A PHYSICAL SPACE APPROACH TO WAVE EQUATION BILINEAR ESTIMATES

SERGIU KLAINERMAN, IGOR RODNIANSKI, AND TERENCE TAO

Dedicated to the memory of Tom Wolff

ABSTRACT. Bilinear estimates for the wave equation in Minkowski space are normally proven using the Fourier transform and Plancherel's theorem. However, such methods are difficult to carry over to non-flat situations (such as wave equations with rough metrics, or connections with non-zero curvature). In this note we describe an alternative physical space approach which relies on vector fields, energy estimates as well as tube localization, splitting into coarse and fine scales, and induction on scales (in the spirit of Wolff [29], [30]).

1. INTRODUCTION

In this paper $n \geq 2$ is an integer, and all implicit constants may depend on n.

We shall call a *free wave* any (possibly vector-valued) solution[1] ϕ to the free wave equation

$$\Box \phi(x, t) = (-\frac{\partial^2}{\partial t^2} + \Delta)\phi(x, t) = 0$$

in $\mathbf{R}^{n+1}$. We write $\phi[t] := (\phi(t), \phi_t(t))$ for the data of ϕ at t. This space is equipped with the energy inner product

$$\langle \phi[t], \psi[t] \rangle_e := \frac{1}{2} \int_{\mathbf{R}^n} \nabla \phi(x, t) \cdot \nabla \overline{\psi}(x, t) + \phi_t(x, t) \overline{\psi}_t(x, t) \ dx,$$

and we write $E(\phi[t]) := \langle \phi[t], \phi[t] \rangle_e$. If ϕ, ψ are free waves, then these quantities are time-independent and we shall simply write them as $\langle \phi, \psi \rangle_e$ and $E(\phi)$ respectively.

If ϕ is a free wave and λ is a dyadic number (i.e. an integer power of two), we say that ϕ has frequency λ if the spatial Fourier transform $\hat{\phi}(\xi, t)$ is supported in the region[2] $|\xi| \sim \lambda$; note that this property is independent[3] of t.

1991 *Mathematics Subject Classification.* 35J10.

[1] To avoid technicalities we shall assume that our waves are Schwartz functions in space.

[2] Here and in the sequel, we use C to denote various constants depending only on n and a parameter ε which will appear later in the paper, and we use $A \lesssim B$ to denote the estimate $A \leq CB$. Finally, we use $A \sim B$ to denote the estimate $A \lesssim B \lesssim A$.

[3] The time independence property is invalid in a curved spacetime. There are however techniques, based on paradifferential calculus, which allow one to circumvent this difficulty, see [12] and [15].

Let λ be any dyadic number. We use the symbol P_λ to denote any Littlewood-Paley projection operators of the form

$$\widehat{P_\lambda \phi}(\xi, t) := \eta(\xi/\lambda)\hat{\phi}(\xi, t)$$

where η is any bump function adapted to the annulus $|\xi| \sim 1$; the exact choice of η may vary from line to line. Note that if ϕ is a free wave, then $P_\lambda \phi$ is a free wave with frequency λ. Also, if ϕ has frequency λ, then we have the reproducing formula

$$\nabla_x^b \phi = \lambda^{b-a} P_\lambda \nabla_x^a \phi \qquad (1)$$

for any $a, b \geq 0$ and some suitable P_λ (possibly tensor-valued). Here ∇_x^a refers to fractional spatial derivatives. By computing the convolution kernel of P_λ we thus obtain the pointwise estimates

$$|\nabla_x^b \phi(x_0)| \lesssim \lambda^{b-a} \int |\nabla_x^a \phi(x_0 + y/\lambda)|(1 + |y|)^{-M} \, dy \qquad (2)$$

for any $M \geq 0$, with the implicit constant depending on a, b, M. (Similar estimates obtain for spacetime derivatives $\nabla_{x,t}$ by using the equation $\Box \phi = 0$). In particular we have the Sobolev inequality

$$\|\phi\|_\infty \lesssim \lambda^{n/2-1} E(\phi)^{1/2} \qquad (3)$$

whenever ϕ is a free wave of frequency λ. In practice, when we wish to estimate a bilinear expression $Q(\phi, \psi)$ in L^2, we shall put the lower frequency wave in L^∞ (using (3) or more sophisticated variants), and the higher frequency wave in L^2 (using energy estimates, possibly involving vector fields).

