The Short Pulse Method and the Formation of Trapped Surfaces in General Relativity

Demetrios Christodoulou

In 1965 Penrose introduced the concept of a *trapped surface*. He defined a trapped surface to be a closed spacelike surface in spacetime, such that an infinitesimal virtual displacement of the surface along either family of future-directed null geodesic normals to the surface leads to a pointwise decrease of the area element. On the basis of this concept, Penrose proved an *incompleteness* theorem. In the light of subsequent work, namely the uniqueness theorem of the maximal development of given initial data by Choquet-Bruhat and Geroch, and the work of Rendall on the characteristic initial value problem, the incompleteness theorem of Penrose may be re-stated as follows:

Let us be given regular characteristic initial data on a complete null geodesic cone C_o of a point o. Let (M^*, g) be the maximal future development of the data on C_o. Suppose that M^* contains a trapped surface. Then (M^*, g) is future null geodesically incomplete.

Now, it is not *a priori* obvious that trapped surfaces are *evolutionary*. That is, it is not obvious whether trapped surfaces can form in evolution starting from initial conditions in which no such surfaces are present. What is more important, the physically interesting problem is the problem where the initial conditions are arbitrarily far from already containing trapped surfaces, and we are asked to follow the long time evolution and show that, under suitable circumstances, trapped surfaces eventually form. Only an analysis of the dynamics of gravitational collapse can achieve this aim.

John Wheeler, my teacher in physics, posed to me the following problem back in 1968: *to establish the formation of black holes in pure general relativity, by the focusing of incoming gravitational waves*. It took me 40 years to solve this problem. The solution is contained the monograph “The Formation of Black Holes in General Relativity”. [Fortunately, Wheeler soon gave me an easier problem to study and I was able to complete my Ph.D. in 1971 at the age of 20, rather than in 2008 at the age of 57.]

I shall now state the simplest version of the theorem on the formation of trapped surfaces in pure general relativity which this monograph establishes. This is the limiting version, where we have an asymptotic characteristic initial value problem with initial data at past null infinity. Denoting by $\underline{u}$ the "advanced time", it is assumed that the initial data are trivial for $\underline{u} \leq 0$.

Let k, l be positive constants, $k > 1$, $l < 1$. Let us be given smooth asymptotic initial data at past null infinity which is trivial for advanced time $\underline{u} \leq 0$. Suppose that the incoming energy per unit solid angle in each direction in the advanced time interval $[0, \delta]$ is not less than $k/8\pi$. Then if δ is suitably small, the maximal development of the data contains a trapped surface S which is diffeomorphic to S^2 and has area

$$\text{Area}(S) \geq 4\pi l^2$$

We remark that by virtue of the scale invariance of the vacuum Einstein equations, the theorem holds with k, l, and δ, replaced by ak, al, and $a\delta$, respectively, for any positive constant a.

The above theorem is obtained through a theorem in which the initial data is given on a complete future null geodesic cone C_o. The generators of the cone are parametrized by an affine parameter s measured from the vertex o and defined so that the corresponding null geodesic vectorfield has projection T at o along a fixed unit future-directed timelike vector T at o. It is assumed that the initial data are trivial for $s \leq r_0$, for some $r_0 > 1$. The boundary of this trivial region is then a round sphere of radius r_0.

The advanced time $\underline{u}$ is then defined along C_o by

$$\underline{u} = s - r_0 \tag{1}$$

The formation of trapped surfaces theorem is similar in this case, the only difference being that the "incoming energy per unit solid angle in each direction in the advanced time interval $[0, \delta]$", a notion defined only at past null infinity, is replaced by the integral

$$\frac{r_0^2}{8\pi} \int_0^\delta e d\underline{u} \tag{2}$$

on the affine parameter segment $[r_0, r_0 + \delta]$ of each generator of C_o.

The function e is an invariant of the conformal intrinsic geometry of C_o, given by:

$$e = \frac{1}{2}|\hat{\chi}|^2_{\not{g}} \qquad (3)$$

where $\not{g}$ is the induced metric on the sections of C_o corresponding to constant values of the affine parameter, and $\hat{\chi}$ is the *shear* of these sections, the trace-free part of their 2nd fundamental form relative to C_o.

