FINITE-TIME DEGENERATION OF HYPERBOLICITY WITHOUT BLOWUP FOR QUASILINEAR WAVE EQUATIONS

JARED SPECK∗†

Abstract. In 3D, we study the Cauchy problem for the wave equation $-\partial_{tt}^2 \Psi + (1 + \Psi)P\Delta \Psi = 0$ for $P \in \{1, 2\}$. We exhibit a form of Tricomi-type degeneracy formation that has not previously been studied in more than one spatial dimension. Specifically, using only energy methods and ODE techniques, we exhibit an open set of data such that $\Psi$ is initially near 0 while $1 + \Psi$ vanishes in finite time. In fact, generic data, when appropriately rescaled, lead to this phenomenon. The solution remains regular in the following sense: there is a high-order $L^2$-type energy, featuring degenerate weights only at the top-order, that remains bounded. When $P = 1$, we show that any $C^1$ extension of $\Psi$ to the future of a point where $1 + \Psi = 0$ must exit the regime of hyperbolicity. Moreover, the Kretschmann scalar of the Lorentzian metric corresponding to the wave equation blows up at those points. Thus, our results show that curvature blowup does not always coincide with singularity formation in the solution variable. Similar phenomena occur when $P = 2$, where the vanishing of $1 + \Psi$ corresponds only to a breakdown in strict hyperbolicity.

The data are compactly supported and are allowed to be large or small as measured by an unweighted Sobolev norm. However, we assume that initially, the spatial derivatives of $\Psi$ are nonlinearly small relative to $|\partial_t \Psi|$, which allows us to treat the equation as a perturbation of the ODE $\frac{d^2}{dt^2} \Psi = 0$. We show that for appropriate data, $\partial_t \Psi$ remains quantitatively negative, which drives the degeneracy formation. Our result complements those of Alinhac and Lindblad, who showed that if the data are small as measured by a Sobolev norm with radial weights, then the solution is global.

Keywords: degenerate hyperbolic, strictly hyperbolic, Tricomi equation, weakly hyperbolic

Mathematics Subject Classification (2010) Primary: 35L80; Secondary: 35L05, 35L72

December 2, 2016

1. Introduction

Many authors have studied model nonlinear wave equations as a means to gain insight into more challenging wave-like quasilinear equations, such as Einstein’s equations of general relativity, the compressible Euler equations without vorticity, and the equations of elasticity. Motivated by the same considerations, in this paper, we study model quasilinear wave equations in three spatial dimensions. In a broad sense, we are interested in finding initial conditions without symmetry assumptions that lead to some kind of stable breakdown. In

∗Massachusetts Institute of Technology, Cambridge, MA, USA. jspeck@math.mit.edu.
†Our results can be generalized to apply to any number of spatial dimensions.
our main results, which we summarize just below equation (1.1.2), we exhibit a type of stable degenerate solution behavior, distinct from blowup, that, to the best of our knowledge, has not previously been studied in the context of quasilinear equations in more than one spatial dimension. Roughly, we show that there exists an open set of data such that certain principal coefficients in the equation vanish in finite time without a singularity forming in the solution. More precisely, the vanishing of the coefficients corresponds to the vanishing of the wave speed, which in turn is tied to other kinds of degeneracies described below. In the case of one spatial dimension, Kato-Sugiyama and Sugiyama have obtained similar results [53, 94–96] using a proof by contradiction that relies on the method of Riemann invariants. However, since the method of Riemann invariants is not applicable in more than one spatial dimension and since we are interested in direct proofs, our approach here is quite different.

Through our study of model problems, we are aiming to develop approaches that might be useful for studying the kinds of degeneracies that might occur in solutions to more physically realistic quasilinear equations. One consideration behind this aim is that there are relatively few breakdown results for quasilinear equations compared to the semilinear case. A second consideration is that many of the techniques that have been used to study semilinear wave equations do not apply in the quasilinear case; see Subsect. 1.4 for further discussion. A third consideration concerns fundamental limitations of semilinear model equations: they are simply incapable of exhibiting some of the most important degeneracies that can occur in solutions to quasilinear equations. In particular, the degeneracy exhibited by the solutions from our main results cannot be exhibited in solutions to semilinear wave equations with principal part equal to the linear wave operator $\Box_m$. As a second example of breakdown that is unique to the quasilinear case, we note that the phenomenon of shock formation, described in more detail at the end of Subsect. 1.6, cannot occur in solutions to semilinear equations since, in the semilinear case, the evolution of characteristics is not influenced in any way by the solution.

In view of the above discussion, it is significant that our analysis has robust features and could be extended to apply to a large class of quasilinear equations. The robustness stems from the fact that our proofs are based only on energy estimates, ODE-type estimates, and the availability of an important monotonic spacetime integral (which we describe below) that arises in the energy estimates. However, rather than formulating a theorem about a general class of equations, we prefer to keep the paper short and to exhibit the main ideas by studying only the model equation (1.1.1) in the cases $P = 1, 2$.

---

\[\Box_m := -\partial_t^2 + \Delta\] denotes the standard linear wave operator corresponding to the Minkowski metric $m = \text{diag}(-1, 1, 1, 1)$ on $\mathbb{R}^{1+3}$.

3In the works on shock formation for quasilinear equations, described in Subsect. 1.6, the intersection of the characteristics is tied to the blowup of some derivative of the solution.

4Of course, we can only hope to treat wave equations whose principal spatial coefficients vanish when evaluated at some finite values of the solution variable; equations such as $-\partial_t^2 \Psi + (1 + \Psi^2)\Delta \Psi = 0$ are manifestly immune to the kind of degeneracies under study here.
1.1. **Statement of the equations and summary of the main results.** Specifically, in the cases $P = 1, 2$, we study the following model Cauchy problem on $\mathbb{R}^{1+3}$:

\[-\partial_t^2 \Psi + (1 + \Psi)^P \Delta \Psi = 0, \quad (1.1.1a)\]

\[(\Psi |_{\Sigma_0}, \partial_t \Psi |_{\Sigma_0}) = (\tilde{\Psi}, \tilde{\Psi}_0), \quad (1.1.1b)\]

where $(x^0 := t, x^1, x^2, x^3)$ are a fixed set of standard rectangular coordinates on $\mathbb{R}^{1+3}$, $\Delta := \sum_{a=1}^{3} \partial^2_a$ is the standard Euclidean Laplacian on $\mathbb{R}^3$, and throughout, $\Sigma := \{t\} \times \mathbb{R}^3 \cong \mathbb{R}^3$. We sometimes denote the spatial coordinates by $x := (x_1^1, x_2^2, x_3^3)$. Note that we can rewrite (1.1.1a) as

\[g^{-1}\alpha\beta(\Psi)\partial_\alpha \partial_\beta \Psi = 0, \quad (1.1.2)\]

This geometric perspective will be useful at various points in our discussion.

We now summarize our results; see Theorem 4.1 and Prop. 4.1 for precise statements.

**Summary of the main results.** There exists an open subset of $H^6(\mathbb{R}^3) \times H^5(\mathbb{R}^3)$, comprising compactly supported initial data $(\hat{\Psi}, \hat{\Psi}_0)$ with $\Psi$, its spatial derivatives, and its mixed space-time derivatives initially satisfying a nonlinear smallness condition compared to $\max_{\Sigma_0}[\hat{\Psi}_0]_-, \|\hat{\Psi}_0\|_{L^\infty}$, such that the coefficient $1 + \Psi$ in (1.1.1a) vanishes at some time $T_* \in (0, \infty)$. In fact, the finite-time vanishing of $1 + \Psi$ always occurs if $\hat{\Psi}_0$ is non-trivial and the data are appropriately rescaled; see Remark 2.3. Moreover,

\[\Psi \in C \left([0, T_*], H^6(\mathbb{R}^3)\right) \cap L^1 \left([0, T_*], H^6(\mathbb{R}^3)\right) \cap C \left([0, T_*], H^5(\mathbb{R}^3)\right)\]

while for any $N < 5$,

\[\partial_t \Psi \in C \left([0, T_*], H^5(\mathbb{R}^3)\right) \cap L^\infty \left([0, T_*], H^5(\mathbb{R}^3)\right) \cap C \left([0, T_*], H^N(\mathbb{R}^3)\right).\]

In addition, the Kretschmann scalar $Riem(g)^{\alpha\beta\gamma\delta}Riem(g)_{\alpha\beta\gamma\delta}$ blows up precisely at points where $1 + \Psi$ vanishes, where $Riem(g)$ denotes the Riemann curvature of $g$. In the case $P = 1$, the solution exits the regime of hyperbolicity at time $T_*$ and thus cannot be continued beyond it as a classical solution to a hyperbolic equation. Similar results hold in the case $P = 2$, the difference being that strict hyperbolicity breaks down when $1 + \Psi$ vanishes but hyperbolicity does not.\(^7\) This leaves open, in the case $P = 2$, the possibility of classically extending the solution past time $T_*$, as we further explain in Remark 1.2.

1.2. **Paper outline.** The remainder of the paper is organized as follows.

- In Subsect. 1.3 we provide some initial remarks expanding upon various aspects of our results.

---

5Throughout we use Einstein’s summation convention.

6Throughout, $[p]_- := \min\{p, 0\}$.

7In the literature, equations exhibiting this kind of degeneracy are often referred to as *weakly hyperbolic*.
In Subsect. 1.4, we mention some techniques that have been used in studying the breakdown of solutions to semilinear equations. As motivation for the present work, we point out some limitations of the semilinear techniques for the study of quasilinear equations.

In Subsect. 1.5, we provide a brief overview of the proof of our main results.

In Subsect. 1.6, we describe some connections between our results and prior work on degenerate hyperbolic PDEs.

In Subsect. 1.7, we summarize our notation.

In Sect. 2, we state our assumptions on the initial data and introduce bootstrap assumptions.

In Sect. 3, we use the bootstrap and data-size assumptions of Sect. 2 to derive a priori pointwise and energy estimates. From the energy estimates, we deduce improvements of the bootstrap assumptions.

In Sect. 4, we use the estimates of Sect. 3 to prove our main results.

1.3. Initial remarks on the main results. As far as we know, there are no prior results in the spirit of our main results in more than one spatial dimension. There are, however, examples in which the Cauchy problem for a quasilinear wave equation has been solved (for suitable data without symmetry assumptions) such that some derivative of the solution blows up but the solution itself remains bounded. One class of such examples comprises shock formation results, which we describe in more detail at the end of Subsect. 1.6. A second example is Luk’s work [68] on the formation of weak null singularities in a family of solutions to the Einstein-vacuum equations. Specifically, he exhibited a stable family of solutions such that Christoffel symbols (which are, roughly speaking, the first derivatives of the solution) blow up along a null boundary while the metric (that is, the solution itself) extends continuously past the null boundary. We stress that the degeneracy we have exhibited in our main results is much less severe than in the above results; there is no blow up in our solutions, except possibly at the top derivative level (due to the degeneracy of the weights in the energy (1.5.1)).

We also point out a connection between our work here and our joint work with Rodnianski [83], in which we proved a stable blowup result (without symmetry assumptions) for solutions to the Einstein-scaler field and Einstein-stiff fluid systems. In that work, the wave speed became, relative to a geometrically defined coordinate system, infinite at the singularity.

Although the infinite wave speed is in the opposite direction of the degeneracy exhibited by our main results (in which the wave speed vanishes), the analysis in [83] shares a key feature with that of the present work: the solution regime studied is such that the time derivatives dominate the evolution. That is, the spatial derivatives remain negligible, all the way up to the degeneracy; see Subsect. 1.6 for further discussion regarding this issue for the solutions under study here. Hence, both [83] and the present article exhibit the stability of ODE-type behavior in some solutions to quasilinear wave equations.

