Singularities and horizons in the collisions of gravitational waves

Ulvi Yurtsever
Theoretical Astrophysics, California Institute of Technology, Pasadena, California 91125
(Received 5 August 1988)

It is well known that when gravitational plane waves propagating and colliding in an otherwise flat background interact they produce singularities. In this paper we explore the structure of the singularities produced in the collisions of arbitrarily polarized gravitational plane waves and we consider the problem of whether (or under what conditions) singularities can be produced in the collisions of *almost-plane* gravitational waves with finite but very large transverse sizes. First we analyze the asymptotic structure of a general arbitrarily polarized colliding plane-wave spacetime near its singularity. We show that the metric is asymptotic to a generalized inhomogeneous-Kasner solution as the singularity is approached. In general, the asymptotic Kasner axes as well as the asymptotic Kasner exponents along the singularity are functions of the spatial coordinate that runs tangentially to the singularity in the non-plane-symmetric direction. It becomes clear that for specific values of these asymptotic Kasner exponents and axes the curvature singularity created by the colliding waves degenerates to a coordinate singularity, and that a nonsingular Killing-Cauchy horizon is thereby obtained. Our analysis proves that these horizons are unstable in the full nonlinear theory against small but generic plane-symmetric perturbations of the initial data, and that in a very precise and rigorous sense, "generic" initial data for colliding arbitrarily polarized plane waves always produce all-embracing, spacelike curvature singularities without Killing-Cauchy horizons. Next we turn to the problem of colliding almost-plane gravitational waves, and by combining the results that we obtain in this paper and in other previous papers with the Hawking-Penrose singularity theorem and the Cauchy stability theorem, we prove that if the initial data for two colliding almost-plane waves are sufficiently close to being exactly plane symmetric across a sufficiently large but bounded region of the initial surface, then their collision must produce spacetime singularities. Although our analysis proves the existence of these singularities rigorously, it does not give any information about either their global structure (e.g., whether they are hidden behind an event horizon) or their local asymptotic behavior (e.g., whether they are of Belinsky-Khalatnikov-Lifshitz generic-mixmaster type).

I. INTRODUCTION AND OVERVIEW

With a short paper[1] published in Nature in 1971, Khan and Penrose announced their discovery of a new exact solution to the vacuum Einstein field equations; it described the interaction between two impulsive, plane-symmetric gravitational waves, propagating and colliding in an otherwise flat background spacetime. The collision was followed by a spacetime region in which the nonlinear interaction between the waves generated a gravitational field qualitatively different from the linear superposition of the two incoming fields. In fact, the spacetime curvature generated by the collision increased without bound along all timelike world lines in the interaction region, and it ultimately diverged to form a spacetime singularity where the observers' world lines reached and terminated in finite proper time. Despite its complicated local and global structure,[2] the physical interpretation of this solution was simple: Each of the two colliding plane waves generated a spacetime geometry in its wake which acted like an infinite, perfectly converging lens,[3] focusing any radiation field which passed through the plane wave while propagating in the opposite direction. When the two plane waves collided, each of them was thus perfectly focused by the other's background geometry; diffraction effects were prevented from counterbalancing this perfect focusing by the global exact plane symmetry of spacetime. As a result, while they propagated through the interaction region the amplitude of the colliding waves grew without bound and ultimately diverged, creating a spacelike curvature singularity which bounded the interaction region in all future directions.

In the nearly two decades since the discovery of the Khan-Penrose[1] solution (and of the simultaneous discovery of other similar solutions by Szekeres[4]), the progress in the search for exact solutions describing colliding plane waves has been phenomenal, with significant contributions by many workers. Recent research in this field has particularly benefited from the carrying over of the inverse-scattering techniques for generating stationary axisymmetric solutions (one spacelike and one timelike Killing vectors) of Einstein's equations to the problem of generating plane-symmetric solutions (two commuting spacelike Killing vectors). For a brief description of the history of these developments and a (necessarily incomplete) list of references, we refer the readers to Refs. 5 and 6 (especially Sec. I of Ref. 5 and Sec. I of Ref. 6) and to the references cited therein.

In our view the greatest significance of the problem of colliding gravitational waves lies not with those aspects of it that are peculiar to specific exact solutions, but rather with its potential to provide insight into some of the broader issues in general relativity (such as cosmic censorship, structure of singularities, ...) which arise naturally in studying the dynamics of fully nonlinear gravitational fields. From this point of view, gravitational-wave collisions can be considered as the vacuum analogues of gravitational collapse, and as such they provide a framework in which issues such as cosmic censorship can be discussed without the undue complications of a specifically chosen nonzero stress-energy tensor. In fact, we contend that among all the issues raised by the last two decades of exact-solution research on colliding plane waves the following two are the most important, and that owing to their inherent generality these issues are not likely to be completely resolved by work on exact solutions alone.

On the one hand, thanks to the work of Chandrasekhar and Xanthopoulos[7] who first discovered this phenomenon, we now know that colliding plane waves do not always create spacelike curvature singularities with a global structure similar to the singularity of the Khan-Penrose solution: for some choices of the incoming plane waves, their collision produces a nonsingular Killing-Cauchy horizon[8] at the points where ordinarily one would expect curvature singularities to form. The spacetime can then be extended smoothly across this horizon (in nonunique ways) to obtain several inequivalent, maximal solutions, which all evolve from the same initial data posed by the incoming, colliding plane waves (breakdown of predictability). It is therefore of fundamental importance to determine (i) under what conditions on the initial data (the incoming plane waves) the collision creates singularities and under what conditions it creates horizons, (ii) what are the local structures of the singularities and horizons thus created, and (iii) whether "generic" initial data (with respect to some appropriate notion of genericity) always produce "pure" spacetime singularities without Killing-Cauchy horizons, i.e., whether any breakdowns in global predictability can occur in "generic" gravitational plane-wave collisions. The issue here is then that of the *structure* of singularities produced by colliding plane waves.

On the other hand, it is natural to raise the issue of whether (or under what conditions) spacetime singularities can be produced by the collisions of gravitational waves which are not exactly plane symmetric, but which have finite but very large transverse "spatial" sizes; i.e., by the collisions of *almost-plane* gravitational waves. This second issue is then that of the *existence* (and possibly also the structure) of singularities created in the collisions of almost-plane gravitational waves.

In a series of two papers published previously in this journal (Refs. 6 and 9), we attempted to resolve the above issues in the special case where the colliding waves had parallel constant-linear polarizations. Thus, in Ref. 6 we showed that the asymptotic structure of a colliding parallel-polarized plane-wave spacetime near its singularity can be completely and explicitly determined in terms of the initial data posed by the incoming waves. Our analysis proved that although Killing-Cauchy horizons can be produced in the collisions of parallel-polarized plane waves, these horizons are unstable in the full nonlinear theory against small but generic plane-symmetric perturbations of the initial data, and that in a very precise sense, "generic" initial data always produce all-embracing, spacelike curvature singularities without Killing-Cauchy horizons. In Ref. 9, we analyzed the collision between two almost-plane gravitational waves whose initial data across a bounded region of the initial surface were identical with the initial data posed by colliding parallel-polarized exactly plane waves, but fell off in an arbitrary way at larger transverse distances. We proved that if this bounded region of exact plane symmetry in the initial surface is sufficiently large, then the collision between the almost-plane waves is guaranteed to produce a spacetime singularity with the same local structure as in an exact plane-wave collision.

The work described in the present paper is a continuation of the work reported in Refs. 6 and 9. The main results of this paper are (i) the generalization of the results of Refs. 6 and 9 to the case where the polarizations of the colliding waves are entirely arbitrary (i.e., neither parallel nor constant linear), and (ii) the proof of a much stronger version of the singularity theorem of Ref. 9; specifically, that if the initial data for two colliding almost-plane waves are *sufficiently close* to being exactly plane symmetric across a sufficiently large but bounded region of the initial surface, then their collision must produce spacetime singularities. Sections II and III and Sec. IV A below describe the above-mentioned generalization of the results of Ref. 6 and Ref. 9, respectively, whereas Sec. IV B is devoted to the new singularity theorem. The five appendixes at the end of the paper deal with a number of issues of a more technical nature that are raised during the course of the analyses in Secs. II–IV. We note, however, that these appendixes (especially Appendixes A, C, and D) contain a large amount of information, some of which might be useful in future research on questions that are left unresolved in this paper. We feel that any serious reading of the paper *must* include at least the three Appendixes A, C, and D.

The more precise plan of this paper is as follows.

In Sec. II A, we give a very brief review of Szekeres's[4] formulation of the field equations and the characteristic initial-value problem for colliding arbitrarily polarized plane waves, in the (u,v,x,y) coordinate system which we call "Rosen-type" and which is tuned to the plane symmetry of the spacetime. This formulation is entirely analogous to the corresponding formulation for the parallel-polarized case which we have discussed in Sec. II A of Ref. 6. Consequently, here we only present the essential facts and formulas that will be needed in later sections, and refer the reader to Sec. II A of Ref. 6 for the details of their derivation and meaning. In this section and throughout the paper, we try to maintain as much parallelism as possible between our presentation here and the presentation in Refs. 6 and 9. For this reason, the readers may find it helpful to carry along and look at these two previous papers[6,9] while reading the present paper.

In Sec. II B, we perform a coordinate transformation to a new (α,β,x,y) coordinate system, in which the field equations and the initial-value problem associated with them take simpler forms. Again the construction and the properties of this new coordinate system are straightforward generalizations of the construction and properties of the (α,β) coordinates discussed in Sec. II B of Ref 6. However, while the field equations for colliding parallel-polarized plane waves (Sec. II B of Ref. 9) reduced in the (α,β) coordinates to a single *linear* hyperbolic equation for which an explicit Riemann function could be found,[4,6] in the general case the simplification achieved by this coordinate change, though substantial, is not as great: The field equations in the (α,β) coordinates reduce to a system of *nonlinear*, coupled hyperbolic partial differential equations (PDE's) for two functions which represent the dimensionless amplitudes for the two independent modes of polarization. Although at present it seems unlikely (because of their high nonlinearity) that an explicit general solution (Riemann function) can be found for these equations, in Appendix C we discuss some interesting and suggestive aspects of this particular system of nonlinear PDE's which might later prove useful in the search for such a general solution. A further disturbing consequence of this fundamental nonlinearity in the field equations for colliding nonparallel-polarized plane waves is that the global existence and uniqueness of their solutions may not be guaranteed. In the parallel-polarized case, it is guaranteed by the linearity of the single nontrivial field equation that there exists a unique, global solution defined throughout the domain of dependence of the initial surface, i.e., throughout the entire interaction region up to the "singularity" $\{\alpha=0\}$ at which either spacetime singularities or Killing-Cauchy horizons form (Secs. II B and III A of Ref. 6). In contrast, the field equations in the nonparallel-polarized case are nonlinear, and it is well known that solutions of nonlinear hyperbolic PDE's do not in general exist globally. This raises the possibility that solutions of the field equations might break down at points which lie *within* the interaction region *before* the "singular" surface $\{\alpha=0\}$, and consequently the possibility that colliding nonparallel-polarized plane waves might create spacetime singularities in the region where $\alpha>0$; such singularities, if present, would not be treatable by analyzing the asymptotic structure of spacetime near $\alpha=0$. Fortunately, however, a careful analysis which we undertake in Appendix A shows that thanks to some very special properties possessed by the field equations, the global existence and uniqueness of their solutions can be proved despite the presence of strong nonlinearities. Therefore, the singularities and horizons created by colliding plane waves always lie on or beyond the surface $\{\alpha=0\}$.

Our discussions in Sec. II B and in Appendix A bring us to the analysis of the asymptotic structure of spacetime near $\alpha=0$. Relying on the results of Appendix B which show that as $\alpha\rightarrow 0$ the spatial-derivative terms in the field equations are asymptotically negligible compared to the α-derivative terms, we begin Sec. III A by studying the ordinary differential equations that are obtained by eliminating the spatial β-derivative terms from the field equations; this allows us to determine the asymptotic behavior of the metric functions near $\alpha=0$. We show that the spacetime metric asymptotically approaches a generalized inhomogeneous Kasner[10] solution as α approaches zero, where the time coordinate t of the asymptotic Kasner spacetime is monotonically related to α, and the Kasner singularity at $t=0$ corresponds to the singularity at $\alpha=0$. We call this asymptotic inhomogeneous-Kasner structure "generalized" because unlike the parallel-polarized case in which the asymptotic Kasner exponents were associated with the fixed set of axes $\{x,y\}$ throughout the singularity (Sec. III A of Ref. 6), here in general the asymptotic Kasner axes are linear combinations of x,y and they vary across the singularity as functions of the spatial coordinate β. Since we do not have a general solution for the field equations in the nonparallel-polarized case, in contrast to Sec. III A of Ref. 6 we cannot in general relate the asymptotic Kasner exponents and/or axes along the singularity to the initial data posed along the wave fronts of the incoming, colliding plane waves. (See, however, Appendix C where one such relation is obtained in a special case.) As in Ref. 6, in general these asymptotic Kasner exponents as well as the asymptotic Kasner axes depend on β, the spacelike coordinate running along the nontrivial spatial (z) direction in the spacetime.

We begin Sec. III B with a discussion of Tipler's theorem,[11,12] which proves that in any vacuum, nonflat plane-symmetric spacetime there must exist either a spacetime singularity (where null geodesics terminate) or a Killing-Cauchy horizon (where the strict plane symmetry of spacetime breaks down). We note that the content of Tipler's theorem is made particularly transparent by our analysis of the asymptotic structure of colliding plane-wave spacetimes. On the one hand, it becomes clear from our discussion in Sec. III A that the asymptotic Kasner exponents and axes (throughout a connected interval in the spatial coordinate β) may take on the values associated with a degenerate Kasner solution. Since a degenerate Kasner spacetime is flat and possesses a Killing-Cauchy horizon at $t=0$ instead of a singularity, it follows that when the asymptotic Kasner exponents for the colliding plane-wave metric are degenerate a nonsingular Killing-Cauchy horizon forms at $\alpha=0$ across which spacetime can be extended smoothly. On the other hand, it is easily seen from the expressions of the Newman-Penrose curvature quantities in the (α,β) coordinates that if the asymptotic Kasner exponents are *nondegenerate*, then $\alpha=0$ is a curvature singularity. Next we observe that when a Killing-Cauchy horizon forms at $\alpha=0$, the spacetime can be extended through it in infinitely many different ways; the geometry beyond the horizon cannot be determined from the initial data posed by the incoming, colliding plane waves. We then briefly recall our earlier work in Ref. 8, where we proved general theorems stating the instability of such Killing-Cauchy horizons in any plane-symmetric spacetime against generic, plane-symmetric perturbations. For the special case of the Killing-Cauchy horizons which occur in collisions of parallel-polarized plane waves, our discussions in Sec. III C of Ref. 6 proved that in fact these instabilities render the set of "all" horizon-producing initial data "nongeneric" with respect to a very precise notion of nongenericity. More specifically, our analysis in Ref. 6 proved that the subset of all initial data which produce at least one connected Killing-Cauchy horizon larger than Planck size is nongeneric within the set of all colliding parallel-polarized plane-wave initial data. Correspondingly, by making use of the discussions in Appendixes A and B, we prove in Sec. III B the generalization of this result (with the same notion of genericity as in Ref. 6) to the case of colliding arbitrarily polarized plane waves. In addition, by introducing a more sophisticated notion of genericity which we describe in greater detail in Appendix D, we prove that the subset of *all* horizon-producing initial data (and not just the subset of those data which produce horizons larger than Planck size) is nongeneric within the set of all initial data for colliding plane waves. We also discuss why we believe that our topological notion of genericity (described in Appendix D) is more appropriate in general relativity than other possible "probabilistic" notions based on measure theory.

In Sec. IV A, using the conclusions we obtained in the previous sections, we prove the generalization of the singularity result that was proved for parallel-polarized colliding almost-plane waves in Sec. II of Ref. 9 to the case of colliding almost-plane waves with arbitrary polarizations. More specifically, we prove that if the initial data posed by two colliding almost-plane gravitational waves are (i) identical with the initial data posed by two colliding exactly plane waves (with arbitrary polarizations) across a bounded but sufficiently large region of the initial surface, and (ii) fall off in an arbitrary way (consistent with the constraint equations) at larger transverse distances, then the collision between the almost-plane waves is guaranteed to produce a spacetime singularity with the same local structure as in an exact plane-wave collision.

In Sec. IV B, we combine the Hawking-Penrose singularity theorem (Ref. 13 and Sec. 8.2 of Ref. 14), the Cauchy stability theorem,[15] and a lemma about the null cones in a nondegenerate Kasner spacetime which we discuss in Appendix E, to prove that the conclusion of the singularity theorem of Sec. IV A about the *existence* of singularities remains valid when the colliding almost-plane waves are not *exactly* plane-symmetric over any region, but are only *approximately* plane-symmetric across their central regions. In other words, we prove that if the initial data for two colliding almost-plane waves are *sufficiently close* to being exactly plane symmetric across a sufficiently large but bounded region of the initial surface, then their collision must produce spacetime singularities. Although our analysis proves the existence of these singularities rigorously, it does not give any information about either their global structure (e.g., whether they are hidden behind an event horizon) or their local asymptotic behavior (e.g., whether they are of Belinsky-Khalatnikov-Lifshitz[10] generic-mixmaster type).

Our notation and other conventions throughout this paper are the same as in Refs. 6 and 9. Equation numbers that refer to equations of Refs. 6 or 9 will be denoted by a prefix 6 or 9; for example, Eq. (6.3.13) and Eq. (9.2.6) refer, respectively, to Eq. (3.13) of Ref. 6 and Eq. (2.6) of Ref. 9.

As in our previous papers,[6,9] here we are concerned exclusively with the collisions of purely gravitational (vacuum) waves. Whether the conclusions of Secs. II and III in this paper remain valid in the presence of matter fields coupled to the colliding plane waves is an interesting and unexplored question.

II. FIELD EQUATIONS FOR COLLIDING GRAVITATIONAL PLANE WAVES

A. Formulation of the problem in the Rosen-type (u,v,x,y) coordinate system

In any plane-symmetric spacetime (see Sec. III B of Ref. 12, or Sec. II of Ref. 8 for a careful definition of plane symmetry), there exists a canonical null tetrad[16] whose construction is described in Sec. III B of Ref. 12. In this null tetrad, which we call the standard tetrad, $\mathbf{l}$ and $\mathbf{n}$ are tangent to the two null geodesic congruences everywhere orthogonal to the plane-symmetry generating Killing vector fields ξ_1 and ξ_2, and $\mathbf{m}$ and its complex conjugate are linear combinations of the ξ_i, $i=1,2$. As is shown by Szekeres in Ref. 4 and discussed briefly in Sec. II A of Ref. 6, the special geometry of a colliding plane-wave spacetime allows us to find a local coordinate system (u,v,x,y) in which $\xi_i=\partial/\partial x^i$ $[(x^1,x^2)\equiv(x,y)]$, and in which the standard tetrad can be brought into the form

$$\mathbf{l}=2e^{M}\frac{\partial}{\partial u}\ ,\qquad \mathbf{n}=\frac{\partial}{\partial v}\ ,$$
$$\mathbf{m}=N_1\frac{\partial}{\partial x}+N_2\frac{\partial}{\partial y}\ , \tag{2.1}$$

with

$$N_1=\frac{1}{\sqrt{2}}e^{(U-V)/2}\sqrt{\cosh W}\times\exp\{\tfrac{1}{2}i[\arcsin(\tanh W)]\}\ ,$$
$$N_2=\frac{i}{\sqrt{2}}e^{(U+V)/2}\sqrt{\cosh W}\times\exp\{-\tfrac{1}{2}i[\arcsin(\tanh W)]\}\ , \tag{2.2}$$

where M, U, V, and W are real functions of u and v only. (Notice the slight phase difference between our choice for N_1 and N_2 here and that in Sec. II A of Ref.6 [Eqs. (6.2.4)]. The only equations in this paper that are affected by this discrepancy are the expressions for the Newman-Penrose curvature quantities [Eqs. (2.12) below] which differ from the corresponding expressions in Ref. 6 [Eqs. (6.2.19)] by factors of 2 or i.) The null tetrad given by Eqs. (2.1) and (2.2) gives rise to the metric

$$g=-e^{-M}du\,dv+e^{-U}[\cosh W(e^{V}dx^2e^{-V}dy^2)-2\sinh W\,dx\,dy]\ . \tag{2.3}$$

Thus, the functions $V(u,v)$ and $W(u,v)$ represent the dimensionless amplitude of the two independent polarization modes in the gravitational radiation field (2.3).

The vacuum Einstein field equations for the metric (2.3) can be written in the form[4]

$$2(U_{,uu}+M_{,u}U_{,u})-U_{,u}{}^2-V_{,u}{}^2\cosh^2W-W_{,u}{}^2=0\ , \tag{2.4a}$$

$$2(U_{,vv}+M_{,v}U_{,v})-U_{,v}{}^2-V_{,v}{}^2\cosh^2W-W_{,v}{}^2=0\ , \tag{2.4b}$$

$$U_{,uv}-U_{,u}U_{,v}=0\ , \tag{2.4c}$$

$$V_{,uv}-\tfrac{1}{2}(U_{,u}V_{,v}+U_{,v}V_{,u})+(V_{,u}W_{,v}+V_{,v}W_{,u})\tanh W=0\ , \tag{2.4d}$$

$$W_{,uv}-\tfrac{1}{2}(U_{,u}W_{,v}+U_{,v}W_{,u})-V_{,u}V_{,v}\sinh W\cosh W=0\ , \tag{2.4e}$$

where the integrability condition for the first two equations is satisfied by virtue of the last three, and yields the remaining field equation

$$M_{,uv}-\tfrac{1}{2}(V_{,u}V_{,v}\cosh^2W-U_{,u}U_{,v})-\tfrac{1}{2}W_{,u}W_{,v}=0\ . \tag{2.5}$$

It is sufficient to solve Eqs. (2.4c)–(2.4e) first and to obtain M by quadrature from the first two equations (2.4a) and (2.4b) afterward, since Eq. (2.5) as well as the integrability condition for Eqs. (2.4a) and (2.4b) are automatically satisfied as a result of Eqs. (2.4c)–(2.4e).