The theorem for a cone C_o is established for any $r_0 > 1$ and the smallness condition on δ is independent of r_0. The domain of dependence, in the maximal development, of the trivial region in C_o is a domain in Minkowski spacetime bounded in the past by the trivial part of C_o and in the future by $\underline{C}_e$, the past null geodesic cone of a point e at arc length $2r_0$ along the timelike geodesic Γ_0 from o with tangent vector T at o. Considering then the corresponding complete timelike geodesic in Minkowski spacetime, fixing the origin on this geodesic to be the point e, the limiting form of the theorem is obtained by letting $r_0 \to \infty$, keeping the origin fixed, so that o tends to the infinite past along the timelike geodesic.

Almost all the work goes into establishing an *existence theorem* for a development of the initial data which extends far enough into the future so that trapped spheres have eventually a chance to form within this development. On the other hand, there is a wealth of information in this existence theorem, which gives us full knowledge of the geometry of spacetime when trapped surfaces begin to form.

I shall now give a brief discussion of the mathematical methods employed. Three methods are used, two of which originated in my work with Klainerman on the stability of the Minkowski spacetime, and the third method, which I call the *short pulse method*, was introduced in the work which we are discussing. I shall begin by summarizing the first two methods.

The first method is peculiar to Einstein's equations, while the second has wider application, and can, in principle, be extended to all Euler-Lagrange systems of partial differential equations of hyperbolic type.

The first method is a way of looking at Einstein's equations which allows estimates for the spacetime curvature to be obtained. Instead of considering the Einstein equations themselves, we considered the Bianchi identities in the form which they assume by virtue of the Einstein equations. We then introduced the general concept of a *Weyl field* W on a 4-dimensional Lorentzian manifold (M, g) to be a 4-covariant tensorfield with the algebraic properties of the Weyl or *conformal* curvature tensor. Given a Weyl field W one can define a left dual ${}^*W$ as well as a right dual W^*, but as a consequence of the algebraic properties of a Weyl field the two duals coincide.

Moreover, ${}^*W = W^*$ is also a Weyl field. A Weyl field is subject to equations which are analogues of Maxwell's equations for the electromagnetic field. These are linear equations, in general inhomogeneous, which we call *Bianchi equations*. They are of the form:

$$\nabla^\alpha W_{\alpha\beta\gamma\delta} = J_{\beta\gamma\delta} \tag{4}$$

the right hand side J, or more generally any 3-covariant tensorfield with the algebraic properties of the right hand side, we call a *Weyl current*.

These equations seem at first sight to be the analogues of only half of Maxwell's equations, but it turns out that they are equivalent to the equations

$$\nabla_{[\alpha} W_{\beta\gamma]\delta\epsilon} = \epsilon_{\mu\alpha\beta\gamma} J^{*\mu}{}_{\delta\epsilon}, \quad J^{*}_{\beta\gamma\delta} = \frac{1}{2} J_{\beta}{}^{\mu\nu} \epsilon_{\mu\nu\gamma\delta} \tag{5}$$

which are analogues of the other half of Maxwell's equations. Here ϵ is the volume 4-form of (M, g). The fundamental Weyl field is the Riemann curvature tensor of (M, g), (M, g) being a solution of the vacuum Einstein equations, and in this case the corresponding Weyl current vanishes, the Bianchi equations reducing to the Bianchi identities.

Given a vectorfield Y and a Weyl field W or Weyl current J there is a "variation" of W and J with respect to Y, a modified Lie derivative $\tilde{\mathcal{L}}_Y W$, $\tilde{\mathcal{L}}_Y J$, which is also a Weyl field or Weyl current respectively. The modified Lie derivative commutes with duality. The Bianchi equations have certain conformal covariance properies which imply the following. If J is the Weyl current associated to the Weyl field W according to the Bianchi equations, then the Weyl current associated to $\tilde{\mathcal{L}}_Y W$ is the sum of $\tilde{\mathcal{L}}_Y J$ and a bilinear expression which is on one hand linear in ${}^{(Y)}\tilde{\pi}$ and its first covariant derivative and other the other hand in W and its first covariant derivative.