In the study of non-linear wave equations it has been realized that bilinear estimates on free waves play a fundamental role. Two sample estimates are given below; for a more thorough discussion of such estimates see [10], [8], [24], [18], [16].

Theorem 1.1. *Let λ, μ be dyadic numbers, and let $\varepsilon > 0$. Then we have the estimate*

$$\|Q(\phi, \psi)\|_{L^2(\mathbf{Q}_R)} \lesssim R^\varepsilon \min(\lambda, \mu)^{\frac{n-1}{2}+\varepsilon} E(\phi)^{1/2} E(\psi)^{1/2} \quad \text{when} \ \ R\min(\lambda, \mu) \gtrsim 1$$

$$R^{1/2} \min(\lambda, \mu)^{\frac{n}{2}} E(\phi)^{1/2} E(\psi)^{1/2} \quad \text{when} \ \ R\min(\lambda, \mu) \lesssim 1 \quad (4)$$

whenever $\mathbf{Q}_R$ is a spacetime cube of some side-length $R > 0$, ϕ and ψ are free waves with frequency λ and μ respectively, and Q is one of the null forms

$$\begin{aligned} Q_0(\phi, \psi) : &= \phi_t \psi_t - \nabla\phi \cdot \nabla\psi, \\ Q_{\alpha\beta}(\phi, \psi) : &= \partial_\alpha \phi \partial_\beta \psi - \partial_\beta \phi \cdot \partial_\alpha \psi. \end{aligned}$$

Theorem 1.2. *Let μ, λ be dyadic numbers, and let ϕ, ψ be free waves of frequency λ. Then we have*

$$\|P_\mu(\phi\psi)\|_{L_t^{q/2} L_x^{r/2}(\mathbf{R}^{n+1})} \lesssim (\frac{\mu}{\lambda})^{n-\frac{4}{q}-\frac{2n}{r}} \lambda^{n-2-\frac{2}{q}-\frac{2n}{r}} E(\phi)^{1/2} E(\psi)^{1/2}$$

whenever (q, r) are an admissible pair of Strichartz exponents in the sense that

$$2 \leq q, r \leq \infty; \quad \frac{1}{q} + \frac{(n-1)/2}{r} \leq \frac{(n-1)/2}{2}; \quad (n, q, r) \neq (3, 2, \infty).$$

Apart from the ε loss in Theorem 1.1, all the exponents are sharp [8]. The ε can be removed; we will discuss this issue later in the paper.

These two theorems apply only to frequency-localized free waves, but one can use Littlewood-Paley decomposition and orthogonality in a standard manner to extend these theorems to arbitrary free waves. More precisely, we have

Corollary 1.3 ("First generation" null form estimate, see [10]). *If Q is a null form and ϕ, ψ are free waves, then*

$$\|Q(\phi, \psi)\|_{L^2_{t,x}([0,1] \times \mathbf{R}^n)} \lesssim E(\langle D \rangle^{\frac{n-1}{2}+\varepsilon} \phi)^{1/2} E(\psi)^{1/2}$$

for any $\varepsilon > 0$, where $\langle D \rangle$ is the Fourier multiplier with symbol $(1 + |\xi|^2)^{1/2}$.

Corollary 1.4 (Bilinear improvement to Strichartz' inequality). *Let ϕ, ψ be free waves, and let (q, r) be an admissible pair of Strichartz exponents. Then we have*

$$\||D|^{-\sigma}(\phi \psi)\|_{L^{q/2}_t L^{r/2}_x(\mathbf{R}^{n+1})} \lesssim E(|D|^{\frac{n-\sigma}{2}-\frac{1}{q}-\frac{n}{r}}\phi)^{1/2} E(|D|^{\frac{n-\sigma}{2}-\frac{1}{q}-\frac{n}{r}}\psi)^{1/2}$$

whenever $\sigma < n - \frac{4}{q} - \frac{2n}{r}$, and $|D|$ is the Fourier multiplier with symbol $|\xi|$.

The derivation of these corollaries from the Theorems is given in the Appendix. We remark that Corollary 1.3 was proven in [10] (with the sharp endpoint $\varepsilon = 0$) by methods heavily reliant on the Fourier transform. Corollary 1.4 was similarly proven using Fourier transform methods in [17], see also [11].

