

# SINGULARITIES IN COSMOLOGY

ROGER PENROSE

*Mathematical Institute, Oxford, U.K.*

**Abstract.** Singularities in space-time can be broadly divided into three classes: past-spacelike (in white holes or the big bang), timelike (naked singularities) and future-spacelike (in black holes or the final recollapse). In a closed Universe, if a simple restriction is made to eliminate timelike singularities, the inference may be drawn that the topology of the Universe is unchanging with time. Thermodynamical considerations lead one to infer that the final singularity of recollapse must differ markedly in structure from the initial big bang. This may plausibly be related to the existence of black holes and the presumed non-existence of white holes.

In this lecture I shall confine myself to making general qualitative remarks about singularities. The previous speaker, Prof. Lifshitz has given an excellent account Belinskii *et al.*, 1974) of the very fine and detailed work (Khalatnikov and Lifshitz, 1963; Belinskii *et al.*, 1970, 1972) carried out over a number of years by the Soviet school. I do not feel able to add to that here.

For the sake of simplicity I shall also confine my remarks to the case of closed universe models only. This is not intended to reflect any bias on my part as to whether I believe the Universe to be open or closed in fact. I have no strong feelings on the matter. It is merely that the statements of the results that I shall be concerned with are much more clear-cut in the case of closed universe models than open ones. Generalizations to the case of open universes are certainly possible and are treated to some extent in my other lecture (Penrose, 1974).

Let me begin by reviewing the main singularity theorem from which we can infer the existence of some form of singularity in a general closed universe model:

**THEOREM (Hawking and Penrose, 1970):** *A space-time which*

- (i) *contains no closed timelike curves.*
- (ii) *satisfies Einstein's equations (without cosmological term) and the energy condition ( $\rho + p_i \geq 0$ ,  $\rho + \sum p_i \geq 0$ ),*
- (iii) *is sufficiently general (i.e.  $t_{[a} R_{b]cd[e} t_{f]} t^c t^d \neq 0$  somewhere along each timelike or null geodesic,  $t^a$  being the tangent vector), and*
- (iv) *contains a closed spacelike hypersurface, cannot be geodesically complete in all timelike and null directions.*

The concept of geodesic incompleteness is not quite the same as that of a physical singularity. One would expect that in the actual universe, the 'reason' for the geodesic incompleteness would be that space-time curvatures become so large that the local physics becomes drastically affected – to the extent that the normal ideas of space and time might break down. But the theorem says nothing about curvatures becoming large. That is its main weakness. Nevertheless, if one applies the theorem to a universe model which is maximally extended, then the association of geodesic incompleteness

with *some* form of physical singularity seems reasonable – even if in particular cases, the singularity might be, say, of the ‘conical’ type with curvatures remaining finite right up to the singularity.

The theorem does not enable us to locate the singular points – it is very negative in this respect. Nor does it supply a definition of something that we could actually call a singular point. In fact various alternative definitions of singular points have been given, each with its own particular merits (cf. Hawking and Ellis, 1973). The definition I shall use here is basically that given by Geroch *et al.* (1972). I shall apply the definition to universe models which are closed in the sense of having no points at infinity. Thus I assume, for simplicity, that no timelike curves exist with infinite proper length.

Let me first indicate in very crude and oversimplified terms how we might expect to classify singularities broadly into three groups:

- (i) past-spacelike (big bang, white hole)
- (ii) timelike (naked singularity)
- (iii) future spacelike (black hole, final singularity).

I am including ‘null’ under the heading of ‘spacelike’ here; it is not hard to make a precise distinction but I shall not bother with it. The essential feature of a past spacelike singularity is that it supplies a past singular end-point to an otherwise past-endless timelike curve. A future spacelike singularity supplies a future singular end-point to an otherwise future-endless timelike curve. A timelike singular point may be thought of as supplying *both* a past end-point to a past-endless timelike curve and a future end-point to a future-endless timelike curve. (However, I shall not quite define things this way.) Two past-endless timelike curves  $\gamma, \eta$  are deemed to have the same (singular) past end-point if and only if they have the same futures (Figure 1). Likewise, two future-endless timelike curves  $\gamma', \eta'$  are deemed to have the same (singular) future end-point if and only if they have the same pasts.