The initial-value problem associated with the field equations (2.4) and (2.5) is best formulated in terms of initial data posed on null (characteristic) surfaces. A natural choice for the initial characteristic surface is the surface made up of the two intersecting null hyperplanes which form the initial wave fronts of the incoming plane waves, and which, by a readjustment of the null coordinates u and v if necessary, can be arranged to be the surfaces $\{u=0\}$ and $\{v=0\}$. The geometry of the resulting characteristic initial-value problem is depicted in Fig. 1. The initial data supplied by the plane wave propagating in the v direction (to the right in Fig. 1) are posed on the $u\geq 0$ portion of the surface $\{v=0\}$, and the initial data supplied by the plane wave propagating in the u direction (to the left in Fig. 1) are posed on the $v\geq 0$ portion of the surface $\{u=0\}$. In region IV, which represents the spacetime before the passage of either plane wave, the geometry is flat and all metric coefficients M, U, V, and W vanish identically. Now recall our discussions in Sec. II A of Ref. 6 about the gauge freedom in the choice of the (u,v,x,y) coordinate system, and about how this freedom manifests itself in the choice of initial data on the characteristic initial surface $\{u=0\}\cup\{v=0\}$. For exactly the same reasons as described in those discussions, here as well as in Ref. 6, the choice of the initial data $\{M(u=0,v),\ M(u,v=0)\}$ for the metric function M is completely arbitrary. As we did in Ref. 6, we will fix this gauge freedom once and for all by posing our initial data so that

$$M(u=0,v)=M(u,v=0)\equiv 0\ . \tag{2.6}$$

After making this gauge choice, it becomes clear from the field equations (2.4) that the initial data on $\{u=0\}\cup\{v=0\}$ are completely determined by only the four freely specifiable functions $V_1(u)\equiv V(u,v=0)$, $W_1(u)\equiv W(u,v=0)$, $V_2(v)\equiv V(u=0,v)$, and $W_2(v)\equiv W(u=0,v)$. In other words, the initial data consist of

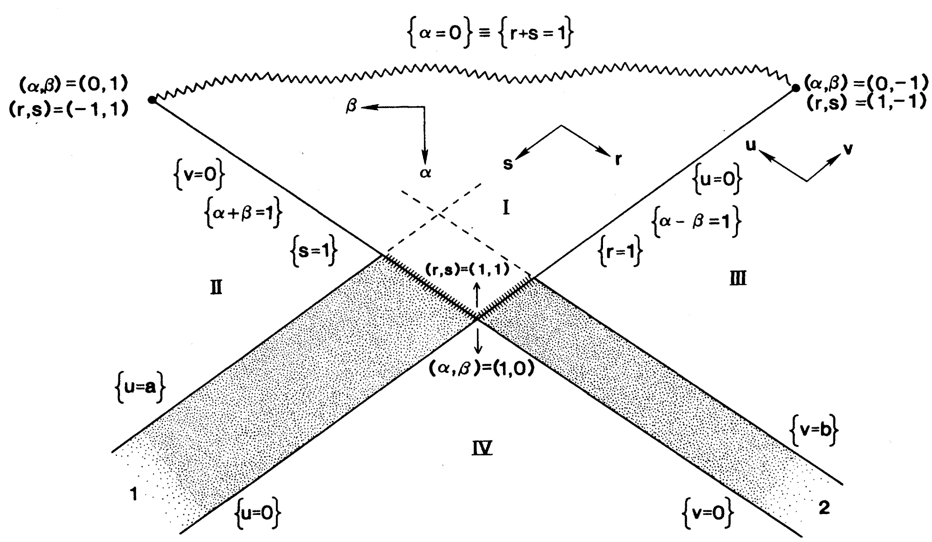

FIG. 1. The two-dimensional geometry of the characteristic initial-value problem for colliding plane waves. The null surfaces $\{u=0\}$ and $\{v=0\}$ are the past wave fronts of the incoming plane waves 1 and 2. Initial data corresponding to waves 1 and 2 are posed, respectively, on the upper portions of the surfaces $\{v=0\}$ and $\{u=0\}$ that are adjacent to the interaction region I. The geometry in region IV is flat, and the geometry in regions II and III is given by the metric describing the incoming waves 1 and 2, respectively. The geometry of the interaction region I is uniquely determined by the solution of the above initial-value problem. The directions in which the various lines of constant coordinates u, v, α, β, r, and s run are also indicated, along with the descriptions of the initial null surfaces in these different coordinate systems.

$$\{V_1(u),\ W_1(u),\ V_2(v),\ W_2(v)\}\ , \tag{2.7}$$

where $V_1(u)$, $W_1(u)$ and $V_2(v)$, $W_2(v)$ are C^1 (and piecewise C^2) functions for $u \geq 0$ and $v \geq 0$, respectively, which are freely specified except for the initial conditions $V_1(u=0)=W_1(u=0)=V_2(v=0)=W_2(v=0)=0$. The remaining functions $U_1(u) \equiv U(u,v=0)$ and $U_2(v) \equiv U(u=0,v)$ which specify the initial values of the metric function $U(u,v)$ are uniquely determined, by the initial data (2.7), through the constraint equations [cf. Eqs. (2.4a) and (2.4b)]

$$2U_{1,uu} - U_{1,u}{}^2 = V_{1,u}{}^2\cosh^2 W_1 + W_{1,u}{}^2\ , \tag{2.8a}$$

$$2U_{2,vv} - U_{2,v}{}^2 = V_{2,v}{}^2\cosh^2 W_2 + W_{2,v}{}^2\ , \tag{2.8b}$$

with the initial conditions $U_1(u=0)=U_2(v=0)=0$, $U_{1,u}(u=0)=U_{2,v}(v=0)=0$. Note that, if we define two new functions $f(u)$ and $g(v)$ by

$$f(u) \equiv e^{-U_1(u)/2}\ , \qquad g(v) \equiv e^{-U_2(v)/2}\ , \tag{2.9}$$

we can express Eqs. (2.8) in the form of "focusing" equations:

$$\frac{f_{,uu}}{f} = -\tfrac{1}{4}(V_{1,u}{}^2\cosh^2 W_1 + W_{1,u}{}^2)\ , \tag{2.10a}$$

$$\frac{g_{,vv}}{g} = -\tfrac{1}{4}(V_{2,v}{}^2\cosh^2 W_2 + W_{2,v}{}^2)\ , \tag{2.10b}$$

with the initial conditions $f(0)=g(0)=1$, $f'(0)=g'(0)=0$. It immediately follows from Eqs. (2.10) and (2.9) that

$$\begin{aligned} &f(u)<1\ , \quad f'(u)<0 \quad \forall\, u>0\ ,\\ &g(v)<1\ , \quad g'(v)<0 \quad \forall\, v>0\ , \end{aligned} \tag{2.11a}$$

$$\begin{aligned} &U_1(u)>0\ , \quad U'_1(u)>0 \quad \forall\, u>0\ ,\\ &U_2(v)>0\ , \quad U'_2(v)>0 \quad \forall\, v>0\ , \end{aligned} \tag{2.11b}$$

as long as the initial data (2.7) are nontrivial for both incoming waves [i.e., as long as neither $V_1(u)$ and $W_1(u)$ nor $V_2(v)$ and $W_2(v)$ are identically zero], and as long as the initial surfaces $\{u=0\}$ and $\{v=0\}$ correspond to the true initial wave fronts of the colliding waves [i.e., as long as either $V_1(u)\neq 0$ or $W_1(u)\neq 0$ and either $V_2(v)\neq 0$ or $W_2(v)\neq 0$ for *all* sufficiently small but positive u and v], both of which conditions we will always assume throughout this paper.

In Secs. III A and III B below, when we discuss the asymptotic structure of the colliding plane-wave spacetime described by Eqs. (2.1)–(2.3), we will need the following equations which express the Newman-Penrose[16] curvature quantities in the null tetrad (2.1) and (2.2) in terms of the metric coefficients M, U, V, and W; the derivation of these equations can be found in Ref. 4:

$$\begin{aligned} \Psi_0 = -2e^{2M}\{&[2V_{,u}W_{,u}\sinh W - V_{,u}(U_{,u}-M_{,u})\cosh W + V_{,uu}\cosh W]\\ &-i[W_{,uu}-(U_{,u}-M_{,u})W_{,u} - V_{,u}{}^2\sinh W\cosh W]\}\ , \end{aligned} \tag{2.12a}$$

$$\Psi_2 = e^{M}[M_{,uv} - i(V_{,v}W_{,u} - V_{,u}W_{,v})\cosh W]\ , \tag{2.12b}$$

$$\begin{aligned} \Psi_4 = -\tfrac{1}{2}\{&[2V_{,v}W_{,v}\sinh W - V_{,v}(U_{,v}-M_{,v})\cosh W + V_{,vv}\cosh W]\\ &+i[W_{,vv}-(U_{,v}-M_{,v})W_{,v} - V_{,v}{}^2\sinh W\cosh W]\}\ , \end{aligned} \tag{2.12c}$$

$$\Psi_1 = \Psi_3 = 0\ . \tag{2.12d}$$

B. Field equations in the (α,β) coordinates

We now construct a new coordinate system in which the field equations and the initial-value problem associated with them take simpler forms. The construction and the properties of this new coordinate system are straightforward generalizations of the construction and properties of the (α,β) coordinates discussed in Sec. II B of Ref. 6. Consequently, here we will be somewhat concise in our presentation and refer the reader to Ref. 6 for details.

Consider the interaction region (region I in Fig. 1) where $u \geq 0$ and $v \geq 0$. This region is the *domain of dependence*[14] of the characteristic initial surface $\{u=0\}\cup\{v=0\}$, on which the initial-value problem defined by Eqs. (2.4) and (2.6)–(2.8) is to be solved. Consider the field equation (2.4c) in the interaction region. It follows from this equation that if we define

$$\alpha(u,v) \equiv e^{-U(u,v)}\ , \tag{2.13}$$

then, throughout the interaction region, $\alpha(u,v)$ satisfies

$$\alpha_{,uv} = 0\ , \tag{2.14}$$

the flat-space wave equation in two dimensions. Equation (2.14) suggests that we define the complementary variable, $\beta(u,v)$, such that

$$\beta_{,u} = -\alpha_{,u}\ , \qquad \beta_{,v} = \alpha_{,v}\ . \tag{2.15}$$

Clearly, the integrability condition for Eqs. (2.15) is satisfied by virtue of Eq. (2.14). The initial-value problem for $\alpha(u,v)$ is easily solved, and when combined with Eq. (2.15), it yields the expressions[6]

$$\alpha(u,v) = e^{-U_1(u)} + e^{-U_2(v)} - 1\ , \tag{2.16a}$$

$$\beta(u,v) = e^{-U_2(v)} - e^{-U_1(u)}\ , \tag{2.16b}$$

which complete the construction of the new variables (α,β). To see that these variables actually define a new coordinate system, note that, by Eqs. (2.16),

$$d\alpha \wedge d\beta = 2U_1'(u)U_2'(v)e^{-[U_1(u)+U_2(v)]}\, du \wedge dv \; . \tag{2.17}$$

Therefore, from Eqs. (2.11b), (2.17), and the inverse function theorem,[17] it follows that the functions (α,β,x,y) constitute a regular coordinate system wherever the coordinate system (u,v,x,y) is regular in the interior of the interaction region, where $u>0$, $v>0$. On the other hand, by Eq. (2.17) and the initial conditions for Eqs. (2.8), the coordinates α,β are singular along the initial null surfaces $\{u=0\}$ and $\{v=0\}$. In other words, the singularities of the coordinate system (α,β,x,y) consist of the singularities of the (u,v,x,y) coordinates (when there are any), and the singularity along the initial characteristic surface $\{u=0\}\cup\{v=0\}$. Since the only place in the interaction region where the coordinates (u,v,x,y) can develop singularities is the "surface" $\{\alpha=0\}$ (see Sec. III A), it follows that the coordinate system (α,β,x,y) covers the domain of dependence of the initial surface $\{u=0\}\cup\{v=0\}$ regularly except for the singularities on $\{u=0\}$ and $\{v=0\}$.

The coordinates (α,β,x,y) enjoy a number of properties which make them useful in studying the field equations for colliding plane waves. We will not list these properties here as they are discussed in detail in Sec. II B of Ref. 6; instead, we will proceed directly with the analysis of the initial-value problem (2.4) and (2.6)–(2.8) in the new coordinate system (α,β,x,y). First we note the transformation rules

$$\partial_u = \alpha_{,u}(\partial_\alpha - \partial_\beta) \; , \tag{2.18a}$$

$$\partial_v = \alpha_{,v}(\partial_\alpha + \partial_\beta) \; , \tag{2.18b}$$

and their inverses

$$\partial_\alpha = \frac{1}{2}\left[\frac{1}{\alpha_{,u}}\partial_u + \frac{1}{\alpha_{,v}}\partial_v\right] , \tag{2.19a}$$

$$\partial_\beta = \frac{1}{2}\left[\frac{1}{\alpha_{,v}}\partial_v - \frac{1}{\alpha_{,u}}\partial_u\right] , \tag{2.19b}$$

which are derived using Eq. (2.15). [For our notation, see the explanations following Eqs. (6.2.31) and (6.2.34) in Ref. 6.] A short computation involving Eqs. (2.18) and (2.19) now gives

$$-du\, dv = \frac{1}{4\alpha_{,u}\alpha_{,v}}(-d\alpha^2 + d\beta^2) \; . \tag{2.20}$$

When inserted into Eq. (2.3) and combined with Eq. (2.13), Eq. (2.20) yields the expression

$$g = \frac{e^{-M}}{4\alpha^2 U_{,u} U_{,v}}(-d\alpha^2 + d\beta^2) + \alpha[\cosh W(e^V dx^2 + e^{-V}dy^2) - 2\sinh W\, dx\, dy] \tag{2.21}$$

for the spacetime metric, which is valid throughout the interaction region (region I in Fig. 1). Next, another short calculation using Eqs. (2.18) and (2.19) together with Eq. (2.14) gives

$$\partial_\alpha{}^2 - \partial_\beta{}^2 = \frac{1}{\alpha_{,u}\alpha_{,v}}\partial_u\partial_v \; . \tag{2.22}$$

Combining Eq. (2.22) with the field equations (2.4d) and (2.4e) and using Eqs. (2.18) and (2.19), we obtain the field equations satisfied by the amplitudes V and W in the (α,β,x,y) coordinate system:

$$V_{,\alpha\alpha} + \frac{1}{\alpha}V_{,\alpha} - V_{,\beta\beta} = 2(V_{,\beta}W_{,\beta} - V_{,\alpha}W_{,\alpha})\tanh W \; , \tag{2.23a}$$

$$W_{,\alpha\alpha} + \frac{1}{\alpha}W_{,\alpha} - W_{,\beta\beta} = (V_{,\alpha}{}^2 - V_{,\beta}{}^2)\sinh W \cosh W \; . \tag{2.23b}$$

To obtain the remaining field equations, we proceed as follows: First we note that after defining a new function P by

$$e^P \equiv 4c\, e^M U_{,u} U_{,v} \; , \tag{2.24}$$

where c is an arbitrary constant having the dimensions of $(\text{length})^2$ [we will fix c later with our normalization condition Eq. (2.28)], we can rewrite the field equations (2.4a) and (2.4b) in the form

$$2P_{,u} = 3U_{,u} + \frac{1}{U_{,u}}(V_{,u}{}^2\cosh^2 W + W_{,u}{}^2) \; , \tag{2.25a}$$

$$2P_{,v} = 3U_{,v} + \frac{1}{U_{,v}}(V_{,v}{}^2\cosh^2 W + W_{,v}{}^2) \; . \tag{2.25b}$$

Combining Eqs. (2.25) with Eqs. (2.18) and using Eq. (2.13) we obtain, after some rearrangements,

$$(2P + 3\ln\alpha)_{,\alpha} = -\alpha[(V_{,\alpha}{}^2 + V_{,\beta}{}^2)\cosh^2 W + W_{,\alpha}{}^2 + W_{,\beta}{}^2] \; , \tag{2.26a}$$

$$(2P + 3\ln\alpha)_{,\beta} = -2\alpha(V_{,\alpha}V_{,\beta}\cosh^2 W + W_{,\alpha}W_{,\beta}) \; . \tag{2.26b}$$

Equations (2.26) suggest that it will be convenient to define the combination $2P+3\ln\alpha$ as a new variable, which, together with the variables V and W, would uniquely determine the metric in the (α,β,x,y) coordinate system. Thus, after first introducing the two "normalization" length scales l_1 and l_2 by the equations

$$l_1 \equiv \frac{1}{2U_{,u}(u_0,v_0)} \; , \quad l_2 \equiv \frac{1}{2U_{,v}(u_0,v_0)} \; , \tag{2.27a}$$

where (u_0,v_0), $u_0>0$, $v_0>0$ is an arbitrary, fixed point in the *interior* of the interaction region, we define a new function $Q(\alpha,\beta)$ by the relation

$$e^{Q/2} \equiv 4l_1 l_2\, e^M\, U_{,u} U_{,v}\, \alpha^{3/2} \; . \tag{2.27b}$$

Using Eqs. (2.27a), we then fix the constant c which occurs in Eq. (2.24):

$$c \equiv l_1 l_2 \; . \tag{2.28}$$

Note that the length scales l_1 and l_2 are determined by Eqs. (2.27a) in a well-defined manner, since by Eqs. (2.13) and (2.16a)

$$U(u,v) = -\ln\alpha(u,v) = -\ln(e^{-U_1(u)} + e^{-U_2(v)} - 1) , \quad (2.29)$$

so that

$$U_{,u}(u,v) = \frac{1}{\alpha(u,v)} U_1'(u) e^{-U_1(u)} ,$$

$$U_{,v}(u,v) = \frac{1}{\alpha(u,v)} U_2'(v) e^{-U_2(v)} ;$$

and, therefore, by Eqs. (2.11b), $U_{,u}(u,v)>0$, $U_{,v}(u,v)>0$ for any point (u,v) in the interior of the interaction region, where $u>0$, $v>0$, and where [as long as (u,v) is in the domain of dependence of the initial surface $\{u=0\}\cup\{v=0\}$] $\alpha(u,v)>0$. It is now easy to obtain the remaining field equations, satisfied by the new variable $Q(\alpha,\beta)$: Combining Eq. (2.27b) with Eqs. (2.28) and (2.24), and then using Eqs. (2.26), we find

$$Q_{,\alpha} = -\alpha[(V_{,\alpha}{}^2 + V_{,\beta}{}^2)\cosh^2 W + W_{,\alpha}{}^2 + W_{,\beta}{}^2] , \quad (2.30a)$$

$$Q_{,\beta} = -2\alpha(V_{,\alpha}V_{,\beta}\cosh^2 W + W_{,\alpha}W_{,\beta}) , \quad (2.30b)$$

where the integrability condition for Eqs. (2.30) is satisfied by virtue of the field equations (2.23) for $V(\alpha,\beta)$ and $W(\alpha,\beta)$.

We now combine Eq. (2.27b) with the expression (2.21) for the metric in the interaction region. This gives us the expression of the interaction region metric in terms of the three unknown variables V, W, and Q. Then, by using the initial value of Q that follows from our normalization conditions Eqs. (2.27), we construct the unique solution $Q(\alpha,\beta)$ of the field equations (2.30) by quadrature. As a result, we obtain the following expressions for the metric and the field equations, valid in the interaction region of any arbitrarily polarized colliding plane-wave spacetime:

$$\begin{aligned} g = {} & e^{-Q(\alpha,\beta)/2}\frac{l_1 l_2}{\sqrt{\alpha}}(-d\alpha^2 + d\beta^2) \\ & + \alpha[\cosh W(\alpha,\beta)(e^{V(\alpha,\beta)} + e^{-V(\alpha,\beta)}dy^2) \\ & -2\sinh W(\alpha,\beta)\,dx\,dy] , \end{aligned} \quad (2.31)$$

where V, W, and Q satisfy the field equations

$$V_{,\alpha\alpha} + \frac{1}{\alpha}V_{,\alpha} - V_{,\beta\beta} = 2(V_{,\beta}W_{,\beta} - V_{,\alpha}W_{,\alpha})\tanh W , \quad (2.32a)$$

$$W_{,\alpha\alpha} + \frac{1}{\alpha}W_{,\alpha} - W_{,\beta\beta} = (V_{,\alpha}{}^2 - V_{,\beta}{}^2)\sinh W\cosh W . \quad (2.32b)$$

$$\begin{aligned} Q(\alpha,\beta) = {} & \int_{C:(\alpha_0,\beta_0)}^{(\alpha,\beta)} \{-\alpha[(V_{,\alpha}{}^2 + V_{,\beta}{}^2)\cosh^2 W + W_{,\alpha}{}^2 + W_{,\beta}{}^2]d\alpha \\ & -2\alpha(V_{,\alpha}V_{,\beta}\cosh^2 W + W_{,\alpha}W_{,\beta})d\beta\} + 2M(\alpha_0,\beta_0) + 3\ln\alpha_0 . \end{aligned} \quad (2.33)$$

Here, $\alpha_0 \equiv \alpha(u_0,v_0)$, $\beta_0 \equiv \beta(u_0,v_0)$, $M(\alpha_0,\beta_0) \equiv M(u_0,v_0)$, and C is *any* (differentiable) curve in the (α,β) plane that starts at the initial point (α_0,β_0), and ends at the field point (α,β) at which Q is to be computed. The result of the integral in Eq. (2.33) depends only on the end points of the curve C, since the integrability condition for Eqs. (2.30) is satisfied by virtue of the field equations (2.32).

Equations (2.31)–(2.33) summarize the initial-value problem for colliding plane waves in a conveniently compact form. The only unknowns that must be found by solving partial differential equations (PDE) are the functions $V(\alpha,\beta)$ and $W(\alpha,\beta)$ which satisfy the *nonlinear* system of coupled hyperbolic PDE (2.32). Once $V(\alpha,\beta)$ and $W(\alpha,\beta)$ are known, Q is determined by the explicit expression (2.33) up to an unknown additive constant, which—by suitably choosing the initial point (u_0,v_0) [or (α_0,β_0)]—can be made arbitrarily small. The only disadvantage of the formalism (2.31)–(2.33) is the coordinate singularity that the (α,β) chart develops on the characteristic initial surface $\{u=0\}\cup\{v=0\}$. This coordinate singularity causes, among other things, the function $Q(\alpha,\beta)$ to be logarithmically divergent (to $-\infty$) on the surfaces $\{u=0\}$ and $\{v=0\}$. However, it is still possible to set up a well-defined initial-value problem for the functions $V(\alpha,\beta)$ and $W(\alpha,\beta)$, using initial data posed on the same characteristic surface $\{u=0\}\cup\{v=0\}$. In addition, since we are primarily interested in the behavior of spacetime near the singular "surface" $\{\alpha=0\}$ well away from the coordinate singularity on the initial null surfaces, the above formalism based on (α,β) coordinates is well suited to our objectives.

To understand how to pose initial data for the field equations (2.32), first note that [cf. Eqs. (2.16)] in the α,β coordinates the initial null surfaces $\{u=0\}$ and $\{v=0\}$ are expressed as (Fig. 1)

$$\{u=0\} \equiv \{\alpha-\beta=1\} , \qquad \{v=0\} \equiv \{\alpha+\beta=1\} . \quad (2.34)$$

Equations (2.34) suggest introducing "characteristic" coordinates

$$r \equiv \alpha - \beta , \qquad s \equiv \alpha + \beta , \quad (2.35)$$

so that the initial null surfaces become (see Fig. 1)

$$\{u=0\} \equiv \{r=1\} , \qquad \{v=0\} \equiv \{s=1\} . \quad (2.36)$$

The initial-value problem for the functions V and W consists of the field equations (2.32), and the initial data on the characteristic initial surface $\{r=1\}\cup\{s=1\}$ given by the freely specifiable functions $V(r,s=1)$, $W(r,s=1)$ and $V(r=1,s)$, $W(r=1,s)$. More precisely, the initial data consist of

$$\{\ V(r,1),\ W(r,1),\ V(1,s),\ W(1,s)\ \}\ , \tag{2.37}$$

where $V(r,1)$, $W(r,1)$ and $V(1,s)$, $W(1,s)$ are C^1 (and piecewise C^2) functions for $r\in(-1,1]$ and $s\in(-1,1]$, respectively, which are freely specified except for the initial conditions $V(r=1,1)=W(r=1,1)=V(1,s=1)=W(1,s=1)=0$.

There is a one-to-one correspondence between the initial data of the form (2.7), and initial data of the form (2.37) for the initial-value problem of colliding plane waves. When initial data are given in the form of Eq. (2.7), i.e., when the functions $V_1(u)$, $W_1(u)$ and $V_2(v)$, $W_2(v)$ are specified, initial data in the form of Eq. (2.37) are uniquely determined in the following way: First, Eqs. (2.8) are solved with the given data $V_1(u)$, $W_1(u)$ and $V_2(v)$, $W_2(v)$, and the functions $U_1(u)$ and $U_2(v)$ are obtained as the unique solutions [cf. the discussion following Eqs. (2.8)]. Then, using the identities [cf. Eqs. (2.16) and Eq. (2.35)]

$$r=2e^{-U_1(u)}-1\ ,\qquad s=2e^{-U_2(v)}-1 \tag{2.38}$$

along the initial null surfaces $\{u=0\}$ and $\{v=0\}$, $u(r)$ and $v(s)$ are defined as the unique solutions to the implicit equations

$$r=2e^{-U_1[u(r)]}-1\ ,\qquad s=2e^{-U_2[v(s)]}-1\ . \tag{2.39}$$

Finally, the initial data $\{V(r,1),\ W(r,1),\ V(1,s),\ W(1,s)\}$ in the form (2.37) are determined uniquely from the data $\{V_1(u),\ W_1(u),\ V_2(v),\ W_2(v)\}$ by

$$V(r,1)=V_1[u=u(r)]\ ,\qquad W(r,1)=W_1[u=u(r)]\ . \tag{2.40}$$

$$V(1,s)=V_2[v=v(s)]\ ,\qquad W(1,s)=W_2[v=v(s)]\ .$$

Conversely, when initial data are given in the form of Eq. (2.37), i.e., when the functions $V(r,1)$, $W(r,1)$ and $V(1,s)$, $W(1,s)$ are specified, initial data in the form of (2.7) are uniquely determined in the following way: First, the differential equations

$$2U_{1,uu}-U_{1,u}{}^2=4e^{-2U_1}\,U_{1,u}{}^2\,\{[V_{,r}(r=2e^{-U_1}-1,1)]^2\cosh^2 W(r=2e^{-U_1}-1,1)+[W_{,r}(r=2e^{-U_1}-1,1)]^2\}\ , \tag{2.41a}$$

$$2U_{2,vv}-U_{2,v}{}^2=4e^{-2U_2}\,U_{2,v}{}^2\,\{[V_{,s}(1,s=2e^{-U_2}-1)]^2\cosh^2 W(1,s=2e^{-U_2}-1)+[W_{,s}(1,s=2e^{-U_2}-1)]^2\}\ , \tag{2.41b}$$

for the functions $U_1(u)$ and $U_2(v)$ are solved with the initial conditions $U_1(u=0)=U_2(v=0)=0$, $U_{1,u}(u=0)=U_{2,v}(v=0)=0$ [cf. Eqs. (2.8)]. Then, using Eqs. (2.39), the initial data $\{V_1(u),\ W_1(u),\ V_2(v),\ W_2(v)\}$ in the form (2.7) are determined uniquely from the data $\{V(r,1),\ W(r,1),\ V(1,s),\ W(1,s)\}$ by

$$\begin{aligned} V_1(u)&=V(r=2e^{-U_1(u)}-1,1)\ ,\\ W_1(u)&=W(r=2e^{-U_1(u)}-1,1)\ ,\\ V_2(v)&=V(1,s=2e^{-U_2(v)}-1)\ ,\\ W_2(v)&=W(1,s=2e^{-U_2(v)}-1)\ . \end{aligned} \tag{2.42}$$

This completes the formulation of the initial-value problem for the system of coupled nonlinear hyperbolic PDE (2.32).

III. ASYMPTOTIC STRUCTURE OF COLLIDING PLANE-WAVE SPACETIMES NEAR $\alpha=0$

A. Singularities and horizons at $\alpha=0$: A generalized inhomogeneous Kasner asymptotic structure

It is clear from the expression (2.31) of the metric that the "surface" $\{\alpha=0\}$ represents some kind of singularity [either a spacetime singularity or (at least) a coordinate singularity] of the colliding plane-wave spacetime. In this section and in Sec. III B, we will study the asymptotic behavior of the colliding plane-wave metric (2.31)–(2.33) near this singularity $\{\alpha=0\}$.