The purpose of this note is to give different proofs of Theorems 1.1, 1.2 and their corollaries which rely on physical space methods instead of the Fourier transform. As such, we believe these arguments will be useful for such situations as quasilinear wave equations with rough data, in which the Fourier transform is difficult to use directly. The techniques used here extend to many other types of bilinear estimates, but we shall restrict ourselves to the above two types of estimates for simplicity.

We briefly outline the ideas used in our arguments:

- **Splitting physical space into coarse scales and fine scales.** If a large cube $\mathbf{Q}_R$ is partitioned into disjoint[4] sub-cubes $\mathbf{Q}_r$, we can split

$$\|\phi\|_{L^p(\mathbf{Q}_R)} = \Big(\sum_{\mathbf{Q}_r : \mathbf{Q}_r \subset \mathbf{Q}_R} \|\phi\|^p_{L^p(\mathbf{Q}_r)} \Big)^{1/p}. \tag{5}$$

  The inner $L^p(\mathbf{Q}_r)$ norm represents the fine scales, while the outer l^p norm represents the coarse scales. The strategy is then to choose r carefully and handle the coarse and fine scales in different ways. This should be compared with Fourier-based methods, which often employ a similar decomposition but in the frequency variable ([17], [24], etc.).
- **Decomposition into wave packets.** In a spacetime cube $\mathbf{Q}_R$, one can split a free wave ϕ into almost orthogonal "wave packets" ϕ_T, each of which are also free waves and are essentially localized in a tube T, which is a suitable neighbourhood of a light ray. The traditional decomposition into wave

---

[4]We always ignore sets of measure zero when we say that two sets are disjoint.

packets, which has its origin in the so called second dyadic decomposition[5] pioneered by Fefferman [7] and Córdoba [6] in connection to the L^p multiplier problem, relies heavily on the Fourier transform.

In our paper we suggest a different construction of wave packets based on a "dynamical" physical space decomposition which avoids Fourier space decompositions and the related method of "stationary phase". This is achieved by localizing our waves at two distinct times. One can use the vectorfield method, instead of stationary phase, to show that the resulting waves concentrate along tubes.

- **Distinguishing between transverse and parallel interactions.** Once ϕ and ψ are decomposed into wave packets, one can then separate into two cases depending on the angle between the two packets. When the angle is small, we say we have a *parallel interaction*, and we expect the cancellation in the null structure to play a role. In the large angle case we have a *transverse interaction*, which we expect to be well localized in physical space. In the original approach to the proof of bilinear estimates the two cases[6] were both treated with the help of the spacetime Fourier transform. In some cases one can use Lorentz transforms to reduce the parallel case to the transverse case: see [27], [26], [29], [30] for further discussion.

  The new twist in our paper is the treatment of parallel interactions, for which the vectorfield method turns out to be very well suited. Together with the dynamic tube localization and the treatment of transversal interactions, based on induction on scales initiated in [3], [27], [26], [29], [30], this allows us to give an almost[7] completely geometric treatment of the first bilinear estimate, see Theorem 1.1

- **Induction on scales (or bootstrap arguments).** To prove an estimate at a large spatial scale R, assume inductively that an estimate has already been obtained at smaller scales r. This idea has been around for some time (see e.g. [3], [27], [25]), but only in the recent work of Wolff [29], [30] was it realized that such techniques can give very sharp estimates.

- **Improved decay estimates.** The original proof of Theorem 1.2 was based on an improved $L^1 - L^\infty$ decay estimate which takes advantage of the size of the Fourier support of the data in a small rectangle in Fourier space.

  Here we take a different approach, first developed by [22], based on a similar phenomenon in physical space. This hinges on the observation that "bump waves", i.e. solution of the free wave equation with initial data localized at frequency one and decreasing rapidly away from the unit ball centered at the origin, are essentially supported on a set of codimension one. In flat space this can be easily seen from the explicit representation of bump waves as Fourier integrals. Due to this fact one can improve the L^∞ decay of bump waves by averaging over a relatively coarse scale in physical space. Here we show that this improved decay property can be derived by vectorfield techniques in the spirit of [12]. Thus, by avoiding the explicit representation of solutions, we

---

[5]These were developed further by Bourgain [3] and later authors ([27], [25], [29], [23], [30], etc.)

[6]or more precisely their Fourier space counterparts.

[7]With the exception of the dyadic decomposition.

give an essentially geometric derivation of Theorem 1.2. This opens the way, we believe, for significant generalizations to quasilinear equations[8].

