cosmological scale.

We are here mainly concerned with analyzing the situation of a black hole. We know that there has to be a singularity somewhere, but in order to get a black hole we have to show that it is surrounded by an event horizon. The cosmic censorship hypothesis asserts just this, essentially that one cannot see the singularity itself from outside. In particular it implies that there is some region that cannot send signals to external infinity. The boundary of this region is the event horizon. We can also apply a theorem given in Stephen's last lecture to this boundary, as the event horizon is the boundary of the past of future null infinity. Thus we know that this boundary

- must be a null surface where it is smooth, generated by null geodesies,

- contains a future-endless null geodesic originating from each point at which it is not smooth,

and that

- the area of spatial cross sections cannot ever decrease with time.

It has also, in effect, been shown (Israel 1967, Carter 1971, Robinson 1975, Hawking 1972) that the asymptotic future limit of such a space-time is the Kerr spacetime. This is a very remarkable result, as the Kerr metric is a very nice exact solution of the Einstein vacuum equations. This argument also relates to the issue of black hole entropy and I shall, in effect, come back to it in the next lecture (Chapter 4).

Accordingly, we indeed have something with a qualitative similarity to the OS solution. There are some modifications—namely, that we end up with the Kerr solution rather than the Schwarzschild solution—but these are relatively minor. The essential picture is rather similar.

However, the precise arguments are based on the cosmic censorship hypothesis. In fact, cosmic censorship is very important, as the whole theory depends upon it, and without it we might see dreadful things instead of a black hole. So we do really have to ask ourselves whether it is true. A long time ago I thought that this hypothesis might be false and I made various attempts to find counterexamples. (Stephen Hawking once claimed that one of the strongest justifications for the cosmic censorship hypothesis was the fact that I had tried and failed to prove that it is wrong—but I think this is a very feeble argument!)

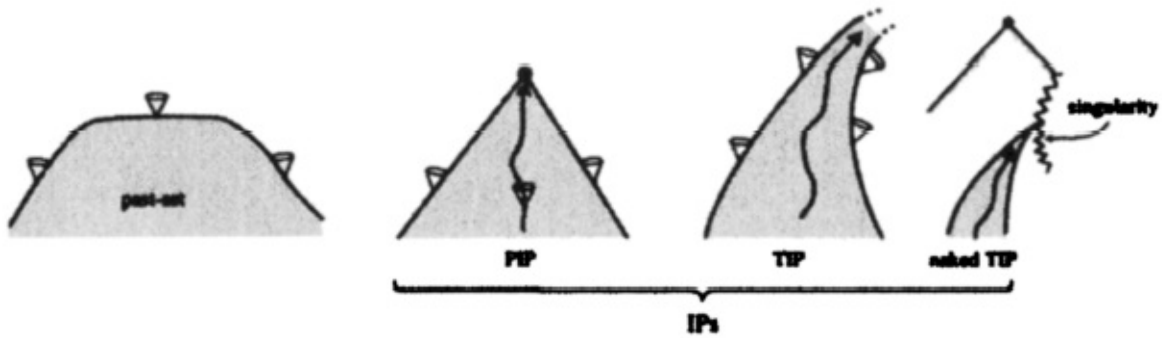

Figure 2.2 Past-sets, PIPs, and TIPs.

I want to discuss cosmic censorship in the context of certain ideas concerning *ideal* points for spacetimes. (These concepts are due to Seifert 1971, and Geroch, Kronheimer, and Penrose 1972). The basic idea is that one should incorporate into the spacetime actual "singular points" and "points at infinity," namely the *ideal points*. Let me first introduce the concept of an IP, i.e., an *indecomposable past-set*. Here a "past-set" is a set which contains its own past, and "indecomposable" means that it cannot be split into two past-sets neither of which contains the other. There is a theorem which tells us that one can also describe any IP as the past of some timelike curve (fig. 2.2).

There are two categories of IP, namely PIPs and TIPs. A PIP is a *proper* IP, i.e., the past of a spacetime point. A TIP is a *terminal* IP, not the past of an actual point in spacetime. TIPs define the future ideal points. Furthermore, one can distinguish TIPs according to whether this ideal point is "at infinity" (in which case there is a timelike curve generating the IP of infinite proper length)—an ∞-TIP—or a *singularity* (in which case every timelike curve generating it has finite proper length)—a singular TIP. Obviously all these concepts can be similarly applied to future-sets rather than to past-sets. In this case we have IFs (indecomposable futures), divided into PIFs and TIFs, the TIFs being subdivided into ∞-TIFs and singular TIFs. Let me also remark that for all this to work we have to assume, in effect, that there are no closed timelike curves—actually a marginally weaker condition: no two points have the same future or the same past.