---

8Specially, the $\Sigma_t$ have constant mean curvature and the spatial coordinates are transported along the unit normal to $\Sigma_t$.

9Note that the effective wave speed for equation (1.1.1a) is $(1 + \Psi)^{P/2}$. 
Remark 1.1 (Remarks on small data). The methods of Alinhac \cite{6} and Lindblad \cite{67} yield that small-data solutions to equation (1.1.1a) exist globally\footnote{The asymptotics of the solution can be distorted compared to the linear case since the equations do not satisfy the null condition.} where the size of the data is measured by Sobolev norms with radial weights. Consequently, if \((\Psi, \Psi_0)\) are compactly supported data to which our main results apply, then for \(\lambda\) sufficiently large, the solution corresponding to the data \((\lambda^{-1}\Psi, \lambda^{-1}\Psi_0)\) is global. On the other hand, our main results apply to data that are allowed to be small in certain unweighted norms, as long as the spatial derivatives are “very small.” How can we reconcile these two competing statements? The answer is that our data assumptions are nonlinear in nature and are not invariant under the rescaling \((\Psi, \Psi_0) \to (\lambda^{-1}\Psi, \lambda^{-1}\Psi_0)\) if \(\lambda\) is too large. We may caricature the situation as follows (see Subsect. 2.3 for more precise assumptions): if \(\epsilon\) is the size of \(\nabla \Psi\) at time 0 (where \(\nabla\) denotes the spatial coordinate gradient) and \(\delta\) is the size of \(\partial_t \Psi\) at time 0, then, roughly speaking, some parts of our proof rely on the assumption that \(\epsilon \leq \delta^k\) for some \(k \geq 2\). The point is that if \(\lambda\) is too large, then the assumption is not satisfied, the reason being that \(\epsilon\) and \(\delta\) both scale like \(\lambda^{-1}\). One can contrast this against Remark 2.3, where we note that a different scaling of the data always leads to our nonlinear smallness assumptions being satisfied.

Remark 1.2 (In what sense it possible to extend the solution past the degeneracy?). It is of interest to know if and when the solutions provided by our main results can be extended, as solutions, past the time of first vanishing of \(1 + \Psi\). As we explain below, there are no prior extension results of this kind for quasilinear equations in more than one spatial dimension, most likely due to the possibility of derivative loss past the degeneracy. In one spatial dimension, Manfrin has obtained results \cite{69,70} that, for \(C^\infty\) initial data, allow one to locally continue the solution to equation (1.1.1a) in the case \(P = 2\) past the time of first vanishing of \(1 + \Psi\). The techniques that he employed heavily rely on the assumption of one spatial dimension; see Subsect. 1.6 for further discussion. We are also aware of a few results \cite{33,39} for quasilinear equations in more than one spatial dimension in which the authors proved local well-posedness for equations featuring a degeneracy related to – but distinct from – the one under study here. However, the degeneracy in those works was created by a “prescribed semilinear factor” rather than a quasilinear-type solution-dependent factor. For this reason, it is not clear that the techniques used in those works is of relevance for trying to extend solutions to (1.1.1a) beyond points where \(1 + \Psi\) vanishes; see the paragraph below equation (1.6.5) for further discussion.

We also note that there are various well-posedness results \cite{22,35,36} for degenerate wave equations of Kirchhoff type. An example of an equation of this type is

\[ -\partial_t^2 \Psi + F \left( \int_{\Omega} |\nabla \Psi|^2 \, dx \right) \Delta \Psi = 0, \tag{1.3.1} \]

where \(\Omega\) a bounded open set in \(\mathbb{R}^n\) and \(F = F(s) \geq 0\) satisfies various technical conditions (with \(F = 0\) corresponding to the degeneracy). However, it remains open whether or not the techniques used in studying Kirchhoff-type equations are of relevance for proving local

\footnote{For example, a careful analysis of the proof of inequality (3.2.1a) yields that the implicit constant in front of the \(\epsilon\) term on the RHS depends on \(\delta_*^{-2}\), where \(\delta_*\) is defined in (2.1.2).}
well-posedness for equation (1.1.1a) (in the case $P = 2$) in regions where $1 + \Psi$ is allowed to vanish.

Despite the lack of results concerning extending the solution to (1.1.1a) past the time of first vanishing of $1 + \Psi$, the Cauchy-Kovalevskaya theorem nonetheless yields a simple (admittedly unsatisfying) result in this direction. Specifically, one can consider initial data $(\Psi, \Psi_0)$ that are analytic in a neighborhood of the origin in $\Sigma_0$ and are such that $1 + \Psi \geq 0$ and such that $1 + \Psi$ vanishes at one or more points. Then, in both of the cases $P = 1$ and $P = 2$, the Cauchy-Kovalevskaya theorem yields a solution to equation (1.1.1a) in a neighborhood $\Omega$ of the origin in $\mathbb{R}^{1+3}$. Moreover, for appropriate initial data, one may ensure that i) $1 + \Psi > 0$ everywhere in $\{t < 0\} \cap \Omega$; ii) $1 + \Psi < 0$ at some points in $\{t > 0\} \cap \Omega$; and iii) $\partial_t \Psi < 0$ everywhere in $\Omega$. These simple inequalities, which are consistent with the behavior of the solutions that we exhibit in our main results, show that the curvature blowup that occurs when $1 + \Psi = 0$ is not always an obstacle to continuing the solution classically.

In the case $P = 1$, the analytic solutions described in the previous paragraph are of little significance for modeling wave propagation beyond the degeneracy in the sense that equation (1.1.1a) becomes elliptic at points where $1 + \Psi < 0$. However, in the case $P = 2$, equation (1.1.1a) maintains its hyperbolicity throughout $\Omega$; only the strict hyperbolicity of the equation breaks down when $1 + \Psi$ vanishes, which corresponds to a wave of zero speed. Hence, it could be that when $P = 2$, the degeneracy is not severe in the sense that the Sobolev-class solutions from our main results might be extendable, as a Sobolev-class solution, past the time of first vanishing of $1 + \Psi$. As we mentioned above, there are currently no results in this direction for quasilinear equations in more than one spatial dimension. The lack of results might be tied to the following key difficulty: the best energy estimates available for linear degenerate hyperbolic wave equations exhibit a loss of derivatives. By this, we roughly mean that the estimates for the linear equation are of the form $\|\Psi\|_{H^N(\Sigma_t)} \lesssim \|\Psi\|_{H^{N+d}(\Sigma_0)} + \|\Psi_0\|_{H^{N+d}(\Sigma_0)}$, where the loss of derivatives $d$ depends in a complicated way on the details of the degeneration of the coefficients in the equation; see Subsect. 1.6 for further discussion. As is described in [33], in some cases, the loss of derivatives is known to be optimal. Since proofs of well-posedness for nonlinear equations rely on estimates for linearized equations, any derivative loss would pose a serious obstacle to extending the solution to equation (1.1.1a) (in the case $P = 2$) past the time of first vanishing of $1 + \Psi$ as a Sobolev-class solution. At the very least, one would need to rely on a method capable of handling a finite loss of derivatives in solutions to quasilinear equations. As is well-known [37], in some cases, it is sometimes possible to handle a finite loss of derivatives using the Nash-Moser framework.

1.4. Remarks on methods used for studying blowup in solutions to semilinear wave equations. Although we are not aware of any other results in the spirit of the present work, there are many results exhibiting the most well-known type of degeneracy that can

---

12 It is of course important for applying the Cauchy-Kovalevskaya theorem that the coefficients in equation (1.1.1a) are analytic functions of $\Psi$.
13 The results of Prop. 4.1 can be extended to show that the Kretschmann scalar blows up when $1 + \Psi$ vanishes for analytic solutions satisfying the above properties i)-iii) in $\Omega$.
14 By degenerate, we mean that the wave equation is allowed to violate strict hyperbolicity at one or more points.
occur in solutions to wave equations in three spatial dimensions: the finite-time blowup of initially smooth solutions. Our main goal in this subsection is to recall some of the most important of these results but, at the same time, to describe some limitations of the proof techniques for the study of more general equations. We will focus only on constructive results, by which we mean that the proofs provide a detailed description of the degeneracy formation and the mechanisms driving it, as in the present work. Constructive results, especially those proved via robust techniques, are clearly desirable if one aims to understand the mechanisms of breakdown in solutions to realistic physical and geometric systems. They are also important if one aims to continue the solution past the breakdown, as is sometimes possible if it is not too severe; see, for example, [15] for a recent result in spherical symmetry concerning weakly locally extending solutions to the relativistic Euler equations past the first shock singularity. Importantly, we will confine our discussion to prior results for semilinear equations since, as we mentioned earlier, aside from the shock formation results described at the end of Subsect. 1.6, most constructive breakdown results for wave equations in three spatial dimensions are blowup results for semilinear equations.

Specifically, most constructive breakdown results for wave equations in three spatial dimensions are blowup results for semilinear equations (or systems) of the form $\Box_m \Psi = f(\Psi, \partial \Psi)$, where $f$ is a smooth nonlinear term. Many important approaches have been developed to prove constructive blowup for such equations, especially for scalar equations with $f = f(\Psi)$ given by a power law and for systems of wave-map type; see, for example, [28, 32, 34, 54, 57, 62, 63, 71, 81, 84]. There are also related results that are conditional in the sense that they do not guarantee that the solution will blow up. Instead they characterize the possible behaviors of the solution by providing information such as i) how the singularity would form if the solution is not global and ii) the structure of the data sets that lead to the various outcomes; see, for example, [55, 58–61, 72–75, 80, 93].

Although the above results and others like them have yielded major advancements in our understanding of the blowup of solutions to semilinear equations, their proofs fundamentally rely on tools that are not typically applicable to quasilinear equations. Here are some important examples, where we are not specific about exactly which semilinear equations have been treated with the stated technique:

- The existence of a conserved energy (which is not available for Einstein’s equations\textsuperscript{17}).

  This allows, among other things, for the application of techniques from Hamiltonian mechanics.

\textsuperscript{15}Constructive proofs of blowup stand, of course, in contrast to proofs of breakdown by contradiction. There are many examples in the literature of proofs of blowup by contradiction for wave-like or wave equations. Two of the most important ones are Sideris’ blowup result [89] (proved by virial identity arguments) for the compressible Euler equations under a polytropic equation of state and John’s proof [51] of breakdown for several classes of semilinear and quasilinear wave equations in three spatial dimensions. See also Levine’s influential work [65], in which he proved a non-constructive blowup result for semilinear wave equations on an abstract Hilbert space.

\textsuperscript{16}Arguably, the most sophisticated blowup results of this type have been proved for nonlinearities that correspond to energy critical equations.

\textsuperscript{17}For asymptotically flat solutions, the ADM mass is conserved. However, in three spatial dimensions, this quantity has thus far proven to be too weak to be of any use in controlling solutions.
Finite-Time Degeneration of Hyperbolicity

- The invariance of the solutions under appropriate scalings (which does not generally occur for the compressible Euler equations).\(^\text{18}\)
- The availability of well-posedness results in low regularity spaces such as the energy space (Lindblad showed \(^\text{66}\) that low regularity well-posedness fails for a large class of quasilinear equations in three spatial dimensions).
- The existence of a non-trivial ground state solution (corresponding to the existence of a soliton solution) and sharp classification results on the possible behaviors of the solution for initial data with energy less than the ground state: either there is finite-time blowup in both time directions or global existence, according to the sign of a functional (for quasilinear equations, there is no known analog of this kind of dichotomy). Moreover, in some cases, there are more complicated classification results available for solutions with energy just above the ground state.
- A characterization of a certain norm of the ground state as a size threshold separating global scattering solutions from ones that can blow up or exhibit other degenerate behavior (again, for quasilinear equations, there is no known analog of this kind of dichotomy).
- A characterization of the ground state as the universal blowup-profile under various assumptions.
- The availability of profile decompositions for sequences bounded in the natural energy space, which allows one to view the sequence as a superposition of linear solutions plus a small error (for quasilinear equations, there is no known analog of this).
- Channel of energy type arguments showing that a portion of the solution propagates precisely at speed one (again, for quasilinear equations, there is no known analog of this phenomenon).
- The possibility of sharply characterizing the spectrum (see, for example, \(^\text{18}\)) of linear operators tied to the dynamics (which, for quasilinear equations in many solution regimes, is exceedingly difficult).