By frequency-localized fundamental solution we mean a free wave of frequency 1 whose initial data is rapidly decreasing away from the origin. It is well known that the L^∞ decay of such a solution in time can be combined with energy estimates to yield Strichartz estimates. However, there are other properties of the solution which are also useful. For instance, the fundamental solution is essentially supported on a set of codimension 1, which means that one can improve the L^∞ decay of the (absolute value of) the fundamental solution by averaging over a relatively coarse scale in physical space. This fact (first used in [22]) can be used to control the low-frequency portion of a bilinear or nonlinear expression. We remark that vector fields methods are well suited for obtaining this type of control on the fundamental solution, even in rough metrics (see e.g. [12]).

In the proof of Theorem 1.1 we shall use the first four ideas, while in the proof of Theorem 1.2 we shall use the first and fifth ideas. To realize these ideas we shall use energy estimates and commutation with special vectorfields in the spirit of [12], see also [13], [14]. It is possible to use other methods (e.g. Fourier integral parametrices) as a substitute for vector field methods, but we believe the vector field approach is better adapted for more non-linear situations (e.g. quasilinear wave equations) in which the Fourier transform is less amenable. This paper is organized as follows. We prove Theorem 1.1 in Sections 2 to 7. In Section 2 we set up the basic machinery of tubes or wave packets, and in Section 3 we combine this with an induction on scales argument to reduce to the parallel interaction case. After some vector fields preliminaries in Section 4 we handle the parallel interaction case in Section 5. Finally, we give two proofs of the key tube localization lemma (one a hybrid of spatial and Fourier localization methods, and the other based entirely on spatial localization) in Sections 6 and 7. After some remarks in Section 8 we then prove Theorem 1.2 in Section 9.

2. Tube localization

A fundamental tool in our proof of Theorem 1.1 is the ability to decompose a free wave into "wave packets" which are concentrated in tubes along null rays. To make this precise we need some notation. We shall be working in the standard spacetime cube

$$\mathbf{Q}_R^{(0)} := \{(x,t) : 0 \le x_i \le R \text{ for all } i = 1, \dots, n; \quad R \le t \le 2R\} \tag{6}$$

---

[8]The recent result of [1] is based on a generalization of the bilinear estimates in Theorem 1.2 to wave equations with rough coefficients. The construction is based on Fourier Integral operators.

for some $R \gtrsim 1$. Note that this cube is some distance away from the initial spatial hypersurface $t = 0$; this will be very convenient when we apply vector field techniques.

We shall also need a small constant $0 < \varepsilon \ll 1$; all our implicit constants may depend on ε.

Definition 2.1. Let $1 \lesssim r \lesssim R$ be scales, let $\omega \in S^{n-1}$ be a direction, and let $x_0 \in \mathbf{R}^n$ be a point in space. We call the set

$$T = T(R, r, \omega, x_0) := \{(x, t) : |x - t\omega - x_0| \leq r; 0 \leq t \leq 3R\}$$

the *tube* with *length R*, *width r*, *velocity ω*, and *initial position x_0*.

We now define the notion of a wave packet localized to a tube. We shall normalize our definition in the situation when the original waves ϕ, ψ are normalized to have energy 1.

Definition 2.2. Let T be a tube and ϕ_T be a free wave with frequency $\mu \gtrsim 1$ and energy $E(\phi_T) \lesssim 1$. We say that ϕ_T is a *wave packet localized to T* if one has the estimates

$$|\nabla_{x,t}\phi_T(x, t)| \lesssim (R + \text{dist}((x, t), T))^{-100n}\mu^{n/2} \tag{7}$$

whenever $R/2 \leq t \leq 5R/2$ and $(x, t) \notin T$, as well as the L^2 variant

$$(\int_{x:|x|\lesssim R,(x,t)\notin T} |\nabla_{x,t}\phi_T|^2 \, dx)^{1/2} \lesssim (R + \text{dist}((0, t), T))^{-100n} \tag{8}$$

for all $R/2 \leq t \leq 5R/2$.

The estimate (7) essentially follows from (8) by (2), perhaps at the cost of replacing T by a tube of twice the radius. (The error term coming from inside the tube can be controlled by (3)). In practice we shall use (7) for low-frequency waves and (8) for high-frequency waves.

The main tube decomposition lemma can now be stated as follows.