Before proceeding with the analysis of asymptotic structure, recall the conclusions of Sec. II B in Ref. 6, where the field equations for colliding *parallel-polarized* plane waves were studied in (α,β) coordinates. [Compare Eqs. (6.2.43) and (6.2.44) with Eqs. (2.31)–(2.33) above.] There the field equations reduced to a single *linear* hyperbolic PDE for $V(\alpha,\beta)$ [Eq. (6.2.44a)], followed by a quadrature for $Q(\alpha,\beta)$ [Eq. (6.2.44b)] similar to Eq. (2.33) above. [The readers can rederive these equations by simply putting $W\equiv 0$ in Eqs. (2.31)–(2.33).] It is well known that, for linear hyperbolic PDE of the kind (6.2.44a), solutions with sufficiently smooth initial data exist globally (see, for example, Secs. 5.2 and 5.3 of Ref. 18 and p. 115 in Sec. 4.2 of Ref. 19). Therefore, it was guaranteed by the linearity of Eq. (6.2.44a) in Ref. 6 that the field equations for V and Q had unique global solutions defined throughout the domain of dependence of the initial surface, i.e., throughout the interaction region $\{\alpha>0\}$. In fact, a general solution (Riemann function[19]) for Eq. (6.2.44a) could be found in closed form [Eq. (6.2.59)], which yielded an explicit representation [Eq. (6.2.60)] of the global solution $V(\alpha,\beta)$ (for $\alpha>0$) in terms of initial data. This assured that the singularities [or Killing-Cauchy horizons (coordinate singularities)] created by colliding parallel-polarized plane waves always lie at or beyond the surface $\{\alpha=0\}$; this surface is in fact the boundary of the domain of dependence, and as Eq. (6.2.43) makes clear, some kind of singularity is always present there.

In contrast with the parallel-polarized case, the field equations (2.32) for arbitrarily polarized colliding plane waves are nonlinear. It is a standard result (see, e.g., Ref. 20 and Sec. VI.6 of Ref. 21) that quasilinear hyperbolic PDE's of the form (2.32) always have unique, *local* solutions, defined in a neighborhood of the initial surface on which regular initial data are posed. On the other hand, it is also well known[22–32] that in general these local solutions do not exist *globally*; i.e., in general solutions of nonlinear hyperbolic PDE's blow up or otherwise break down in finite time within the interior of their domain of dependence. [A particularly lucid example of this breakdown-in-finite-time phenomenon for solutions of nonlinear hyperbolic PDE's is discussed by Klainerman, following his Eq. (13) in Ref. 23.] We also note in this connection that thanks to the recent work of Klainerman,[23,28,31] Shatah,[25] Sideris,[29] Klainerman and Ponce,[27] and Christodoulou,[32] it is now known that for initial data which are sufficiently "small" in some appropriate sense, solutions of nonlinear hyperbolic PDE's of the kind (2.32) *do* exist globally, i.e., throughout the domain of dependence of the initial surface. (See Appendix A for a somewhat more detailed discussion of this point.) In any case, as we have also discussed in the Introduction (Sec. I), if the global existence of solutions with arbitrary (not necessarily small) initial data were false for the field equations (2.32), then this would have the disturbing consequences that (i) colliding nonparallel-polarized plane waves might create singularities in the interior of the interaction region where $\alpha>0$, and (ii) these singularities, if present, would not be analyzable by studying the asymptotic spacetime structure near $\alpha=0$. Therefore, before the asymptotic-structure analysis of this section can be relied on to fully describe the singularity structure of colliding plane-wave spacetimes, it is necessary to have a proof that solutions of Eqs. (2.32) exist globally for *all* initial data.

Obviously, one way to prove this global existence result would be to obtain a general solution (Riemann function[19]) for Eqs. (2.32), in the same way as the Riemann function [Eq. (6.2.59)] of Ref. 6 yielded the explicit expression (6.2.60) of the solution V in terms of initial data, and thus provided a constructive proof for the global existence of V in the parallel-polarized case. It seems unlikely, however, that such a general solution can be found for the nonlinear system (2.32); hence the global existence of solutions for (2.32) must be proved using nonconstructive arguments. Indeed, such a nonconstructive proof *can* be provided, as we discuss in detail in Appendix A, thanks to some very special properties possessed by the field equations [especially the existence of the positive-definite conserved energy form Eq. (A18)]. Thus, our discussions in Appendix A prove that the singularities and Killing-Cauchy horizons (see below) created by colliding plane waves always lie at or beyond $\{\alpha=0\}$; no singularities ever occur in the interior of the interaction region where $\alpha>0$. [Incidentally, Appendix A also proves as a special case that the global solution (6.2.60) for $V(\alpha,\beta)$ coupled with $W\equiv 0$ is the unique solution of Eqs. (2.32) corresponding to initial data (2.37) with $W(r,1)=W(1,s)\equiv 0$; i.e., colliding plane waves which are initially parallel polarized remain parallel polarized everywhere after they scatter each other.] Furthermore, in Appendix B we use the results of Appendix A to prove that the spatial (β) derivative terms in the field equations (2.32) are asymptotically negligible compared to the timelike (α) derivative terms as the singularity $\{\alpha=0\}$ is approached. As we will heavily rely on these results in the discussions below, we suggest to those readers who desire greater logical completeness that they read Appendixes A and B at this point, before proceeding with the rest of Secs. III A and III B.

Since as $\alpha\to 0$ the β derivative terms in Eqs. (2.32) are asymptotically negligible compared to the α-derivative terms (Appendix B), the asymptotic behaviors of V and W near $\alpha=0$ are identical with those of the solutions of the ordinary differential equations

$$V_{,\alpha\alpha}+\frac{1}{\alpha}V_{,\alpha}=-2V_{,\alpha}W_{,\alpha}\tanh W\ , \tag{3.1a}$$

$$W_{,\alpha\alpha}+\frac{1}{\alpha}W_{,\alpha}=V_{,\alpha}{}^2\sinh W\cosh W \tag{3.1b}$$

obtained from Eqs. (2.32) by ignoring all terms with β derivatives.

Consider first Eq. (3.1a) for V. Dividing both sides by $V_{,\alpha}$ and integrating, we obtain

$$\ln|\alpha V_{,\alpha}|+2\ln(\cosh W)=C\ , \tag{3.2}$$

which immediately yields

$$V_{,\alpha}=\frac{C}{\alpha\cosh^2 W}\ . \tag{3.3}$$

[Here and henceforth C will stand for an arbitrary (indefinite) constant.] Clearly, the constant C in Eq. (3.3) will in general depend on β. Thus, we rename the constant C of Eq. (3.3) as $\epsilon_1(\beta)$, and then apply a further integration to obtain

$$V(\alpha,\beta)=\epsilon_1(\beta)\int\frac{d\alpha}{\alpha\cosh^2 W}+\delta_1(\beta)+H_1(\alpha,\beta)\ , \tag{3.4a}$$

where

$$\lim_{\alpha\to 0}H_1(\alpha,\beta)\equiv 0\ . \tag{3.4b}$$

Equations (3.4) determine the asymptotic behavior of $V(\alpha,\beta)$ once the asymptotic behavior of W is known.

To find the asymptotic behavior of $W(\alpha,\beta)$, consider Eq. (3.1b) for W and insert into it the expression for $V_{,\alpha}$ given by Eq. (3.3); this yields

$$\frac{1}{\alpha}(\alpha W_{,\alpha})_{,\alpha}=\frac{\epsilon_1{}^2}{\alpha^2\cosh^3 W}\sinh W\ . \tag{3.5}$$

Multiplying both sides of Eq. (3.5) by $2\alpha^2 W_{,\alpha}$ and integrating once after collecting all terms on the left-hand side, we obtain

$$(\alpha W_{,\alpha})^2+\frac{\epsilon_1{}^2}{\cosh^2 W}=C\equiv\epsilon_2{}^2\ , \tag{3.6}$$

where we have renamed the β-dependent constant C as $\epsilon_2(\beta)$. We will always assume, without loss of generality, that by convention $\epsilon_2 \geq 0$. Equation (3.6) can then be rewritten in the form

$$\pm \frac{d\alpha}{\alpha} = \frac{dW}{[\epsilon_2^2 - (\epsilon_1^2/\cosh^2 W)]^{1/2}} = \frac{1}{\epsilon_2} \frac{\cosh W\, dW}{[\cosh^2 W - (\epsilon_1^2/\epsilon_2^2)]^{1/2}} . \tag{3.7}$$

The integration of Eq. (3.7) is elementary, and it yields the following two possibilities for the asymptotic behavior of $W(\alpha,\beta)$ near $\alpha=0$:

$$W(\alpha,\beta) = \delta_2(\beta)\,\alpha^{\epsilon_2(\beta)} + H_2(\alpha,\beta) \tag{3.8a}$$

in which case $[\epsilon_1(\beta)/\epsilon_2(\beta)]^2$ must equal 1,

$$W(\alpha,\beta) = \pm\,\epsilon_2(\beta)\ln\alpha + \delta_2(\beta) + H_2(\alpha,\beta) \tag{3.8b}$$

in which case $[\epsilon_1(\beta)/\epsilon_2(\beta)]^2$ is arbitrary, where

$$\lim_{\alpha\to 0} H_2(\alpha,\beta) \equiv 0 . \tag{3.8c}$$

Combining Eqs. (3.8) with Eqs. (3.4), we find that there are three and only three distinct possible asymptotic behaviors for V and W near $\alpha=0$. We can express these three possible cases in the following final form.

Case (a). In this case $[\epsilon_1(\beta)/\epsilon_2(\beta)]^2$ must equal 1, and $\epsilon_2(\beta)>0$:

$$\begin{aligned} V(\alpha,\beta) &= \epsilon_1(\beta)\ln\alpha + \delta_1(\beta) + H_1(\alpha,\beta) , \\ W(\alpha,\beta) &= \delta_2(\beta)\,\alpha^{\epsilon_2(\beta)} + H_2(\alpha,\beta) . \end{aligned} \tag{3.9a}$$

Case (b). In this case $[\epsilon_1(\beta)/\epsilon_2(\beta)]^2$ is arbitrary, and $\epsilon_2(\beta)>0$:

$$\begin{aligned} V(\alpha,\beta) &= \frac{2\epsilon_1(\beta)}{\epsilon_2(\beta)} e^{\pm 2\delta_2(\beta)} \alpha^{2\epsilon_2(\beta)} \\ &\quad + \delta_1(\beta) + H_1(\alpha,\beta) , \\ W(\alpha,\beta) &= \pm\,\epsilon_2(\beta)\ln\alpha + \delta_2(\beta) + H_2(\alpha,\beta) . \end{aligned} \tag{3.9b}$$

Case (c). In this case $\epsilon_2(\beta)=\epsilon_1(\beta)=0$:

$$\begin{aligned} V(\alpha,\beta) &= \delta_1(\beta) + H_1(\alpha,\beta) , \\ W(\alpha,\beta) &= \delta_2(\beta) + H_2(\alpha,\beta) . \end{aligned} \tag{3.9c}$$

In all three cases (a)–(c) above the terms $H_i(\alpha,\beta)$ have the general form ($i \equiv 1,\ 2$)

$$H_i(\alpha,\beta) = \sum_{k=2}^{\infty} c^{(i)}{}_k(\beta)\,\alpha^k + \sum_{k=2,\ l=1}^{\infty} d^{(i)}{}_{kl}(\beta)\,\alpha^k(\ln\alpha)^l . \tag{3.10}$$

[Equation (3.10) follows from the expressions (3.3) and (3.6) for $V_{,\alpha}$ and $W_{,\alpha}$. In fact, Eqs. (3.3) and (3.6) constrain the form of $H_i(\alpha,\beta)$ even further than Eq. (3.10), and we will use these extra constraints below in deriving the asymptotic form of $Q(\alpha,\beta)$ near $\alpha=0$.]

The asymptotic behavior of the metric function $Q(\alpha,\beta)$ is obtained by combining Eqs. (3.9) and (3.10) with the field equation (2.33). The final result can be described as

$$\text{Case (a):}\quad Q(\alpha,\beta) = -\,\epsilon_1{}^2(\beta)\ln\alpha + \mu(\beta) + L(\alpha,\beta) , \tag{3.11a}$$

$$\text{Case (b):}\quad Q(\alpha,\beta) = -\,\epsilon_2{}^2(\beta)\ln\alpha + \mu(\beta) + L(\alpha,\beta) , \tag{3.11b}$$

$$\text{Case (c):}\quad Q(\alpha,\beta) = \mu(\beta) + L(\alpha,\beta) , \tag{3.11c}$$

where

$$\lim_{\alpha\to 0} L(\alpha,\beta) \equiv 0 , \tag{3.12}$$

but $L(\alpha,\beta)$ does not necessarily have the general form (3.10).

With Eqs. (3.9)–(3.12), we have completed our analysis of the asymptotic forms of the metric functions V, W, and Q near $\alpha=0$; at this point readers might find it useful to compare Eqs. (3.9)–(3.12) with the corresponding Eqs. (6.3.4)–(6.3.7) of Ref. 6 for the parallel-polarized case.

Now we are ready to analyze the asymptotic behavior of the arbitrarily polarized colliding plane-wave metric (2.31) near the singular surface $\{\alpha=0\}$. We first note that the x-y part of the metric (2.31), when considered as a two-dimensional symmetric tensor field on $\{u=\text{const},\ v=\text{const}\}$ sections, is positive definite and nondegenerate;[33] i.e., it is a Euclidean metric. (That this must be the case becomes clear when one recalls that by the definition of plane symmetry[8,9] the Killing vectors $\partial/\partial x$ and $\partial/\partial y$ must span a spacelike two-dimensional plane in each tangent space. Only asymptotically, as $\alpha\to 0$, can this 2-plane become null.) Consequently, it is possible to diagonalize the x-y part of the metric by using two spacelike, orthonormal 1-forms defined throughout the interaction region. When this is done, we find that the metric (2.31) can be brought into the diagonal form

$$g = e^{-Q(\alpha,\beta)/2}\frac{l_1 l_2}{\sqrt{\alpha}}(d\alpha^2 + d\beta^2) + \alpha(\omega^1\otimes\omega^1 + \omega^2\otimes\omega^2) \tag{3.13a}$$

with the orthogonal spacelike 1-forms

$$\omega^1 = \frac{e^{\hat{V}/2}}{(2\sinh\hat{V})^{1/2}}\left[P\,dx - \frac{\sinh W}{P}\,dy \right] , \tag{3.13b}$$

$$\omega^2 = \frac{e^{-\hat{V}/2}}{(2\sinh\hat{V})^{1/2}}\left[\frac{\sinh W}{P}\,dx + P\,dy \right] , \tag{3.13c}$$

where,

$$P \equiv (\sinh\hat{V} + \sinh V\cosh W)^{1/2} , \tag{3.13d}$$

$$\hat{V} \equiv \ln[\cosh V\cosh W + (\cosh^2 V\cosh^2 W - 1)^{1/2}] . \tag{3.13e}$$

A short computation using Eqs. (3.13) shows that when considered as functions of the variables V and W, the 1-forms ω^1 and ω^2 are discontinuous at $W=0$; the 1-form ω^1 (as well as ω^2) tends to two different limits as $W\to 0$ depending on whether $W\to +0$ or $W\to -0$. In contrast, the tensor field $\omega^1\otimes\omega^1+\omega^2\otimes\omega^2$ depends on V and W smoothly; in fact

$$\omega^1\otimes\omega^1+\omega^2\otimes\omega^2\to e^{V}dx^2+e^{-V}dy^2 \quad \text{as } W\to\pm 0\ .$$

Therefore, the discontinuities in the dependence of ω^i on V and W are unimportant when analyzing the asymptotic structure of the spacetime geometry (3.13a) near $\alpha=0$.

We now combine Eqs. (3.13) with Eqs. (3.9)–(3.12), and obtain the following final results for the asymptotic form of the metric (2.31) as $\alpha\to 0$.

Case (a). In this case $[\epsilon_1(\beta)/\epsilon_2(\beta)]^2$ must equal 1, and $\epsilon_2(\beta)>0$:

$$g(\beta)\sim l_1l_2e^{-\mu(\beta)/2}\alpha^{[\epsilon_1^2(\beta)-1]/2}(-d\alpha^2+d\beta^2)+e^{\delta_1(\beta)}\alpha^{1+\epsilon_1(\beta)}dx^2+e^{-\delta_1(\beta)}\alpha^{1-\epsilon_1(\beta)}dy^2\ . \tag{3.14a}$$

Case (b). In this case $[\epsilon_1(\beta)/\epsilon_2(\beta)]^2$ is arbitrary and $\epsilon_2(\beta)>0$:

$$\begin{aligned} g(\beta)&\sim l_1l_2e^{-\mu(\beta)/2}\alpha^{[\epsilon_2^2(\beta)-1]/2}(-d\alpha^2+d\beta^2)+\alpha(\omega^1\otimes\omega^1+\omega^2\otimes\omega^2)\ ,\\ \omega^1(\beta)&\sim e^{\mp\delta_2(\beta)/2}\alpha^{-\epsilon_2(\beta)/2}(e^{\delta_1(\beta)/2}dx\pm e^{-\delta_1(\beta)/2}dy)\ ,\\ \omega^2(\beta)&\sim\frac{e^{\pm\delta_2(\beta)/2}}{\cosh\delta_1(\beta)}\alpha^{\epsilon_2(\beta)/2}(\mp e^{-\delta_1(\beta)/2}dx+e^{\delta_1(\beta)/2}dy)\ . \end{aligned}\tag{3.14b}$$

Case (c). In this case $\epsilon_2(\beta)=\epsilon_1(\beta)=0$:

$$\begin{aligned} g(\beta)&\sim l_1l_2e^{-\mu(\beta)/2}\alpha^{-1/2}(-d\alpha^2+d\beta^2)+\alpha(\omega^1\otimes\omega^1+\omega^2\otimes\omega^2)\ ,\\ \omega^1&\sim\frac{s(\beta)}{[s^2(\beta)-1]^{1/2}}\left[q(\beta)\,dx-\frac{\sinh\delta_2(\beta)}{q(\beta)}dy\right],\\ \omega^2&\sim\frac{1}{[s^2(\beta)-1]^{1/2}}\left[\frac{\sinh\delta_2(\beta)}{q(\beta)}dx+q(\beta)\,dy\right], \end{aligned}\tag{3.14c}$$

$$q(\beta)\equiv\left[\frac{s^2(\beta)-s(\beta)}{2s(\beta)}+\sinh\delta_1(\beta)\cosh\delta_2(\beta)\right]^{1/2},$$

$$s(\beta)\equiv\cosh\delta_1(\beta)\cosh\delta_2(\beta)+[\cosh^2\delta_1(\beta)\cosh^2\delta_2(\beta)-1]^{1/2}\ .$$

In all three cases (3.14a)–(3.14c), the asymptotic structure of the metric is generalized inhomogeneous Kasner. The following equations are derived from Eqs. (3.14) in order to express this inhomogeneous-Kasner structure more precisely [compare also Eqs. (6.3.14)–(6.3.19) of Ref. 6].

Case (a):

$$g(\beta)\sim-\frac{16\,l_1l_2e^{-\mu(\beta)/2}}{[\epsilon_1^{\,2}(\beta)+3]^2}dt^2+l_1l_2e^{-\mu(\beta)/2}t^{2p_3}d\beta^2+e^{\delta_1(\beta)}t^{2p_1}dx^2+e^{-\delta_1(\beta)}t^{2p_2}dy^2\ , \tag{3.15a}$$

where

$$t\equiv\alpha^{[\epsilon_1^{\,2}(\beta)+3]/4}\ , \tag{3.15b}$$

and

$$p_3(\beta)=\frac{\epsilon_1^{\,2}(\beta)-1}{\epsilon_1^{\,2}(\beta)+3}\ ,\quad p_1(\beta)=\frac{2[1+\epsilon_1(\beta)]}{\epsilon_1^{\,2}(\beta)+3}\ ,\quad p_2(\beta)=\frac{2[1-\epsilon_1(\beta)]}{\epsilon_1^{\,2}(\beta)+3}\ . \tag{3.15c}$$

Case (b):

$$g(\beta)\sim-\frac{16\,l_1l_2e^{-\mu(\beta)/2}}{[\epsilon_2^{\,2}(\beta)+3]^2}dt^2+l_1l_2e^{-\mu(\beta)/2}t^{2p_3}d\beta^2+e^{\mp\delta_2(\beta)}t^{2p_1}dX_{(\beta)}^{\,2}+\frac{e^{\pm\delta_2(\beta)}}{\cosh^2\delta_1(\beta)}t^{2p_2}dY_{(\beta)}^{\,2}\ , \tag{3.16a}$$

where

$$t\equiv\alpha^{[\epsilon_2^{\,2}(\beta)+3]/4}\ , \tag{3.16b}$$

and

$$p_3(\beta)=\frac{\epsilon_2^{\,2}(\beta)-1}{\epsilon_2^{\,2}(\beta)+3}\ ,\quad p_1(\beta)=\frac{2[1+\epsilon_2(\beta)]}{\epsilon_2^{\,2}(\beta)+3}\ ,\quad p_2(\beta)=\frac{2[1-\epsilon_2(\beta)]}{\epsilon_2^{\,2}(\beta)+3}\ , \tag{3.16c}$$

$$X_{(\beta)}\equiv(e^{\delta_1(\beta)/2}x\pm e^{-\delta_1(\beta)/2}y)\ ,\quad Y_{(\beta)}\equiv(\mp e^{-\delta_1(\beta)/2}x+e^{\delta_1(\beta)/2}y)\ . \tag{3.16d}$$

Case (c):

$$g(\beta)\sim-\tfrac{16}{9}l_1l_2e^{-\mu(\beta)}dt^2+l_1l_2e^{-\mu(\beta)/2}t^{2p_3}d\beta^2+\frac{s^2(\beta)}{s^2(\beta)-1}t^{2p_1}dX_{(\beta)}^{\,2}+\frac{1}{s^2(\beta)-1}t^{2p_2}dY_{(\beta)}^{\,2} \tag{3.17a}$$

where

$$t\equiv\alpha^{3/4}\ , \tag{3.17b}$$

and

$$p_3(\beta)=-\tfrac{1}{3}\ ,\quad p_1(\beta)=\tfrac{2}{3}\ ,\quad p_2(\beta)=\tfrac{2}{3}\ , \tag{3.17c}$$

$$X_{(\beta)} \equiv \left[q(\beta)x - \frac{\sinh\delta_2(\beta)}{q(\beta)} y \right] ,$$
$$Y_{(\beta)} \equiv \left[\frac{\sinh\delta_2(\beta)}{q(\beta)} x + q(\beta) y \right] . \tag{3.17d}$$

Equations (3.15)–(3.17) demonstrate that at a fixed value of β the asymptotic limit of the spacetime metric (2.31) has the general form of a vacuum Kasner[10] solution:

$$g = -a\,dt^2 + bt^{2p_3}d\beta^2 + ct^{2p_1}dX^2 + dt^{2p_2}dY^2 , \tag{3.18}$$

where a, b are constants having the dimensions of (length)^2, c, d are dimensionless constants, t, β are dimensionless coordinates, and the exponents p_k, $k=1,2,3$ in all cases satisfy the Kasner relations [cf. Eqs. (3.15c), (3.16c), and (3.17c)]

$$p_1(\beta) + p_2(\beta) + p_3(\beta) = p_1{}^2(\beta) + p_2{}^2(\beta) + p_3{}^2(\beta) = 1 . \tag{3.19}$$

The coordinates X, Y are asymptotically constant linear combinations [cf. Eqs. (3.16d) and (3.17d)] of the spacelike (Killing) coordinates x and y that determine the asymptotic Kasner axes along which the exponents p_1 and p_2 are defined (the exponent p_3 is always associated with the β axis). In fact, it becomes clear from Eqs. (3.15)–(3.17) that in general these asymptotic Kasner axes (defined by the coordinates $X_{(\beta)}$, $Y_{(\beta)}$), like the Kasner exponents $p_k(\beta)$, depend on the spatial coordinate β across the singularity: hence the rationale for our use of the term "generalized inhomogeneous Kasner" to describe the asymptotic structures (3.15)–(3.17).[6]

If all of the exponents p_k are different from 1 [or equivalently by Eqs. (3.19) all are different from 0], then the Kasner spacetime (3.18) possesses a curvature singularity at $t=0$. (For a brief description of the geometry of the Kasner solution see Sec. III A of Ref. 6.) It follows that when $p_k(\beta)$ are similarly all different from 0 in any of the three cases (a)–(c) [Eqs. (3.15)–(3.17)], the colliding plane-wave spacetime (2.31) possesses a curvature singularity at $(\alpha=0,\beta)$. Conversely, when any of the p_k in Eq. (3.18) is equal to 1 (in which case both other exponents are zero), the metric (3.18) is flat (a degenerate Kasner solution[6]) and $\{t=0\}$ is a nonsingular Killing-Cauchy horizon[8] in the Kasner spacetime. Similarly, we claim that if any of the two exponents $p_1(\beta)$, $p_2(\beta)$ is identically equal to 1 across an interval (β_1,β_2) [the exponent $p_3(\beta)$ can never equal 1, see Eqs. (3.15c) and (3.16c)], then the surface $\{\alpha=0,\ \beta_1<\beta<\beta_2\}$ is a Killing-Cauchy horizon for the colliding plane-wave spacetime (2.31). More precisely, we claim the following.

(i) In case (a), the surface $\{\alpha=0,\ \beta_1<\beta<\beta_2\}$ is a Killing-Cauchy horizon if and only if

$$|\epsilon_1(\beta)| \equiv 1 \qquad \forall\ \beta \in (\beta_1,\beta_2) . \tag{3.20a}$$

In this case, the spacelike Killing vector that becomes null on the horizon[8] is either $\partial/\partial x$ (when $\epsilon_1 \equiv +1$) or $\partial/\partial y$ [when $\epsilon_1(\beta) \equiv -1$].