Although the above methods are impressively powerful within their domain of applicability, since they do not seem to apply to quasilinear equations, we believe that it is important to develop new methods for studying the kinds of breakdown that can occur in the quasilinear case. It is for this reason that have we have chosen to study the model wave equations \(^\text{1.1.1a}\).

1.5. Brief overview of the analysis. As we mentioned above, the solutions that we study are such that \(\Psi, \nabla \Psi_0\) (where \(\nabla\) denotes the spatial coordinate gradient), and sufficiently many of their spatial derivatives are “nonlinearly small” (in appropriate norms) compared to \(|\Psi_0| := |\min\{\Psi_0, 0\}|\) and \(\frac{1}{\|\Psi_0\|_{L^\infty}}\). A key aspect of our work is that we are able to propagate the smallness, long enough for the coefficient \(1 + \Psi\) in equation \(^\text{1.1.1a}\) to vanish. Put differently, our main results show that under the smallness assumptions, the solution

\(^{18}\)In particular, the fluid equation of state does not generally enjoy any useful scaling transformation properties.

\(^{19}\)It is conceivable that channel of energy type results might hold for certain quasilinear wave equations in various solution regimes, since channel of energy type arguments seem to be somewhat stable under perturbations.
to (1.1.1a) behaves in many ways like a solution to the second-order ODE \( \frac{d^2}{dt^2} \Psi = 0 \). The reason that \( \Psi \) vanishes for the first time is that \( \partial_t \Psi \) is sufficiently negative at one or more spatial points (a condition which persists by the previous remarks). To control solutions, we use the (non-conserved) energy\(^{20}\)

\[
\mathcal{E}_{[2,5]}(t) := \sum_{k'=2}^{5} \int_{\Sigma_t} |\partial_t \nabla^k \Psi|^2 + (1 + \Psi)^P |\nabla \nabla^k \Psi|^2 + |\nabla^k \Psi|^2 \, dx.
\]  

(1.5.1)

We avoid using low-order energies corresponding to \( k' = 0,1 \) in (1.5.1) because for the solution regime under consideration, such energies would contain terms that are allowed to be large, and we prefer to work only with small energies. Hence, to control the low-order derivatives of \( \Psi \), we derive ODE-type estimates that rely in part on the energy estimates for its higher derivatives and Sobolev embedding. Analytically, the main challenge is that the vanishing of \( 1 + \Psi \) leads to the degeneracy of the top-order spatial derivative terms in (1.5.1), which makes it difficult to control some top-order error integrals in the energy estimates.

To close the energy estimates, we exploit the following monotonicity, which is available due to our assumptions on the data:

\( \partial_t \Psi \) is quantitatively strictly negative in a neighborhood of points where \( 1 + \Psi \) is close to 0.

This quantitative negativity yields, in our energy identities, the spacetime error integral

\[
\int_{s=0}^{t} \int_{\Sigma_s} (\partial_t \Psi)(1 + \Psi)^{P-1} |\nabla \nabla^k \Psi|^2 \, dx \, ds,
\]  

(1.5.2)

which has a “friction-type” sign in regions where \( 1 + \Psi \) is close to 0 but positive; see the spacetime integral on LHS (3.3.3). It turns out that the availability of this spacetime integral compensates for the degeneracy of the energy (1.5.1) and yields integrated control over the spatial derivatives up to top order; \textbf{this is the key to closing the proof.}

1.6. \textbf{Comparing with and contrasting against other results for degenerate hyperbolic equations.} For solutions such that \( 1 + \Psi \) is near 0, equation (1.1.1a) may be viewed as a “nearly degenerate” quasilinear hyperbolic PDE. For this reason, the proof of our main results has ties to some prior results on degenerate hyperbolic PDEs, which we now discuss. In one spatial dimension, various aspects of degenerate hyperbolic PDEs have been explored in the literature, such as the branching of singularities \([1, 7, 10]\), uniqueness of solutions for equations that are hyperbolic in one region but can change type \([88]\), and conditions that are \textit{necessary} for well-posedness \([98]\). However, in the rest of this subsection, we will discuss only well-posedness results since they are most relevant in the context of our main results.

In one spatial dimension, there are many results on well-posedness, in various function spaces, for linear degenerate wave equations for the form

\[-\partial_t^2 \Psi + a(t, x) \partial_x^2 \Psi + b(t, x) \partial_x \Psi + c(t, x) \partial_t \Psi = f(t, x),\]  

(1.6.1)

where \( a(t, x) \geq 0 \) and \( a(t, x) = 0 \) corresponds to degeneracy via a breakdown in strict hyperbolicity. For example, if the coefficients \( a(t, x), b(t, x), \) and \( c(t, x) \) are \textit{analytic and}

\(^{20}\)Throughout, \( dx := dx^1 dx^2 dx^3 \) denotes the standard Euclidean volume form on \( \Sigma_t \).
verify certain technical assumptions, then it is known that equation (1.6.1) is well-posed for $C^\infty$ data; see also for similar results. There are also results on well-posedness for degenerate semilinear equations. For example, in [24], the authors used a Nash-Moser argument to prove $C^\infty$ local well-posedness for semilinear equations of the form

$$-\partial_t^2 \Psi + a(t, x)\partial_x^2 \Psi = f(t, x, \Psi, \partial_t \Psi, \partial_x \Psi),$$

where $a(t, x) \geq 0$ is analytic and $a$ and $f$ verify appropriate technical assumptions. We clarify that in contrast to our work here, in the above works, the authors solved the equation in a full spacetime neighborhood of the degeneracy.

A serious limitation of the above results is that techniques relying on analyticity assumptions are of little use for studying quasilinear Cauchy problems with Sobolev-class data, such as the problems we consider here. Fortunately, well-posedness results for degenerate linear equations that do not rely on analyticity assumptions are also known; see, for example, [23, 40–43, 79, 97] and the references within. We note in particular that [40–43, 79, 97] provide Sobolev estimates for the solution in terms of a Sobolev norm of the data, with a finite loss of derivatives. We also mention the related works [11, 12] (see also the references within), in which the author proves well-posedness results (in $C^\infty$ and Gevrey spaces) for linear wave equations with two kinds of degeneracy: i) breakdown in strict hyperbolicity (corresponding to the vanishing of certain coefficients) and ii) the blowup of the time derivatives of certain coefficients in the wave equation. We also mention the works [46, 47] in which the authors obtain necessary and sufficient conditions for Gevrey space well-posedness of linear degenerate hyperbolic equations.

More relevant for our work here are Manfrin’s aforementioned proofs [69, 70] (see also the related work [13]) of $C^\infty$ well-posedness for various degenerate quasilinear wave equations, including those of the form

$$-\partial_t^2 \Psi + \Psi^{2k} a(t, x, \Psi)\partial_x^2 \Psi = f(t, x, \Psi),$$

where $k \geq 1$ is an integer and $a(t, x, \Psi)$ is uniformly bounded from above and from below, strictly away from 0 (and $\Psi = 0$ corresponds to the degeneracy). More precisely, for $C^\infty$ initial data, Manfrin used weighted energy estimates and Nash-Moser estimates to prove local well-posedness for solutions to (1.6.3). Note that his results apply to our model equation (1.1.1a) in the case $P = 2$. However, like the other results that we have discussed so far in Subsect. 1.6, Manfrin’s results have not been extended to more than one spatial dimension. The main reason, presumably, is that the energy estimate weights are complicated to construct and are based on dividing spacetime into various regions with the help of “separating functions” adapted to the degeneracy; these constructions do not easily generalize to the case of more than one spatial dimension.

To further explain these results and their connection to our work here, we consider the simple Tricomi-type equation

$$-\partial_t^2 \Psi + a(t)\Delta \Psi = 0,$$
where \( a(t) \geq 0 \). It is known \(^{[17]}\) that in one spatial dimension, the linear equation \((1.6.4)\) can be ill-posed\(^{21}\) even if \( a = a(t) \) is \( C^{\infty} \). Hence, it should not be taken for granted that we can (for suitable data) solve equation \((1.1.1a)\) all the way up to the time of first vanishing of \( 1 + \Psi \). Roughly, what can go wrong in an attempt to solve the linear equation \((1.6.4)\) is that \( a(t) \) can be highly oscillatory near a point \( t_0 \) with \( a(t_0) = 0 \). In fact, in the example from \(^{[17]}\), \( a(t) \) oscillates infinitely many times near \( t_0 \). This generates, in the basic energy identity, an uncontrollable term involving the ratio \( \frac{\partial}{\partial t} a(t) / a(t) \) and leads to ill-posedness in domains \([A, B) \times \mathbb{R}\) when \( t_0 \in [A, B) \).

In all of the aforementioned well-posedness results, the technical conditions imposed on the coefficients rule out the infinite oscillatory behavior from \(^{[17]}\) that led to ill-posedness. To provide a more concrete example, we note that in one spatial dimension, Han derived \(^{[38]}\) degenerate energy estimates for linear wave equations of the form
\[
-\partial_t^2 \Psi + a(t, x) \partial_x^2 \Psi + b_0(t, x) \partial_t \Psi + b(t, x) \partial_x \Psi + c(t, x) \Psi = f(t, x),
\]
where the coefficients verify certain technical conditions, including, roughly speaking, that \( a(t, x) \geq 0 \) behaves like \( t^n + c_{m-1}(x) t^{m-1} + \cdots + c_1(x) t + c_0(x) \). Even though \( a \) is allowed to vanish at some points, it does not exhibit highly oscillatory behavior in the \( t \) direction. In \(^{[40]}\), similar results were derived in \( n \geq 1 \) spatial dimensions.

As we have mentioned, there are only a few works that address well-posedness for degenerate quasilinear equations exhibiting features that have anything in common with equation \((1.1.1a)\) near the degeneracy \( 1 + \Psi = 0 \). A notable example is Dreher’s PhD thesis \(^{[33]}\), in which he proved local well-posedness results in Sobolev spaces for several classes of quasilinear hyperbolic PDEs in any number of dimensions with various kinds of space and time degeneracies. However, a key difference between the equations studied in \(^{[33]}\) and our work is that the degeneracies in \(^{[33]}\) were “prescribed” in the sense that they were caused only by degenerate semilinear factors that explicitly depend on the time and space variables. That is, if one deletes the degenerate semilinear factors, then one obtains a strictly hyperbolic PDE for which local well-posedness follows from standard techniques. Dreher made technical assumptions on the degenerate semilinear factors that were sufficient for proving well-posedness. In contrast, the degeneracy caused by \( 1 + \Psi = 0 \) in equation \((1.1.1a)\) is “purely quasilinear” in nature. The following model equation in one spatial dimension gives a sense of the kinds of prescribed degeneracy treated by Dreher in \(^{[33]}\):
\[
-\partial_t^2 \Psi + t^2 f(\Psi) \partial_x^2 \Psi = 0,
\]
where \( f \) is smooth and verifies \( f(\Psi) > 0 \). Note that the absence of strict hyperbolicity in a neighborhood of \( \Sigma_0 \) is not caused by the quasilinear factor \( f(\Psi) \), but rather by the semilinear factor \( t^2 \). A related example, coming from geometry, is the aforementioned work \(^{[39]}\), in which the authors proved the existence of local \( C^k \) embeddings of surfaces of non-negative Gaussian curvature into \( \mathbb{R}^3 \). The quasilinear system of PDEs studied there degenerated at points where the Gauss curvature of the surface vanishes. As in Dreher’s work \(^{[33]}\), the degeneracy was “prescribed” in the sense that it was caused by the Gauss curvature (which

\(^{21}\)In \(^{[17]}\), which addressed solutions in one spatial dimension, the authors exhibited a smooth function \( a(t) \geq 0 \) with \( a(t_0) = 0 \) for some \( t_0 > 0 \) and data such that there is no distributional solution to \((1.6.4)\) extending past time \( t_0 \).
is “known”). Hence, the authors were free to make technical assumptions on the Gauss curvature to ensure the local well-posedness of the PDE system.