How can we describe naked singularities and the cosmic censorship hypothesis in this framework? First of all, the cosmic censorship hypothesis should not exclude the big bang (since otherwise cosmologists would be in big trouble). Now, things always come out of the big bang and never fall into it. Thus, we might try to define a naked singularity as something that a timelike curve can both enter and exit from. Then the big bang problem is automatically taken care of. It does not count as naked. In this framework we can define a *naked* TIP as a TIP that is contained in a PIP. This is an essentially local definition, i.e., we do not require the observer to be at infinity. It turns out (Penrose 1979) that the exclusion of naked TIPs is the same condition in a spacetime if we replace "past" by "future" in this definition (exclusion of naked TIFs). The hypothesis that such naked TIPs (or, equivalently, TIFs) do not occur in generic spacetimes is called the *strong cosmic censorship* hypothesis. Its intuitive meaning is that a singular point (or infinite point)—the TIP in question—cannot simply "appear" in the middle of a spacetime in such a way that it is "visible" at some finite point—the vertex of the PIP in question. It is sensible that the observer needn't be at infinity since in a given spacetime we might not know whether there actually is an infinity. Furthermore, if the strong cosmic censorship hypothesis were violated we could, at a finite time, observe a particle actually falling into a singularity, where the rules of physics would cease to hold (or else reaching infinity, which is about as bad). We can also express the *weak cosmic censorship* hypothesis in this language: we just have to replace PIP by ∞-TIP.

The strong cosmic censorship hypothesis implies that a generic spacetime with matter, subject to reasonable equations of state (for example, vacuum), can be extended to one that is free of naked singularities (naked singular TIPs). It turns out (Penrose 1979) that the exclusion of naked TIPs is equivalent to global hyperbolicity, or that the spacetime is the whole domain of dependence of some Cauchy surface (Geroch 1970). We note that this formulation of the strong cosmic censorship is manifestly symmetric in time: we can interchange future and past if we interchange IPs and IFs.

In general, we need additional conditions to rule out *thunderbolts*. By a thunderbolt we mean a singularity which reaches null infinity, destroying the spacetime as it goes (cf. Penrose 1978, fig. 7). This need not violate cosmic censorship as stated. There exist stronger versions of cosmic censorship which take care of this (Penrose 1978, condition CC4).

So let us come back to the question whether cosmic censorship is true. First of all, let us note that it is probably not true in quantum gravity. In particular, exploding black holes (about which Stephen Hawking will explain more later) result in situations where cosmic censorship seems to be violated.

In classical general relativity there are various results in both directions. In one attempt to disprove cosmic censorship I derived certain inequalities which would hold if cosmic censorship were true (Penrose 1973). In fact, they did turn out to be true (Gibbons 1972)—and this seems to give support to the idea that something like cosmic censorship should hold. On the negative side there are some special examples (which, however, violate the genericity condition) and some sketchy numerical evidence that is subject to various objections. There are, furthermore, some indications about which I have only learned very recently—in fact, Gary Horowitz only mentioned them to me yesterday—that some of the aforementioned inequalities do not hold if the cosmological constant is positive. Personally I have always believed that the cosmological constant should be zero, but it would be very interesting if cosmic censorship depended upon it being, say, nonpositive. In particular, there might be an intriguing relationship between the nature of the singularities and the nature of infinity. Infinity is spacelike if the cosmological constant is positive, but null if it is zero. Correspondingly, singularities might sometimes turn out to be timelike (which means naked, i.e., violating cosmic censorship) if the cosmological constant is positive, but perhaps singularities cannot be timelike (i.e., satisfying cosmic censorship) if it is zero.