Lemma 2.3 (Tube decomposition lemma). *Let ϕ be a free wave with energy $E(\phi) = 1$ and frequency $\mu \gtrsim 1$, and let $R \gtrsim 1$ and $R^{1/2+\varepsilon} \lesssim r \lesssim R$. Then there exists a collection $\mathbf{T}$ of tubes of length R and width r, and to each tube T in $\mathbf{T}$ there exists a wave packet ϕ_T localized to T with frequency μ, such that we have the decomposition*

$$\phi = \sum_{T \in \mathbf{T}} \phi_T + \phi_{error} \tag{9}$$

where the error term is a free wave which satisfies $E(\phi_{error}) \ll E(\phi)$. Also, we have the Bessel inequality

$$\sum_{T \in \mathbf{T}} E(\phi_T) \lesssim \tag{10}$$

and the almost orthogonality relationship

$$E(\sum_{T \in \mathbf{T}'} \phi_T) \lesssim R^{C\varepsilon} \sum_{T \in \mathbf{T}'} E(\phi_T) + R^{-100n} \tag{11}$$

for any subset $\mathbf{T}'$ of $\mathbf{T}$. Finally, to control error terms we require the technical property that given any cube $\mathbf{Q}_R$, the number of tubes in $\mathbf{T}$ which intersect $\mathbf{Q}_R$ is $O(R^{10n})$.

At a heuristic level one should imagine that $\phi_{error} = 0$, and that the ϕ_T are an orthogonal decomposition of ϕ.

This type of lemma has been used by many authors. For the closely related restriction problem for the Fourier transform, these wave packets first appear in the work of Fefferman, Knapp, and Córdoba, and were developed substantially further by Bourgain [3], [4], Wolff [29], [30], and others (e.g. [23], [28]). In these papers the tube decomposition is achieved by a decomposition in both physical and frequency space. We shall give two proofs of this lemma in Sections 6, 7. The first proof is similar to the previous arguments but only uses the Fourier transform at the level of the initial data rather than in spacetime, and uses commutation with vector fields to control the localization. The second proof is based purely on physical space methods, and achieves localization to a tube by localizing the wave to a disk at two different times (see somewhat related "slices" approach to the Kakeya problem in e.g. [5]).

In the high frequency case $\mu \gg 1$ the condition $r \gtrsim R^{1/2+\varepsilon}$ can be relaxed to $r \gtrsim R^{1/2+\varepsilon}\mu^{-1/2+\varepsilon}$, but we are unable to exploit this for our bilinear estimates because our argument needs ϕ and ψ to be decomposed into tubes of equal width.

Assume Lemma 2.3 for the moment. In the next section we see how this lemma can be combined with an induction on scales argument to reduce the bilinear estimate (4) to the "parallel interaction case". In the section after that we show how commutation with vector fields can be used to treat these parallel interactions, thus completing the proof of (4).

3. INDUCTION ON SCALES

We now begin the proof of Theorem 1.1. The idea is to use Lemma 2.3 and an induction on scales (or bootstrap) argument to reduce (4) to the parallel interaction case.

We first give some preliminary reductions. Fix λ and μ; without loss of generality we may assume that $\lambda \leq \mu$. By rescaling we may take $\lambda = 1$.

The case $R \lesssim 1$ is easy to dispose of by Hölder and the Sobolev inequality (3):

$$\|Q(\phi,\psi)\|_{L^2(\mathbf{Q}_R)} \lesssim R\|\nabla_{x,t}\phi\|_{L_t^\infty L_x^\infty}\|\nabla_{x,t}\psi\|_{L_t^\infty L_x^2} \lesssim R^{\frac{1}{2}}E(\phi)^{1/2}E(\psi)^{1/2}. \tag{12}$$

Thus we may take $R \gg 1$.

We now set up the bootstrap argument. For any $\alpha > 0$, let $B(\alpha)$ denote the statement that the bilinear estimate

$$\|Q(\phi, \psi)\|_{L^2(\mathbf{Q}_R)} \lesssim R^\alpha E(\phi)^{1/2} E(\psi)^{1/2} \qquad (B_\alpha)$$

holds for all free waves ϕ, ψ, all $R \gg 1$ with frequency 1 and μ respectively, and all spacetime cubes $\mathbf{Q}_R$; the implicit constant can depend on α and the implicit constants used to define such notions as "ψ has frequency μ". Our objective is to prove $B(\alpha)$ holds for all $\alpha > 0$.