I should like to indicate why I feel that this method of identifying singular points is rather natural from the physical point of view. Let us consider the big bang singularity of the normal cosmological models. If we think of the initial singularity as a *single* point then we have the unnatural situation that because of the existence of particle horizons, this point gives rise to an infinity of causally disconnected regions at the next instant. It seems more natural to think of the singularity as a three-dimensional spacelike surface. (This picture can be obtained by a conformal rescaling of the other one (cf. Penrose, 1968).) Each point of this surface has a distinct domain of influence, the ‘future’ of that point (Figure 2). The above definition of singular points leads naturally to such a three-dimensional description of the big bang singularity. Each point of the singular set is directly associated with the region of space-time that it can influence.

Let us see why, from this point of view, the normal black hole ( $r=0$ ) singularity of the Schwarzschild solution must be regarded as future-spacelike. The situation is depicted in Figure 3. All future-endless curves inside the event horizon must hit the singularity and no timelike curve can leave it. This singularity turns out also to be three-dimensional like that of the big bang. Timelike curves entering the singularity

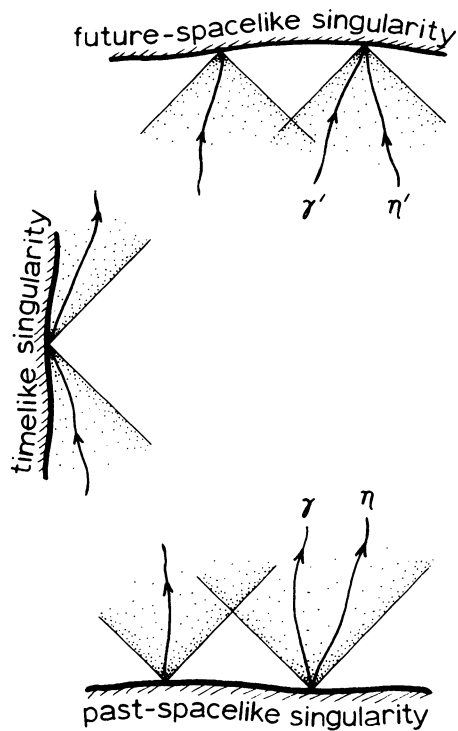


Fig. 1.

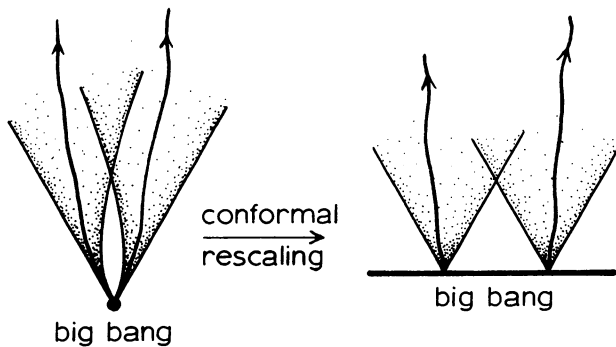


Fig. 2.

from different directions or at different 'times' will have distinct pasts. The space-like character of this singularity is perhaps at variance with one's initial intuition since there is clearly a sense in which it persists with time. Nevertheless its local spacelike nature seems inescapable. I shall indicate later a possible criterion for distinguishing the black hole type of singularity from the more cosmological sort. The situation for