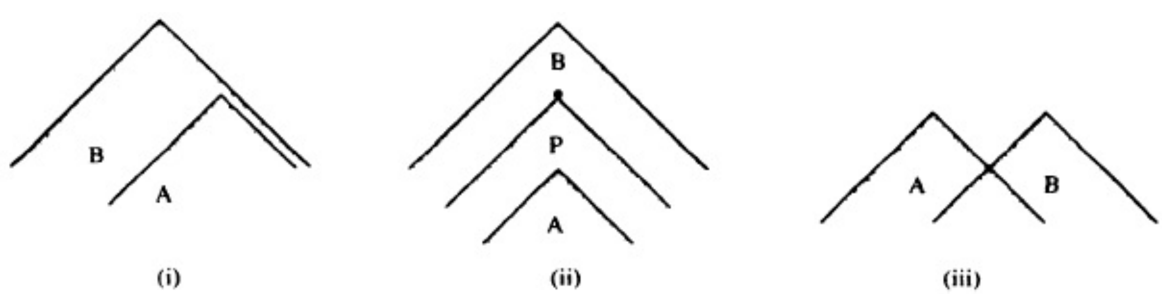

Figure 2.3 Causal relations between IPs: (i) *A* causally precedes *B;* (ii) *A* chronologically precedes *B;* (iii) *A* and *B* are spacelike separated.

To discuss the timelike or spacelike nature of singularities, let me explain the causal relations between IPs. Generalizing causality between points, we can say that an IP *A* causally precedes an IP *B*, if $A \subset B$; and *A* chronologically precedes B, if there is a PIP *P* such that $A \subset P \subset B$. We call *A* and *B* spacelike separated if neither causally precedes the other (fig. 2.3).

Strong cosmic censorship can then be expressed as saying that generic singularities are never timelike. Spacelike (or null) singularities can be either of past or future type. Hence, if strong cosmic censorship holds, singularities fall into two classes:

(P) Past types, defined by TIFs.
(F) Future type, defined by TTPs.

Naked singularities could unite the two possibilities into one, as a naked singularity would be a TIP and a TIF at the same time. Therefore it is really a consequence of cosmic censorship that these classes are separate. Typical examples of class (F) are singularities in black holes and the big crunch (if it exists), and of class (P) the big bang and possibly white holes (if they exist). I do not actually believe that the big crunch is likely to happen (for ideological reasons that I shall come to in the final lecture), and white holes are very much more unlikely because they disobey the second law of thermodynamics.

Perhaps the two types of singularity satisfy completely different laws. Maybe the quantum gravity laws for them should indeed be quite different. *I* think that Stephen Hawking disagrees here with me [SWH: "Yes!"], but I regard the following as evidence for this proposal:

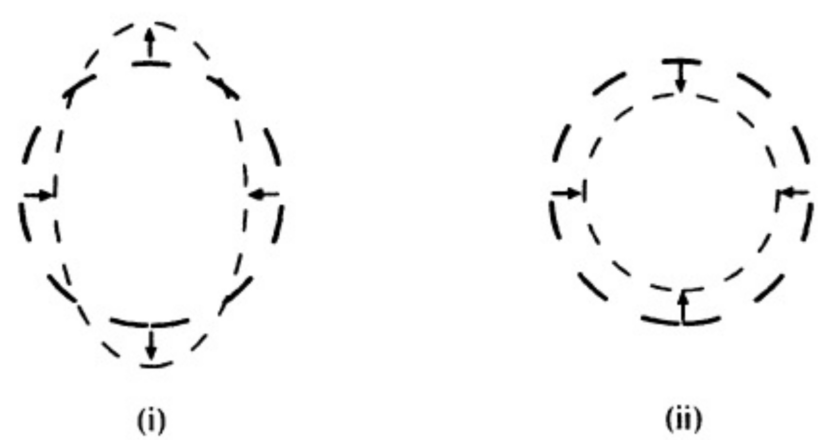

Figure 2.4 The acceleration effects of spacetime curvature: (i) the tidal distortion due to Weyl curvature; (ii) the volume-decreasing effect of Ricci curvature.

(1) The second law of thermodynamics.
(2) The observations of the early universe (e.g., COBE), which indicate that it was very uniform.
(3) The existence of black holes (virtually observed).

From (1) and (2) it can be argued that the big bang singularity was extremely uniform, and from (1) that it is free of white holes (for white holes violently disobey the second law of thermodynamics). Thus, very different laws must hold for the singularities of black holes (3). To describe this difference more precisely, recall that the spacetime curvature is described by the Riemann tensor R_{abcd}, which is the sum of the Weyl tensor C_{abcd} (describing the tidal distortions, which are volume preserving to first order) and a part equivalent to the Ricci tensor R_{ab} (times the metric g_{cd}, with indices appropriately scrambled), which describes volume-decreasing distortions (fig. 2.4).