(ii) In case (b), the surface $\{\alpha=0,\ \beta_1<\beta<\beta_2\}$ is a Killing-Cauchy horizon if and only if

$$\epsilon_2(\beta) \equiv 1 , \qquad \delta_1(\beta) \equiv \text{const} \equiv \delta_1 \qquad \forall\ \beta \in (\beta_1,\beta_2) . \tag{3.20b}$$

In this case, the spacelike Killing vector that becomes null on the horizon is

$$\frac{\partial}{\partial Y_{(\beta)}} \equiv \frac{1}{2\cosh\delta_1} \left[\mp e^{-\delta_1/2} \frac{\partial}{\partial x} + e^{\delta_1/2} \frac{\partial}{\partial y} \right] . \tag{3.21}$$

In case (c), $(\alpha=0,\beta)$ is always a curvature singularity since the exponents $p_k(\beta)$ are all different from zero [Eqs. (3.17c)].

To prove the above claims (i) and (ii), we proceed exactly as we did in Ref. 6: First, we obtain the expressions of the Newman-Penrose curvature quantities (2.12) in the (α,β) coordinates. This can be done in precisely the same way as that explained in Sec. III B of Ref. 6; it gives (note that as in Ref. 6 the quantity $\alpha_{,\upsilon}$ in Eqs. (3.22) below is finite and nonvanishing as $\alpha\to 0$ [cf. Eqs. (2.16a) and (2.11b)]; consequently it can be regarded as a constant when analyzing the asymptotic behaviors of Ψ_2, Ψ_0, and Ψ_4 near the singularity)

$$\Psi_2 = \frac{e^{Q(\alpha,\beta)/2}}{8l_1l_2}\alpha^{1/2}\left[Q_{,\alpha\alpha} - Q_{,\beta\beta} - \frac{1}{\alpha^2} - 4i\,(V_{,\beta}W_{,\alpha} - V_{,\alpha}W_{,\beta})\cosh W \right] , \tag{3.22a}$$

$$\begin{aligned}\Psi_0 = -\frac{1}{8l_1{}^2l_2{}^2}\frac{e^{Q(\alpha,\beta)}\alpha}{\alpha_{,\upsilon}{}^2}\Bigg\{ &\cosh W \left[\tfrac{1}{2}(V_{,\alpha} - V_{,\beta}) \left[Q_{,\alpha} - Q_{,\beta} + \frac{3}{\alpha} \right] + V_{,\alpha\alpha} + V_{,\beta\beta} - 2V_{,\alpha\beta} \right] \\ &+ 2\sinh W (V_{,\alpha} - V_{,\beta})(W_{,\alpha} - W_{,\beta}) \\ &- i \Bigg[\tfrac{1}{2}(W_{,\alpha} - W_{,\beta}) \left[Q_{,\alpha} - Q_{,\beta} + \frac{3}{\alpha} \right] + W_{,\alpha\alpha} + W_{,\beta\beta} - 2W_{,\alpha\beta} \\ &- (V_{,\alpha}{}^2 + V_{,\beta}{}^2 - 2V_{,\alpha}V_{,\beta})\sinh W\cosh W \Bigg] \Bigg\} ,\end{aligned} \tag{3.22b}$$

$$\Psi_4 = -\tfrac{1}{2}\alpha_{,v}{}^2 \Bigg\{ \cosh W \left[\tfrac{1}{2}(V_{,\alpha}+V_{,\beta}) \left[Q_{,\alpha}+Q_{,\beta}+\frac{3}{\alpha} \right] + V_{,\alpha\alpha}+V_{,\beta\beta}+2V_{,\alpha\beta} \right]$$
$$+ 2\sinh W (V_{,\alpha}+V_{,\beta})(W_{,\alpha}+W_{,\beta})$$
$$+i \Bigg[\tfrac{1}{2}(W_{,\alpha}+W_{,\beta}) \left[Q_{,\alpha}+Q_{,\beta}+\frac{3}{\alpha} \right] + W_{,\alpha\alpha}+W_{,\beta\beta}+2W_{,\alpha\beta}$$
$$-(V_{,\alpha}{}^2+V_{,\beta}{}^2+2V_{,\alpha}V_{,\beta})\sinh W \cosh W \Bigg] \Bigg\} . \tag{3.22c}$$

Next, we replace $Q_{,\alpha}$ and $Q_{,\beta}$ in Eqs. (3.22) with their values in terms of V and W given by Eqs. (2.30), and then substitute for V and W their asymptotic limits Eqs. (3.9) and (3.10) where the coefficients $c^{(i)}{}_k$ and $d^{(i)}{}_{kl}$ are obtained in terms of ϵ_1, ϵ_2, δ_1, δ_2 upon inserting Eqs. (3.9) into the field equations (2.32) [compare Eqs. (6.3.38) of Ref. 6]. Inspection of the resulting asymptotic expressions for the curvature quantities yields the following conclusions [compare Eqs. (6.3.33)–(6.3.35) of Ref. 6].

(i) The surface $\{\alpha=0,\ \beta_1<\beta<\beta_2\}$ is a (connected) Killing-Cauchy horizon if and only if one of the two conditions (3.20a) or (3.20b) is satisfied throughout (β_1,β_2). When such a Killing-Cauchy horizon $\mathcal{S}$ forms, the curvature quantities Ψ_2, Ψ_0, and Ψ_4 are finite and well behaved (but in general nonzero) through $\mathcal{S}$ as $\alpha\to 0$ at any $\beta\in(\beta_1,\beta_2)$.

(ii) Suppose the point $p\equiv(\alpha=0,\ \beta=\beta_0)$ does not belong to a Killing-Cauchy horizon, i.e., suppose there is no interval (β_1,β_2) containing β_0 throughout which one of the conditions (3.20a) or (3.20b) is satisfied. Then p corresponds to a curvature singularity of the colliding plane-wave spacetime except when one of the following is true at p: In case (a), $\epsilon_1(\beta_0)=\pm 1$, $\epsilon_1'(\beta_0)=\epsilon_1''(\beta_0)=0$. In case (b), $\epsilon_2(\beta_0)=1$, $\epsilon_2'(\beta_0)=\epsilon_2''(\beta_0)=\delta_1'(\beta_0)=\delta_1''(\beta_0)=0$. Although under any one of the above circumstances p is not a curvature singularity (Ψ_2, Ψ_0, and Ψ_4 are finite as $\alpha\to 0$ *at* $\beta=\beta_0$), it still corresponds to a spacetime singularity since there is no topological neighborhood around p which is completely free of neighboring curvature singularities (cf. the assumption that p does not belong to a Killing-Cauchy horizon).

It has become clear in this section that the asymptotic behavior of a general colliding plane-wave spacetime near its singularity is completely characterized by the four functions $\epsilon_1(\beta)$, $\epsilon_2(\beta)$, $\delta_1(\beta)$, and $\delta_2(\beta)$. In contrast with Ref. 6 where the corresponding functions $\epsilon(\beta)$ and $\delta(\beta)$ in the parallel-polarized case could be expressed explicitly in terms of initial data [Eqs. (6.3.13) and (6.3.12b)], here such expressions cannot be found in general due to the absence of a Riemann function for Eqs. (2.32). Consequently it is not in general possible to relate the asymptotic Kasner axes and exponents along the singularity $\alpha=0$ to the initial data (2.37) posed along the wave fronts of the incoming plane waves. In Appendix C, when we discuss some intriguing aspects of the field equations (2.32) which might some day prove useful in the search for a Riemann function, we also indicate an interesting special case in which one of the asymptotic structure functions can be expressed explicitly in terms of the initial data posed by the colliding waves [Eq. (C6)].

B. Instability and nongenericity of the Killing-Cauchy horizons that occur at $\alpha=0$

Our analysis in the previous section proved that whenever the "surface" $\{\alpha=0\}$ is free of Killing-Cauchy horizons, it represents a curvature singularity of the colliding plane-wave spacetime (2.31). In fact, that this must be true in general in *any* plane-symmetric spacetime is the content of a singularity theorem due to Tipler.[11] (A discussion of this theorem emphasizing its relevance to Killing-Cauchy horizons as well as to singularities can be found in Sec. III B of Ref. 12.) More precisely, Tipler's theorem proves that any nonflat, plane-symmetric spacetime in which the null convergence condition[14] holds is either null-geodesically incomplete or possesses a region where its strict plane symmetry[8,12] breaks down; i.e., the spacetime either contains singularities (where null geodesics terminate) or Killing horizons (where at least one of the plane-symmetry-generating spacelike Killing vectors becomes null).

The horizons $\mathcal{S}$ that occur in colliding plane-wave spacetimes are Killing horizons since as discussed in Sec. III A [Eqs. (3.20) and (3.21)] there exists a spacelike, constant (hence Killing) linear combination of the Killing vectors $\partial/\partial x$ and $\partial/\partial y$ which becomes null on $\mathcal{S}$. As a consequence, on $\mathcal{S}$ the Rosen-type coordinates (u,v,x,y) [and also the coordinates (α,β,x,y)] break down, developing coordinate singularities similar to those developed by (t,β,X,Y) at the surface $\{t=0\}$ of the *degenerate* (flat) Kasner solution (3.18). As another consequence of this breakdown of strict plane symmetry, the past-directed null generators of $\mathcal{S}$ (which are tangent to the Killing direction that becomes null on $\mathcal{S}$) fail to intersect the initial characteristic surface $\mathcal{N}\equiv\{u=0\}\cup\{v=0\}$; i.e., $\mathcal{S}$ is outside the domain of dependence[14] $D^+(\mathcal{N})$ of $\mathcal{N}$. In fact, it is easy to see that $\mathcal{S}$ constitutes precisely the future boundary of $D^+(\mathcal{N})$; more precisely, the Killing horizon $\mathcal{S}$ is at the same time a future *Cauchy horizon* for the initial characteristic surface $\mathcal{N}$.

It is well known that spacetime can be smoothly extended across the Killing-Cauchy horizon $\mathcal{S}$ in infinitely many different ways. The geometry of spacetime beyond $\mathcal{S}$ is not uniquely determinable by the initial data posed on $\mathcal{N}$; global predictability breaks down. Since these causal properties of the horizons $\mathcal{S}$ and their implications were discussed extensively in Sec. III C of Ref. 6 (see also Fig. 2 of Ref. 6), we will not repeat those discussions here. We will only note, as a particularly relevant implication of the breakdown of predictability, that the occurrence of horizons in the collisions of gravitational plane waves might appear to diminish the predictive power of Tipler's singularity theorem: If a horizon forms, existence of singularities cannot be proved; in fact when horizons are present the existence of singularities is false: there are examples[34] of exact solutions for nonvacuum colliding plane waves which have everywhere-nonsingular extensions beyond their Killing-Cauchy horizons.

We also recall our discussion in Ref. 6 of the strong cosmic censorship conjecture,[35,36] and of how, when suitably restricted to plane-symmetric spacetimes, it predicts the instability of the Killing-Cauchy horizons $\mathcal{S}$. These instabilities are also discussed extensively in the literature: On the one hand, there are examples of exact colliding plane-wave solutions whose horizons are destroyed and replaced by singularities when matter fields are introduced into the spacetime;[37] on the other hand, there are general theorems proving the linearized instability of arbitrary Killing-Cauchy horizons in plane-symmetric spacetimes,[8] and of compact Killing horizons in a general spacetime.[38] In fact, for the special case of the Killing-Cauchy horizons which occur in collisions of parallel-polarized plane waves, our discussions in Sec. III C of Ref. 6 prove that the instabilities render the set of horizon-producing initial data "nongeneric" with respect to a very precise notion of nongenericity. More specifically, our analysis in Ref. 6 proves that the subset of all initial data which produce at least one connected Killing-Cauchy horizon larger than Planck size is nongeneric within the set of all colliding parallel-polarized plane-wave initial data. Correspondingly, by making use of the results of Appendixes A and B and of Sec. III A, we will prove below the generalization of this result (with the same notion of genericity as in Ref. 6) to the case of colliding arbitrarily polarized plane waves. In addition, by using a more sophisticated notion of genericity described in detail in Appendix D, we will prove that the subset of *all* horizon-producing initial data (and not just the subset of those data which produce horizons larger than Planck size) is nongeneric within the set of all initial data for colliding plane waves. We will also discuss why we believe that our topological notion of genericity (described in Appendix D) is more appropriate in general relativity than other possible "probabilistic" notions based on measure theory. Note that these results (i) fully restore the predictive power of Tipler's singularity theorem: *generic* gravitational plane-wave collisions always produce "pure" spacetime singularities without Killing-Cauchy horizons, and (ii) similarly yield a proof of "plane-symmetric" strong cosmic censorship:[35,36] generic plane-symmetric gravitational initial data always evolve into *inextendible* globally hyperbolic maximal developments. [To be more precise, our analysis proves (ii) only within the class of plane-symmetric metrics which can be brought into the form (2.3); this class includes (but is larger than) all metrics which are flat in some open set somewhere in spacetime.[4,8]]

To prove our results on the nongenericity of plane-symmetric Killing-Cauchy horizons, we proceed as follows. We first make the space D of all initial data in the form (2.37),

$$D \equiv \{ p \mid p \equiv [V(r,1),\ W(r,1),\ V(1,s),\ W(1,s)] \} , \tag{3.23}$$

a Banach space[39] completed under the norm (say)

$$\| p \| \equiv \left[\int_{-1}^{1} [\, |V(r,1)|^2 + |W(r,1)|^2 \,]\, dr + \int_{-1}^{1} [\, |V(1,s)|^2 + |W(1,s)|^2 \,]\, ds \right]^{1/2}$$

(the precise choice of the norm is immaterial). Similarly, the space F of all asymptotic structure functions,

$$F \equiv \{ f \mid f \equiv [\epsilon_1(\beta),\ \delta_1(\beta),\ \epsilon_2(\beta),\ \delta_2(\beta)] \} , \tag{3.24}$$

can be made a Banach space after completion with respect to the norm

$$\| f \| \equiv \left[\int_{-1}^{1} [\, |\epsilon_1(\beta)|^2 + |\delta_1(\beta)|^2 + |\epsilon_2(\beta)|^2 + |\delta_2(\beta)|^2 \,]\, d\beta \right]^{1/2}$$

(again the precise choice of the norm is unimportant). The vector space structures in both D and F are defined pointwise; thus, under the above norms both D and F are isomorphic to standard L^2 spaces.[39] We also construct the space $A \equiv \{ q \mid q \equiv [f, \sigma(\beta)] \}$ of all possible asymptotic behaviors. Here $f \in F$, and $\sigma(\beta)$ is a function with values in the (discrete) flag set $\{a,\ +b,\ -b,\ c\}$; the flag $\sigma(\beta)$ determines which of the four possible asymptotic behaviors described by the structure functions f and Eqs. (3.9a)–(3.9c) is actually assumed by (V,W) near $\alpha=0$ and at β. Obviously, the function $\sigma(\beta)$ is not continuous in general; however it can be assumed to be Lebesgue measurable[40] on $(-1,1)$. Also, in order to have each point of A correspond to a distinct asymptotic behavior, we impose the restrictions that $\sigma(\beta) = c$ if and only if $\epsilon_1(\beta) = \epsilon_2(\beta) = 0$ and that $\sigma(\beta) = a$ only if $|\epsilon_1(\beta)| = |\epsilon_2(\beta)|$, *or* $\delta_2(\beta) = 0$, for all $q \in A$, $q = [f, \sigma(\beta)]$. We make A a complete metric space by introducing the distance function

$$\begin{aligned} d(q,q') \equiv \| f - f' \| + \int_{-1}^{1} (1 - \delta_{\sigma(\beta)\sigma'(\beta)}) \times \{ & \delta_{\sigma(\beta)a}\,\delta_{\sigma'(\beta)c}\,[\, |\epsilon_1(\beta)| + |\delta'_2(\beta)| \,] + \delta_{\sigma(\beta)c}\,\delta_{\sigma'(\beta)a}\,[\, |\epsilon'_1(\beta)| + |\delta_2(\beta)| \,] \\ & + \delta_{\sigma(\beta)\pm b}\,\delta_{\sigma'(\beta)c}\,[\, |\epsilon_2(\beta)| + |\epsilon_1(\beta)| \,] + \delta_{\sigma(\beta)c}\,\delta_{\sigma'(\beta)\pm b}\,[\, |\epsilon'_2(\beta)| + |\epsilon'_1(\beta)| \,] \\ & + \delta_{\sigma(\beta)a}\,\delta_{\sigma'(\beta)\pm b}[\, |\epsilon'_2(\beta)| + |\delta'_2(\beta)| \,] + \delta_{\sigma(\beta)\pm b}\,\delta_{\sigma'(\beta)a}\,[\, |\epsilon_2(\beta)| + |\delta_2(\beta)| \,] \}\, d\beta , \end{aligned}$$

where $q \equiv [f, \sigma(\beta)]$, $q' \equiv [f', \sigma'(\beta)]$, $\delta_{\sigma \pm b} \equiv \delta_{\sigma + b} + \delta_{\sigma - b}$, $\delta_{\sigma\sigma'}$ denotes the Kronecker delta symbol, and the integral over β is the Lebesgue integral with respect to the standard Lebesgue measure on $(-1,1)$.[40] This elaborate structure of the distance function d is introduced in order to make sure that q approaches q' $[d(q,q') \to 0]$ if and only if the asymptotic behavior described by q approaches that described by q' [cf. Eqs. (3.9)].

By the global existence and uniqueness of solutions of the field equations (2.32) (Sec. III A and Appendixes A and B), there exists a well-defined map

$$\mathscr{E} : D \to A \quad . \tag{3.25}$$

To every $p \in D$, the map $\mathscr{E}$ assigns the unique $q \in A$ that determines the asymptotic behavior near $\alpha=0$ of the global solution which evolves from p. Moreover, it follows from the global well posedness[18,19] of the initial-value problem for (V,W) that $\mathscr{E}$ is a continuous map. [By "global well posedness," we mean the property that solutions of the initial-value problem carry the initial data on a hypersurface Σ_1 onto the data induced on a future hypersurface Σ_2 in a continuous way, i.e., the property that solutions on compact subsets of $D^+(\Sigma_1)$ depend continuously on their initial values on Σ_1. Once global existence and uniqueness of solutions are proved (Appendix A), global well posedness follows from standard arguments; see Ref. 18, Ref. 21, and Sec. 4.2 of Ref. 19.] We claim that the map $\mathscr{E}$ has an inverse

$$\mathscr{E}^{-1} : A \to D$$

which is also continuous. To see this, note that given $q \in A$, $q \equiv [f, \sigma(\beta)]$, we can determine a unique solution (V,W) in the following way: Using the structure functions $\epsilon_1(\beta)$, $\delta_1(\beta)$, $\epsilon_2(\beta)$, $\delta_2(\beta)$ provided by f, we determine the asymptotic limit (3.9) for V and W. [The ambiguity as to which Eq. (3.9) to use will be resolved by the flag $\sigma(\beta)$.] Inserting these expressions (3.9) and (3.10) of V and W into the field equations (2.32), we can compute all the coefficients $c^{(i)}{}_k$ and $d^{(i)}{}_{kl}$ of Eq. (3.10) in terms of f; this yields an asymptotic solution for (V,W). On a spacelike surface in the vicinity of $\alpha=0$, this asymptotic solution induces well-posed initial data, and by global existence and uniqueness (Appendix A) these data can be evolved back onto the initial surface where they induce the desired initial data $p = \mathscr{E}^{-1}(q) \in D$. Clearly, by this construction $\mathscr{E}(p)=q$ and $\mathscr{E}^{-1}[\mathscr{E}(p)]=p$, thus, $\mathscr{E}^{-1}$ is a genuine inverse for $\mathscr{E}$. Again by arguments based on global well posedness of the initial-value problem for (V,W), $\mathscr{E}^{-1}: A \to D$ is a continuous map. Thus, $\mathscr{E}: D \to A$ is a homeomorphism.

In the parallel-polarized $(W \equiv 0)$ case of Ref. 6, the homeomorphism $\mathscr{E}$ is known in explicit form. There, D is the Banach space of all data of the form $[V(r,1), V(1,s)]$, A is the Banach space of all pairs $[\epsilon(\beta), \delta(\beta)]$ [which in the general case correspond to $\epsilon_1(\beta)$ and $\delta_1(\beta)$], and $\mathscr{E}$ is the *linear* map $D \to A$ given by the integral equations (6.3.13) and (6.3.12b). (Note that in this case $A \equiv F$; i.e., no flags σ are necessary to distinguish between different cases of asymptotic behavior [in other words, in this case $\sigma(\beta) \equiv a$ and $\epsilon_2(\beta) = \delta_2(\beta) \equiv 0$, cf. Eqs. (6.3.4)–(6.3.7)].) The inverse of $\mathscr{E}$, $\mathscr{E}^{-1}$, is defined by solving these integral equations for $V(r,1)$ and $V(1,s)$ given $q=[\epsilon(\beta),\delta(\beta)]$. Both $\mathscr{E}$ and $\mathscr{E}^{-1}$ are linear continuous (bounded) maps. Therefore $\mathscr{E}: D \to A$ is a continuous Banach space isomorphism. Note that the construction of an asymptotic solution V from given q is explicitly carried out in Ref. 6 via Eqs. (6.3.7) and (6.3.38).

Now we return to the general (arbitrarily polarized) case, and for each $\delta > 0$ we define a subset H_δ of A as

$$\begin{aligned} H_\delta \equiv \; & \{[f,\sigma(\beta)] \in A \mid \text{there exists at least one connected subinterval of length } \geq \delta \\ & \quad \text{in } (-1,1) \text{ across which } \epsilon_1(\beta) \equiv \pm 1\} \\ \cup \; & \{[f,\sigma(\beta)] \in A \mid \text{there exists at least one connected subinterval of length } \geq \delta \\ & \quad \text{in } (-1,1) \text{ across which } \epsilon_2(\beta) \equiv 1 \text{ and } \delta_1(\beta) \equiv \text{const}\} \quad . \end{aligned} \tag{3.26}$$

By Eqs. (3.20), if $p \in D$ is such that the evolution of p creates at least one connected Killing-Cauchy horizon of β-length $\geq \delta$, then p must belong to $\mathscr{E}^{-1}(H_\delta)$. (See Fig. 2 of Ref. 6.) Clearly, H_δ is a nongeneric subset in the sense of Ref. 6: H_δ is closed and its complement is dense in A. Since $\mathscr{E}$ is a homeomorphism, this implies that $\mathscr{E}^{-1}(H_\delta)$ is nongeneric in D for each $\delta > 0$. Taking $\delta = \delta_P \equiv l_P/\sqrt{l_1 l_2}$ where l_P is the Planck length, this proves that the set of all initial data in D which create at least one connected Killing-Cauchy horizon of larger than Planck size is a nongeneric subset [since it is contained in the nongeneric subset $\mathscr{E}^{-1}(H_{\delta_P})$]. By the same arguments as in Sec. III C of Ref. 6, this is equivalent to proving the full nonlinear instability of the Killing-Cauchy horizons at $\alpha=0$ against generic, plane-symmetric perturbations of the initial data.

Now, assuming that the reader has read through Appendix D, we consider the nongenericity of the set of *all* horizon-producing initial data. We introduce the subset

$$W \equiv \mathscr{E}^{-1}\left[\bigcup_{\delta>0} H_\delta\right] = \bigcup_{\delta>0} \mathscr{E}^{-1}(H_\delta)$$

of D. The set of *all* horizon-producing initial data, $W_H \subset D$, is contained in W : $W_H \subset W$ [cf. Eqs. (3.20) and (3.26)]. Since $\mathscr{E}: D \to A$ is a homeomorphism, we have the following: (i) each $\mathscr{E}^{-1}(H_\delta)$ is a closed set with empty interior, (ii) $\mathscr{E}^{-1}(H_{\delta_2}) \supset \mathscr{E}^{-1}(H_{\delta_1})$ [since $H_{\delta_2} \supset H_{\delta_1}$ by Eq. (3.26)] whenever $\delta_2 < \delta_1$, and (iii) $\bigcup_{\delta>0} \mathscr{E}^{-1}(H_\delta) = W$. As the Banach space D, being a complete metric space, is a Baire space, (i)–(iii) imply that the subset $W \subset D$ is thin in the sense of Appendix D. Therefore, by the definition of nongeneric subsets given in Appendix D, the subset $W_H \subset W \subset D$ of *all* horizon-producing initial data is nongeneric within the space of all plane-symmetric initial data D.

Finally, we make a few remarks on the use of the intuitive notion of genericity in theoretical physics. When physicists use the adjective "generic" they may be referring to any one of two fundamentally different intuitive notions, although the distinction is often not stated explicitly. One of these notions has an essentially *probabilistic* nature: Suppose a system (or a person/observer) chooses a set of parameters (initial conditions, integration constants, model parameters, . . .) out of a continuum of possibilities, and suppose there is evidence that in general the choice is made *at random*. Then the physicists' notions of "nongeneric choice" or "nongeneric outcome" would nicely correspond to the mathematical notion of "measure zero";[40] i.e., a nongeneric choice ($\equiv$ a choice with zero probability) would be one that belongs to a subset of measure zero within the set of all choices. The second notion, on the other hand, has a *constructive* nature: Suppose there is a system or a theoretical model that is to be constructed out of a continuum of possibilities; an initial-value problem is a nice example of such a model. Here "genericity" is the issue of whether the model continues to "behave" in the same way when it is perturbed slightly away from its original form, i.e., the issue of whether the model is constructible in practice (compare the concept of "structural stability" in the theory of dynamical systems[41]). Consequently, genericity in this case is best formulated mathematically as a *topological* condition since the fundamental notions involved in its intuitive description are notions of "neighborhood," e.g. notions such as "slightly perturbed," "nearby," and "stable." [In fact, the probabilistic and topological concepts of genericity are *not* compatible with each other mathematically; for example (as has been pointed out to us by Geroch[42]), the unit interval admits topological homeomorphisms under which closed nowhere-dense subsets with zero Lebesgue measure are carried onto closed nowhere-dense subsets with positive measure.] It is our view that the notion of genericity that is appropriate in general relativity, and in any other similar dynamical-evolution context, is the second *topological* notion as opposed to the more common probabilistic one. We hope that the specific topological concept of genericity discussed in Appendix D will find other useful applications in relativity besides the application that we have described in this section.