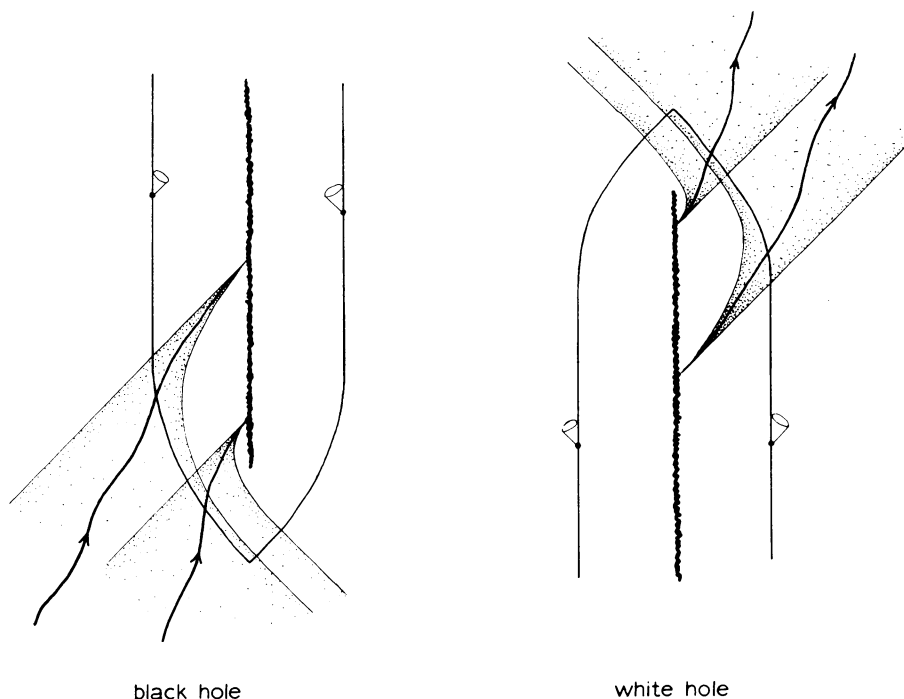


Fig. 3.

white holes is just the time-reverse of that for black holes. Again the singularity is three-dimensional and spacelike – at least, in the spherically symmetrical case.

The situation arising with a naked singularity can be formulated as follows. Suppose there is a point  $p$  in the space-time and a past-endless timelike curve  $\gamma$  whose future is contained in the future of  $p$  (Figure 4). Since  $\gamma$  is past-endless and since points at infinity have been excluded, we must assign a singular past end-point  $q$  to  $\gamma$ . But the future of  $q$  (being the future of  $\gamma$ ) lies in the future of  $p$ . Thus the singular point  $q$  itself may be reasonably thought of as lying to the (causal) future of  $p$ . Thus  $q$  is both in the future of some observer (say  $p$ ) and in the past of another (the points of  $\gamma$ ). So we may regard  $q$  as a *timelike* singular point. Indeed, since  $\gamma$  lies to the future of  $p$ , we may choose some observer (timelike curve) who starts at  $p$  and travels to some point  $r$  on  $\gamma$ . Then the singular point  $q$  lies to his future when he is at  $p$  and to his past when he is at  $r$ . This is the situation of a *naked singularity*. We may, for example, envisage a gravitational collapse taking place, where the state is initially non-singular but where the system evolves into a singular configuration, the resulting singularity being, unlike that in a black hole, visible to observers at large distances. Such singularities are called *naked*, and it is customary to rule out their existence by a hypothesis: the hypothesis of *cosmic censorship* (cf. Penrose, 1974). In fact, it turns out that if we adopt the form of this hypothesis that excludes singular points  $q$  of just the type considered above, then the space-time must be *globally hyperbolic*. This means (cf. Geroch, 1970) that

the entire space-time can be evolved from a single Cauchy hypersurface – indeed that the topology of the Universe must be unchanging for all time. For a closed universe (no points at infinity) we may regard the condition of global hyperbolicity as equivalent to cosmic censorship. The condition is actually time-symmetrical: exactly the same condition arises if, in the above,  $q$  had been taken as the singular future end-point of a future-endless timelike curve  $\gamma'$  and  $r$  as a point whose past contains the past of  $\gamma'$ .