Here we denote by ${}^{(Y)}\tilde{\pi}$ the *deformation tensor* of Y, namely the trace-free part of the Lie derivative of the metric g with respect to Y. This measures the rate of change of the conformal geometry of (M, g) under the flow generated by Y. From the fundamental Weyl field, the Riemann curvature tensor of (M, g), and a set of vector fields $Y_1, ..., Y_n$ which we call *commutation fields*, derived Weyl fields of up to any given order m are generated by the repeated application of the operators $\tilde{\mathcal{L}}_{Y_i} : i = 1, ..., n$.

A basic requirement on the set of commutation fields is that it spans the tangent space to M at each point. The Weyl currents associated to these derived Weyl fields are then determined by the deformation tensors of the commutation fields.

Given a Weyl field W there is a 4-covariant tensorfield $Q(W)$ associated to W, which is symmetric and trace-free in any pair of indices. It is a quadratic expression in W, analogous to the Maxwell energy-monentum-stress tensor for the electromagnetic field.

We call $Q(W)$ the *Bel-Robinson tensor* associated to W, because it coincides with the tensor discovered by Bel and Robinson in the case of the fundamental Weyl field, the Riemann curvature tensor of a solution of the vacuum Einstein equations.

The Bel-Robinson tensor has a remarkable positivity property:

$Q(W)(X_1, X_2, X_3, X_4)$ is non-negative for any tetrad X_1, X_2, X_3, X_4 of future directed non-spacelike vectors at a point. Moreover, the divergence of $Q(W)$ is a bilinear expression which is linear in W and in the associated Weyl current J.

Given a Weyl field W and a triplet of future directed non-spacelike vectorfields X_1, X_2, X_3, which we call *multiplier fields* we define the *energy-momentum density* vectorfield $P(W; X_1, X_2, X_3)$ associated to W and to the triplet X_1, X_2, X_3 by:

$$P(W; X_1, X_2, X_3)^\alpha = -Q(W)^\alpha_{\beta\gamma\delta} X_1^\beta X_2^\gamma X_3^\delta \tag{6}$$

Then the divergence of $P(W; X_1, X_2, X_3)$ is the sum of $-(\text{div}Q(W))(X_1, X_2, X_3)$ and a bilinear expression which is linear in $Q(W)$ and in the deformation tensors of X_1, X_2, X_3.

The divergence theorem in spacetime, applied to a domain which is a development of part of the initial hypersurface, then expresses the integral of the 3-form dual to $P(W; X_1, X_2, X_3)$ on the future boundary of this domain, in terms of the integral of the same 3-form on the past boundary of the domain, namely on the part of the initial hypersurface, and the spacetime integral of the divergence. The boundaries being *achronal* - that is, no pair of points on each boundary can be joined by a timelike curve - the integrals are integrals of non-negative functions, by virtue of the positivity property of $Q(W)$.

For the set of Weyl fields of order up to m which are derived from the fundamental Weyl field, the Riemann curvature tensor of (M, g), the divergences are determined by the deformation tensors of the commutation fields and their derivatives up to order m, and by the deformation tensors of the multiplier fields. And the integrals on the future boundary give control of all the derivatives of the curvature up to order m. This is how estimates for the spacetime curvature are obtained, once a suitable set of multiplier fields and a suitable set of commutation fields have been provided.

This is precisely where the second method comes in. This method constructs the required sets of vectorfields by using the geometry of the two parameter foliation of spacetime by the level sets of two functions. These two functions, in the first realization of this method, where the *time function* t, the level sets of which are maximal spacelike hypersurfaces H_t of vanishing total linear momentum, and the *optical function* u, which we may think of as "retarded time", the level sets of which are outgoing null hypersurfaces C_u. These where chosen so that density of the foliation of each H_t by the traces of the C_u, that is, by the surfaces of intersection $S_{t,u} = H_t \bigcap C_u$, which are diffeomorphic to S^2, tends to 1 as $t \to \infty$.