IV. SINGULARITIES IN THE COLLISIONS OF ALMOST-PLANE GRAVITATIONAL WAVES

A. A singularity result for colliding almost-plane waves whose initial data are exactly plane symmetric across a sufficiently large region of the initial surface

The content and derivation of the results of this section are so much in parallel with those of Sec. II in Ref. 9 that here we will give only the precise statements of the main conclusions, and brief comments about their derivation. To put the material of this section in proper context, we recommend that readers consult the detailed discussions in Secs. I and II of Ref. 9.

In this paper, as in Ref. 9, we will define an almost-plane wave as a gravitational wave spacetime[12] on which there exist (i) a local coordinate system (u,v,x,y), and (ii) a length scale L_T that characterizes the variation in the x, y directions of the components of geometric quantities, such that (iii) throughout the intersection of a suitable partial Cauchy surface Σ with the wave's *central region* (which has the form $\mathcal{C}=\{|x|<L_T, |y|<L_T, u,v\}$), the metric components and other quantities are very nearly equal to the corresponding quantities for an exact plane-wave spacetime; and (iv) the curvature components fall off to zero arbitrarily (but in a manner consistent with the constraint equations on Σ) as $x^2+y^2 \to \infty$ at constant u and v. When we consider two almost-plane gravitational waves colliding on an otherwise flat background, we will always assume that the central regions of the two waves collide with each other. Then [at least in some neighborhood of the characteristic initial surface $\mathcal{N}=\mathcal{N}_1 \cup \mathcal{N}_2$ formed by the initial wave fronts $\mathcal{N}_1$, $\mathcal{N}_2$ of the colliding waves (Fig. 1)], it is possible[9,12] to set up a local coordinate system in which the conditions (ii)–(iv) above are satisfied for *both* colliding waves simultaneously; but possibly with different transverse length scales $(L_T)_1$ and $(L_T)_2$. In this coordinate system, the initial data supplied by the almost-plane wave 1 and posed on the initial null surface $\mathcal{N}_2$ are very nearly equal, throughout $\mathcal{C}_1 \cap \mathcal{N}_2$, to the initial data posed by a corresponding exact plane wave 1; and the initial data supplied by the almost-plane wave 2 and posed on the initial null surface $\mathcal{N}_1$ are very nearly equal, throughout $\mathcal{C}_2 \cap \mathcal{N}_1$, to the initial data supplied by a corresponding exact plane wave 2. The fundamental problem of colliding almost-plane gravitational waves is then to determine whether (or under what conditions on the initial data) the evolution of these data produces spacetime singularities.

The following lemma is proved in exactly the same way as Lemma 1 of Ref. 9; its derivation uses only the result (Secs. III A and III B) that the asymptotic limit of a *generic* colliding plane-wave metric is an inhomogeneous (nondegenerate) Kasner solution. Restricted to the parallel-polarized case, this fact was also the only ingredient in the proof of Lemma 1 of Ref. 9.

Lemma 1: The intersection $J^-(q) \cap \mathcal{N}$ between the initial surface $\mathcal{N} \equiv \mathcal{N}_1 \cup \mathcal{N}_2$ (Fig. 1), and the causal past $J^-(q)$ of any (generic[9]) point q in the interaction region of a *generic*, arbitrarily polarized colliding plane-wave spacetime is a compact set, whose transverse ($\equiv x,y$) dimensions approach finite limits (i.e., remain bounded from above) as the point q approaches the singularity at $\alpha=0$.

In fact, when the point q has a β value sufficiently far away from the edge points $\beta=+1$ and $\beta=-1$ (e.g., for $-\frac{1}{2}<\beta<\frac{1}{2}$), β remains approximately constant along the past-directed null geodesics from q which extend farthest in the x, y directions; hence, the asymptotic limit (3.18) of the metric (with β-dependent coefficients a, b, c, and

d) remains a good approximation along these geodesics. Furthermore, the coordinates x, y are constant linear combinations of $X_{(\beta)}$, $Y_{(\beta)}$, and in general at least one of the coefficients in each combination is of order 1 whereas the other may be small compared to 1 [cf. Eqs. (3.16d) and (3.17d)]. Therefore, for such a point q approaching $\alpha=0$ at, say, $-\frac{1}{2}<\beta<\frac{1}{2}$, we can estimate the limits of the maximum transverse (coordinate) dimensions of $J^-(q)\cap\mathcal{N}$ by the quantities [compare Eqs. (9.2.4) of Ref. 9]

$$L_x(\beta)=2\left[\frac{a(\beta)}{c(\beta)}\right]^{1/2}\frac{1}{1-p_1(\beta)} \tag{4.1a}$$

and

$$L_y(\beta)=2\left[\frac{a(\beta)}{d(\beta)}\right]^{1/2}\frac{1}{1-p_2(\beta)}, \tag{4.1b}$$

where the constants $a(\beta)$, $c(\beta)$, $d(\beta)$, and the exponents $p_1(\beta)$, $p_2(\beta)$ are found upon comparing Eq. (3.18) with either Eq. (3.15a), Eq. (3.16a), or Eq. (3.17a), depending on whether the asymptotic behavior of the metric is described by case (a), case (b), or case (c), respectively. [Compare Eqs. (9.2.5) of Ref. 9.]

As in Ref. 9, Lemma 1 can be rephrased in the following equivalent form.

Lemma 1 (second version): In a generic colliding (arbitrarily polarized) plane-wave spacetime, the singularity $\{\alpha=0\}$ represents a future c boundary,[14] whose (generic) "points" [which are "terminal indecomposable past sets" (Sec. 6.8 of Ref. 14)] intersect the initial surface $\mathcal{N}$ in subsets with compact closure. In other words, *unless* the colliding plane-wave solution possesses Killing-Cauchy horizons at $\{\alpha=0\}$ destroying its global hyperbolicity [which can only occur for "nongeneric" initial data (Sec. III B)], the (generic) points of the singularity $\{\alpha=0\}$ (when they are considered as points on the future causal boundary of spacetime) can be regarded as part of the domain of dependence $D^+(\mathcal{N})$ of the initial surface $\mathcal{N}$.

The following result was discussed and proved in Sec. II of Ref. 9 (see Lemma 2 and Fig. 4 of Ref. 9).

Lemma 2: Let $(\mathcal{M},g)$ be a spacetime and Σ be a partial Cauchy surface in $(\mathcal{M},g)$ on which gravitational initial data [whose development gives the metric on $D^+(\Sigma)$] are posed. Let $S\subset\Sigma$ be a closed subset, and $\mathcal{U}\subset\Sigma$ be an open subset containing S (Fig. 4 of Ref. 9). Suppose that the initial data on Σ are replaced with a new set of initial data which coincide with the original data throughout $\mathcal{U}$. Then, unless a spacetime singularity forms and penetrates into $D^+(S)$ from outside $D^+(S)$, the new solution coincides with the old solution throughout $D^+(S)$. Here $D^+(S)$ denotes the domain of dependence of S with respect to the original metric and coincides with the domain of dependence of S with respect to the new metric.

Now, introducing the quantity L defined by

$$L\equiv\inf_{-1/2<\beta<+1/2}\max\left[L_x(\beta),L_y(\beta)\right], \tag{4.2}$$

and combining Lemma 2 with the second version of Lemma 1, it becomes clear that we have obtained a proof for the following singularity theorem.

Theorem 1: Let the initial data for two colliding almost-plane gravitational waves be identical to the initial data for two colliding arbitrarily polarized exact plane waves throughout a region $\mathcal{C}$ in the initial surface of the form $\mathcal{C}=\{|x|\leq L_T,|y|\leq L_T\}$. Let the corresponding initial data for this plane-symmetric portion be generic so that the maximal development of the complete plane-symmetric data produces "pure" spacetime singularities at $\alpha=0$ without Killing-Cauchy horizons (Sec. III B). Let these plane-symmetric initial data be represented by the point $p\equiv[V(r,1),W(r,1),V(1,s),W(1,s)]$ in the space D. Compute the image $[f,\sigma(\beta)]\equiv\mathcal{E}(p)\in A$ of p under the map $\mathcal{E}$ defined by Eq. (3.25) (see Sec. III B for notation). Using $[f,\sigma(\beta)]$, construct the quantities $L_x(\beta)$ and $L_y(\beta)$ defined by Eqs. (4.1), and the quantity L defined by Eq. (4.2). Then, if $L_T\gg L$, the evolution of the almost-plane-symmetric data produces spacetime singularities; i.e., the colliding almost-plane waves create spacetime singularities.

Clearly, singularities which are guaranteed to exist by the above theorem will have a local structure that is precisely the same as the structure of the plane-symmetric singularities; i.e., locally these singularities will be of generalized inhomogeneous Kasner type.

Consider now the physically interesting regime where the colliding almost-plane waves both have amplitudes small compared to unity: $h_1\ll 1$, $h_2\ll 1$. (This means that both $V(r,1)$, $V(1,s)$ and $W(r,1)$, $W(1,s)$ are small compared to 1; more precisely, the typical amplitude h for a general plane wave is defined by $h^2\equiv h_+{}^2+h_\times{}^2$, where h_+ and $h_\times$ are the typical magnitudes of V and W, respectively [cf. Eqs. (2.10)].) By Eqs. (6.3.12) and (6.3.13) of Ref. 6 and by the continuity of the map $\mathcal{E}$ [Eq. (3.25)], the quantities $\epsilon_i(\beta)$ and $\delta_i(\beta)$ ($i\equiv 1, 2$) are small compared to 1 in this case. Therefore, if we can choose the initial point (u_0,v_0) [Eqs. (2.27a)] in such a way that the quantity $\mu(\beta)$ is also smaller than or of order unity, then by Eqs. (4.1), (4.2), and (3.15)–(3.17) we could conclude that $L\sim\sqrt{l_1 l_2}$. In fact, as demonstrated in the Appendix of Ref. 9, such a choice is possible: if we fix u_0 and v_0 such that $\lambda_1\ll u_0\ll f_1$, and $\lambda_2\ll v_0\ll f_2$ (where f_1, f_2 are the first focal lengths and λ_1, λ_2 are the typical wavelengths of the colliding waves), then the point (u_0,v_0) belongs to a domain in the interaction region where (i) gravity is weak (since $u_0\ll f_1$ and $v_0\ll f_2$), so that U and the constant additive terms in Eq. (2.33) are small compared to unity, and (ii) the integration path in Eq. (2.33) is sufficiently far away (since $u_0\gg\lambda_1$ and $v_0\gg\lambda_2$) from the coordinate singularities on the initial null surfaces $\{u=0\}$ and $\{v=0\}$ (Sec. II B), so that the contribution to $\mu(\beta)$ from the integrand in Eq. (2.33) (which diverges towards the coordinate singularities on these initial null surfaces) is of order unity [Eqs. (3.11)]. Moreover, with this choice for (u_0,v_0), Eqs. (2.27a) give

$$l_1\sim f_1\,,\qquad l_2\sim f_2\,. \tag{4.3}$$

Since by the above arguments $\mu(\beta)$ is of order 1, when combined with Eqs. (4.1), (4.2), and (3.15)–(3.17) Eqs. (4.3) finally yield the following order-of-magnitude estimate for L, valid for colliding almost-plane waves with small amplitudes:

$$L \sim \sqrt{f_1 f_2}\ . \tag{4.4}$$

Therefore, by Theorem 1, if the colliding almost-plane waves have small initial amplitudes and are exactly plane symmetric across a region of size $L_T \gg \sqrt{f_1 f_2}$ over the initial surface, then their collision produces singularities. These singularities have the same (inhomogeneous-Kasner) local structure as the singularities produced by the exact-plane-wave collision.

B. Singularities produced by colliding almost-plane waves with arbitrary initial data: An existence theorem

In this section we will prove that the conclusions of Theorem 1 (Sec. IV A) about the *existence* of singularities in almost-plane-wave collisions remain valid when the colliding waves are only approximately (but not exactly) plane symmetric throughout their central regions. More precisely, we will prove that if p is a choice of gravitational initial data on $\mathcal{N}$ that satisfies the conditions of Theorem 1 with $L_T \gg L$, then there exists a neighborhood $\mathcal{W}$ of p, open within the space of all gravitational initial data on $\mathcal{N}$, such that the Cauchy development of any data in $\mathcal{W}$ produces spacetime singularities. (For a still more precise statement see below.) Note that a proof of this statement would immediately follow if we could prove that the solutions on $D^+(\mathcal{N})$ depended *uniformly continuously* on the initial data on $\mathcal{N}$. This is in general false, however, because general theorems which assert the continuous dependence of solutions on initial data (such as the Cauchy stability theorem, see, e.g., Sec. 7.6 of Ref. 14) are valid with respect to the *compact-open* topology [i.e., the open topology based on convergence on *compact* subsets of $D^+(\mathcal{N})$], and not with respect to the *open* topology [i.e., the open topology based on (uniform) convergence on $D^+(\mathcal{N})$] on the spaces of all initial data on $\mathcal{N}$ and all four-metrics on $D^+(\mathcal{N})$. [We will denote by $\mathcal{D}$ and $\mathcal{G}$ these spaces of all (vacuum) initial data on $\mathcal{N}$ and all Lorentz metrics on $D^+(\mathcal{N})$, respectively, both topologized with the compact-open topology. The space $\mathcal{D}$ should not be confused with the Banach space D of all *plane-symmetric* vacuum data on $\mathcal{N}$ (Sec. III B).] To see more intuitively why uniform-continuous dependence on initial data fails, recall (i) that singularities can be thought of as points "at infinity," and (ii) that when the initial data p are slightly perturbed their development cannot remain uniformly close to the original solution all the way to infinity (i.e., all the way up to the singularities). The main content of the singularity theorem of this section lies in showing how to get around this failure of uniform-continuous dependence in the specific case of colliding almost-plane gravitational waves.

We first list three Lemmas whose corollaries will directly lead to the proof of our singularity theorem.

Lemma 3: In a nondegenerate Kasner spacetime [Eq. (3.18)], the future null cone $\dot{J}^+(q)$ of any point q starts to reconverge near the singularity $\{t=0\}$, i.e., on each future-directed null geodesic from q the convergence $\hat{\theta}$ (Sec. 4.2 of Ref. 14) of the null generators of $\dot{J}^+(q)$ becomes negative near $t=0$.

The proof of Lemma 3 is given in Appendix E.

Corollary 1: Let $p \in \mathcal{D}$ denote a choice of vacuum initial data on $\mathcal{N}$ that describes colliding almost-plane gravitational waves, and let p satisfy the conditions of Theorem 1 with $L_T \gg L$. Then, for every point q in the Cauchy development of p that lies sufficiently close to the singularity whose existence is guaranteed by Theorem 1, the future null cone $\dot{J}^+(q)$ of q starts to reconverge near $\{\alpha=0\}$; i.e., on each future null geodesic from q the convergence $\hat{\theta}$ of null generators of $\dot{J}^+(q)$ becomes negative near $\alpha=0$.

This corollary follows immediately from Lemma 3, Theorem 1, and the result (Secs. III A and III B) that the asymptotic singularity structure of a generic colliding plane-wave spacetime is of inhomogeneous nondegenerate Kasner type.

Lemma 4: Let $p \in \mathcal{D}$ be vacuum initial data which satisfy the conditions of Theorem 1 with $L_T \gg L$. Then p has an open neighborhood $\mathcal{W}$ in $\mathcal{D}$ such that for any $d \in \mathcal{W}$ the maximal Cauchy development of d contains points q whose future null cones $\dot{J}^+(q)$ start to reconverge.

Proof: In the maximal development of p we can find a compact region $\mathcal{K}$ containing at least some of the points q whose null cones reconverge according to Corollary 1. Furthermore, for at least one such point q, we can obviously also arrange (without destroying the compactness of $\mathcal{K}$) that $\mathcal{K}$ contains a spherical section through the null cone $\dot{J}^+(q)$ of q at which the convergence $\hat{\theta}$ of each null generator of $\dot{J}^+(q)$ is negative. Clearly (since the topology on $\mathcal{G}$ is the compact-open topology), the maximal development of p has an open neighborhood $\mathcal{U}$ in the space of all metrics $\mathcal{G}$, such that these properties of the compact region $\mathcal{K}$ and the point $q \in \mathcal{K}$ continue to hold under any metric on $\mathcal{K}$ that comes from $\mathcal{U}$. The Einstein map, which assigns to every initial data in $\mathcal{D}$ its maximal Cauchy development in $\mathcal{G}$, is continuous by the Cauchy stability theorem.[15] Therefore, the inverse image of $\mathcal{U}$ under the Einstein map is an open subset $\mathcal{W}$ of $\mathcal{D}$, and it is easy to see that this $\mathcal{W}$ satisfies the properties required by the lemma.

Lemma 5 [Hawking-Penrose singularity theorem[13,14]]: Spacetime is causal-geodesically incomplete if (i) $R_{\mu\nu}K^\mu K^\nu \geq 0$ for every nonspacelike vector $\mathbf{K}$, (ii) the causal genericity condition (condition 4.4.5 of Ref. 14) is satisfied, (iii) the chronology condition holds (there are no closed timelike curves), and (iv) there exists a point q such that on every future directed null geodesic from q the convergence $\hat{\theta}$ of the generators of $\dot{J}^+(q)$ becomes negative.

Lemma 5 is stated and proved as Theorem 8.2.2 in Ref. 14.

The following singularity theorem is now obtained as a direct corollary of Lemma 5.

Theorem 2: Let $p \in \mathcal{D}$ be vacuum initial data which satisfy the conditions of Theorem 1 with $L_T \gg L$. Let $\mathcal{W} \subset \mathcal{D}$ be that open neighborhood of p in $\mathcal{D}$ whose existence and properties are demonstrated in Lemma 4. Then, for any $d \in \mathcal{W}$ one of the following is true.

(a) The maximal Cauchy development of d is a maximal (inextendible[14]) spacetime. In this case, this unique inextendible spacetime satisfies conditions (i) (since the maximal development is vacuum), (ii) [cf. Eqs. (3.22)], (iii) (since the maximal development is globally hyperbolic), and (iv) (since Lemma 4 holds for the neighborhood $\mathcal{W}$) of Lemma 5, and therefore it is causal-geodesically incomplete (singular).

(b) The maximal (W^4-) Cauchy development[14,43] of d is bounded by shock waves through which spacetime is extendible but not in a smooth (W^4) way [here W^k denotes the space of metrics which belong to the Sobolev spaces $W^k(\mathcal{V})$ for all spacetime regions $\mathcal{V} \subset \mathcal{M}$ with smooth boundary and compact closure; for details see Secs. 7.4 and 7.6 of Ref. 14]. It is generally believed[14,43] (although not yet proved) that in this case there will be an extension of the maximal development *through* the shock waves, which is uniquely determined by the initial data d and for which conditions (i) and (iii) of Lemma 5 are satisfied. If this is the case, then by the Cauchy stability theorem and the choice of $\mathcal{W}$ conditions (ii) and (iv) will also hold. Thus, if the extension is maximal (i.e., if no Cauchy horizons are encountered), then it will be an inextendible causal-geodesially incomplete (singular) spacetime by Lemma 5.

(c) The maximal (Cauchy) development of d obtained as in (a) [or the maximal development-extension obtained by maximally applying (b)] is bounded by Cauchy horizons; thereby it is extendible. [Note that these Cauchy horizons (if they occur) have nothing to do with the *Killing-Cauchy* horizons (Secs. III A and III B) which are excluded *a priori* by the assumption (Theorem 1) that the central plane-symmetric portion of the initial data p are generic.] In this case, those extensions beyond the Cauchy horizon(s) for which conditions (i) and (iii) of Lemma 5 are everywhere satisfied [note that conditions (ii) and (iv) are always satisfied for any extension] will give maximal spacetimes which are causal-geodesically incomplete (singular) by Lemma 5. For those extensions beyond the Cauchy horizon(s) which violate conditions (i) or (iii) of Lemma 5, the incompleteness of the extended (maximal) spacetime cannot be proved.

On the other hand, if the strong cosmic censorship hypothesis[35,36] holds, then the outcome (c) above is "nongeneric," and, hence, we get the following corollary.

Corollary: If the strong cosmic censorship conjecture[35,36] holds (at least in vacuum) and $\mathcal{W} \subset \mathcal{D}$ is chosen as in Theorem 2, then the unique maximal (inextendible) spacetime obtained from the maximal Cauchy development of any "generic" initial data $d \in \mathcal{W}$ is causal-geodesically incomplete (singular).

Combined with Eq. (4.4), Theorem 2 can be rephrased (roughly) as saying that if two colliding almost-plane waves with small initial amplitudes are sufficiently close to being exactly plane symmetric across a region of size $L_T \gg \sqrt{f_1 f_2}$ on the initial surface, then their collision produces spacetime singularities. Note that the theorem does not give any quantitative information about the "size" of the open neighborhood $\mathcal{W}$ (cf. Lemma 4); i.e., it does not indicate with what degree of accuracy the initial data of the colliding waves must approximate exact plane symmetry in order to produce singularities. Likewise, although the theorem proves the existence of the singularities rigorously, it does not give any information about either their global structure (e.g., whether they are hidden behind an event horizon[9]) or their local asymptotic behavior (e.g., whether they are of Belinsky-Khalatnikov-Lifshitz[10] generic-mixmaster type).

ACKNOWLEDGMENTS

I would like to thank Valeria Ferrari and Basilis Xanthopoulos for informative discussions. This work was partially supported by the National Science Foundation under Grant No. AST 85-14911.

APPENDIX A: PROOF OF GLOBAL EXISTENCE AND UNIQUENESS FOR SOLUTIONS OF THE FIELD EQUATIONS FOR COLLIDING PLANE WAVES

In this appendix we will study the field equations

$$V_{,\alpha\alpha} + \frac{1}{\alpha} V_{,\alpha} - V_{,\beta\beta} = 2(V_{,\beta} W_{,\beta} - V_{,\alpha} W_{,\alpha}) \tanh W , \tag{2.32a}$$

$$W_{,\alpha\alpha} + \frac{1}{\alpha} W_{,\alpha} - W_{,\beta\beta} = (V_{,\alpha}{}^2 - V_{,\beta}{}^2) \sinh W \cosh W \tag{2.32b}$$

for colliding arbitrarily polarized plane waves. We will prove that for *any* smooth initial data

$$\{\ V(r,1),\ W(r,1),\ V(1,s),\ W(1,s)\ \} , \tag{2.37}$$

the solution (V,W) of the initial-value problem (2.32) and (2.37) exists globally and is unique throughout the domain of dependence $D^+(\mathcal{N}) = \{\alpha - \beta \leq 1,\ \alpha + \beta \leq 1\} \cap \{\alpha > 0\}$ of the characteristic initial surface

$$\mathcal{N} \equiv \{r = 1,\ -1 < s \leq 1\} \cup \{s = 1, -1 < r \leq 1\} ,$$
$$r \equiv \alpha - \beta , \qquad s \equiv \alpha + \beta \tag{A1}$$

on which the smooth initial data (2.37) are posed. Notice that here we regard the problem (2.32) and (2.37) as a hyperbolic initial-value problem defined on an *ordinary Euclidean space* R^2, rather than as a problem defined on the interaction region of a four-dimensional Lorentzian colliding plane-wave spacetime (2.31). In this formulation, the Euclidean R^2 on which (2.32) and (2.37) are to be solved is determined simply by the Euclidean coordinates (α,β) [or (r,s)]; the geometry of this Euclidean space and of the characteristic initial-value problem (2.32), (2.37), and (A1) are described in Fig. 2 (cf. also Fig. 1).