It is perhaps hard to visualize how the topology of the Universe can be unchanging in a situation where there are black holes, and possibly also white holes, where new black holes may be forming, perhaps old white holes disappearing, and where distinct black holes may be congealing into one. In fact, the Cauchy hypersurface must have the property that it passes 'underneath' all the black hole singularities and 'above' all the white hole singularities. The situation is indicated in Figure 5. The Cauchy hypersurface remains spacelike inside the black holes' horizons owing to the 'tipping' effect on the light cones.

It should be observed that our hypotheses have not ruled out the possibility of white holes. Indeed, it is difficult to distinguish, on purely qualitative grounds, between a white hole singularity and the big bang itself. An argument has been given by Zel'dovich (1974) to the effect that white holes ought spontaneously to decay in a very short period of time owing to particle creation effects at or near the singularity. It should be emphasized that the distinction between black and white holes which seems to occur here is closely bound up with the statistical nature of time-directivity and to the asymmetric behaviour of radiation. We recall that radioactive nuclei are observed to *decay*,

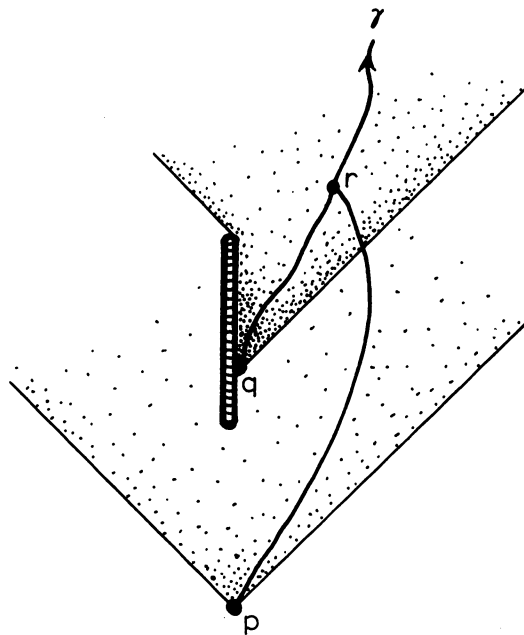


Fig. 4.

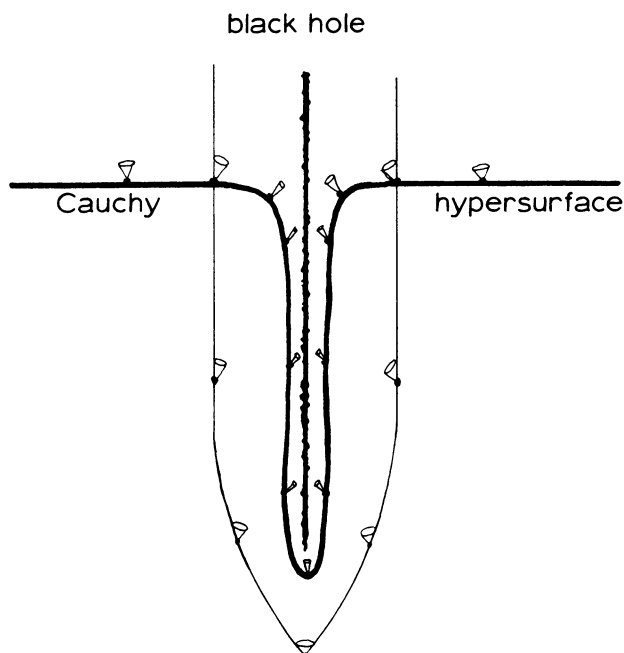


Fig. 5.

accompanied by the emission of radiation to infinity. But we do not expect such nuclei spontaneously to assemble themselves while absorbing radiation sent in from infinity. The local physical laws are the same in each case, but the type of boundary conditions that one allows are quite different for the two cases.