From (12) we have the estimate $B(\frac{1}{2})$. The claim will now follow from bootstrapping the following Proposition.

Proposition 3.1. *If $B(\alpha)$ holds for some $\alpha > 0$, then $B((1-c)\alpha + C\varepsilon)$ holds for all $\varepsilon > 0$, where $c > 0$ is an absolute constant depending only on n.*

Indeed, from Proposition 3.1 it follows that the infimum of all $\alpha > 0$ for which $B(\alpha)$ holds must be zero.

It remains to prove the Proposition. Fix $\alpha > 0$, $\varepsilon > 0$, and assume that $B(\alpha)$ already holds. Fix free waves ϕ, ψ and a scale $R \gg 1$, and a cube $\mathbf{Q}_R$. We may assume that we have the normalization $E(\phi) = E(\psi) = 1$. We shall take advantage of spacetime translation invariance to set $\mathbf{Q}_R = \mathbf{Q}_R^{(0)}$, where $\mathbf{Q}_R^{(0)}$ was defined in (6); note in particular that the cube $\mathbf{Q}_R$ is a distance R away from the initial spatial hypersurface $t = 0$. Our objective is to prove that

$$\|Q(\phi, \psi)\|_{L^2(\mathbf{Q}_R^{(0)})} \lesssim R^{(1-c)\alpha + C\varepsilon}. \qquad (13)$$

If one applies the hypothesis $B(\alpha)$ directly to this cube $\mathbf{Q}_R^{(0)}$ one only gets a bound of R^α. To improve upon this we shall only apply $B(\alpha)$ at smaller scales than R, and use transversality to sum efficiently (in the spirit of [29]).

We let $R^{1/2+\varepsilon} \ll r \ll R$ be an intermediate dyadic scale between $\sqrt{R}$ and R to be chosen later; r will eventually be fairly close to $\sqrt{R}$. We now use Lemma 2.3 to write

$$\phi = \sum_{T \in \mathbf{T}} \phi_T + \phi_{error}; \quad \psi = \sum_{T' \in \mathbf{T}'} \psi_{T'} + \psi_{error}$$

where $\mathbf{T}$, $\mathbf{T}'$ are collections of tubes with length R and width r, and for each $T \in \mathbf{T}$, ϕ_T is a wave packet localized to T obeying (10), (11). Similarly for the $\psi_{T'}$.

The ϕ_{error}, ψ_{error} are small in energy and their contribution can be iterated away by the usual argument[9]. We can therefore split the remaining portions of the null form $Q(\phi, \psi)$ into a sum over interactions between wave packets:

$$\sum_{T \in \mathbf{T}} \sum_{T' \in \mathbf{T}'} Q(\phi_T, \psi_{T'}). \qquad (14)$$

---

[9]Specifically, one assumes that $B(\alpha)$ holds with an explicit constant A, possibly depending on R, and use the small energy of ϕ_{error} and ψ_{error} to bound the contribution of these terms by at most $A/10$. At the end one can absorb the $A/10$ into the left hand side.

By (7) and Hölder one may easily dispose of those tubes T for which $\operatorname{dist}(T, \mathbf{Q}_R^{(0)}) \gtrsim$ R. Similarly for T'. Thus we shall assume in the sequel that $\operatorname{dist}(T, \mathbf{Q}_R^{(0)}), \operatorname{dist}(T', \mathbf{Q}_R^{(0)}) \lesssim$ R for all $T \in \mathbf{T}$ and $T' \in \mathbf{T}'$. In particular the cardinality of $\mathbf{T}$ and $\mathbf{T}'$ is only $O(R^{10n})$.

We shall need an intermediate separation scale $r \ll \rho_0 \ll R$ to be chosen later; eventually ρ_0 will be close to r. We will assume that ρ_0 is a dyadic number.

Informally, the strategy will be as follows. We wish to divide the above sum into *parallel interactions* $|x_0(T) - x_0(T')| \leq \rho_0$ and *transverse interactions* $|x_0(T) - x_0(T')| \geq \rho_0$, where $x_0(T)$ denotes the initial position of T. The parallel interactions should be small thanks to the cancellation in the null form, and we will use vector field methods to exploit this. For the transverse interactions, we do not exploit the null structure, but instead we use transversality to ensure that $T \cap T'$ is localized to a cube of length $\ll R$, so that when we apply the hypothesis $B(\alpha)$ we shall gain in the α exponent. Of course we still have to sum over all tubes T, T', but it turns out that we have enough orthogonality in (10), (11) to achieve this.