It was clear that the two functions did not enter the problem on equal footing. The optical function u played a much more important role. This is due to the fact that the problem involved outgoing waves reaching future null infinity, and it is the outgoing family of null hypersurfaces C_u which follow these waves. The role of the family of maximal spacelike hypersurfaces H_t was to obtain a suitable family of sections of each C_u, the family $S_{t,u}$ corresponding to a given u, and to serve as a means by which, in the proof of the existence theorem, the method of continuity can be applied.

The geometric entities describing the two parameter foliation of spacetime by the $S_{t,u}$ are estimated in terms of the spacetime curvature. This yields estimates for the deformation tensors of the multiplier fields and the commutation fields in terms of the spacetime curvature, thus connecting with the first method.

A variant of the second method is obtained if we place in the role of the time function t another *optical function* $\underline{u}$, which we may think of as "advanced time", the level sets of which are incoming null hypersurfaces. A two parameter family of surfaces diffeomorphic to S^2, the "wave fronts", are then obtained, namely the intersections of this incoming family with the outgoing family of null hypersurfaces.

In the work under discussion, the roles of the two optical functions are reversed, because we are considering incoming rather than outgoing waves, and it is the incoming null hypersurfaces $\underline{C}_{\underline{u}}$, the level sets of $\underline{u}$, which follow these waves. However, in this work, taking the other function to be the conjugate optical function u is not merely a matter of convenience, but it is essential for what we wish to achieve. This is because the C_u, the level sets of u, are here, like the initial hypersurface C_o itself, future null geodesic cones with vertices on the timelike geodesic Γ_0, and the trapped spheres which eventually form are sections $S_{\underline{u},u} = \underline{C}_{\underline{u}} \bigcap C_u$ of "late" C_u, everywhere along which those C_u have negative expansion.

We now come to the *short pulse method*. This method is a method of treating the focusing of incoming waves, and like the second method it is of wider application. Its point of departure resembles that of the short wavelength or geometric optics approximation, in so far as it depends on the presence of a small length, but thereafter the two approaches diverge.

The short pulse method is a method which, in the context of Euler-Lagrange systems of partial differential equations of hyperbolic type, allows us to establish an existence theorem for a development of the initial data which is large enough so that interesting things have a chance to occur within this development, if a nonlinear system is involved. One may ask at this point: what does it mean for a length to be small in the context of the vacuum Einstein equations? For, the equations are scale invariant. Here *small* means *by comparison to the area radius of the trapped sphere to be formed*.

With initial data on a complete future null geodesic cone C_o, as explained above, which are trivial for $s \leq r_0$, we consider the restriction of the initial data to $s \leq r_0 + \delta$. In terms of the advanced time $\underline{u}$, we restrict attention to the interval $[0, \delta]$, the data being trivial for $\underline{u} \leq 0$. The retarded time u is set equal to $u_0 = -r_0$ at o and therefore on C_o, which is then also denoted C_{u_0}. Also, $u - u_0$ is defined along Γ_0 to be one half the arc length from o. This determines u everywhere. The development whose existence we want to establish is that bounded in the future by the spacelike hypersurface H_{-1} where $\underline{u} + u = -1$ and by the incoming null hypersurface $\underline{C}_\delta$. We denote this development M_{-1}.

We define L and $\underline{L}$ to be the future directed null vectorfields the integral curves of which are the generators of the C_u and $\underline{C}_{\underline{u}}$, parametrized by $\underline{u}$ and u respectively, so that

$$Lu = \underline{L}\underline{u} = 0, \quad L\underline{u} = \underline{L}u = 1 \qquad (7)$$

The flow Φ_τ generated by L defines a diffeomorphism of $S_{\underline{u},u}$ onto $S_{\underline{u}+\tau,u}$, while the flow $\underline{\Phi}_\tau$ generated by $\underline{L}$ defines a diffeomorphism of $S_{\underline{u},u}$ onto $S_{\underline{u},u+\tau}$.