When the implications of this distinction between past and future are examined in the setting of a closed universe, one is inevitably confronted with the question as to whether there is an essential difference between the past and future space-time singularities. The 'blame' for any statistical difference in the boundary conditions in the past and future gets pushed back into the singularities themselves. One possibility for the distinction between past and future is that the very slight violations of time-reversal invariance in local physics which *are* observed (Casella, 1968) might play some crucial and greatly magnified role in the high curvature regimes which neighbour the singularities themselves. Another possibility that is sometimes considered (although I personally find it hard to take too seriously – particularly when examined in relation to black holes) is that the statistical time's arrow will somehow reverse itself when the Universe – assumed closed – reaches maximum expansion. Setting aside such possibilities, and still retaining, for simplicity, the picture of a closed universe, we are driven to ask how a possible difference in singularity structure between the big bang and the final recollapse might in some way be related to the statistical time's arrow.

We can look at this question in a somewhat different way. Let us imagine that

some appropriate definition of entropy can be found which refers to the whole Universe and includes gravitational effects (i.e. it does not refer just to systems of particles on a given curved background, (cf. Tolman, 1934), but the curvature itself must contribute appropriately to the entropy). In the early stages of the big bang one has a picture of something resembling thermal equilibrium – at least that is how calculations are done. So we envisage that the entropy (ignoring possible gravitational curvature contributions) is somehow almost at its maximum for the given ‘size’ that the Universe has at that time. As the Universe expands, the entropy increases, but lags behind the possible theoretical maximum that might be allowed for a universe of that size (Figure 6). Then the Universe recontracts and the theoretical maximum entropy is reduced with it. But since the entropy of the Universe has been increasing all the time, the Universe cannot reach a state resembling the compact system in thermal equilibrium in which it appeared to start out. Somehow the gravitational irregularities – apparently contributing positively to the entropy – must have grown. The model is now presumably riddled with black holes, or something similar which can be defined for a closed universe. One seems driven to some sort of concept of black hole entropy – or, at least, some sort of positive entropy residing in curvature irregularities.

In this connection one is reminded of the concept, due to Beckenstein (1973) (and the related ideas due to Bardeen *et al.*, 1974) of a *black hole entropy*  $S_{bh}$  defined by the formula

$$S_{bh} = \frac{\eta kc^3}{\hbar G} A,$$

where  $\eta$  is some numerical constant of order unity,  $k$  is Boltzman’s constant, and

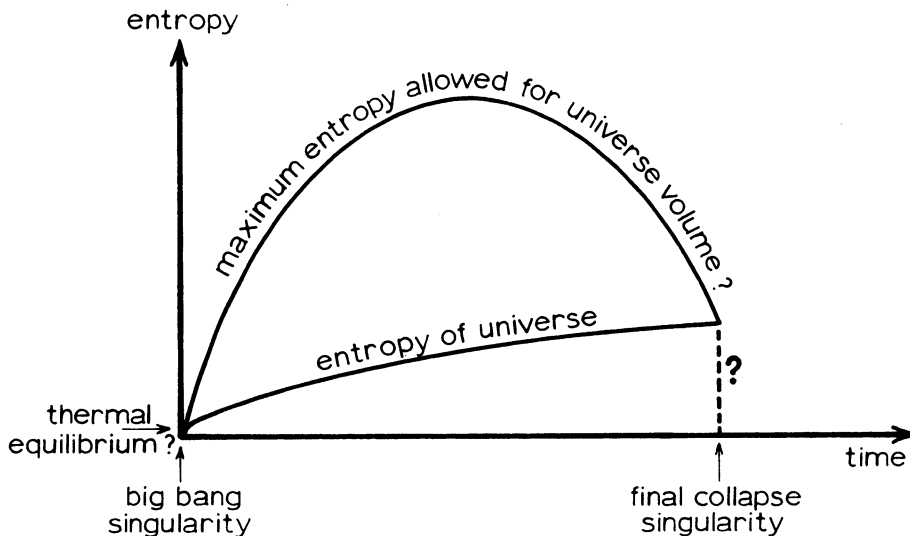


Fig. 6.

where  $A$  is the surface area of the black hole. This definition is intended for use only in asymptotically flat space-times and the formula has been the subject of some controversy. I do not propose to take sides on the issue here. I only mention it as one possible line of approach which has some bearing on the difficult and somewhat nebulous questions that I am attempting to raise.