Before we actually prove the global existence and uniqueness of solutions for the initial-value problem (2.32), (2.37), and (A1), we will first describe how this problem can be transformed into an equivalent problem in ordinary four-dimensional Minkowski spacetime. It will turn out that the results of this appendix and also of Appendix B below are much easier to obtain for this Minkowski-space problem than the original initial-value problem described above. To explain how this equivalent

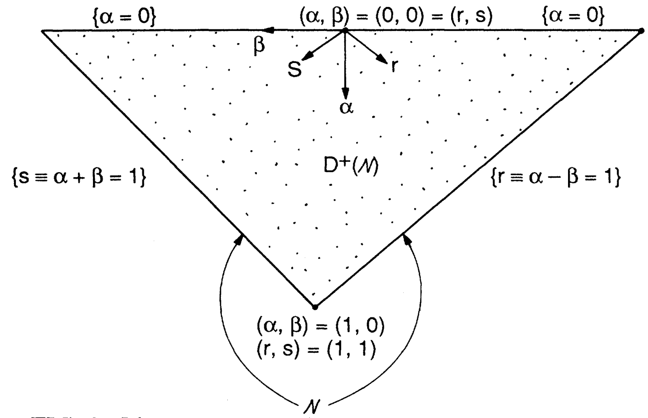

FIG. 2. The geometry of the initial-value problem described by Eqs. (2.32), (2.37), and (A1). The problem is posed in the ordinary Euclidean space R^2 determined by the coordinates (α,β). The characteristic initial surface $\mathcal{N}$ is given by $\mathcal{N} \equiv \{r=1,\ -1<s\le 1\} \cup \{s=1,\ -1<r\le 1\}$, where $r\equiv\alpha-\beta$ and $s\equiv\alpha+\beta$. The domain of dependence $D^+(\mathcal{N})$ is given by $D^+(\mathcal{N})=\{\alpha-\beta\le 1,\ \alpha+\beta\le 1\}\cap\{\alpha>0\}$.

problem arises, we first introduce a "fiducial" four-dimensional spacetime with the metric

$$g_M \equiv -d\alpha^2+d\beta^2+\alpha^2 d\xi^2+d\eta^2\ , \tag{A2}$$

and we consider the invariant wave equations

$$\Box V \equiv V^{;\mu}{}_{;\mu} = -2g_M(\nabla V,\nabla W)\tanh W \equiv -2V^{;\mu}W_{;\mu}\tanh W\ , \tag{A3a}$$

$$\Box W \equiv W^{;\mu}{}_{;\mu} = g_M(\nabla V,\nabla V)\sinh W\cosh W \equiv V^{;\mu}V_{;\mu}\sinh W\cosh W\ , \tag{A3b}$$

defined on this fiducial background (A2). When written explicitly in terms of the (α,β,ξ,η) coordinates, Eqs. (A3) take the form

$$-V_{,\alpha\alpha}-\frac{1}{\alpha}V_{,\alpha}+V_{,\beta\beta}+\frac{1}{\alpha^2}V_{,\xi\xi}+V_{,\eta\eta} = -2\left[V_{,\beta}W_{,\beta}+\frac{1}{\alpha^2}V_{,\xi}W_{,\xi}+V_{,\eta}W_{,\eta} - V_{,\alpha}W_{,\alpha}\right]\tanh W\ , \tag{A4a}$$

$$-W_{,\alpha\alpha}-\frac{1}{\alpha}W_{,\alpha}+W_{,\beta\beta}+\frac{1}{\alpha^2}W_{,\xi\xi}+W_{,\eta\eta} = \left[V_{,\beta}{}^2+\frac{1}{\alpha^2}V_{,\xi}{}^2+V_{,\eta}{}^2 - V_{,\alpha}{}^2\right]\sinh W\cosh W\ , \tag{A4b}$$

and when compared with Eqs. (2.32) they immediately show that the solutions $V(\alpha,\beta)$, $W(\alpha,\beta)$ of the field equations (2.32) correspond *precisely* to the (ξ,η)-independent solutions (V,W) of the invariant wave equations (A3). The advantage of introducing the fiducial spacetime (A2) now becomes clear after one realizes that the metric (A2) is in fact *flat*: By introducing the new coordinates

$$T=-\alpha\cosh\xi\ ,\qquad X=-\alpha\sinh\xi\ ,\qquad Y=\eta\ ,\qquad Z=\beta \tag{A5}$$

in terms of which

$$\alpha=(T^2-X^2)^{1/2}\ ,\qquad \beta=Z\ ,\qquad \xi=\operatorname{arctanh}(X/T)\ ,\qquad \eta=Y\ , \tag{A6}$$

we find that

$$g_M \equiv -d\alpha^2+d\beta^2+\alpha^2 d\xi^2+d\eta^2 = -dT^2+dX^2+dY^2+dZ^2\ . \tag{A7}$$

In fact, a short computation using Eqs. (A5) and (A6) gives

$$\frac{\partial}{\partial\alpha}=\frac{1}{\alpha}\left[T\frac{\partial}{\partial T}+X\frac{\partial}{\partial X}\right]\ ,\qquad \frac{\partial}{\partial\beta}=\frac{\partial}{\partial Z}\ ,\qquad \frac{\partial}{\partial\xi}=X\frac{\partial}{\partial T}+T\frac{\partial}{\partial X}\ ,\qquad \frac{\partial}{\partial\eta}=\frac{\partial}{\partial Y}\ . \tag{A8}$$

Therefore, the spacetime (A2) is precisely the wedge $\{|T|>|X|,\ T<0\}$ in Minkowski space, and (α,β,ξ,η) are the usual wedge coordinates, tuned to the plane-symmetric structure on the wedge that arises due to the presence of the Killing vectors $\partial/\partial Y=\partial/\partial\eta$ (which generates translations) and $X\partial/\partial T+T\partial/\partial X=\partial/\partial\xi$ [which generates (spacelike) Lorentz boosts] (see Sec. I of Ref. 8 for a more detailed discussion of the geometry of this wedge region). The invariant wave equations (A3) can now be rewritten in the form

$$\Box V = -2\nabla V\cdot\nabla W\tanh W\ , \tag{A9a}$$

$$\Box W = (\nabla V)^2\sinh W\cosh W\ , \tag{A9b}$$

where $(\nabla V)^2\equiv\nabla V\cdot\nabla V$, and $\Box$ and $\cdot$ denote the usual wave operator and the usual Lorentzian inner product on Minkowski spacetime, respectively. The term "invariant wave equations" for Eqs. (A9) or (A3) expresses the fact that if (V,W) is any solution to Eqs. (A9) then $(V\circ\phi,\ W\circ\phi)$ is also a solution, where ϕ is any isometry (i.e., any Poincaré transformation) on the (flat) spacetime (A7); that is, isometries of the spacetime leave the solutions invariant. This in particular implies that solutions of Eqs. (A9) are mapped onto solutions under translations along ξ and η [$\equiv$ boosts along X and translations along Y; see Eqs. (A8)].

Notice that we have now obtained a complete reformulation of the initial-value problem Eqs. (2.32), (2.37), and (A1): (i) Instead of the solutions $V(\alpha,\beta)$, $W(\alpha,\beta)$ of Eqs. (2.32), we deal with the plane-symmetric [$\equiv(\xi,\eta)$-independent] solutions of the invariant wave equations (A9) on the Minkowski wedge $\{|T|>|X|,\ T<0\}$. We write these nonlinear wave equations (A9) in the form

$$V_{,kk}-V_{,TT} = -2(V_{,k}W_{,k}-V_{,T}W_{,T})\tanh W\ , \tag{A10a}$$

$$W_{,kk} - W_{,TT} = (V_{,k}V_{,k} - V_{,T}V_{,T})\sinh W \cosh W \,, \tag{A10b}$$

where $x^k \equiv x^1, x^2, x^3 \equiv X, Y, Z$, and we adopt the summation convention that repeated spacelike orthonormal indices $k, l, m, \ldots$ are summed over regardless of whether or not they are contracted. (ii) Instead of posing the initial data for (V, W) in the form (2.37) and (A1), we pose plane-symmetric [$\equiv$ (ξ,η)-independent] initial data for Eqs. (A9) [or equivalently for Eqs. (A10)] on the characteristic initial surface

$$\begin{aligned}\mathcal{C} \equiv &\{(T^2-X^2)^{1/2}-Z=1,\ T<0, 0<(T^2-X^2)\le 1\}\\ &\cup \{(T^2-X^2)^{1/2}+Z=1, T<0, 0<(T^2-X^2)\le 1\}\ .\end{aligned} \tag{A11}$$

The surface $\mathcal{C}$ is a null hypersurface in the fiducial Minkowski spacetime (A7); in fact $\mathcal{C}$ is generated by null geodesics that are orthogonal to the spacelike two-surface $\mathcal{Z} \equiv \{\alpha = (T^2 - X^2)^{1/2} = 1,\ \beta = Z = 0\}$ inside the Minkowski wedge, i.e., by those null generators of $\dot{J}^+(\mathcal{Z})$ that have their past end points on $\mathcal{Z}$. [The readers can see without much difficulty that in the three-dimensional Minkowski space where the Y dimension is absent, $\mathcal{C}$ (where it is a two-dimensional hypersurface $\equiv {}^{(3)}\mathcal{C}$) would be made up of two symmetrically configured half-null-cones intersecting each other at $\mathcal{Z}$; the apex of each half-null-cone would lie on the crease $\{X = T = 0\}$ of the horizon $\{|T| = |X|,\ T \le 0\}$. The surface $\mathcal{C}$ in the four-dimensional case (a three-dimensional null hypersurface) is obtained by just sweeping this two-dimensional ${}^{(3)}\mathcal{C}$ through spacetime parallel to the Y direction.] The two-dimensional (with Z and Y directions suppressed) geometry of this initial-value problem is depicted in Fig. 3. From the invariant character of the nonlinear wave equations (A9), it immediately follows that once we prove the global existence and uniqueness of solutions for Eqs. (A9) with arbitrary initial data posed on an arbitrary initial surface in Minkowski spacetime, this would automatically prove the global existence and uniqueness of solutions for the initial-value problem (2.32), (2.37), and (A1). In particular, when plane-symmetric [$\equiv$ (ξ,η)-independent] initial data for (V, W) are posed on $\mathcal{C}$, the unique global solution (V, W) of the above initial-value problem (A9)–(A11) would be everywhere independent of (ξ,η) (i.e., it would be everywhere plane symmetric); these functions $V(\alpha,\beta)$ and $W(\alpha,\beta)$ would therefore constitute the unique global solution of Eqs. (2.32) corresponding to initial data (2.37) that have the same functional form as the data posed on $\mathcal{C}$ [expressed in (α,β) or (r,s) coordinates].

The introduction of the fiducial four-dimensional Minkowski space (A7) has transformed the problem (2.32), (2.37), and (A1) into a problem in ordinary flat spacetime. [Note that this fiducial flat space (A7) is entirely "fictitious"; i.e., there is no geometric relationship between the spacetime (A7) and the colliding plane-wave spacetime (2.31).] More specifically, by embedding the two-dimensional hyperbolic initial-value problem (2.32), (2.37), and (A1) in a higher-dimensional flat space (from where it is recovered under the restriction of plane symmetry), we have eliminated the singular terms involving $1/\alpha$ from Eqs. (2.32) [compare Eqs. (2.32) with Eqs. (A9)]. The focusing effect described by these singular terms of Eqs. (2.32) has been transformed, in the new formulation (A9)–(A11), into the geometric effect of the exact plane symmetry imposed on the initial data. More precisely, the domain of dependence of the new initial surface $\mathcal{C}$ [Eq. (A11)] is (cf. Fig. 3)

$$D^+(\mathcal{C}) = \{\,|T| > |X|,\ T < 0\,\} \cap J^+(\mathcal{C})\ . \tag{A12}$$

In particular, the horizon $\{|T| = |X|,\ T \le 0\} \equiv \{\alpha = 0\}$ of the Minkowski wedge is the future Cauchy horizon $H^+(\mathcal{C})$ of $\mathcal{C}$ [more precisely, $H^+(\mathcal{C}) = \{|T| = |X|,\ T \le 0\} \cap J^+(\mathcal{C})$]; in fact the region $J^-(q) \cap \mathcal{C}$ becomes unboundedly large in the ξ direction as any arbitrary point q of the wedge approaches the horizon (Fig. 3). As a result, when the initial data posed on $\mathcal{C}$ have a plane-symmetric [(ξ,η)-independent] structure, the data "seen" by any field point q become infinitely extended in the ξ direction as q approaches the horizon $\{\alpha=0\}$. This effect in the formalism (A9)–(A11) is the geometric counterpart of the focusing effect caused by the singular $1/\alpha$ terms in Eqs. (2.32). In particular, it now becomes very clear why the solutions (V, W) of Eqs. (2.32) in general

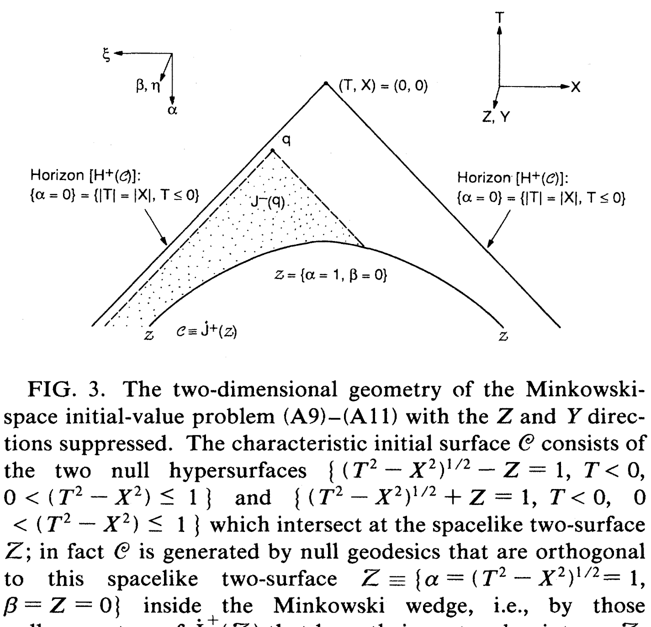

FIG. 3. The two-dimensional geometry of the Minkowski-space initial-value problem (A9)–(A11) with the Z and Y directions suppressed. The characteristic initial surface $\mathcal{C}$ consists of the two null hypersurfaces $\{(T^2 - X^2)^{1/2} - Z = 1,\ T < 0,\ 0 < (T^2 - X^2) \le 1\}$ and $\{(T^2 - X^2)^{1/2} + Z = 1,\ T < 0,\ 0 < (T^2 - X^2) \le 1\}$ which intersect at the spacelike two-surface $\mathcal{Z}$; in fact $\mathcal{C}$ is generated by null geodesics that are orthogonal to this spacelike two-surface $\mathcal{Z} \equiv \{\alpha = (T^2 - X^2)^{1/2} = 1,\ \beta = Z = 0\}$ inside the Minkowski wedge, i.e., by those null generators of $\dot{J}^+(\mathcal{Z})$ that have their past end points on $\mathcal{Z}$. The domain of dependence of the initial surface $\mathcal{C}$ is $D^+(\mathcal{C}) = \{\,|T| > |X|,\ T < 0\} \cap J^+(\mathcal{C})$, and the horizon $\{|T| = |X|,\ T \le 0\} \equiv \{\alpha = 0\}$ of the Minkowski wedge is the future Cauchy horizon $H^+(\mathcal{C})$ of $\mathcal{C}$. The region $J^-(q) \cap \mathcal{C}$ becomes unboundedly large in the ξ direction as any arbitrary point q of the wedge approaches the horizon. As a result, when the initial data posed on $\mathcal{C}$ have a plane-symmetric [(ξ,η)-independent] structure, the data "seen" by any field point q become infinitely extended in the ξ direction as q approaches the horizon $\{\alpha=0\}$. This effect in the formalism (A9)–(A11) is the geometric counterpart of the focusing effect caused by the singular $1/\alpha$ terms in Eqs. (2.32).

develop singularities at $\alpha=0$ (Sec. III A): The global existence of solutions for the initial-value problem (A9)–(A11) (which we will prove below) guarantees that (V,W) are smooth throughout the domain of dependence $D^+(\mathcal{C})$ of the initial surface $\mathcal{C}$, but not necessarily on $\mathcal{C}$'s Cauchy horizon $\{|T|=|X|,\ T\geq 0\}\equiv\{\alpha=0\}$ where the field points are influenced by an infinitely large sector of the initial data (Fig. 3).

In the remaining paragraphs of this appendix we will explain how the global existence and uniqueness of solutions for the system (A9)–(A11) are proved. We remark that the above-discussed specific technique of "resolving" the singularities (i.e., the $1/\alpha$ terms) of the system (2.32), (2.37), and (A1) by embedding it into a higher-dimensional problem [Eqs. (A9)–(A11)] might prove useful more generally, i.e., in studying other PDE's with similar singular coefficients. (Note also that this technique is quite similar to the well-known method of "resolution of singularities" frequently used in the qualitative theory of ordinary differential equations; see, for example, Refs. 41 and 44.)

We now turn to the proof of global existence for Eqs. (A9). The proof of *local existence* (LE) *and uniqueness* for any nonlinear hyperbolic system of the kind Eqs. (A9) is standard and can be found, among other places, in Sec. VI.6 of Ref. 21, and in Refs. 20, 22, 23, and 26. This local result can be stated as follows.

LE: Let Σ be any regular partial Cauchy surface (or a characteristic initial surface consisting of two intersecting null surfaces) in Minkowski space, and let $\{V_0,\ \dot{V}_0,\ W_0,\ \dot{W}_0\}$ ($\{V_0,\ W_0\}$) be regular initial data for Eqs. (A9) on Σ. Then, there exist a neighborhood $\mathcal{U}$ of Σ, and unique functions (V,W) defined on $\mathcal{U}$ which satisfy Eqs. (A9) on $\mathcal{U}$ and which induce the given initial data on Σ. If the data and Σ are C^∞, then (V,W) are C^∞ on $\mathcal{U}$.

In general, *global existence* for a *nonlinear* system of hyperbolic PDE's of the kind Eqs. (A9) is false; see Refs. 22–32 and Secs. I and III A of this paper. Global existence (GE) means, in more precise terms, the following.

GE: Let Σ be any regular partial Cauchy surface in Minkowski space, and let $\{V_0,\ \dot{V}_0,\ W_0,\ \dot{W}_0\}$ be regular initial data for Eqs. (A9) on Σ. Then, there exist unique functions (V,W) defined *throughout the domain of dependence* $D^+(\Sigma)$ *of* Σ, which satisfy Eqs. (A9) on $D^+(\Sigma)$ and which induce the given initial data on Σ. If the data and Σ are C^∞, then (V,W) are C^∞ on $D^+(\Sigma)$.

From the recent work of Klainerman,[23,28,31] Shatah,[25] Sideris,[29] Klainerman and Ponce,[27] and Christodoulou,[32] we know that nonlinear wave equations of the type Eqs. (A9) have global solutions for *small* initial data. More precisely we have the following.

GE for small initial data: Let Σ be any regular partial Cauchy surface and d be regular initial data for Eqs. (A9) on Σ. If d is small, i.e., if the Sobolev norm $\|d\|$ of $\{V_0,\ \dot{V}_0,\ W_0,\ \dot{W}_0\}$ in some suitable Sobolev space[21,39] $W^{k,2}(\Sigma)$ is sufficiently small, then the conclusions of GE above are true for Σ and d.

Now, in order to prove GE for Eqs. (A9) for arbitrary Σ and arbitrary initial data, it is sufficient to prove the following reduced global existence (RGE) result.

RGE: Let arbitrary regular initial data d for Eqs. (A9) be posed on $\Sigma\equiv\{T=0\}$, and let d be compact supported in an open ball $S_0\subset\Sigma$ in Σ. [More precisely, $S_0\equiv\{(X_i,0)\,|\,(X_i-Y_i)(X_i-Y_i)<R^2\}$, for some fixed $(Y_i,0)$ in $\Sigma\equiv\{T=0\}$, and $R>0$.] Then, solutions (V,W) exist which are defined and satisfy Eqs. (A9) throughout the interior $D^+(S_0)$ of the null cone $H^+(\overline{S_0})$, and which coincide with the data d on Σ. These functions (V,W) are *unique*, and they are C^∞ in $D^+(S_0)$ if the initial data d are C^∞.

For Eqs. (A9), RGE implies the more general GE because the characteristics are independent of the specific solution (V,W): the characteristics of Eqs. (A9) are always fixed to be the null hypersurfaces of Minkowski spacetime. Thus, given an arbitrary partial Cauchy surface Σ and arbitrary data d on it, for any point $q\in D^+(\Sigma)$ we can apply the construction described in Fig. 4(a), and introduce a $\{T=0\}$ surface [with some suitable Lorentz coordinates $(T,\ X,\ Y,\ Z)$] in the vicinity of the compact region $J^-(q)\cap\Sigma$. This reduces the global existence problem for Σ to the problem of RGE, provided the data on Σ can be transferred onto $\{T=0\}$ by means of LE. If this fails, then we iteratively apply the construction described in Fig. 4(a) to the points of $\{T=0\}$ [Fig.

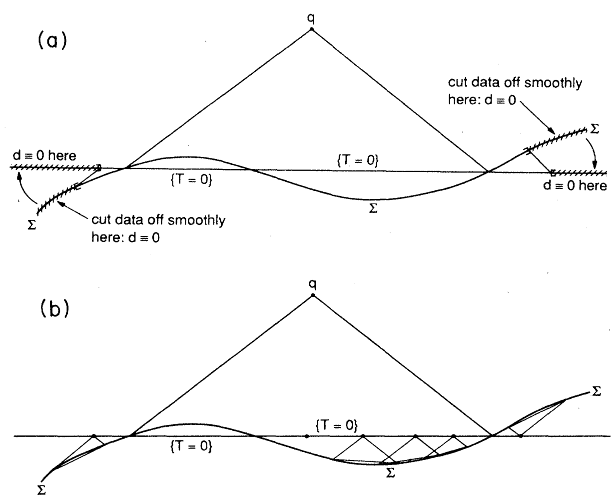

FIG. 4. (a) If reduced global existence (RGE) holds for Eqs. (A9), then this suffices to prove general global existence (GE) (see the precise formulations given in the text): Given an arbitrary partial Cauchy surface Σ and arbitrary data d on it, for any point $q\in D^+(\Sigma)$ we can introduce a $\{T=0\}$ surface [with some suitable Lorentz coordinates $(T,\ X,\ Y,\ Z)$] in the vicinity of the compact region $J^-(q)\cap\Sigma$. This reduces the global existence problem for Σ to the problem of RGE, provided the data on Σ can be transferred onto $\{T=0\}$ by means of local existence (LE). (b) If this fails, then we iteratively apply the construction described in (a) to the points of $\{T=0\}$, and we continue this iteration until the new smaller $\{T=0\}$ surfaces fall into that small neighborhood of Σ on which local existence is guaranteed by LE. Tracing our steps backwards by means of RGE after this last step is achieved, we see that the data on Σ can indeed be transferred to the first $\{T=0\}$ surface depicted in (a).

4(b)], and we continue this iteration until the new smaller $\{T=0\}$ surfaces fall into that small neighborhood of Σ on which local existence is guaranteed by LE. Tracing our steps backwards by means of RGE after this last step is achieved, we see that the data on Σ can indeed be transferred to the *first* $\{T=0\}$ surface depicted in Fig. 4(a).

Remark 1: Once RGE and hence GE are proved as we will do below, then it follows that GE also holds when Σ is a characteristic initial surface consisting of two intersecting null hypersurfaces. This is because for a characteristic Σ and d we can apply the construction described in Fig. 5, and transfer the data d on Σ onto a spacelike hypersurface Σ' which lies in that neighborhood of Σ where local existence is guaranteed by LE. Since global existence and uniqueness hold for Σ' and d', they consequently hold for Σ and d (see Fig. 5).

Remark 2: Here we will prove only the existence part of RGE; once existence is proved, global uniqueness follows from standard arguments as in Refs. 22 and 26.

Proof of RGE for Eqs. (A9): This proof uses three fundamental ingredients.

(i) Conserved positive-definite energy form for Eqs. (A9): One of the most intriguing and special properties of Eqs. (A9) is that they can be derived from a simple Lagrangian. Introducing the Lagrange density

$$\mathcal{L} \equiv -\tfrac{1}{2}\cosh^2 W(V^{,\mu}V_{,\mu}) - \tfrac{1}{2}W^{,\mu}W_{,\mu} , \tag{A13}$$

where $x^\mu \equiv x^0, x^1, x^2, x^3 \equiv T, X, Y, Z$ and greek indices $\mu, \nu, \rho, \ldots$ take the values 0, 1, 2, 3, it is easily seen that the Euler-Lagrange equations

$$\begin{aligned}\frac{\partial}{\partial x^\mu}\frac{\partial\mathcal{L}}{\partial V_{,\mu}} - \frac{\partial\mathcal{L}}{\partial V} &= 0 ,\\ \frac{\partial}{\partial x^\mu}\frac{\partial\mathcal{L}}{\partial W_{,\mu}} - \frac{\partial\mathcal{L}}{\partial W} &= 0 ,\end{aligned}\tag{A14}$$

when applied to $\mathcal{L}$ of Eq. (A13), yield precisely the nonlinear invariant wave equations (A9a) and (A9b). Consequently, we can define a conserved stress-energy tensor

$$T_{\mu\nu} \equiv \eta_{\mu\nu}\mathcal{L} - \frac{\partial\mathcal{L}}{\partial V^{,\mu}}V_{,\nu} - \frac{\partial\mathcal{L}}{\partial W^{,\mu}}W_{,\nu} , \tag{A15}$$

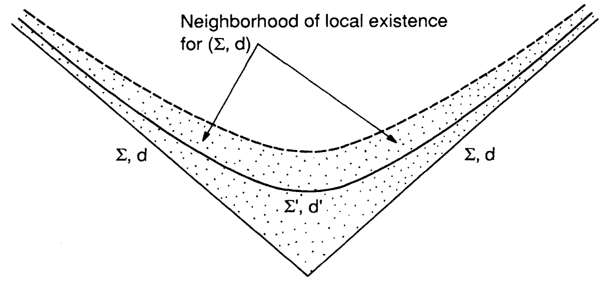

FIG. 5. If global existence for Eqs. (A9) is proved for spacelike initial surfaces, then it also holds when Σ is a characteristic initial surface consisting of two null hypersurfaces that intersect transversally: Given a characteristic surface Σ and data d posed on it, there is a neighborhood (dotted region) of Σ where local existence is guaranteed (by LE; see text). We can find a spacelike initial surface Σ' that lies entirely in this neighborhood, and thereby transfer the data d posed on Σ onto new data d' posed on Σ'. If global existence and uniqueness hold for Σ' and d', then they also hold for Σ and d.

which satisfies

$$T^{\mu\nu}{}_{,\nu} \equiv 0 . \tag{A16}$$

When combined with Eq. (A13), Eq. (A15) gives

$$\begin{aligned}T_{\mu\nu} = {}& \cosh^2 W\, V_{,\mu}V_{,\nu} + W_{,\mu}W_{,\nu}\\ &- \tfrac{1}{2}\eta_{\mu\nu}(\cosh^2 W\, V^{,\rho}V_{,\rho} + W^{,\rho}W_{,\rho}) .\end{aligned}\tag{A17}$$

Therefore, the positive-definite energy form

$$\begin{aligned}T_{TT} = {}& \tfrac{1}{2}[\cosh^2 W(V_{,k}V_{,k} + V_{,T}{}^2)\\ &+ W_{,k}W_{,k} + W_{,T}{}^2]\end{aligned}\tag{A18}$$

has the *conservation* property

$$T_{TT,T} - T_{Tk,k} = 0 . \tag{A19}$$

Consequently, when compact-supported initial data for (V, W) are posed on the initial surface $\{T = \text{const} \equiv \tau\}$, the positive-definite conserved energy form Eq. (A18) satisfies

$$\frac{\partial}{\partial\tau}\int_{\{T=\tau\}} T_{TT}\, d^3X = 0 = \frac{\partial}{\partial\tau}\int_{\{T=\tau\}} \tfrac{1}{2}[\cosh^2 W(V_{,T}{}^2 + V_{,k}V_{,k}) + W_{,T}{}^2 + W_{,k}W_{,k}]\, d^3X \equiv 0 \tag{A20}$$

for all $\tau \geq T$.