We now give another example of prior work that is more closely connected to our main results. In [85], the authors proved local well-posedness in homogeneous Sobolev spaces on domains of the form $[0,T) \times \mathbb{R}^n$ for semilinear Tricomi equations of the form

$$-\partial_t^2 \Psi + t^p \Delta \Psi = f(\Psi),$$

(1.6.7)

where $P \in \mathbb{N}$ and $f$ is a nonlinearity such that $f$ and $f'$ obey certain $P,n-$dependent power-law growth bounds at $\infty$. See [86,87] for related results. Note that the coefficient $t^p$ in (1.6.7) does not oscillate; once again, this is the key difference compared to the ill-posedness result for equation (1.6.4) mentioned above. As we described in Subsect. 1.5, for the data under consideration, equation (1.6.7) is a good model for equation (1.1.1a) in the sense that the degenerating coefficient $(1 + \Psi)^p$ in (1.1.1a) behaves in some ways, when $1 + \Psi$ is small, like the coefficient $t^p$ (near $t = 0$) in (1.6.7).

In view of the above discussion, one should not expect to be able to solve equation (1.1.1a) all the way up to points with $1 + \Psi = 0$ unless one makes assumptions on the data that preclude highly oscillatory behavior in regions where $1 + \Psi$ is small. In this article, we avoid the oscillatory behavior by exploiting the relative largeness of $[\partial_t \Psi]_-$ (consistent with Remark 1.1) and the relative smallness of $\partial_t^2 \Psi$ in regions where $1 + \Psi$ is small, which are present at time 0 and which we propagate; see the estimates (3.2.2) and (3.3.5c). As we have mentioned, the relative largeness of $[\partial_t \Psi]_-$ can be viewed as a kind of monotonicity in the problem. One might say that this monotonicity makes up for the lack of remarkable structure in (1.1.1a), including that it is not an Euler-Lagrange equation, its solutions admit no known coercive conserved quantities, and there are no definite-signed nonlinearities. As we described in Subsect. 1.5, this monotonicity yields an important signed spacetime integral that we use to close the energy estimates; see the spacetime integral on LHS (3.3.4). The largeness of $[\partial_t \Psi]_-$ is connected to so-called Levi-type conditions that have appeared in the literature. Roughly, a Levi condition is a quantitative relationship between the sizes of various coefficients in the equation and their derivatives. As an example, we note that in their study [24] of well-posedness for equation (1.6.2) with analytic coefficients, the authors studied linearized equations of the form (1.6.1) under the Levi condition $|b(t,x)| \lesssim |a(t,x)| + |\partial_t \sqrt{a}(t,x)|$; the Levi condition allowed them, for the linearized equation, to construct suitable weights for the energy estimates (even at points where $a$ vanishes), which were sufficient for proving well-posedness. In the problems under study here, the largeness of $[\partial_t \Psi]_-$ at points with $1 + \Psi = 0$ can be viewed as a Levi-type condition for the coefficient $(1 + \Psi)^p$ in equation (1.1.1a), which allows us to handle various error terms that arise when we derive energy estimates for the solution’s higher derivatives.

The degenerate energy estimates featured in our proofs have some features in common with the foundational works [2,3,14] of Alinhac and Christodoulou on the formation of shock singularities in solutions to quasilinear wave equations in two or three spatial dimensions; see also the follow-up works [16,25,27,90,92] and the survey article [44]. In those works, the authors constructed a dynamic geometric coordinate system that degenerated in a precise

---

The key point is that since our solutions are such that $\partial_t \Psi < 0$ when $1 + \Psi = 0$, it follows that $1 + \Psi$ behaves, to first order, linearly in $t$ near points where it vanishes.
fashion\textsuperscript{23} as the shock formed. Consequently, relative to the geometric coordinates, the solution remains rather smooth\textsuperscript{24} which was a key fact used to control error terms. A crucial feature of the proofs is that the energy estimates\textsuperscript{25} contained weights that vanished at the shock, which is in analogy with the vanishing of the weight $(1 + \Psi)^p$ in (1.5.1) at the degeneracy. A second crucial feature of the proofs of shock formation is that they relied on the fact that the weight has a \textit{quantitatively strictly negative time derivative} in a neighborhood of points where it vanishes. This yields a critically important monotonic spacetime integral that is in analogy with the one (1.5.2) that we use to control various error terms in the present work.

We close this subsection by noting that the degeneracy that we encounter in our study of equation (1.1.1) is related to - but distinct from - a particular kind of absence of strict hyperbolicity that has been studied in the context of the compressible Euler equations for initial data verifying the physical vacuum condition; see, for example, \cite{19,21,48,49}. The key difference between these works and ours is that in \cite{19,21,48,49}, the degeneracies occurred along the fluid-vacuum boundary in spacelike directions rather than a timelike one. In particular, the degeneracy was already present at time 0. More precisely, at time 0, the fluid density vanished a certain rate, meaning that the density’s derivative in the (spacelike) normal direction to the vacuum boundary satisfied a quantitative signed condition. It turns out that this condition yields a signed integral in the energy identities that is essential for closing the energy estimates. The signed integral exploited in those works is in analogy with the integral (1.5.2), but the integrals in \cite{19,21,48,49} were available because the solution’s (spacelike) normal derivative had a sign, which is in contrast to the sign of the timelike derivative $\partial_t \Psi$ exploited in the present work. With the help of the signed integral, the authors of \cite{19,21,48,49} were able to prove degenerate energy estimates with weights that vanished at a certain rate in the normal direction to the vacuum boundary. Ultimately, these degenerate estimates allowed them to prove local well-posedness in Sobolev spaces with degenerate weights.

1.7. Notation. In this subsection, we summarize some notation that we use throughout.

- $\{x^\alpha\}_{\alpha=0,1,2,3}$ denotes the standard rectangular coordinates on $\mathbb{R}^{1+3} \simeq \mathbb{R} \times \mathbb{R}^3$. $\partial_\alpha := \frac{\partial}{\partial x^\alpha}$ denotes the corresponding coordinate partial derivative vector fields. $x^0 \in \mathbb{R}$ is the time coordinate and $\vec{x} := (x^1, x^2, x^3) \in \mathbb{R}^3$ are spatial coordinates.
- We sometimes use the alternate notation $x^0 = t$ and $\partial_0 = \partial_t$.
- Greek “spacetime” indices such as $\alpha$ vary over 0, 1, 2, 3 and Latin “spatial” indices such as $a$ vary over 1, 2, 3. We use Einstein’s summation convention in that repeated indices are summed over their respective ranges.

\textsuperscript{23}In essence, the authors straightened out the characteristics via a solution-dependent change of coordinates.

\textsuperscript{24}The high-order geometric energies were allowed to blow up at the shock, which led to enormous technical complications in the proofs. Note that this possible high-order energy blowup is distinct than the formation of the shock.

\textsuperscript{25}There are many shock-formation results for solutions to quasilinear equations in one spatial dimension, with important contributions coming from Riemann \cite{82}, Oleinik \cite{78}, Lax \cite{64}, Klainerman and Majda \cite{56}, John \cite{50–52}, and many others. However, those results are based exclusively on the method of characteristics and hence, unlike in the case of two or more spatial dimensions, the proofs do not rely on energy estimates.
• We raise and lower indices with $g^{-1}$ and $g$ respectively (not with the Minkowski metric!).
• We sometimes omit the arguments of functions appearing in pointwise inequalities. For example, we sometimes write $|f| \leq C \epsilon$ instead of $|f(t, \underline{x})| \leq C \epsilon$.
• $\nabla^k \Psi$ denotes the array of $k$th-order derivatives of $\Psi$ with respect to the rectangular spatial coordinate vectorfields. For example, $\nabla \Psi := (\partial_1 \Psi, \partial_2 \Psi, \partial_3 \Psi)$.
• $|\nabla \leq \Psi| := \sum_{k' = 0}^k |\nabla_k' \Psi|$.
• $|\nabla[a,b] \Psi| := \sum_{k' = a}^b |\nabla_k' \Psi|$.
• $H^N(\Sigma_t)$ denotes the standard Sobolev space of functions on $\Sigma_t$ with corresponding norm $\|f\|_{H^N(\Sigma_t)} := \left\{ \sum_{a_1 + a_2 + a_3 \leq N} \int_{\mathbb{R}^3} |\partial_1^a \partial_2^a \partial_3^a f(t, \underline{x})|^2 \, dx \right\}^{1/2}$. In the case $N = 0$, we use the notation “$L^2$” in place of “$H^0$.”
• $L^\infty(\Sigma_t)$ denotes the standard Lebesgue space of functions on $\Sigma_t$ with corresponding norm $\|f\|_{L^\infty(\Sigma_t)} := \text{ess sup}_{\underline{x} \in \mathbb{R}^3} |f(t, \underline{x})|$.
• We sometimes write $O(B)$ to denote a quantity $A$ with the following property: there exists a constant $C > 0$ such that $|A| \leq C|B|$.

2. ASSUMPTIONS ON THE INITIAL DATA AND BOOTSTRAP ASSUMPTIONS

In this short section, we state our assumptions on the data $(\Psi|_{\Sigma_0}, \partial_t \Psi|_{\Sigma_0}) := (\Psi_0, \dot{\Psi}_0)$ for the model equation (1.1.1a) and formulate bootstrap assumptions that are convenient for studying the solution. We also show that there exist data verifying the assumptions.

2.1. Assumptions on the data. We assume that the initial data are compactly supported and verify the following size assumptions:

$$
\|\nabla \leq 4 \dot{\Psi}\|_{L^\infty(\Sigma_0)} + \|\nabla^{[1,3]} \dot{\Psi}_0\|_{L^\infty(\Sigma_0)} + \|\nabla \dot{\Psi}\|_{H^0(\Sigma_0)} + \|\nabla^2 \dot{\Psi}_0\|_{H^3(\Sigma_0)} \leq \dot{\epsilon}, \quad \|\dot{\Psi}_0\|_{L^\infty(\Sigma_0)} \leq \dot{\delta},
$$

(2.1.1)

where $\dot{\epsilon}$ and $\dot{\delta}$ are two data-size parameters that we will discuss below (roughly, $\dot{\epsilon}$ will have to be small for our proofs to close). In our analysis, we will essentially propagate the above size assumptions all the way up until the time of breakdown in hyperbolicity. A possible exception can occur for the top space derivatives of $\Psi$, which are are not able to control uniformly in $L^2(\Sigma_t)$ due to the presence of degenerate weights in our energy (see Def. [3.1]).

Before we can proceed, we must first introduce the crucial parameter $\dot{\delta}_*$ that controls the time of first breakdown in hyperbolicity; our analysis shows that for $\dot{\epsilon}$ sufficiently small, the time of first breakdown is $(1 + O(\dot{\epsilon}))\delta_*^{-1}$; see also Remark [2.1].