The positive function Ω defined by

$$g(L, \underline{L}) = -2\Omega^2 \tag{8}$$

may be thought of as the inverse density of the double null foliation. We denote by $\hat{L}$ and $\underline{\hat{L}}$ the corresponding normalized future directed null vectorfields

$$\hat{L} = \Omega^{-1} L, \ \ \underline{\hat{L}} = \Omega^{-1} \underline{L}, \ \text{ so that } g(\hat{L}, \underline{\hat{L}}) = -2 \tag{9}$$

The first step is the analysis of the equations along the initial hypersurface C_{u_0}. The analysis is particularly clear and simple because of the fact that C_{u_0} is a null hypersurface, so we are dealing with the characteristic initial value problem and there is a way of formulating the problem in terms of free data which are not subject to any constraints. The full set of data which includes all the curvature components and their transversal derivatives, up to any given order, along C_{u_0}, is then determined by integrating ordinary differential equations along the generators of C_{u_0}. We show that the free data may be described as a 2-covariant symmetric positive definite tensor density m, of weight -1 and unit determinant, on S^2, depending on $\underline{u}$.

This is of the form:

$$m = \exp \psi \tag{10}$$

where ψ is a 2-dimensional symmetric trace-free matrix valued "function" on S^2, depending on $\underline{u} \in [0, \delta]$, and transforming under change of charts on S^2 in such a way so as to make m a 2-covariant tensor density of weight -1. The transformation rule is particularly simple if stereographic charts on S^2 are used.

Then there is a function O defined on the intersection of the domains of the north and south polar stereographic charts on S^2, with values in the 2-dimensional symmetric orthogonal matrices of determinant -1 such that in going from the north polar chart to the south polar chart or vise-versa, $\psi \mapsto \tilde{O}\psi O$ and $m \mapsto \tilde{O}mO$.

The crucial ansatz of the short pulse method is the following. We consider an arbitrary smooth 2-dimensional symmetric trace-free matrix valued "function" ψ_0 on S^2, depending on $s \in [0,1]$, which extends smoothly by 0 to $s \leq 0$, and we set:

$$\psi(\underline{u},\vartheta) = \frac{\delta^{1/2}}{|u_0|}\psi_0\left(\frac{\underline{u}}{\delta},\vartheta\right), \quad (\underline{u},\vartheta) \in [0,\delta] \times S^2 \tag{11}$$

The analysis of the equations along C_{u_0} then gives, for the components of the spacetime curvature along C_{u_0}:

$$\begin{aligned}
\sup_{C_{u_0}} |\alpha| &\leq O_2(\delta^{-3/2}|u_0|^{-1}) \\
\sup_{C_{u_0}} |\beta| &\leq O_2(\delta^{-1/2}|u_0|^{-2}) \\
\sup_{C_{u_0}} |\rho|, \sup_{C_{u_0}} |\sigma| &\leq O_3(|u_0|^{-3}) \\
\sup_{C_{u_0}} |\underline{\beta}| &\leq O_4(\delta|u_0|^{-4}) \\
\sup_{C_{u_0}} |\underline{\alpha}| &\leq O_5(\delta^{3/2}|u_0|^{-5}) \qquad (12)
\end{aligned}$$

Here $\alpha, \underline{\alpha}$ are the trace-free symmetric 2-covariant tensorfields on each $S_{\underline{u},u}$ defined by:

$$\alpha(X,Y) = R(X,\hat{L},Y,\hat{L}),\ \underline{\alpha}(X,Y) = R(X,\underline{\hat{L}},Y,\underline{\hat{L}}) \tag{13}$$

for any pair of vectors X, Y tangent to $S_{\underline{u},u}$ at a point, $\beta, \underline{\beta}$ are the 1-forms on each $S_{\underline{u},u}$ defined by:

$$\beta(X) = \frac{1}{2}R(X,\hat{L},\underline{\hat{L}},\hat{L}),\ \underline{\beta}(X) = \frac{1}{2}R(X,\underline{\hat{L}},\underline{\hat{L}},\hat{L}) \tag{14}$$