(ii) Energy inequality for Eqs. (A9): If the initial-value problem for Eqs. (A9) is posed as in the statement of RGE (see above), then, combined with the positive definiteness of T_{TT}, Eq. (A20) yields (consult Fig. 6 for a description of the relevant geometry)

$$\begin{aligned}&\int_{S_T} \tfrac{1}{2}[\cosh^2 W(V_{,T}{}^2 + V_{,k}V_{,k}) + W_{,T}{}^2 + W_{,k}W_{,k}]\, d^3X\\ &\qquad\leq \int_{S_0} \tfrac{1}{2}[\cosh^2 W(V_{,T}{}^2 + V_{,k}V_{,k}) + W_{,T}{}^2 + W_{,k}W_{,k}]\, d^3X ,\end{aligned}\tag{A21}$$

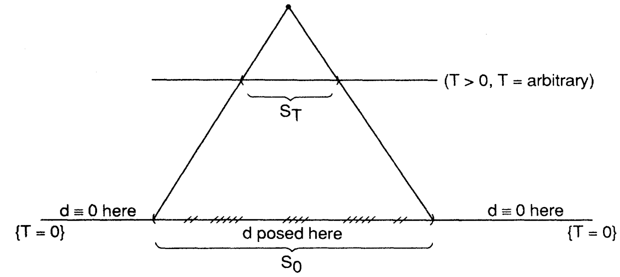

FIG. 6. The geometry of the energy inequality (A21). Initial data d are posed on $\{T=0\}$ and are compact supported in the open ball S_0. The domain of dependence $D^+(S_0)$ of S_0 is the interior of the null cone $H^+(\overline{S_0})$, and S_τ denotes the compact set $\{T=\tau\}\cap D^+(S_0)$.

for all $T>0$. [Here S_τ denotes the compact set $\{T=\tau\}\cap D^+(S_0)$ (Fig. 6).]

(iii) Independence of the characteristics of Eqs. (A9) from the solutions: As we have noted before, for any solution (V,W) the characteristic surfaces of Eqs. (A9) are fixed to be the null hypersurfaces of Minkowski spacetime; i.e., they are independent of the solution.

Now, the proof of RGE follows from the following arguments.

The conservation property Eq. (A20) of the energy form Eq. (A18) implies that the $W^{1,2}$ Sobolev norm of the initial data is conserved; hence the solution does not deteriorate in the L^2 sense. However, this fact by itself is not sufficient to prove RGE: the estimates for the "life-span" of solutions of nonlinear hyperbolic PDE's in general depend on the norm of the data d in higher-order Sobolev spaces than $W^{1,2}$; e.g., they depend on the norm $\|d\|$ in $W^{k,2}(S_0)$ where $k \geq 5$. (See Refs. 22, 23, 26, and 27.) Nevertheless, the (standard) argument outlined in the following paragraph [which uses all three ingredients (i)–(iii) above] suffices to prove RGE.

When the initial data posed on S_0 are analytic, it follows from the Cauchy-Kovalewski theorem[18] that there exists a local analytic solution, determined by an explicit, convergent power series. As is shown in Ref. 20, the fact (iii) above and the energy inequality (A21) imply that in fact this unique analytic solution exists *globally* throughout $D^+(S_0)$. Now, for smooth but nonanalytic initial data d, one approximates d by a series of analytic data d_n; $d_n \to d$ as $n\to\infty$. When combined with (iii), the energy inequality (A21) then shows that the corresponding global analytic solutions (V_n, W_n) in $D^+(S_0)$ *converge* (in $W^{1,2}$) to a smooth global solution (V,W); these limits of V_n and W_n throughout $D^+(S_0)$ constitute the unique global solution of Eqs. (A9) with initial data d. The most crucial step of this proof lies in showing the convergence of the series of analytic solutions (V_n, W_n) throughout $D^+(S_0)$; the energy inequality (A21) is essential for doing so. For the details, consult Ref. 20, Ref. 30, and Sec. VI.5 of Ref. 21.

APPENDIX B: PROOF THAT THE SPATIAL-DERIVATIVE TERMS IN THE FIELD EQUATIONS FOR COLLIDING PLANE WAVES ARE ASYMPTOTICALLY NEGLIGIBLE NEAR $\alpha = 0$

In this appendix, we will prove that the (global) solutions $V(\alpha,\beta)$, $W(\alpha,\beta)$ of the field equations (2.32) have the same asymptotic behavior near $\alpha=0$ as the solutions of the ordinary differential equations

$$V_{,\alpha\alpha}+\frac{1}{\alpha}V_{,\alpha}+2V_{,\alpha}W_{,\alpha}\tanh W=0\,, \tag{3.1a}$$

$$W_{,\alpha\alpha}+\frac{1}{\alpha}W_{,\alpha}-V_{,\alpha}{}^2\sinh W\cosh W=0 \tag{3.1b}$$

which are obtained from Eqs. (2.32) by ignoring all terms with β derivatives. As in Appendix A, we will find that working exclusively with the standard problem (2.32), (2.37), and (A1) is not terribly useful, and we will work instead with the equivalent plane-symmetric [$\equiv(\xi,\eta)$-independent] Minkowski-space initial-value problem given by Eqs. (A9)–(A11).

We begin by introducing the following differential operators $\Lambda_\alpha{}^{(i)}[V,W]$ and $\Lambda_\beta{}^{(i)}[V,W]$ ($i\equiv 1,2$), which are well-behaved *throughout* the Minkowski spacetime $\mathcal{M}$ of Eq. (A7) and which act on smooth functions (V,W) defined on $\mathcal{M}$:

$$\begin{aligned}\Lambda_\alpha{}^{(1)}[V,W]\equiv{}&(T\partial_T+X\partial_X)^2V-(X\partial_T+T\partial_X)^2V\\&+2[(T\partial_T+X\partial_X)V\,(T\partial_T+X\partial_X)W\\&\quad-(X\partial_T+T\partial_X)V\\&\quad\times(X\partial_T+T\partial_X)W]\tanh W\,,\end{aligned} \tag{B1a}$$

$$\begin{aligned}\Lambda_\beta{}^{(1)}[V,W]\equiv{}&-[V_{,YY}+V_{,ZZ}\\&+2(V_{,Y}W_{,Y}+V_{,Z}W_{,Z})\tanh W]\,,\end{aligned} \tag{B1b}$$

$$\begin{aligned}\Lambda_\alpha{}^{(2)}[V,W]\equiv{}&(T\partial_T+X\partial_X)^2W-(X\partial_T+T\partial_X)^2W\\&+\{[(X\partial_T+T\partial_X)V]^2\\&\quad-[(T\partial_T+X\partial_X)V]^2\}\\&\times\sinh W\cosh W\,,\end{aligned} \tag{B2a}$$

$$\begin{aligned}\Lambda_\beta{}^{(2)}[V,W]\equiv{}&-[W_{,YY}+W_{,ZZ}\\&-(V_{,Y}{}^2+V_{,Z}{}^2)\sinh W\cosh W]\,,\end{aligned} \tag{B2b}$$

where ∂_{x^μ} denotes the differential operator $\partial/\partial x^\mu$. Comparing Eqs. (B1) and (B2) with Eqs. (A4) and using Eqs. (A8), it is easy to see that *throughout the open wedge region* $\Lambda\equiv\{|T|>|X|,\ T<0\}$ in Minkowski space (where $\alpha>0$), the differential operators $\Lambda_\alpha{}^{(i)}[V,W]$ and $\Lambda_\beta{}^{(i)}[V,W]$ satisfy

$$\Lambda_\alpha^{(1)}[V,W] = \alpha^2 \left[V_{,\alpha\alpha} + \frac{1}{\alpha} V_{,\alpha} - \frac{1}{\alpha^2} V_{,\xi\xi} + 2\left[V_{,\alpha} W_{,\alpha} - \frac{1}{\alpha^2} V_{,\xi} W_{,\xi} \right] \times \tanh W \right],$$

$$\Lambda_\beta^{(1)}[V,W] = -[V_{,\beta\beta} + V_{,\eta\eta} + 2(V_{,\beta} W_{,\beta} + V_{,\eta} W_{,\eta})\tanh W],$$

$$\Lambda_\alpha^{(2)}[V,W] = \alpha^2 \left[W_{,\alpha\alpha} + \frac{1}{\alpha} W_{,\alpha} - \frac{1}{\alpha^2} W_{,\xi\xi} + \left[\frac{1}{\alpha^2} V_{,\xi}{}^2 - V_{,\alpha}{}^2 \right] \times \sinh W \cosh W \right],$$

$$\Lambda_\beta^{(2)}[V,W] = -[W_{,\beta\beta} + W_{,\eta\eta} - (V_{,\beta}{}^2 + V_{,\eta}{}^2)\sinh W \cosh W]. \tag{B3}$$

It therefore becomes clear from Eqs. (A4) and (B3) that throughout the wedge Λ the invariant wave equations (A9) for V and W can be written in the form

$$\frac{1}{\alpha^2} \Lambda_\alpha^{(i)}[V,W] + \Lambda_\beta^{(i)}[V,W] = 0. \tag{B4}$$

On the other hand, if we introduce the differential operators

$$\mathcal{L}_\alpha^{(1)}[V,W] \equiv V_{,\alpha\alpha} + \frac{1}{\alpha} V_{,\alpha} + 2 V_{,\alpha} W_{,\alpha} \tanh W, \tag{B5a}$$

$$\mathcal{L}_\beta^{(1)}[V,W] \equiv -2 V_{,\beta} W_{,\beta} \tanh W - V_{,\beta\beta}, \tag{B5b}$$

$$\mathcal{L}_\alpha^{(2)}[V,W] \equiv W_{,\alpha\alpha} + \frac{1}{\alpha} W_{,\alpha} - V_{,\alpha}{}^2 \sinh W \cosh W, \tag{B6a}$$

$$\mathcal{L}_\beta^{(2)}[V,W] \equiv V_{,\beta}{}^2 \sinh W \cosh W - W_{,\beta\beta}, \tag{B6b}$$

which are well-behaved *throughout the open wedge* Λ but which are *singular* (in fact undefined) outside it, then we can rewrite the field equations (2.32) in the form

$$\mathcal{L}_\alpha^{(i)}[V,W] + \mathcal{L}_\beta^{(i)}[V,W] = 0, \tag{B7}$$

with the additional restriction that the solutions V and W must be plane symmetric, i.e., independent of (ξ,η).

Now consider given plane-symmetric initial data $\{V_0{}^{(\infty)}, W_0{}^{(\infty)}\}$ posed on the initial surface $\mathcal{C}$ of Eq. (A11) (see also Fig. 3). (The rationale for our notation will become clear in a moment.) For any $L > 0$, we construct a new set of initial data on $\mathcal{C}$ by the relations

$$V_0{}^{(L)} \equiv V_0{}^{(\infty)} f^{(L)}(\xi^2+\eta^2), \quad W_0{}^{(L)} \equiv W_0{}^{(\infty)} f^{(L)}(\xi^2+\eta^2), \tag{B8a}$$

where $f^{(L)}(u)$ is a family of smooth functions in $C^\infty(R)$ satisfying (for each $L > 0$)

$$f^{(L)}(u) = 1 \quad \text{for } u \le L^2, \qquad f^{(L)}(u) = 0 \quad \text{for } u \ge 4L^2, \qquad \frac{d}{du} f^{(L)}(u) \le 0 \quad \forall u \in R. \tag{B8b}$$

In other words, the initial data $\{V_0{}^{(L)}, W_0{}^{(L)}\}$ are obtained by smoothly cutting off the plane-symmetric initial data $\{V_0{}^{(\infty)}, W_0{}^{(\infty)}\}$ at a distance $2L$ in the ξ and η directions. [The existence of smooth functions f with the properties (B8b) is a well-known result in elementary analysis; see, e.g., Lemma 1.10 of Ref. 17.] By Appendix A, for each $L > 0$ there exists a global solution $(V^{(L)}, W^{(L)})$ of Eqs. (A9) [or equivalently of Eqs. (B4)] which is defined throughout the wedge Λ and which evolves from the initial data (B8) on $\mathcal{C}$. We claim that for any *finite* $L > 0$, these solutions $(V^{(L)}, W^{(L)})$ are in fact smooth and well behaved *on and across the Cauchy horizon* $H^+(\mathcal{C}) = \{|T| = |X|, T \le 0\} = \{\alpha = 0\}$ of $\mathcal{C}$. To see this, consider the construction depicted in Fig. 7: This figure describes how we build a new initial surface Π by (i) choosing an $R > 0$ with $R > 2L$, (ii) adjoining a smooth spacelike hypersurface Σ to the initial surface $\mathcal{C}$ at the cylindrical cross section $\mathcal{C} \cap \{\xi^2+\eta^2=R^2\}$ *through* $\mathcal{C}$, and finally (iii) discarding the portion of $\mathcal{C}$ that remains in the past of Σ (Fig. 7). [Note that the geometry described in Fig. 7 is fully symmetric in the ξ and η directions; consequently, the three-dimensional picture of Π with the β ($\equiv Z$) direction suppressed can be obtained by rotating Fig. 7 around the T axis.] On the new initial surface Π, we pose new initial data d for (V,W) by leaving the data as they are on $\mathcal{C}$ (i.e., $\{d \text{ on } \mathcal{C}\} \equiv \{V_0{}^{(L)}, W_0{}^{(L)}\}$) and by putting $d \equiv 0$ on Σ. Inspection of Fig. 7 makes it clear that throughout the subset of the wedge Λ that corresponds to the dotted region in Fig. 7, the global solution of the initial-value problem (Π, d) for Eqs. (A9) [or for Eqs. (B4)] is precisely equal to the solution $(V^{(L)}, W^{(L)})$. Moreover, it is also obvious from Fig. 7 that the domain of dependence of Π includes the horizon $H^+(\mathcal{C})$ as well as the region that lies beyond the horizon. Therefore, since by Appendix A the solution of the initial-value problem (Π, d) exists smoothly throughout $D^+(\Pi)$, we conclude that the solution $(V^{(L)}, W^{(L)})$ of the problem $(\mathcal{C}, d^{(L)})$ is also smooth at and across the horizon $H^+(\mathcal{C})$.

The following identities are now easily derived from Eqs. (B3), (B5), (B6), and (B8):

$$\lim_{L\to\infty} V^{(L)} = V^{(\infty)}, \qquad \lim_{L\to\infty} W^{(L)} = W^{(\infty)}, \tag{B9}$$

$$\mathcal{L}_\alpha^{(i)}[V^{(\infty)}, W^{(\infty)}] = \lim_{L\to\infty} \frac{1}{\alpha^2} \Lambda_\alpha^{(i)}[V^{(L)}, W^{(L)}], \tag{B10a}$$

$$\mathcal{L}_\beta^{(i)}[V^{(\infty)}, W^{(\infty)}] = \lim_{L\to\infty} \Lambda_\beta^{(i)}[V^{(L)}, W^{(L)}]. \tag{B10b}$$

By Eq. (B4), the nonlinear wave equations (A9) satisfied by $V^{(L)}$ and $W^{(L)}$ can be written in the form

$$\Lambda_\alpha{}^{(i)}[V^{(L)}, W^{(L)}] = -\alpha^2 \Lambda_\beta{}^{(i)}[V^{(L)}, W^{(L)}] . \tag{B11}$$

Since $\Lambda_\beta{}^{(i)}[V,W]$ are smooth differential operators well behaved throughout Minkowski spacetime, and since by the above paragraph $V^{(L)}$ and $W^{(L)}$ are also well behaved on and across the horizon $H^+(\mathcal{C}) = \{\alpha = 0\}$, Eq. (B11) proves that

$$\Lambda_\alpha{}^{(i)}[V^{(L)}, W^{(L)}] \to 0 \text{ asymptotically as } \alpha \to 0 . \tag{B12}$$

Moreover, it is clear from Eqs. (B1a), (B2a), and (B3) that the operators $\Lambda_\alpha{}^{(i)}[V,W]$ are *not* multiples of α^2; i.e., they *cannot* be written in the form $\alpha^2\mathcal{P}^{(i)}[V,W]$ where $\mathcal{P}^{(i)}[V,W]$ are *smooth* operators throughout the Minkowski spacetime $\mathcal{M}$. Therefore, it follows from Eqs. (B11) and (B12) that the asymptotic behaviors of the solutions $(V^{(L)}, W^{(L)})$ as $\alpha \to 0$ are the same as those of the solutions $(V_{\text{as}}{}^{(L)}, W_{\text{as}}{}^{(L)})$ of

$$\Lambda_\alpha{}^{(i)}[V_{\text{as}}{}^{(L)}, W_{\text{as}}{}^{(L)}] \equiv 0 . \tag{B13}$$

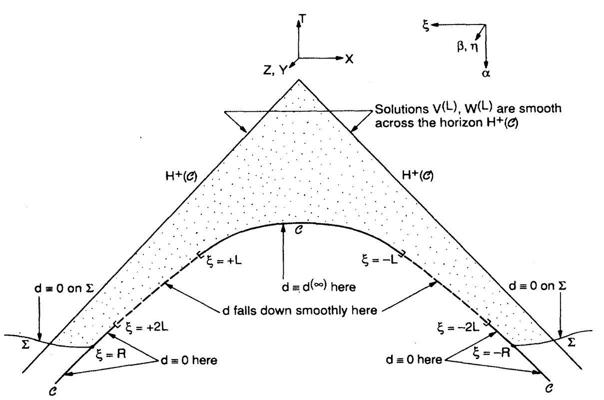

FIG. 7. Geometry of the initial-value problem for Eqs. (A9) where the initial data given by Eqs. (B8) are posed on the characteristic surface $\mathcal{C}$ (see Fig. 3). The initial data $\{V_0{}^{(L)}, W_0{}^{(L)}\}$ [Eqs. (B8)] are obtained by smoothly cutting off the plane-symmetric initial data $\{V_0{}^{(\infty)}, W_0{}^{(\infty)}\}$ at a distance $2L$ in the ξ and η directions. To prove that the solution $(V^{(L)}, W^{(L)})$ that evolves from these data is smooth across the horizon $H^+(\mathcal{C})$, a new initial surface Π is constructed by (i) choosing an $R > 0$ with $R > 2L$, (ii) adjoining a smooth spacelike hypersurface Σ to the initial surface $\mathcal{C}$ at the cylindrical cross section $\mathcal{C} \cap \{\xi^2 + \eta^2 = R^2\}$ *through* $\mathcal{C}$, and finally (iii) discarding the portion of $\mathcal{C}$ that remains in the past of Σ. On the new initial surface Π, new initial data d for (V,W) are posed by leaving the data as they are on $\mathcal{C}$ (i.e., $\{d \text{ on } \mathcal{C}\} \equiv \{V_0{}^{(L)}, W_0{}^{(L)}\}$) and by putting $d \equiv 0$ on Σ. Throughout the subset of the Minkowski wedge that corresponds to the dotted region, the global solution of the initial-value problem (Π, d) for Eqs. (A9) [or for Eqs. (B4)] is precisely equal to the solution $(V^{(L)}, W^{(L)})$. Moreover, the domain of dependence of Π includes the horizon $H^+(\mathcal{C})$. Since by Appendix A the solution of the initial-value problem (Π, d) exists smoothly throughout $D^+(\Pi)$, we conclude that the solution $(V^{(L)}, W^{(L)})$ of the problem $(\mathcal{C}, d^{(L)})$ is also smooth at and across the horizon $H^+(\mathcal{C})$.

On the other hand, in the wedge region Λ where $\alpha > 0$, we can rewrite Eq. (B13) (trivially) as

$$\frac{1}{\alpha^2} \Lambda_\alpha{}^{(i)}[V_{\text{as}}{}^{(L)}, W_{\text{as}}{}^{(L)}] \equiv 0 . \tag{B14}$$

Taking the limit of Eq. (B14) as $L \to \infty$ and using Eqs. (B9) and (B10), we obtain

$$\begin{aligned} 0 &\equiv \lim_{L\to\infty} \frac{1}{\alpha^2} \Lambda_\alpha{}^{(i)}[V_{\text{as}}{}^{(L)}, W_{\text{as}}{}^{(L)}] \\ &= \mathcal{L}_\alpha{}^{(i)}[V_{\text{as}}{}^{(\infty)}, W_{\text{as}}{}^{(\infty)}] \equiv 0 . \end{aligned} \tag{B15}$$

When compared with Eqs. (B5a) and (B6a), Eq. (B15) proves our claim that the solutions $(V^{(\infty)}, W^{(\infty)})$ of the field equations (2.32) have the same asymptotic behavior near $\alpha = 0$ as the solutions of the ordinary differential equations (3.1).

APPENDIX C: SOME REMARKS ON THE FIELD EQUATIONS FOR COLLIDING NONPARALLEL-POLARIZED PLANE WAVES

In this appendix, we will describe some interesting equivalent formulations of the field equations (2.32) for arbitrarily polarized colliding plane waves; we hope that some of these alternative forms might eventually prove useful in the search for a general solution of Eqs. (2.32).

For the first reformulation, we introduce a 1-form $\Theta(\alpha,\beta)$ by the relation

$$\Theta \equiv \cosh^2 W \, dV . \tag{C1}$$

Denoting the α, β components of Θ by Θ_α and Θ_β, respectively (that is, putting $\Theta = \Theta_\alpha d\alpha + \Theta_\beta d\beta$), we can then express the field equations (2.32) purely in terms of Θ and the function $W(\alpha,\beta)$:

$$\frac{1}{\alpha}(\alpha\Theta_\alpha)_{,\alpha} - \Theta_{\beta,\beta} = 0 , \tag{C2a}$$

$$W_{,\alpha\alpha} + \frac{1}{\alpha} W_{,\alpha} - W_{,\beta\beta} = (\Theta_\alpha{}^2 - \Theta_\beta{}^2) \frac{\sinh W}{\cosh^3 W} , \tag{C2b}$$

where Eqs. (C2a) and (C2b) are to be solved subject to the *auxiliary condition*

$$d\Theta = 2\tanh W \, dW \wedge \Theta . \tag{C2c}$$

Now consider the special case determined by the ansatz

$$d\Theta \equiv 0 , \tag{C3}$$

which is equivalent to $dW \wedge dV \equiv 0$, and which is in turn equivalent to the existence of a functional relationship between V and W. The class of solutions that obey the condition (C3) includes all parallel-polarized ($W \equiv 0$) solutions, as well as solutions (V,W) that one obtains from parallel-polarized metrics (2.31) by effecting a constant linear transformation on the coordinates x and y, thereby introducing an artificial cross-polarization component W. However, the special class (C3) is clearly larger than the class of these essentially parallel-polarized solutions. In any case, if by utilizing the condition (C3) we introduce a new function $\tilde{V}(\alpha,\beta)$ that satisfies

$$d\tilde{V} \equiv \Theta = \cosh^2 W\, dV \;, \tag{C4}$$

then the field equations (C2) can be rewritten in terms of the two functions $\tilde{V}$ and W in the form

$$\tilde{V}_{,\alpha\alpha} + \frac{1}{\alpha}\tilde{V}_{,\alpha} - \tilde{V}_{,\beta\beta} = 0 \;, \tag{C5a}$$

$$W_{,\alpha\alpha} + \frac{1}{\alpha}W_{,\alpha} - W_{,\beta\beta} = (\tilde{V}_{,\alpha}{}^2 - \tilde{V}_{,\beta}{}^2)\frac{\sinh W}{\cosh^3 W} \;, \tag{C5b}$$

where Eqs. (C5a) and (C5b) must be solved subject to the *auxiliary condition*

$$d\tilde{V} \wedge dW \equiv 0 \;. \tag{C5c}$$

The solution of the linear equation (C5a) can be found explicitly in terms of initial data; see Sec. II B of Ref. 6, especially Eqs. (6.2.44a) and (6.2.60). In fact, it becomes clear from Eq. (C5a) that in this special case [Eq. (C3)] we can express the asymptotic structure function $\epsilon_1(\beta)$ (Sec. III A) explicitly in terms of the initial data for $\tilde{V}(\alpha,\beta)$: Combining Eqs. (3.3) and (3.4a) with Eq. (C4), and comparing Eq. (C5a) with Eqs. (6.2.44a), (6.3.7), and (6.3.13) of Ref. 6, we obtain

$$\begin{aligned}\epsilon_1(\beta) = &\frac{1}{\pi}\frac{1}{\sqrt{1+\beta}}\int_\beta^1 [(1+s)^{1/2}\tilde{V}(1,s)]_{,s} \times \left[\frac{s+1}{s-\beta}\right]^{1/2} ds \\ &+ \frac{1}{\pi}\frac{1}{\sqrt{1-\beta}}\int_{-\beta}^1 [(1+r)^{1/2}\tilde{V}(r,1)]_{,r} \times \left[\frac{r+1}{r+\beta}\right]^{1/2} dr \;.\end{aligned} \tag{C6}$$

Returning now to the general case (C2), we note that the field equation (C2a) for Θ can be rewritten as

$$(\alpha\Theta_\alpha)_{,\alpha} = (\alpha\Theta_\beta)_{,\beta} \;. \tag{C7}$$

Equation (C7) implies that there exists a function $S(\alpha,\beta)$ that satisfies

$$S_{,\alpha} = \alpha\Theta_\beta \;, \qquad S_{,\beta} = \alpha\Theta_\alpha \;, \tag{C8}$$

and in turn, Eqs. (C8) can be expressed in the equivalent form

$$\Theta \equiv \frac{1}{\alpha} * dS \;, \tag{C9}$$

where the Hodge-star[17] operator $*$ is defined with respect to the two-dimensional flat metric $(-d\alpha^2 + d\beta^2)$. In terms of the two functions $S(\alpha,\beta)$ and $W(\alpha,\beta)$, the field equations (2.32) [or equivalently Eqs. (C2)] can now be rewritten in the alternative form

$$d\left[\frac{1}{\alpha} * dS\right] = \frac{2\tanh W}{\alpha} dW \wedge * dS \;, \tag{C10a}$$

$$d(\alpha * dW) = -\frac{\sinh W}{\alpha\cosh^3 W} dS \wedge * dS \;, \tag{C10b}$$

with no auxiliary conditions.