In the standard cosmological models (due to Friedmann, Lemaître, Robertson, and Walker; see, for example, Rindler 1977) the big bang has vanishing Weyl tensor. (There is also a converse to this, proved by R.P.A.C. Newman, in which a universe with an initial singularity of a conformally regular type with vanishing Weyl tensor must, if suitable equations of state hold, be an FLRW universe; see Newman 1993.) On the other hand, black/white hole singularities have (in the generic case) diverging Weyl tensor. This suggests the following:

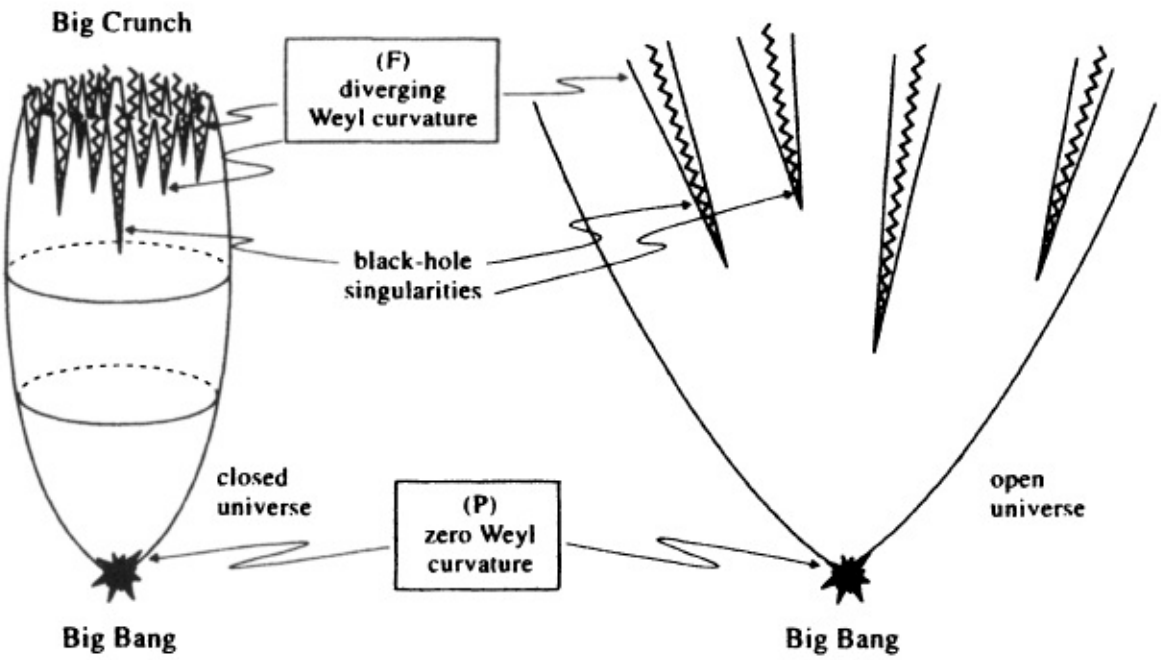

Figure 2.5 The Weyl curvature hypothesis: initial singularities (big bang) are constrained to have vanishing Weyl curvature whereas at final singularities, the Weyl curvature is expected to diverge.

Weyl Curvature Hypothesis

- Initial-type (P) singularities are constrained to have vanishing Weyl tensor.
- Final-type (F) singularities are not constrained.

This is closely in agreement with what one sees. If the universe is closed, the final singularity (the big crunch) will have diverging Weyl tensor, in an open universe the created black holes also have diverging Weyl tensor (see fig. 2.5).

Further support for this hypothesis comes from the fact that the constraint that the early universe was fairly smooth and free of white holes reduces the phase space in the early universe by a factor of at least

$$10^{10^{123}}$$

(This figure is the allowable phase space volume for a black hole of 10^{80} baryons, as follows from the Bekenstein-Hawking black hole entropy formula—Bekenstein 1972, Hawking 1975—and the universe has at least this much matter.)

Thus there should be a law which forces this rather unlikely result to happen! The Weyl curvature hypothesis would provide a law of this kind.

Question: Do you think that quantum gravity removes singularities?

Answer: I don't think it can be quite like that. If it were like that, the big bang would have resulted from a previously collapsing phase. We must ask how that previous phase could have had such a low entropy. This picture would sacrifice the best chance we have of explaining the second law. Moreover, the singularities of collapsing and expanding universes would have to be somehow joined together, but they seem to have very different geometries. A true theory of quantum gravity should replace our present concept of spacetime at a singularity. It should give a clear-cut way of talking about what we call a singularity in classical theory. It shouldn't be simply a nonsingular spacetime, but something drastically different.