In Figure 7 I have tried to depict how an irregular closed universe full of black holes might look from the space-time point of view, where I am assuming that the cosmic censorship principle holds and that the model is free of white holes. But from the point of view of the topological or causal structure, all the ‘stalactites’ which represent black hole singularities could be straightened out and the final singularity might appear indistinguishable from the initial singularity. How, in fact, are we to make precise, in mathematical terms, the very real distinction between the initial and final singularities? I shall just give one possible criterion, for what it is worth, and then leave the matter there.

One virtue of global hyperbolicity is that any two points with a timelike separation have a well-defined maximum time interval between them, this being the length of a maximal timelike geodesic connecting them (cf. Penrose, 1972). This time interval is actually a continuous function of the pair of points. We can extend this concept to apply also to the singular points. We can then ask, for each *given* past singular point, what is the maximum time-interval to future singular points. We find that this maximum varies little as the past singular point is varied. In a corresponding way, we can ask for the maximum interval from past singular points to some *given* future

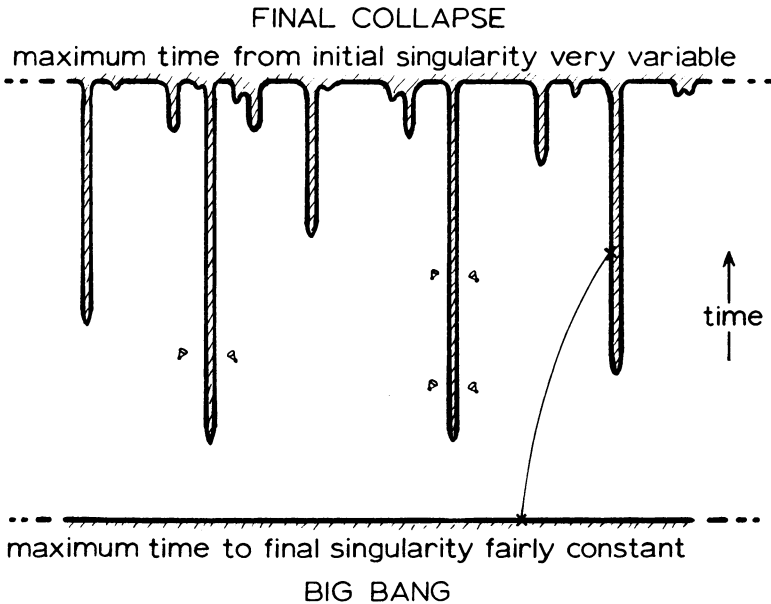


Fig. 7.



singular point. We now find that this maximum varies very greatly as the future singular point is varied. The interval will be small for a singular point on a black hole singularity which occurs early in the Universe's history but it will be large for a singular point which only results from the final recollapse of the entire universe. How one could precisely relate such a time-asymmetry in the singularity structure to the statistical asymmetry of the details of the physical universe is a very open question, however.