APPENDIX D: A MORE SOPHISTICATED FORMULATION OF THE NOTION OF NONGENERICITY IN AN ARBITRARY BAIRE SPACE

Recall the simple definition that we introduced in Sec. III C of Ref. 6 to describe the nongenericity of a subset in an arbitrary Banach space. According to this definition, a subset is nongeneric if it is closed and has a dense complement, i.e., if it is a closed subset with empty interior. Although this notion of genericity is both intuitively plausible and broad enough to describe the nongenericity of larger-than-Planck-size Killing-Cauchy horizons in colliding plane-wave spacetimes (Sec. III B), it is too naive even to identify the set Q of rational numbers as a nongeneric subset within the real line R Similarly, it fails to describe the nongenericity of the subset $\cup_{\delta>0} H_\delta$ of *all* horizon-producing initial data within the Banach space of all initial data for colliding plane waves (Sec. III B). Clearly, a more sophisticated generalization of the above notion of nongenericity is needed to avoid these drawbacks; in this appendix we will describe such a generalization. Just as the above notion of genericity applies not only to a Banach space but more generally to arbitrary topological spaces, so also here we will formulate our generalization for a broad class of topological spaces called *Baire spaces* (see the definitions below). Any complete metric space (hence in particular any Banach space) is a Baire space; thus our notions would be applicable to most function spaces that arise naturally in mathematical physics. In the following, we will omit the full proofs of many of the standard results that we use; more detailed discussions on these results can be found in any textbook on general topology, e.g., in Ref. 45.

We first review some of the basic definitions: A topological space X is called "of the first category" if X is the union of *countably many* closed subsets with empty interiors; otherwise, X is called of the second category. These definitions apply to a subset $S \subset X$ by regarding S as a topological space under the topology induced from X. (Thus: $Q \subset R$ is of the first category; {irrational numbers} $\subset R$ is of the second category.) The space X is said to be a *Baire space*[45] if every nonempty open subset of X is of the second category. It is not very difficult to prove[45] that X is a Baire space if and only if for every countable collection of nonempty closed subsets $\{A_n \subset X\}$ with empty interiors, $\cup_{n=1}^{\infty} A_n \subset X$ is a subset with empty interior. (Thus, Q is *not* a Baire space; R is a Baire space.) A fundamental result[45] is that *every complete metric space is a Baire space*.

Our definition of "thin" subsets: Let B be a Baire space (or more specifically a complete metric space). A subset $S \subset B$ is called *thin* if and only if there exists a family of subsets $\{H_\delta \subset B\}$ with the following properties (here $\delta > 0$ ranges over all positive real numbers): (i) For each $\delta > 0$, H_δ is a closed subset with empty interior in B; (ii) if $\delta_2 < \delta_1$, then $H_{\delta_2} \supset H_{\delta_1}$; (iii) $\cup_{\delta>0} H_\delta = S$.

In particular, if $S \subset B$ is a closed subset with empty interior then it is thin: just take $H_\delta \equiv S$ for all $\delta > 0$. Hence the notion of "thin" subsets generalizes the naive notion of nongenericity that we introduced in Ref. 6. In fact, this is an intuitively plausible generalization: It follows from the properties (i)–(iii) that the thin subset S is essentially the "limit" as $\delta \to 0$ of the "nongeneric" subsets H_δ; therefore, intuitively a thin subset is just the "limit" of a continuous family of subsets which are all nongeneric in the sense of Ref. 6. Some of the other properties that thin subsets have according to the above definition are described in the following paragraph.

The first important property is the following alternative characterization: A subset $S \subset B$ in a Baire space is thin if and only if there exists a countable family $\{A_n \subset B\}$ of closed subsets of B, each with empty interior, such that $S = \cup_{n=1}^{\infty} A_n$. (To prove the if part, given the countable family $\{A_n\}$ of closed subsets with empty interiors satisfying $\cup_{n=1}^{\infty} A_n = S$, put $H_\delta \equiv \cup_{n=1}^{[1/\delta]} A_n$, where $[1/\delta]$ denotes the smallest integer $\geq 1/\delta$. The family $\{H_\delta\}$ satisfies property (i) since B is a Baire space; the other properties (ii) and (iii) are satisfied by construction. To prove the only if part, given the family $\{H_\delta\}$ satisfying properties (i)–(iii), put $A_n \equiv H_{1/n}$.) As a consequence, the subset Q of rationals is thin in R, whereas the subset of irrational numbers is *not* thin. Also, if $S \subset B$ is thin and $P \subset S$ is closed *in* S, then P is a thin subset in B. Notice that our notion of a thin subset is essentially different from the notion of a subset of the first category: A thin subset is not necessarily of the first category (any closed subset with empty interior in a complete metric space is thin but not of the first category), and conversely a subset of the first category is not necessarily thin [the subset $S \subset R^2$ given by $S \equiv \{(x,y) \in R^2 \mid 0<x<1,\ x \text{ is irrational},\ 0 \leq y \leq 1,\ y \text{ is rational}\}$ is of the first category but not thin in R^2]. Nevertheless, it follows from the above alternative characterization of thin subsets that just as the subsets of the first category of a Baire space have empty interiors, so also its thin subsets have empty interiors; in other words the complement of any thin subset is dense in B.

Although it presents a more general alternative to our older, more naive concept of a nongeneric subset, the notion of a thin subset is nevertheless inappropriate as a concept of nongenericity. The reason is that subsets of a thin set are not necessarily thin unless they are closed (see above), whereas intuitively one expects that any subset of a nongeneric set should itself be nongeneric. To satisfy this requirement and at the same time to preserve the remaining plausible characteristics of "thinness," we therefore adopt the following most straightforward derivative of the notion of a thin subset as our generalized concept of nongenericity.

Our notion of nongeneric subsets: A subset $P \subset B$ of a Baire space B is called *nongeneric* if and only if P is contained in a thin subset of B.

It becomes obvious that any subset of a nongeneric subset is itself nongeneric. It also follows that although a nongeneric subset is not necessarily of the first category, any subset of the first category in a Baire space is nongeneric.

APPENDIX E: PROOF OF THE LEMMA THAT FUTURE NULL CONES IN A NONDEGENERATE KASNER SPACETIME START TO RECONVERGE

In this appendix we will prove the following result which is used in the proof of Theorem 2 in Sec. IV B.

Lemma 3: In a nondegenerate Kasner spacetime [Eq. (3.18)], the future null cone $\dot{J}^+(q)$ of any point q starts to reconverge near the singularity $\{t=0\}$; i.e., on each future-directed null geodesic from q the convergence $\hat{\theta}$ (Sec. 4.2 of Ref. 14) of the null generators of $\dot{J}^+(q)$ becomes negative near $t=0$.

Consider a general nondegenerate Kasner spacetime with the metric (3.18):

$$g = -a\,dt^2 + bt^{2p_3}d\beta^2 + ct^{2p_1}dx^2 + dt^{2p_2}dy^2 , \tag{E1}$$

where a, b are positive constants having the dimensions of $(\text{length})^2$, c, d are dimensionless positive constants, t, β are dimensionless coordinates, and the exponents p_k, $k=1,2,3$ satisfy the Kasner relations

$$p_1 + p_2 + p_3 = p_1{}^2 + p_2{}^2 + p_3{}^2 = 1 . \tag{E2}$$

It follows from Eqs. (E2) that if the metric (E1) is nondegenerate [i.e., if all exponents p_k are different from 0 (or equivalently if all exponents are different from 1)], then precisely two exponents are strictly positive and precisely one is strictly negative. Thus we will assume, without loss of generality, that

$$1 > p_1 \geq p_2 > 0 , \quad -1 < p_3 < 0 . \tag{E3}$$

[In fact Eqs. (E2) imply that $p_3 > -(\sqrt{5}-1)/2$ in this case, but we will not use this sharper inequality below.] Consider now an arbitrary point q in the Kasner spacetime (E1) with coordinates t_0, x_0, y_0, β_0, $t_0 > 0$. We will explore the behavior of the future null cone $\dot{J}^+(q)$ of this point q; in particular, we are interested in evaluating the asymptotic behavior (as $t \to 0$) of the convergence $\hat{\theta}$ for the null geodesics which generate $\dot{J}^+(q)$. Let the integrals of motion $g(\gamma_*, \partial/\partial x)$, $g(\gamma_*, \partial/\partial y)$, and $g(\gamma_*, \partial/\partial\beta)$ (associated with the Killing vector fields $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial\beta$) along a future-directed null geodesic γ from q be denoted by

$$g(\gamma_*, \partial/\partial x) \equiv C_x , \qquad g(\gamma_*, \partial/\partial y) \equiv C_y , \\ g(\gamma_*, \partial/\partial\beta) \equiv C_\beta . \tag{E4}$$

Then, a short computation shows that as functions of the time coordinate t, $t \leq t_0$, (i) the coordinates $x(t)$, $y(t)$, and $\beta(t)$ of any point $q(t)$ along the null geodesic γ are given by

$$x(t)=x_0+\int_t^{t_0}\frac{C_x}{cs^{2p_1}}\times\frac{\sqrt{a}}{\left[\dfrac{C_\beta^2}{bs^{2p_3}}+\dfrac{C_x^2}{cs^{2p_1}}+\dfrac{C_y^2}{ds^{2p_2}}\right]^{1/2}}ds\ , \tag{E5a}$$

$$y(t)=y_0+\int_t^{t_0}\frac{C_y}{d\,s^{2p_2}}\times\frac{\sqrt{a}}{\left[\dfrac{C_\beta^2}{bs^{2p_3}}+\dfrac{C_x^2}{cs^{2p_1}}+\dfrac{C_y^2}{ds^{2p_2}}\right]^{1/2}}ds\ , \tag{E5b}$$

$$\beta(t)=\beta_0+\int_t^{t_0}\frac{C_\beta}{bs^{2p_3}}\times\frac{\sqrt{a}}{\left[\dfrac{C_\beta^2}{bs^{2p_3}}+\dfrac{C_x^2}{cs^{2p_1}}+\dfrac{C_y^2}{ds^{2p_2}}\right]^{1/2}}ds\ , \tag{E5c}$$

and (ii) the tangent vector $\gamma_*(t)$ to the null geodesic γ is given by

$$\gamma_*(t)=\left[\frac{C_x}{ct^{2p_1}}\right]\frac{\partial}{\partial x}+\left[\frac{C_y}{dt^{2p_2}}\right]\frac{\partial}{\partial y}+\left[\frac{C_\beta}{b\,t^{2p_3}}\right]\frac{\partial}{\partial\beta}-\frac{1}{\sqrt{a}}\left[\frac{C_\beta^2}{bt^{2p_3}}+\frac{C_x^2}{ct^{2p_1}}+\frac{C_y^2}{dt^{2p_2}}\right]^{1/2}\frac{\partial}{\partial t}\ . \tag{E6}$$

(Note that $\partial/\partial t$ is a past-directed timelike vector.)

In the following, we will assume for simplicity that $p_1>p_2$ [cf. Eqs. (E3)]. After trivial modifications, all arguments below are also valid for the case $p_1=p_2$.

Consider a null geodesic generator of $\dot{J}^{+}(q)$ along which $C_x\neq 0$. It follows from Eq. (E6) that, asymptotically as $t\to 0$,

$$\gamma_*(t)\sim\frac{C_x}{ct^{2p_1}}\frac{\partial}{\partial x}-\frac{|C_x|}{\sqrt{ac}}\frac{1}{t^{p_1}}\frac{\partial}{\partial t} \tag{E7}$$

along such a generator. Now recall that given any null hypersurface $\mathcal{S}$ like $\dot{J}^{+}(q)$, the tangent vectors $\gamma_*(t)$ of the null geodesic generators of $\mathcal{S}$ define a null, geodesic vector field *on* $\mathcal{S}$. If this vector field on $\mathcal{S}$ is extended to *any* vector field $\mathbf{V}$ which is null (but not necessarily geodesic outside $\mathcal{S}$) on a neighborhood of $\mathcal{S}$, then the divergence of $\mathbf{V}$ restricted to $\mathcal{S}$, $(\nabla\cdot\mathbf{V})|_{\mathcal{S}}$, is equal to the convergence $\hat{\theta}$ of $\mathcal{S}$'s null generators; i.e., the quantity $(\nabla\cdot\mathbf{V})|_{\mathcal{S}}$ is independent of the null extension $\mathbf{V}$ and equals $\hat{\theta}$. (For a proof of this well-known fact see Sec. 4.2 of Ref. 14.) Thus, consider the null, geodesic vector field γ_* on $\dot{J}^{+}(q)$ defined by those generators of $\dot{J}^{+}(q)$ which lie in the vicinity of our generator with $C_x\neq 0$; all these neighboring generators similarly have $C_x\neq 0$. Applying the general formula (valid in a coordinate basis)

$$\nabla\cdot\mathbf{V}=\frac{1}{\sqrt{-g}}(\sqrt{-g}\ V^\mu)_{,\mu} \tag{E8}$$

to *any* null extension $\mathbf{V}$ of this field γ_*, we find that Eq. (E7) implies, asymptotically as $t\to 0$,

$$\hat{\theta}(t)=\nabla\cdot\mathbf{V}(t)\sim\frac{C_{x,x}}{c}\frac{1}{t^{2p_1}}-\frac{|C_x|}{\sqrt{ac}}(1-p_1)\frac{1}{t^{1+p_1}} \tag{E9}$$

along our generator, provided $C_{x,x}$ is finite as $t\to 0$. On the other hand, it is obvious that in the vicinity of any generator with $C_x\neq 0$ we can find a null extension $\mathbf{V}$ of γ_* which satisfies $C_{x,x}\equiv 0$. [To see this, observe that the vector field $\partial/\partial x$ intersects $\dot{J}^{+}(q)$ transversally in the vicinity of such a generator. Also, although one might worry about the terms of the form $C_{y,y}/t^{2p_2}$ and $C_{\beta,\beta}/t^{2p_3}$ in $\nabla\cdot\mathbf{V}(t)$ which are not included in Eq. (E9), it similarly follows that whenever $C_y\neq 0$ and $C_\beta\neq 0$ one can find an extension with $C_{y,y}\equiv C_{\beta,\beta}\equiv 0$ and thus make these terms identically zero. On the other hand, a straightforward application of the arguments we present below shows that along the generators on which $C_y=0$ or $C_\beta=0$ the quantities $C_{y,y}(t)$ and $C_{\beta,\beta}(t)$ remain finite as $t\to 0$.] Therefore, it follows from Eq. (E9) that *along any generator of $\dot{J}^{+}(q)$ with $C_x\neq 0$ the convergence $\hat{\theta}$ diverges to $-\infty$ as $t\to 0$.*

Now consider a generator of $\dot{J}^{+}(q)$ along which $C_x=0$ but $C_y\neq 0$. It is easy to see that on such a generator we have, instead of Eq. (E9),

$$\hat{\theta}(t)=\nabla\cdot\mathbf{V}(t)\sim\frac{C_{x,x}}{c}\frac{1}{t^{2p_1}}-\frac{|C_y|}{\sqrt{ad}}(1-p_2)\frac{1}{t^{1+p_2}}\ . \tag{E10}$$

Now, by using Eq. (E5a), we can actually compute the asymptotic behavior of the quantity $C_{x,x}(t)$ *along* this generator on which $C_x=0$ and $C_y\neq 0$. Differentiating both sides of Eq. (E5a) with respect to x and putting $C_x=0$, we obtain

$$1=\int_t^{t_0}\frac{C_{x,x}(t)}{c\,s^{2p_1}}\frac{\sqrt{a}}{\left[\dfrac{C_\beta^2}{bs^{2p_3}}+\dfrac{C_y^2}{ds^{2p_2}}\right]^{1/2}}ds\ . \tag{E11}$$

The asymptotic $(t\to 0)$ limit of Eq. (E11) is easily computed; it gives

$$1\sim\int_t^{t_0}C_{x,x}(t)\frac{\sqrt{ad}}{c|C_y|}s^{p_2-2p_1}ds\ . \tag{E12}$$

After evaluating the integral in Eq. (E12) and combining the result with Eq. (E10), we reach the following final conclusions: (i) When $p_2-2p_1+1<0$,

$$C_{x,x}(t)\sim(2p_1-p_2-1)\frac{c|C_y|}{\sqrt{ad}}\frac{1}{t^{1+p_2-2p_1}} \tag{E13a}$$

and

$$\hat{\theta}(t)\sim\frac{2(p_1-1)}{\sqrt{ad}}|C_y|\frac{1}{t^{1+p_2}}\to-\infty\ . \tag{E13b}$$

(ii) When $p_2 - 2p_1 + 1 > 0$,

$$C_{x,x}(t) \sim (p_2 - 2p_1 + 1)\frac{c|C_y|}{\sqrt{ad}}\frac{1}{t_0^{p_2-2p_1+1}} \tag{E14a}$$

and

$$\hat{\theta}(t) \sim \frac{|C_y|}{\sqrt{ad}}\left[\frac{p_2 - 2p_1 + 1}{t_0^{p_2-2p_1+1}\,t^{2p_1}} - \frac{1-p_2}{t^{1+p_2}}\right] \to -\infty \,. \tag{E14b}$$

(iii) When $p_2 - 2p_1 + 1 = 0$,

$$C_{x,x}(t) \sim \frac{c|C_y|}{\sqrt{ad}}\frac{1}{\ln(t_0/t)} \tag{E15a}$$

and

$$\hat{\theta}(t) \sim \frac{|C_y|}{\sqrt{ad}}\left[\frac{1}{t^{2p_1}\ln(t_0/t)} - \frac{1-p_2}{t^{2p_1}}\right] \to -\infty \,. \tag{E15b}$$

Consequently, our overall conclusion is that *along any generator of $\dot{J}^{+}(q)$ with $C_x = 0$ and $C_y \neq 0$ the convergence $\hat{\theta}$ diverges to $-\infty$ as $t \to 0$.*

For a generator of $\dot{J}^{+}(q)$ along which $C_x = C_y = 0$ but $C_\beta \neq 0$, we have

$$\hat{\theta}(t) = \nabla\cdot\mathbf{V}(t) \sim \frac{C_{x,x}}{c}\frac{1}{t^{2p_1}} + \frac{C_{y,y}}{d}\frac{1}{t^{2p_2}} - \frac{|C_\beta|}{\sqrt{ab}}(1-p_3)\frac{1}{t^{1+p_3}} \,. \tag{E16}$$

The quantities $C_{x,x}$ and $C_{y,y}$ of Eq. (E16) can be computed along this generator in exactly the same way as before, i.e., by differentiating Eqs. (E5a) and (E5b) with respect to x and y, respectively, and then putting $C_x = C_y = 0$. Evaluating the asymptotic forms of the resulting integrals and proceeding in precisely the same manner as we did in Eqs. (E11)–(E15), we obtain the conclusion that *along any generator of $\dot{J}^{+}(q)$ with $C_x = C_y = 0$ and $C_\beta \neq 0$ the convergence $\hat{\theta}$ diverges to $-\infty$ as $t \to 0$.*

Combined with the two previous conclusions, this last result completes the proof of Lemma 3.

[1]K. Khan and R. Penrose, Nature (London) **229**, 185 (1971).
[2]R. A. Matzner and F. J. Tipler, Phys. Rev. D **29**, 1575 (1984).
[3]R. Penrose, Rev. Mod. Phys. **37**, 215 (1965).
[4]P. Szekeres, J. Math. Phys. **13**, 286 (1972).
[5]U. Yurtsever, Phys. Rev. D. **37**, 2790 (1988).
[6]U. Yurtsever, Phys. Rev. D. **38**, 1706 (1988).
[7]S. Chandrasekhar and B. Xanthopoulos, Proc. R. Soc. London **A410**, 311 (1987).
[8]U. Yurtsever, Phys. Rev. D **36**, 1662 (1987).
[9]U. Yurtsever, Phys. Rev. D **38**, 1731 (1988).
[10]E. Kasner, Am. J. Math. **43**, 217 (1921); E. M. Lifshitz and I. M. Khalatnikov, Usp. Fiz. Nauk **80**, 391 (1964) [Sov. Phys. Usp. **6**, 495 (1963)]; C. W. Misner, Phys. Rev. Lett. **22**, 1071 (1969); V. A. Belinsky and I. M. Khalatnikov, Zh. Eksp. Teor. Fiz. **57**, 2163 (1969) [Sov. Phys. JETP **30**, 1174 (1970)]; V. A. Belinsky, I. M. Khalatnikov, and E. M. Lifshitz, Adv. Phys. **19**, 525 (1970).
[11]F. J. Tipler, Phys. Rev. D **22**, 2929 (1980).
[12]U. Yurtsever, Phys. Rev. D. **37**, 2803 (1988).
[13]S. W. Hawking and R. Penrose, Proc. R. Soc. London **A314**, 529 (1970).
[14]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge University Press, Cambridge, England, 1973).
[15]A. Fischer and J. E. Marsden, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England 1979); see also Sec. 7.6 of Ref. 14.
[16]E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[17]F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups* (Springer, New York, 1983).
[18]F. John, *Partial Differential Equations* (Springer, New York, 1982).
[19]P. Garabedian, *Partial Differential Equations* (Chelsea, New York, 1986).
[20]J. Leray, *Hyperbolic Differential Equations* (The Institute of Advanced Study, Princeton, NJ, 1953).
[21]Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1982).
[22]F. John, Commun. Pure Appl. Math. **29**, 649 (1976).
[23]S. Klainerman, Commun. Pure Appl. Math. **33**, 43 (1980).
[24]T. Kato, Commun. Pure Appl. Math. **33**, 501 (1980).
[25]J. Shatah, J. Diff. Eqs. **46**, 409 (1982).
[26]F. John, Commun. Pure Appl. Math. **36**, 1 (1983).
[27]S. Klainerman and G. Ponce, Commun. Pure Appl. Math. **36**, 133 (1983).
[28]S. Klainerman, Commun. Pure Appl. Math. **36**, 325 (1983).
[29]T. C. Sideris, J.Diff. Eqs. **52**, 378 (1984).
[30]W. Strauss, An. Acad. Brasil Cienc. **42**, 645 (1970).
[31]S. Klainerman, Commun. Pure Appl. Math. **38**, 321 (1985).
[32]D. Christodoulou, Commun. Pure Appl. Math. **39**, 267 (1986).
[33]The eigenvalues of the matrix (in x, y coordinates) of the x-y part of (2.31) are $\alpha e^{\pm\hat{V}}$, where $\hat{V}$ is defined in Eq. (3.13e).
[34]B. Xanthopoulos (private communication).
[35]R. Penrose, in *General Relativity: An Einstein Centenary Survey* (Ref. 15).
[36]F. J. Tipler, C. J. S. Clarke, and G. F. R. Ellis, in *Einstein Centenary Volume*, edited by A. Held and P. Bergmann (Plenum, NewYork, 1979).
[37]S. Chandrasekhar and B. Xanthopoulos, Proc. R. Soc. London **A414**, 1 (1987).
[38]V. Moncrief and J. Isenberg, Commun. Math. Phys. **89**, 387 (1983); J. Math. Phys. **26**, 1024 (1985).
[39]K. Yosida, *Functional Analysis* (Springer, Berlin, 1985).
[40]E. Phillips, *An Introduction to Analysis and Integration Theory* (Dover, New York, 1984).
[41]J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, New York, 1983).
[42]R. Geroch (private communication).
[43]Y.Choquet-Bruhat and J. York, in *Einstein Centenary Volume* (Ref. 36).
[44]V. I. Arnold, *Geometric Methods in the Theory of Ordinary Differential Equations* (Springer, New York, 1983).
[45]S. Willard, *General Topology* (Addison-Wesley, Reading, MA, 1970).