**Definition 2.1** (The parameter that controls the time of breakdown in hyperbolicity). We define the data-dependent parameter $\dot{\delta}_*$ as follows:

$$
\dot{\delta}_* := \max_{\Sigma_0} |\dot{\Psi}_0|_-. \quad (2.1.2)
$$
Remark 2.1 (The relevance of $\hat{\delta}_s$). The solutions that we study are such that $\Psi \sim 0$ and $\partial_t^2 \Psi \sim 0$ (throughout the evolution). Hence, by the fundamental theorem of calculus, we have $\partial_t \Psi(t, \xi) \sim \Psi_0(\xi)$ and $1 + \Psi(t, \xi) \sim 1 + t \Psi_0(\xi)$. From this last expression, we see that $1 + \Psi$ is expected to vanish for the first time at approximately $t = \hat{\delta}_s^{-1}$. See Lemma 3.1 for the precise statements.

2.2. Bootstrap assumptions. To prove our main results, it is convenient to rely on a set of bootstrap assumptions, which we provide in this subsection.

The size of $T_{(\text{Boot})}$. We assume that $T_{(\text{Boot})}$ is a bootstrap time with

$$0 < T_{(\text{Boot})} \leq 2 \hat{\delta}_s^{-1}.$$  \hfill (2.2.1)

The assumption (2.2.1) gives us a sufficient margin of error to prove that finite-time degeneration of hyperbolicity occurs, as we explained in Remark 2.1.

Degeneracy has not yet occurred. We assume that for $(t, \xi) \in [0, T_{(\text{Boot})}) \times \mathbb{R}^3$, we have

$$0 < 1 + \Psi(t, \xi).$$  \hfill (2.2.2)

$L^\infty$ bootstrap assumptions. We assume that for $t \in [0, T_{(\text{Boot})})$, we have

$$\|\Psi\|_{L^\infty(\Sigma_t)} \leq 2 \hat{\delta}_s^{-1} \delta + \varepsilon^{1/2},$$  \hfill (2.2.3a)

$$\|\partial_t \Psi\|_{L^\infty(\Sigma_t)} \leq \hat{\delta} + \varepsilon^{1/2},$$  \hfill (2.2.3b)

$$\|\nabla^{[1, 3]} \Psi\|_{L^\infty(\Sigma_t)}, \|\partial_t \nabla^{[1, 3]} \Psi\|_{L^\infty(\Sigma_t)}, \|\partial_t^2 \nabla^{\leq 1} \Psi\|_{L^\infty(\Sigma_t)} \leq \varepsilon,$$  \hfill (2.2.3c)

where $\varepsilon > 0$ is a small bootstrap parameter; we describe our smallness assumptions in the next subsection.

Remark 2.2 (The solution remains compactly supported in space). From (2.2.3a), we deduce that the wave speed $(1 + \Psi)^{p/2}$ associated to equation (1.1.1a) remains uniformly bounded for $(t, \xi) \in [0, T_{(\text{Boot})}) \times \mathbb{R}^3$. Hence, there exists a large, data-dependent ball $B \subset \mathbb{R}^3$ such that $\Psi(t, \xi)$ vanishes for $(t, \xi) \in [0, T_{(\text{Boot})}) \times B^c$.

2.3. Smallness assumptions. For the rest of the article, when we say that “$A$ is small relative to $B$,” we mean that $B > 0$ and that there exists a continuous increasing function $f : (0, \infty) \to (0, \infty)$ such that $A \leq f(B)$. In principle, the functions $f$ could always be chosen to be polynomials with positive coefficients or exponential functions. However, to avoid lengthening the paper, we typically do not specify the form of $f$.

Throughout the rest of the paper, we make the following relative smallness assumptions. We continually adjust the required smallness in order to close our estimates.

- The bootstrap parameter $\varepsilon$ from Subsect. 2.2 is small relative to $\hat{\delta}^{-1}$, where $\hat{\delta}$ is the data-size parameter from (2.1.1).
- $\varepsilon$ is small relative to the data-size parameter $\hat{\delta}_s$ from (2.1.2).

The first assumption will allow us to control error terms that, roughly speaking, are of size $\varepsilon \hat{\delta}^k$ for some integer $k \geq 0$. The second assumption is relevant because the expected degeneracy-formation time is approximately $\hat{\delta}_s^{-1}$ (see Remark 2.1); the assumption will allow

\footnote{Here “$A \sim B$” imprecisely indicates that $A$ is well-approximated by $B$.}
us to show that various error products featuring a small factor $\varepsilon$ remain small for $t \leq 2\delta_*^{-1}$, which is plenty of time for us to show that $1 + \Psi$ vanishes.

Finally, we assume that

$$\varepsilon^{3/2} \leq \hat{\varepsilon} \leq \varepsilon, \quad (2.3.1)$$

where $\hat{\varepsilon}$ is the data smallness parameter from $[2.1.1]$.

2.4. Existence of data. It is easy to construct data verifying the smallness assumptions stated in Subsect. 2.3. For example, we may start with any smooth compactly supported data $(\hat{\Psi}, \hat{\Psi}_0)$ such that $\min_{\mathbb{R}^3} \hat{\Psi}_0 < 0$. We then consider the one-parameter family

$$\left( (\lambda) \hat{\Psi}(x), (\lambda) \hat{\Psi}_0(x) \right) := \left( \lambda^{-1} \hat{\Psi}(x), \hat{\Psi}_0(\lambda^{-1} x) \right).$$

One may check that for $\lambda > 0$ sufficiently large, all of the size assumptions of Subsect. 2.3 are verified. The proof relies on the simple identities

$$\nabla^k (\lambda) \hat{\Psi}(x) = \lambda^{-1} (\nabla^k \hat{\Psi})(x), \quad (2.4.1a)$$

$$\nabla^k (\lambda) \hat{\Psi}_0(x) = \lambda^{-k} (\nabla^k \hat{\Psi}_0)(\lambda^{-1} x)$$

and

$$\left\| \nabla^k (\lambda) \hat{\Psi} \right\|_{L^2(\Sigma_0)} = \lambda^{-1} \left\| \hat{\Psi} \right\|_{L^2(\Sigma_0)}, \quad (2.4.2a)$$

$$\left\| \nabla^k (\lambda) \hat{\Psi}_0 \right\|_{L^2} = \lambda^{3/2-k} \left\| \hat{\Psi}_0 \right\|_{L^2(\Sigma_0)}, \quad (2.4.2b)$$

Remark 2.3 (Degeneracy occurs for solutions launched by any appropriately rescaled non-trivial rescaled data). The discussion in Subsect. 2.4 can easily be used to show that if $\hat{\Psi}$ is non-trivial, then one always generates data to which our results apply by considering the rescaled data $\left( (\lambda) \hat{\Psi}, (\lambda) \hat{\Psi}_0 \right)$ with $\lambda$ sufficiently large. More precisely, if $\min_{\mathbb{R}^3} \hat{\Psi}_0 = 0$, then we must have $\max_{\mathbb{R}^3} \hat{\Psi}_0 > 0$; in this case, the degeneracy in the solution generated by the rescaled data occurs in the past rather than the future.

3. A priori estimates

In this section, we use the data-size assumptions and the bootstrap assumptions of Sect. 2 to derive a priori estimates for the solution. This is the main step in the proof our results.

3.1. Conventions for constants. In our estimates, the explicit constants $C > 0$ and $c > 0$ are free to vary from line to line. These constants are allowed to depend on the data-size parameters $\delta$ and $\delta_*^{-1}$ from Subsect. 2.1. However, the constants can be chosen to be independent of the parameters $\hat{\varepsilon}$ and $\varepsilon$ whenever $\hat{\varepsilon}$ and $\varepsilon$ are sufficiently small relative to $\delta^{-1}$ and $\delta_*$ in the sense described in Subsect. 2.3. For example, under our conventions, we have that $\delta_*^{-2} \varepsilon = O(\varepsilon)$.
3.2. Pointwise estimates. In this subsection, we derive pointwise estimates for $\Psi$ and the inhomogeneous terms in the commuted wave equation.

We start with a simple lemma that provides sharp pointwise estimates for $\Psi$ and $\partial_t \Psi$.

**Lemma 3.1 (Pointwise estimates for $\Psi$ and $\partial_t \Psi$).** Under the data-size assumptions Subsect. 2.1, the bootstrap assumptions of Subsect. 2.2, and the smallness assumptions of Subsect. 2.3, the following pointwise estimates hold for $(t, \vec{x}) \in [0, T_{(\text{Boot})}) \times \mathbb{R}^3$:

\[
\begin{align*}
\Psi(t, \vec{x}) &= t\hat{\Psi}_0(\vec{x}) + O(\varepsilon), \quad (3.2.1a) \\
\partial_t \Psi(t, \vec{x}) &= \hat{\Psi}_0(\vec{x}) + O(\varepsilon). \quad (3.2.1b)
\end{align*}
\]

**Proof.** To derive (3.2.1b), we first use the bootstrap assumptions to deduce $\| (1 + \Psi) P \Delta \Psi \|_{L^\infty(\Sigma_t)} \leq C\varepsilon$. Hence, from equation (1.1.1a), we deduce the pointwise bound $|\partial_t^2 \Psi| \leq C\varepsilon$. From this estimate and the fundamental theorem of calculus, we conclude the desired bound (3.2.1b).

The bound (3.2.1a) then follows from the fundamental theorem of calculus, (3.2.1b), and the small-data bound $\|\hat{\Psi}\|_{L^\infty(\Sigma_0)} \leq \hat{\varepsilon} \leq \varepsilon$. \qed

The next proposition captures the monotonicity that is present at points where $1 + \Psi$ is small. It is of critical importance for the energy estimates.

**Proposition 3.2 (Monotonicity near the degeneracy).** Under the data-size assumptions Subsect. 2.1, the bootstrap assumptions of Subsect. 2.2, and the smallness assumptions of Subsect. 2.3, the following statements hold for $(t, \vec{x}) \in [0, T_{(\text{Boot})}) \times \mathbb{R}^3$:

\[
\Psi(t, \vec{x}) \leq -\frac{1}{2} \implies \partial_t \Psi(t, \vec{x}) \leq -\frac{1}{8} \delta^*_s.
\]

Moreover,

\[
\partial_t \Psi(t, \vec{x}) \geq 0 \implies \Psi(t, \vec{x}) \geq -C\varepsilon. \quad (3.2.3)
\]

**Proof.** To prove (3.2.2), we first use the estimates (3.2.1a) and (3.2.1b) to deduce that $\Psi(t, \vec{x}) = t\hat{\Psi}_0(\vec{x}) + O(\varepsilon)$. Hence, if $\Psi(t, \vec{x}) \leq -\frac{1}{2}$, then $t\partial_t \Psi(t, \vec{x}) \leq -\frac{1}{4}$. Recalling that $0 \leq t < 2\delta^{-1}_s$ (see (2.2.1)), we conclude (3.2.2).

The estimate (3.2.3) then follows as a simple consequence of the bound $\Psi(t, \vec{x}) = t\partial_t \Psi(t, \vec{x}) + O(\varepsilon)$ proved above. \qed

We now derive pointwise estimates for the inhomogeneous terms in the commuted wave equation.