### References

- Bardeen, J. M., Carter, B., and Hawking, S. W.: 1974, to appear.
- Beckenstein, J. D.: 1973, *Phys. Rev.* **D7**, 2333.
- Belinskii, V. A., Khalatnikov, I. M., and Lifshitz, E. M.: 1970, *Adv. Phys.* **19**, 523.
- Belinskii, V. A., Khalatnikov, I. M., and Lifshitz, E. M.: 1972, *Soviet Phys. JETP* **62**, 1606.
- Belinskii, V. A., Khalatnikov, I. M., and Lifshitz, E. M.: 1974, this volume p. 261.
- Casella, R. S.: 1968, *Phys. Rev. Letters* **21**, 1128.
- Geroch, R.: 1970, *J. Math. Phys.* **11**, 437.
- Geroch, R., Kronheimer, E. H., and Penrose, R.: 1972, *Proc. Roy. Soc. London* **A327**, 545.
- Hawking, S. W. and Ellis, G. F. R.: 1973, *The Large Scale Structure of Space-Time*, Cambridge Univ. Press.
- Hawking, S. W. and Penrose, R.: 1970, *Proc. Roy. Soc. London* **A314**, 529.
- Khalatnikov, I. M. and Lifshitz, E. M.: 1963, *Adv. Phys.* **12**, 185.
- Penrose, R.: 1968, in C. M. DeWitt-Morette and J. A. Wheeler (eds.), *Battelle Rencontres*, Benjamin, New York.
- Penrose, R.: 1972, *Techniques of Differential Topology in Relativity*, S.I.A.M., Philadelphia.
- Penrose, R.: 1974, in C. M. DeWitt-Morette (ed.), 'Gravitational Radiation and Gravitational Collapse', *IAU Symp.* **64**, 82.
- Tolman, R. C.: 1934, *Relativity, Thermodynamics and Cosmology*, Clarendon Press, Oxford.
- Zel'dovich, Ya.: 1974, *Proc. IAU Symp. 64 on Gravitational Radiation and Gravitational Collapse*, Warsaw, 5-8 September 1973.

### DISCUSSION

*Zel'dovich*: What changes occur in the case when  $\Omega < 1$  and the Universe is open? In this case black holes also form but most of the geodesics do not end in these black holes but go to the future infinity.

*Penrose*: I discussed the question of open models in my contribution to Symposium No. 64 in Warsaw. The corresponding considerations for open models are similar to those for the closed models provided, in the former case, we consider points at infinity to be in the same category as singular points.

*Misner*: Suppose white holes are admitted. Consider the following possibility. The initial singularity contains a small white hole which will not explode out into the Universe until a late time. Before that time it has fallen within the horizon of a large black hole. Where now can the Cauchy hypersurface be fitted in?

*Penrose*: There is no difficulty about the Cauchy hypersurface entering the event horizon of the black hole and then leaving it again, since this hypersurface is spacelike. It can approach the singularity as closely as you like and then slip underneath it.

*Misner*: You referred to Zel'dovich's arguments in Symposium No. 64 to support the hypothesis that white holes may be avoided. The theoretical necessity for pair creation seems inevitable, but I do not see that this controls the epoch at which the white hole ejects matter into the external universe. Since time translations  $t \rightarrow t - t_0$  in the external Schwarzschild field correspond to space translations along the past spacelike singularity in the white hole (Schwarzschild  $r \rightarrow 0$ ), it would seem that some inhomogeneity feature of the pair creation process fixes the white hole explosion epoch, and Zel'dovich did not estimate such an inhomogeneity parameter.

*Novikov*: The Schwarzschild singularity is spacelike. Pair creation does not change this property. But pair creation near the spacelike singularity changes the initial conditions which determine the evolution of the space-time from the singularity to the future. It must change the sewing together of the singularity region and the external space. (The external space is in the absolute future from the singularity.) It can

control the epoch at which the white hole ejects matter into the external space. My calculations are not finished yet, however.

*Zel'dovich*: I don't agree with Prof. Penrose when he said that the entropy of the Universe might decrease when a black or white hole forms. The entropy of the Universe increases with time and this law does not contradict the fact that we cannot measure the entropy of matter which has already fallen into black holes. To put it in other words, we cannot consider a man to be destroyed when he has turned round the corner of a house and we don't see him.