and ρ, σ are the functions on each $S_{\underline{u},u}$ defined by:

$$\rho = \frac{1}{4} R(\underline{\hat{L}}, \hat{L}, \underline{\hat{L}}, \hat{L}), \ \frac{1}{2} R(X, Y, \underline{\hat{L}}, \hat{L}) = \sigma \not{\epsilon}(X, Y) \quad (15)$$

for any pair of vectors X, Y tangent to $S_{\underline{u},u}$ at a point, $\not{\epsilon}$ being the area form of $S_{\underline{u},u}$. The symbol $O_k(\delta^p |u_0|^r)$ means the product of $\delta^p |u_0|^r$ with a non-negative non-decreasing continuous function of the C^k norm of ψ_0 on $[0,1] \times S^2$. The pointwise magnitudes of tensors on $S_{\underline{u},u}$ are with respect to the induced metric $\not{g}$, which is positive definite, the surfaces being spacelike.

One should focus on the dependence on δ of the right hand sides of 12. This displays what we may call the *short pulse hierarchy*. And this hierarchy is *nonlinear*. For, if only the linearized form of the equations was considered, a different hierarchy would be obtained: the exponents of δ in the first two of 12 would be the same, but the exponents of δ in the last three of 12 would instead be $1/2, 3/2, 5/2$, respectively.

A question that immediately comes up when one ponders the ansatz 11, is why is the “amplitude” of the pulse proportional to the square root of the “length” of the pulse? (the factor $|u_0|^{-1}$ is the standard decay factor in 3 spatial dimensions, the square root of the area of the wave fronts). Where does this relationship come from? Obviously, there is no such relationship in a linear theory.

The answer is that it comes from our desire to form trapped surfaces in the development M_{-1}. If a problem involving the focusing of incoming waves in a different context was the problem under study, for example the formation of electromagnetic shocks by the focusing of incoming electromagnetic waves in a nonlinear medium, the relationship between length and amplitude would be dictated by the desire to form such shocks within our development.

The short pulse hierarchy is the key to the existence theorem as well as to the trapped surface formation theorem. We must still outline however in what way do we establish that the short pulse hierarchy is preserved in evolution. This is of course the main step of the short pulse method. What we do is to reconsider the first two methods previously outlined in the light of the short pulse hierarchy.

Let us revisit the first method. We take as multiplier fields the vectorfields L and K, where

$$K = u^2 \underline{L} \tag{16}$$

As already mentioned above, we take the initial data to be trivial for $\underline{u} \leq 0$ and as a consequence the spacetime region corresponding to $\underline{u} \leq 0$ is a domain in Minkowski spacetime. We may thus confine attention to the non-trivial region $\underline{u} \geq 0$. We denote by M'_{-1} this non-trivial region in M_{-1}.

For each of the Weyl fields to be specified below, we define the energy-momentum density vectorfields

$$\overset{(n)}{P}(W) \;:\; n = 0, 1, 2, 3 \tag{17}$$

where:

$$\begin{aligned}
\overset{(0)}{P}(W) &= P(W; L, L, L) \\
\overset{(1)}{P}(W) &= P(W; K, L, L) \\
\overset{(2)}{P}(W) &= P(W; K, K, L) \\
\overset{(3)}{P}(W) &= P(W; K, K, K)
\end{aligned} \tag{18}$$

As commutation fields we take L, S, defined by:

$$S = u\underline{L} + \underline{u}L, \tag{19}$$

and the three rotation fields $O_i \ : i = 1, 2, 3$. The latter are defined according to the second method as follows. In the Minkowskian region we introduce rectangular coordinates $x^\mu \ : \mu = 0, 1, 2, 3$, taking the x^0 axis to be the timelike geodesic Γ_0. In the Minkowskian region, in particular on the sphere S_{0,u_0}, the O_i are the generators of rotations about the $x^i \ : i = 1, 2, 3$ spatial coordinate axes. The O_i are then first defined on C_{u_0} by conjugation with the flow of L and then in spacetime by conjugation with the flow of $\underline{L}$.