**Lemma 3.3 (Pointwise estimates for the inhomogeneous terms).** Let $\Psi$ be a solution to the wave equation (1.1.1a). For $k = 2, 3, 4, 5$ and $P = 1, 2$, consider following wave equation obtained by commuting (1.1.1a) with $\nabla^k$:

\[
-\partial_t^2 \nabla^k \Psi + (1 + \Psi)^P \Delta \nabla^k \Psi = F^{(k)}. \quad (3.2.4)
\]

Under the data-size assumptions Subsect. 2.1, the bootstrap assumptions of Subsect. 2.2, and the smallness assumptions of Subsect. 2.3, the following pointwise estimates hold $(t, \vec{x}) \in$ 

\footnote{We do not bother to state the precise form of $F^{(k)}$ here.}
Let \([0, T_{\text{Boot}}) \times \mathbb{R}^3\):
\[
|F^{(k)}| \leq C_{\varepsilon} |\nabla^{[2, k+1]} \Psi|, \quad (P = 1), \tag{3.2.5}
\]
\[
|F^{(k)}| \leq C_{\varepsilon}(1 + \Psi)|\nabla^{k+1} \Psi| + \varepsilon |\nabla^{[2, k]} \Psi|, \quad (P = 2). \tag{3.2.6}
\]

**Proof.** We first consider the case \(P = 1\). Commuting \((1.1.1a)\) with \(\nabla^k\), we compute that
\[
|F^{(k)}| \leq C \sum_{a+b \leq k+2} |\nabla^a \Psi| |\nabla^b \Psi|. \tag{3.2.5}
\]
The desired estimate \((3.2.5)\) then follows as a simple consequence of this bound and the bootstrap assumptions. The proof of \((3.2.6)\) is similar, the difference being that when \(P = 2\), we have the bound
\[
|F^{(k)}| \leq C(1 + \Psi)|\nabla^{k+1} \Psi| + C \sum_{a+b \leq k+2} |\nabla^a \Psi| |\nabla^b \Psi|. \tag{3.2.6}
\]

### 3.3. Energy estimates
We will use the following energy, which corresponds to between two and five commutations of the wave equation with \(\nabla\), in order to control solutions.

**Definition 3.1 (The energy).** We define
\[
\mathcal{E}_{[2,5]}(t) := \sum_{k'=2}^{5} \int_{t=0}^{t} \int_{\Sigma_t} |\partial_t \nabla^{k'} \Psi|^2 + (1 + \Psi)^P |\nabla \nabla^{k'} \Psi|^2 + |\nabla^{k'} \Psi|^2 d\xi. \tag{3.3.1}
\]

We now provide the basic energy identity verified by solutions.

**Lemma 3.4 (Basic energy identity).** Let \(\Psi\) be a solution to the wave equation \((1.1.1a)\). Let \(\mathcal{E}_{[2,5]}\) be the energy defined in \((3.3.1)\) and let \(F^{(k)}\) be the inhomogeneous term from \((3.2.4)\). Then for \(P = 1, 2\), we have the following energy identity:
\[
\mathcal{E}_{[2,5]}(t) = \mathcal{E}_{[2,5]}(0) + P \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} (\partial_\tau \Psi)(1 + \Psi)^{P-1} |\nabla \nabla^{k'} \Psi|^2 d\xi ds \tag{3.3.2}
\]
\[
- P \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} (1 + \Psi)^{P-1} (\nabla \Psi) \cdot (\nabla \nabla^{k'} \Psi)(\partial_\tau \nabla^{k'} \Psi) d\xi ds
\]
\[
- 2 \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} (\partial_\tau \nabla^{k'} \Psi)F^{(k')} d\xi ds + 2 \sum_{k'=2}^{5} \int_{s=0}^{t} \int_{\Sigma_s} (\partial_\tau \nabla^{k'} \Psi)(\nabla^{k'} \Psi) d\xi ds.
\]

**Proof.** The identity \((3.3.2)\) is standard and can verified by taking the time derivative of both sides of \((3.3.1)\), using equation \((3.2.4)\) for substitution, integrating by parts over \(\Sigma_t\), and then integrating the resulting identity in time. \(\square\)

With the help of Lemma 3.4, we now derive an inequality verified by the energy \(\mathcal{E}_{[2,5]}\).

**Proposition 3.5 (Integral inequality for the energy).** Let \(\mathcal{E}_{[2,5]}\) be the energy defined in \((3.3.1)\). Let \(1_{[-1 < \Psi \leq -\frac{1}{2}]}\) be the characteristic function of the spacetime subset \(\{(t, x) \mid -1 < \Psi(t, x) \leq -\frac{1}{2}\}\) and define \(1_{\{-\frac{1}{2} < \Psi\}}\) analogously. Under the data-size assumptions Subsect. 2.1, the bootstrap assumptions of Subsect. 2.2, and the smallness assumptions...
of Subsect. 2.3, the following integral inequality holds for $P = 1, 2$ and $t \in [0, T_{(\text{Boot})})$:

$$
E_{[2,5]}(t) + \frac{P}{8} \delta_s \sum_{k'=2}^5 \int_s^t \int_{\Sigma_s} 1_{\{-1 < \Psi \leq -\frac{1}{2}\}} (1 + \Psi)^{P-1} |\nabla \nabla k' \Psi|^2 \, dx \, ds
$$

\hspace{1cm}

(3.3.3)

\hspace{1cm}

\leq E_{[2,3]}(0) + C \sum_{k'=2}^5 \int_s^t \int_{\Sigma_s} |\partial_t \nabla k' \Psi|^2 \, dx \, ds + C \sum_{k'=2}^5 \int_{t=0}^s \int_{\Sigma_s} |\nabla k' \Psi|^2 \, dx \, ds

\hspace{1cm}

+ C \sum_{k'=2}^5 \int_s^t \int_{\Sigma_s} 1_{\{-1 < \Psi \leq -\frac{1}{2}\}} (1 + \Psi)^{2(P-1)} |\nabla \nabla k' \Psi|^2 \, dx \, ds.

Proof. We must bound the terms appearing in the energy identity (3.3.2). We give the proof only the case $P = 1$ since the case $P = 2$ can be handled using similar arguments. We start by bounding the first sum on RHS (3.3.2); this is the only one that requires careful treatment. We split the integration domain $[0, t] \times \mathbb{R}^3$ into two pieces: a piece in which $-1 < \Psi \leq -\frac{1}{2}$ and a piece in which $\Psi > -\frac{1}{2}$. To bound the first piece, we use the estimate (3.2.2) to deduce that whenever $-1 < \Psi \leq -\frac{1}{2}$, the integrand verifies $(\partial_t \Psi)|\nabla \nabla k' \Psi|^2 \leq \frac{1}{8} \delta_s |\nabla \nabla k' \Psi|^2$. We may therefore bring all of the corresponding integrals over to LHS (3.3.3) as positive integrals, as is indicated there. To bound the second piece, we use the estimate (3.2.1b) to bound $\partial_t \Psi$ in $L^\infty$ by $\leq C$, which allows us to bound the integrand by $\leq C |\nabla \nabla k' \Psi|^2$. It follows that since $\Psi > -\frac{1}{2}$ (by assumption), the integrals under consideration are bounded by the third sum on RHS (3.3.3).

To bound the second sum on RHS (3.3.2), we first use the bootstrap assumption (2.2.3c) to bound the integrand factor $\nabla \Psi$ in $L^\infty$ by $\leq \varepsilon$. Thus, using Young’s inequality, we bound the terms under consideration by $\leq$ the sum of the first, third, and fourth sums on RHS (3.3.3).

To bound the third sum on RHS (3.3.2), we use (3.2.5) and Young’s inequality to bound the integrand by $\leq C \varepsilon \sum_{k'=2}^5 |\partial_t \nabla k' \Psi|^2 + C \sum_{k'=2}^6 |\nabla k' \Psi|^2$. It is easy to see that the corresponding integrals are bounded by $\leq$ RHS (3.3.3).

Finally, using Young’s inequality, we bound last sum on RHS (3.3.2) by the first two sums on RHS (3.3.3). \qed

In the next corollary, we use Prop. 3.5 to derive our main a priori energy estimates. We also derive improvements of the bootstrap assumptions (2.2.3a)-(2.2.3c).

Corollary 3.6 (A priori energy estimates and improvement of the bootstrap assumptions). Let $\delta_s$ be the data-size parameter from (2.1.2) and let $1_{\{-1 < \Psi \leq -\frac{1}{2}\}}$ be the characteristic function of the spacetime subset $\{(t, x) \mid -1 < \Psi(t, x) \leq -\frac{1}{2}\}$. There exists a constant $C > 0$ such that under the data-size assumptions Subsect. 2.1, the bootstrap assumptions of Subsect. 2.2 and the smallness assumptions of Subsect. 2.3, the following a
priori energy estimate holds for $P = 1, 2$ and $t \in [0, T_{\text{Boot}})$:

$$
\mathcal{E}_{[2,5]}(t) + \frac{P}{16} \delta_s \sum_{k' = 2}^{5} \int_{s=0}^{t} \int_{\Sigma_{s}} 1_{\{-1 < \Psi \leq -\frac{1}{4}\}} |\nabla \nabla^{k'} \Psi|^2 \, dx \, ds \leq C \delta^2. \tag{3.3.4}
$$

Moreover, we have the following estimates, which are a strict improvement of the bootstrap assumptions (2.2.3a)-(2.2.3c) for $\delta$ sufficiently small:

$$
\|\Psi\|_{L^\infty(\Sigma_t)} \leq 2 \delta_s^{-1} \delta + C \delta, \quad \tag{3.3.5a}
\|\partial_t \Psi\|_{L^\infty(\Sigma_t)} \leq \delta + C \delta, \quad \tag{3.3.5b}
\|\nabla^{[1,3]} \Psi\|_{L^\infty(\Sigma_t)}, \|\partial_t \nabla^{[1,3]} \Psi\|_{L^\infty(\Sigma_t)}, \|\partial_t^2 \nabla^{\leq 1} \Psi\|_{L^\infty(\Sigma_t)} \leq C \delta. \quad \tag{3.3.5c}
$$

Proof. We give the proof only the case $P = 1$ since the case $P = 2$ can be handled using similar arguments. To obtain (3.3.4), we first note that for $\varepsilon$ sufficiently small relative to $\delta_s$, we can absorb the last sum on RHS (3.3.3) into the second term on the LHS. Moreover, since $1 + \Psi \geq 1/2$ on $\{-1/2 < \Psi\}$, we have the bound $\int_{\Sigma_{t}} 1_{\{-1/2 < \Psi\}} |\nabla \nabla^{k'} \Psi|^2 \, dx \leq C \mathcal{E}_{[2,5]}(s)$ for the integrand in the next-to-last sum on RHS (3.3.3). The remaining $\Sigma_t$ integrals are easily seen to be $\leq C \mathcal{E}_{[2,5]}(s)$. Also using the data bound $\mathcal{E}_{[2,5]}(0) \leq C \delta^2$, which follows from our data assumptions (2.2.1), we obtain

$$
\mathcal{E}_{[2,5]}(t) + \frac{1}{16} \delta_s \sum_{k' = 2}^{5} \int_{s=0}^{t} \int_{\Sigma_{s}} 1_{\{-1 < \Psi \leq -\frac{1}{4}\}} |\nabla \nabla^{k'} \Psi|^2 \, dx \, ds \leq C \delta^2 + c \int_{s=0}^{t} \mathcal{E}_{[2,5]}(s) \, ds. \tag{3.3.6}
$$

From (3.3.6), Gronwall's inequality, and (2.2.1), we conclude

$$
\mathcal{E}_{[2,5]}(t) + \frac{1}{16} \delta_s \sum_{k' = 2}^{5} \int_{s=0}^{t} \int_{\Sigma_{s}} 1_{\{-1 < \Psi \leq -\frac{1}{4}\}} |\nabla \nabla^{k'} \Psi|^2 \, dx \, ds \leq C \exp^{ct} \delta^2 \leq C \delta^2,
$$

which is the desired bound (3.3.4).