The Weyl fields which we consider are, besides the fundamental Weyl field R, the Riemann curvature tensor, the following derived Weyl fields

$$\begin{aligned} \text{1st order:} \quad & \tilde{\mathcal{L}}_L R, \ \tilde{\mathcal{L}}_{O_i} R : i = 1,2,3, \ \tilde{\mathcal{L}}_S R \\ \text{2nd order:} \quad & \tilde{\mathcal{L}}_L \tilde{\mathcal{L}}_L R, \ \tilde{\mathcal{L}}_{O_i} \tilde{\mathcal{L}}_L R : i = 1,2,3, \\ & \tilde{\mathcal{L}}_{O_j} \tilde{\mathcal{L}}_{O_i} R : i,j = 1,2,3, \\ & \tilde{\mathcal{L}}_{O_i} \tilde{\mathcal{L}}_S R : i = 1,2,3, \ \tilde{\mathcal{L}}_S \tilde{\mathcal{L}}_S R \end{aligned} \tag{20}$$

We assign to each Weyl field the index l according to the number of $\tilde{\mathcal{L}}_L$ operators in the definition of W in terms of R. We then define total 2nd order energy-momentum densities

$$\overset{(n)}{P}_2 : \; n = 0, 1, 2, 3 \tag{21}$$

as the sum of $\delta^{2l} \overset{(n)}{P}(W)$ over all the above Weyl fields in the case $n = 3$, all the above Weyl fields except those whose definition involves the operator $\tilde{\mathcal{L}}_S$ in the cases $n = 0, 1, 2$.

We then define the total 2nd order energies $\overset{(n)}{E}_2(u)$ as the integrals on the C_u and the total 2nd order fluxes $\overset{(n)}{F}_2(\underline{u})$ as the integrals on the $\underline{C}_{\underline{u}}$, of the 3-forms dual to the $\overset{(n)}{P}_2$. Of the fluxes only $\overset{(3)}{F}_2(\underline{u})$ plays a role in the problem. Finally, with the exponents q_n : $n = 0, 1, 2, 3$ defined by:

$$q_0 = 1, \quad q_1 = 0, \quad q_2 = -\frac{1}{2}, \quad q_3 = -\frac{3}{2}, \quad (22)$$

according to the short pulse hierarchy, we define the quantities

$$\overset{(n)}{\mathcal{E}}_2 = \sup_u \left(\delta^{2q_n} \overset{(n)}{E}_2 (u) \right) \ : \ n = 0, 1, 2, 3,$$

$$\overset{(3)}{\mathcal{F}}_2 = \sup_{\underline{u}} \left(\delta^{2q_3} \overset{(3)}{F}_2 (\underline{u}) \right) \tag{23}$$

The objective then is to obtain bounds for these quantities in terms of the initial data.

This requires properly estimating the deformation tensor of K, as well as the deformations tensors of L, S and the $O_i : i = 1, 2, 3$ and their derivatives of up to 2nd order. In doing this, the short pulse method meshes with the second method previously described.

The estimates of the error integrals, namely the integrals of the absolute values of the divergences of the $\overset{(n)}{P}_2$, then yield inequalities for the quantities 23. These inequalities contain, besides the initial data terms

$$\overset{(n)}{D} = \delta^{2q_n} \overset{(n)}{E}_2 (u_0) \ : \ n = 0, 1, 2, 3, \qquad (24)$$

terms of $O(\delta^p)$ for some $p > 0$, which are innocuous, as they can be made less than or equal to 1 by subjecting δ to a suitable smallness condition, *but they also contain terms of $O(1)$ which are nonlinear in the quantities 23.*

From such a nonlinear system of inequalities, no bounds can in general be deduced, because here, in constrast with the work on the stability of Minkowski spacetime, the initial data quantities are allowed to be arbitrarily large. However a fortunate circumstance occurs: our system of inequalities is *reductive*. That is, the inequalities, taken in proper sequence, reduce to a sequence of sublinear inequalities, thus allowing us to obtain the sought for bounds.