The estimates (3.3.5c) for $\nabla^{[1,3]} \Psi$ and $\partial_t \nabla^{[1,3]} \Psi$ then follow from (3.3.4) and the Sobolev embedding result $H^2(\mathbb{R}^3) \hookrightarrow L^\infty(\mathbb{R}^3)$. Next, we take up to one $\nabla$ derivative of equation (1.1.1a) and use the already obtained $L^\infty$ estimates and the bootstrap assumptions to obtain the bound $\|\partial_t^2 \nabla^{\leq 1} \Psi\|_{L^\infty(\Sigma_t)} \leq C \delta$ stated in (3.3.5c). Using this bound, the fundamental theorem of calculus, and the data assumptions $\|\nabla \Psi_0\|_{L^\infty(\Sigma_0)} \leq \delta$ and $\|\Psi_0\|_{L^\infty(\Sigma_0)} \leq \delta$, we obtain the bounds $\|\partial_t \nabla \Psi\|_{L^\infty(\Sigma_t)} \leq C \delta$ and $\|\partial_t \Psi\|_{L^\infty(\Sigma_t)} \leq \delta + C \delta$, which in particular yields (3.3.5b). Using a similar argument based on the already obtained bound $\|\partial_t \nabla \Psi\|_{L^\infty(\Sigma_t)} \leq C \delta$, we deduce $\|\nabla \Psi\|_{L^\infty(\Sigma_t)} \leq C \delta$. Similarly, from the already obtained bound $\|\partial_t \Psi\|_{L^\infty(\Sigma_t)} \leq \delta + C \delta$, the fundamental theorem of calculus, the initial data bound $\|\Psi\|_{L^\infty} \leq \delta$, and the fact that $0 \leq t < T_{\text{Boot}} \leq 2 \delta_s^{-1}$, we deduce $\|\Psi\|_{L^\infty(\Sigma_t)} \leq 2 \delta_s^{-1} \delta + C \delta$, that is, (3.3.5a). We have thus proved the corollary. □
4. The main results

We now derive our main results: Theorem 4.1 and Prop. 4.1.

**Theorem 4.1 (Stable finite-time degeneracy).** Let $(\hat{\Psi}, \hat{\Psi}_0) \in H^6(\mathbb{R}^3) \times H^5(\mathbb{R}^3)$ be non-trivial compactly supported initial data for the wave equation (1.1.1a) with $P \in \{1, 2\}$ and let $\Psi$ denote the corresponding solution. Let

$$M(t) := \min_{\Sigma_t} \{1 + \Psi\}.$$  \hfill (4.0.7)

Let $\hat{\epsilon}, \hat{\delta},$ and $\hat{\delta}_*$ be the data-size parameters from (2.1.1)-(2.1.2) and assume that $\hat{\delta} > 0$ and $\hat{\delta}_* > 0.$ Note that if $\hat{\epsilon}$ is sufficiently small, then $M(0) = 1 + \mathcal{O}(\hat{\epsilon}) > 0.$ If $\hat{\epsilon}$ is sufficiently small relative to $\hat{\delta}^{-1}$ and $\hat{\delta}_*$ in the sense of Subsect. 2.3, then the following conclusions hold.

**Breakdown in hyperbolicity precisely at time $T_*:$** There exists a $T_*$ verifying

$$T_* = (1 + \mathcal{O}(\hat{\epsilon}))\hat{\delta}_*^{-1}$$  \hfill (4.0.8)

such that the solution exists classically on the slab $[0, T_*) \times \mathbb{R}^3$ and such that we have the following estimate for $0 \leq t < T_*:

$$M(t) > 0.$$  \hfill (4.0.9)

Moreover,

$$\lim_{t \uparrow T_*} M(t) = 0.$$  \hfill (4.0.10)

**Regularity properties on $[0, T_*) \times \mathbb{R}^3:$** On the slab $[0, T_*) \times \mathbb{R}^3,$ the solution verifies the energy bounds (3.3.4), the $L^\infty$ estimates (3.3.5c), and the pointwise estimates (3.2.1a)-(3.2.1b) (with $C\hat{\epsilon}$ on the RHS in place of $\epsilon$ in the latter two estimates). Moreover,

$$\Psi \in C \left([0, T_*), H^6(\mathbb{R}^3)\right) \cap L^\infty \left([0, T_*), H^5(\mathbb{R}^3)\right), \hfill (4.0.11a)$$

$$\partial_t \Psi \in C \left([0, T_*), H^5(\mathbb{R}^3)\right) \cap L^\infty \left([0, T_*), H^5(\mathbb{R}^3)\right). \hfill (4.0.11b)$$

**Regularity properties on $[0, T_*] \times \mathbb{R}^3:$** $\Psi$ extends to a classical solution on the closed slab $[0, T_*] \times \mathbb{R}^3$ enjoying the following regularity properties: for any $N < 5,$ we have

$$\Psi \in C \left([0, T_*], H^5(\mathbb{R}^3)\right) \cap L^1 \left([0, T_*], H^6(\mathbb{R}^3)\right), \hfill (4.0.12a)$$

$$\partial_t \Psi \in C \left([0, T_*], H^N(\mathbb{R}^3)\right) \cap L^\infty \left([0, T_*], H^5(\mathbb{R}^3)\right). \hfill (4.0.12b)$$

In particular, the $L^\infty$ estimates (3.3.5c) and the pointwise estimates (3.2.1a)-(3.2.1b) (with $C\hat{\epsilon}$ on the RHS in place of $\epsilon$ in these estimates) hold on $[0, T_*] \times \mathbb{R}^3.$

**Description of the breakdown along $\Sigma_{T_*}:$$$ The set

$$\Sigma_{T_*}^{\text{Degeneracy}} := \{q \in \Sigma_{T_*} \mid 1 + \Psi(q) = 0\}$$  \hfill (4.0.13)

is non-empty and we have the estimate

$$\sup_{\Sigma_{T_*}^{\text{Degeneracy}}} \partial_t \Psi \leq -\frac{1}{8} \hat{\delta}_*.$$  \hfill (4.0.14)
In particular, in the case \( P = 1 \), the hyperbolicity of the wave equation breaks down on \( \Sigma_{T_*}^{\text{Degeneracy}} \) in the following sense: if \( q \in \Sigma_{T_*}^{\text{Degeneracy}} \), then any \( C^1 \) extension of \( \Psi \) to any spacetime neighborhood \( \Omega_q \) containing \( q \) necessarily contains points \( p \) such that equation \( (1.1.1a) \) is elliptic at \( \Psi(p) \). In contrast, in the case \( P = 2 \), only the strict hyperbolicity of equation \( (1.1.1a) \) breaks down for the first time at \( T_* \).

Proof. Let \( T_{\text{(Boot)}} > 0 \) be any time subject to \( (2.2.1) \) such that the solution exists classically on the slab \([0, T_{\text{(Boot)}}) \times \mathbb{R}^3\) and verifies the bootstrap assumptions of Subsect. 2.2 with \( \varepsilon = C_* \delta \), where \( C_* \) is described just below. Specifically, by standard local well-posedness (see, for example, [45]), if \( \delta \) is sufficiently small and \( C_* > 1 \) is sufficiently large (note that this is consistent with the assumed inequalities \( (2.3.1) \)), then there exists such a \( T_{\text{(Boot)}} > 0 \). Next, we state the following standard continuation result, which can be proved, for example, by making straightforward modifications to the proof of [45] Theorem 6.4.11: the solution can be classically continued to a slab of the form \([0, T_{\text{(Boot)}} + \Delta) \times \mathbb{R}^3 \) (for some \( \Delta > 0 \)) as long as \( \inf_{t \in [0, T_{\text{(Boot)}})} \mathcal{M}(t) > 0 \) and \( \sup_{t \in [0, T_{\text{(Boot)}})} \left\{ \| \Psi \|_{L^\infty(\Sigma_t)} + \sum_{\alpha=0}^3 \| \partial^\alpha \Psi \|_{L^\infty(\Sigma_t)} \right\} < \infty \). The a priori estimates \( (3.3.5a)-(3.3.5c) \) ensure that \( \sup_{t \in [0, T_{\text{(Boot)}})} \left\{ \| \Psi \|_{L^\infty(\Sigma_t)} + \sum_{\alpha=0}^3 \| \partial^\alpha \Psi \|_{L^\infty(\Sigma_t)} \right\} < \infty \). Hence, we deduce that we can extend the solution until \( \mathcal{M} \) vanishes or until time \( 2\delta_*^{-1} \) (recall our assumption \( (2.2.1) \)), whichever happens first. We denote this “maximal time” by \( T_* \). Clearly the energy bounds \( (3.3.4) \) hold on \([0, T_*) \) and, since the a priori estimates \( (3.3.5a)-(3.3.5c) \) show that the bootstrap assumptions hold with \( \delta \) replaced by \( C_* \delta \), it follows that the pointwise estimates \( (3.2.1a)-(3.2.1b) \) hold on \([0, T_*) \times \mathbb{R}^3 \) with \( \delta \) replaced by \( C_* \delta \). Moreover, from definition \( 2.1.2 \) and the estimate \( (3.2.1a) \) (with \( \delta \) replaced by \( C_* \delta \)), we see that in fact, \( T_* = \delta_*^{-1} + O(\delta) = (1 + O(\delta))\delta_*^{-1} < 2\delta_*^{-1} \) and \( \mathcal{M}(T_*) = 0 \).

In the rest of this proof, we sometimes silently use the simple fact that \( \Psi \in H^1(\Sigma_t) \) for \( t \in [0, T_*) \). This fact does not follow from the energy estimates \( (3.3.4) \), but instead follows from \( (3.3.5a)-(3.3.5c) \) and the compactly supported (in space) nature of the solution (since the energy does not control \( \Psi \) itself or \( \nabla \Psi \)). To proceed, we easily conclude from the definition of \( \mathcal{E}_{[2,5]}(t) \) and the fact that the estimate \( (3.3.4) \) holds on \([0, T_*) \) that \( \partial_t \Psi \in L^\infty(\Sigma_t, H^3(\mathbb{R}^3)) \) as stated in \( (4.0.12b) \). Also, this fact trivially implies the corresponding statement in \( (4.0.11b) \), where the closed time interval is replaced with \([0, T_*) \). The facts that \( \Psi \in C \left([0, T_*) \times H^N(\mathbb{R}^3) \right) \) and \( \partial_t \Psi \in C \left([0, T_*) \times H^5(\mathbb{R}^3) \right) \) (as stated in \( (4.0.12a) \) and \( (4.0.12b) \)) are standard results that can be proved using energy estimates and simple facts from functional analysis. We omit the details and instead refer the reader to [91] Section 2.7.5. We note that in proving these facts, it is important that \( \mathcal{M}(t) > 0 \) on \([0, T_*) \), which implies that standard techniques for strictly hyperbolic equations can be used. To obtain the fact that \( \Psi \in L^1([0, T_*], H^6(\mathbb{R}^3)) \) stated in \( (4.0.12a) \) we use the fact that the energy bounds \( (3.3.4) \) hold on \([0, T_*) \) (in particular for the spacetime integral term on the LHS). The fact that \( \Psi \in C \left(\left([0, T_*], H^N(\mathbb{R}^3) \right) \right) \) (as stated in \( (4.0.12a) \)) is a simple consequence of the fundamental theorem of calculus and the already proven fact that \( \partial_t \Psi \in L^\infty(\Sigma_t, H^3(\mathbb{R}^3)) \). To obtain that for \( N < 5 \), we have \( \partial_t \Psi \in C \left(\left([0, T_*], H^N(\mathbb{R}^3) \right) \right) \) (as stated in \( (4.0.12b) \)), we first use the fact that \( \Psi \in C \left([0, T_*], H^6(\mathbb{R}^3) \right) \) to obtain \( \partial_t \Psi \in C \left([0, T_*], H^3(\mathbb{R}^3) \right) \). Hence, from the fundamental theorem of calculus, we obtain \( \partial_t \Psi \in C \left([0, T_*], H^3(\mathbb{R}^3) \right) \). From this fact and the fact that \( \partial_t \Psi \in L^\infty([0, T_*], H^5(\mathbb{R}^3)) \), we obtain, by interpolating between \( H^3 \) and \( H^5 \), the desired conclusion \( \partial_t \Psi \in C \left([0, T_*], H^N(\mathbb{R}^3) \right) \).
Next, we note that definition 2.1.2 and the estimate (3.2.1a) (which holds on \([0, T_\ast] \times \mathbb{R}^3\) by virtue of the fact that \(\Psi \in C ([0, T_\ast], H^5(\mathbb{R}^3)) \subset C ([0, T_\ast], C^3(\mathbb{R}^3))\)) yield that \(\Sigma \) is non-empty. Moreover, from (3.2.2) and the fact that \(\partial_t \Psi \in C ([0, T_\ast], H^{4,3}(\mathbb{R}^3)) \subset C ([0, T_\ast], C^3(\mathbb{R}^3))\), we find that the estimate (4.0.14) holds on \([0, T_\ast] \times \mathbb{R}^3\). Hence, in the case \(P = 1\), if \(q \in \Sigma_{\text{Degeneracy}}\), then any \(C^1\) extension of \(\Psi\) to a neighborhood of \(q\) contains points \(p\) such that \(1 + \Psi(p) < 0\), which renders equation (1.1.1a) elliptic. This is in contrast to the case \(P = 2\) in the sense that equation (1.1.1a) is hyperbolic for all values of \(\Psi\).

Theorem 4.1 yields that \(\Psi\) remains regular, all the way up to the time \(T_\ast\). However, as the next proposition shows, a type of invariant blowup does in fact occur at time \(T_\ast\) in both the cases \(P = 1, 2\). The blowup is tied to the Riemann curvature of the metric \(g\).

**Proposition 4.1 (Blowup of the Kretschmann scalar).** Let \(g = g(\Psi)\) denote the spacetime metric defined in (1.1.2) and let \(\text{Riem}(g)\) denote the Riemann curvature tensor\(^{28}\) of \(g\). Under the assumptions and conclusions of Theorem 4.1, we have the following estimate for the Kretschmann scalar \(\text{Riem}(g)^{\alpha\beta\gamma\delta}\text{Riem}(g)_{\alpha\beta\gamma\delta}\) on \([0, T_\ast] \times \mathbb{R}^3\):

\[
\text{Riem}(g)^{\alpha\beta\gamma\delta}\text{Riem}(g)_{\alpha\beta\gamma\delta} = \frac{15}{2} \frac{\partial_t (\Psi)^4}{(1 + \Psi)^4} + O \left( \frac{\dot{\epsilon}}{(1 + \Psi)^3} \right), \quad (P = 1), \tag{4.0.15a}
\]

\[
\text{Riem}(g)^{\alpha\beta\gamma\delta}\text{Riem}(g)_{\alpha\beta\gamma\delta} = 60 \frac{\partial_t (\Psi)^4}{(1 + \Psi)^4} + O \left( \frac{\dot{\epsilon}}{(1 + \Psi)^3} \right), \quad (P = 2). \tag{4.0.15b}
\]

In particular, \(\text{Riem}(g)^{\alpha\beta\gamma\delta}\text{Riem}(g)_{\alpha\beta\gamma\delta}\) is smooth for \(0 \leq t < T_\ast\), while by (3.2.2) and (4.0.15a) - (4.0.15b), at time \(T_\ast\), \(\text{Riem}(g)^{\alpha\beta\gamma\delta}\text{Riem}(g)_{\alpha\beta\gamma\delta}\) blows up precisely on the subset \(\Sigma_{\text{Breakdown}}\) defined in (4.0.13).

**Proof.** We prove only (4.0.15a) since the proof of (4.0.15b) is similar. The identities in this proof rely on the form of the metric (1.1.2). We first note the following simple decomposition formula, which relies on the standard symmetry and anti-symmetry properties of the Riemann curvature tensor:

\[
\text{Riem}(g)^{\alpha\beta\gamma\delta}\text{Riem}(g)_{\alpha\beta\gamma\delta} = \text{Riem}(g)^{\alpha\beta}_{cd}\text{Riem}(g)_{\alpha\beta}^{cd} + 4\text{Riem}(g)^{\alpha\beta}_{00}\text{Riem}(g)_{00}^{\alpha\beta} - 4g_{cc}g_{dd}g^{bl}\text{Riem}(g)_{bb}^{cd}\text{Riem}(g)_{dd}^{cd}. \tag{4.0.16}
\]

Next, we let \(g\) denote the first fundamental form of \(\Sigma_t\) relative to \(g\), that is, \(g_{ij} = g_{ij} = (1 + \Psi)^{-P}\delta_{ij}\) for \(i, j = 1, 2, 3\), where \(\delta_{ij}\) denotes the standard Kronecker delta. We also let

\[
k^i_j := - (g^{-1})^{ia} \left( \frac{1}{2} \mathcal{L}_{\partial_t} g_{aj} \right) = \frac{1}{2} \{ \partial_t \ln (1 + \Psi) \} \delta^i_j \tag{4.0.17}
\]

denote the (type \([1, 1]\)) second fundamental form of \(\Sigma_t\) relative to \(g\), where \(\mathcal{L}_{\partial_t}\) denotes Lie differentiation with respect to the vectorfield \(\partial_t\) and \(\delta^i_j\) denotes the standard Kronecker delta. Standard calculations based on the Gauss and Codazzi equations for the Lorentzian

---

\(^{28}\)Our sign convention for curvature is \(D_{\alpha}D_{\beta}X_{\mu} - D_{\beta}D_{\alpha}X_{\mu} = \text{Riem}(g)_{\alpha\beta\mu\nu}X^{\nu}\), where \(D\) denotes the Levi-Civita connection of \(g\) and \(X\) is an arbitrary vectorfield.
Finite-Time Degeneration of Hyperbolicity

manifold \((\mathbb{R}^{1+3}, g)\) yield (see, for example, [83, Appendix A]) the identities

\[
\begin{align*}
\text{Riem}(g)_{ab}^{\ cd} &= k^c_a k^d_b - k^d_a k^c_b + \Delta_{ab}^{\ cd}, \\
\text{Riem}(g)_{a0}^{\ cd} &= (\partial_t \ln(1 + \Psi)) k^c_a + k^c_e k^e_a + \Delta_{a0}^{\ cd}, \\
\text{Riem}(g)_{0b}^{\ cd} &= \Delta_{0b}^{\ cd},
\end{align*}
\]

where the error terms are

\[
\begin{align*}
\Delta_{ab}^{\ cd} &:= \text{Riem}(g)_{ab}^{\ cd}, \\
\Delta_{a0}^{\ cd} &:= -\frac{1}{1 + \Psi} \partial_t ((1 + \Psi) k^c_a), \\
\Delta_{0b}^{\ cd} &:= (g^{-1})^{ce} \partial_e (k^e_b) - (g^{-1})^{de} \partial_e (k^e_b) \\
&\quad + (g^{-1})^{ce} \Gamma^{d}_{\ f} k^f_b - (g^{-1})^{ce} \Gamma^{d}_{\ f} k^f_e - (g^{-1})^{de} \Gamma^{c}_{\ f} k^f_b + (g^{-1})^{de} \Gamma^{c}_{\ f} k^f_e,
\end{align*}
\]

and

\[
\Gamma^{i}_{\ jk} := \frac{1}{2} (g^{-1})^{ai} \left\{ \partial_j g_{ak} + \partial_k g_{ja} - \partial_a g_{jk} \right\}
\]

are the Christoffel symbols\(^{29}\) of \(g\). In (4.0.21), \(\text{Riem}(g)\) denotes the Riemann curvature tensor of \(g\). We note that in deriving (4.0.19) and (4.0.22), we have used the simple identity

\[-\partial_t (k^c_a) = (\partial_t \ln(1 + \Psi)) k^c_a - \frac{1}{1 + \Psi} \partial_t ((1 + \Psi) k^c_a).
\]

We will use the estimates of Theorem 4.1 to show that

\[
\begin{align*}
\Delta_{ab}^{\ cd} &= \mathcal{O}(\delta) \frac{1}{1 + \Psi}, \\
\Delta_{a0}^{\ cd} &= \mathcal{O}(\delta), \\
\Delta_{0b}^{\ cd} &= \mathcal{O}(\delta) \frac{1}{1 + \Psi}.
\end{align*}
\]

The desired bound (4.0.15a) then follows from (4.0.16), (4.0.17), (4.0.18), (4.0.19), (4.0.20), (4.0.21)-(4.0.27), and straightforward calculations.

It remains for us to prove (4.0.25)-(4.0.27). To prove (4.0.25), we first use (4.0.24) and (3.3.5a)-(3.3.5c) to deduce

\[
\Gamma^{i}_{\ jk} = \mathcal{O}(\delta) \frac{1}{1 + \Psi}, \quad \partial_t \Gamma^{i}_{\ jk} = \mathcal{O}(\delta) \frac{1}{(1 + \Psi)^2}.
\]

Since \(\text{Riem}(g)_{ab}^{\ cd}\) has the schematic structure \(\text{Riem}(g)_{ab}^{\ cd} = g^{-1} \partial T + g^{-1} \Gamma \cdot \Gamma\) (where \(\partial T\) denotes the gradient with respect to the spatial coordinates), we deduce from (4.0.28) and the simple estimate \((g^{-1})^{ij} = \mathcal{O}(1)(1 + \Psi)\) that \(\text{Riem}(g)_{ab}^{\ cd} = \mathcal{O}(\delta) \frac{1}{1 + \Psi}\), which yields (4.0.25).

To prove (4.0.26), we first use equation (4.0.17), equation (1.1.1a), and the estimates (3.3.5a) and (3.3.5c) to deduce \(\partial_t ((1 + \Psi) k^c_a) = \frac{1}{2} \partial_t^2 \Psi \delta^c_a = \frac{1}{2} (1 + \Psi) \Delta \Psi \delta^c_a = \mathcal{O}(\delta)(1 + \Psi)\). From this

\(^{29}\) Our index conventions for the Christoffel symbols are different than the ones used in many works on differential geometry.
bound and (4.0.22), we conclude (4.0.26). To prove (4.0.27), we first use equation (4.0.17) and the estimates (3.3.5a)-(3.3.5c) to deduce
\[ k^i_j = O(1) \frac{1}{1 + \Psi^i}, \quad \partial_t k^i_j = O(\epsilon) \frac{1}{(1 + \Psi)^2}. \] (4.0.29)
From (4.0.28), (4.0.29), and the simple estimate \( (g^{-1})^{ij} = O(1)(1 + \Psi) \) we conclude (4.0.27). This completes the proof of the proposition.

\[\square\]

Acknowledgments

JS would like to thank Yuuskue Sugiyama for bringing the references \([53, 94–96]\) to his attention and Michael Dreher for pointing out the references \([69, 70]\).

References

[18] Ovidiu Costin, Min Huang, and Wilhelm Schlag, *On the spectral properties of $L^\pm$ in three dimensions*, Nonlinearity 25 (2012), no. 1, 125–164. MR2864380


[38] Qing Han, Jia-Xing Hong, and Chang-Shou Lin, *Local isometric embedding of surfaces with nonpositive gaussian curvature*, J. Differential Geom. 63 (200301), no. 3, 475–520.


Finite-Time Degeneration of Hyperbolicity


[83] Igor Rodnianski and Jacob Sterbenz, On the formation of singularities in the critical $O(3)$-$\sigma$-model, Ann. of Math. (2) 172 (2010), no. 1, 187–242. MR2680419


[95] Yuusuke Sugiyama, Global existence of solutions to some quasilinear wave equation in one space dimension, Differential Integral Equations 26 (201305), no. 5/6, 487–504.