We now turn to the details. For technical reasons we have to take a little more care in dividing the wave packet interactions into parallel and transverse interactions in order for the transverse interactions to decouple nicely (cf. [27]). Specifically, we shall use a bilinear Whitney decomposition of the initial data.

The initial positions x_0 of the tubes in $\mathbf{T}$, $\mathbf{T}'$ can all be contained in a spatial cube K of dyadic side-length CR. Let $\mathbf{K}$ denote the set of all dyadic sub-cubes κ of K whose side-length $l(\kappa)$ is at least $\rho_0/2$. Without loss of generality we may assume that the initial positions x_0 of tubes in $\mathbf{T}$, $\mathbf{T}'$ never lie on the boundary $\partial \kappa$ of any cube $\kappa \in \mathbf{K}$.

If $\kappa, \kappa' \in \mathbf{K}$ are two such sub-cubes of the same side-length, we say that κ and κ' are *close* if κ is not adjacent to κ', but the parent of κ is adjacent to the parent of κ'. We write $\kappa \sim \kappa'$ to denote the statement that κ and κ' are close. Note that if $\kappa \sim \kappa'$, then $\operatorname{dist}(\kappa, \kappa')$ is comparable to the side-length of κ or κ'.

For each $\kappa \in \mathbf{K}$, define the sets

$$\mathbf{T}(\kappa) := \{T \in \mathbf{T} : x_0(T) \in \kappa\} \text{ and } \mathbf{T}'(\kappa) := \{T' \in \mathbf{T}' : x_0(T') \in \kappa\};$$

these represent the tubes emanating from the cube κ. A simple stopping time argument reveals that for any $T \in \mathbf{T}$ and $T' \in \mathbf{T}'$, exactly one of the following statements is true:

- (Transverse interaction) There exists a close pair of cubes $\kappa \sim \kappa'$ such that $T \in \mathbf{T}(\kappa)$ and $T' \in \mathbf{T}'(\kappa')$.
- (Parallel interaction) There exists cubes κ, κ' of side-length $l(\kappa) = \rho_0$ which are either equal or adjacent such that $T \in \mathbf{T}(\kappa)$ and $T' \in \mathbf{T}'(\kappa')$. (We shall write $\kappa \approx \kappa'$ in this case.

We can therefore split the null form (14) into the *transverse interaction component*

$$\sum_{\kappa,\kappa' \in \mathbf{K}:\kappa \sim \kappa'} Q(\phi_\kappa, \psi_{\kappa'}) \tag{15}$$

and the *parallel interaction component*

$$\sum_{\kappa,\kappa' \in \mathbf{K}:\kappa \approx \kappa', l(\kappa)=l(\kappa')=\rho_0} Q(\phi_\kappa, \psi_{\kappa'}) \tag{16}$$

where

$$\phi_\kappa := \sum_{T \in \mathbf{T}(\kappa)} \phi_T \text{ and } \psi_\kappa := \sum_{T' \in \mathbf{T}'(\kappa)} \psi_{T'}.$$

Informally, ϕ_κ represents the portion of ϕ emanating from κ, and similarly for ψ_κ.

Consider first the contribution of the transverse component (15). Observe that there are only $O(\log R)$ possible values for the side-length $l(\kappa)$. Since we are allowing ourselves to lose an R^ε in the final estimate (13), it suffices by the triangle inequality to show that

$$\| \sum_{\kappa,\kappa' \in \mathbf{K}:\kappa \sim \kappa'; l(\kappa)=l(\kappa')=\rho} Q(\phi_\kappa, \psi_{\kappa'}) \|_{L^2(\mathbf{Q}_R^{(0)})} \lesssim R^{\alpha-\sigma+C\varepsilon}$$

for some $\sigma > 0$ and for all $\rho_0 \leq \rho \leq 1$.

Fix ρ. We use the triangle inequality to pull the sum in κ, κ' outside the norm. Observe that for each κ there are only $O(1)$ cubes κ' which are close to κ, and vice versa. By Cauchy-Schwarz and Bessel's inequality (10) it thus suffices to show that

$$\|Q(\phi_\kappa, \psi_{\kappa'})\|_{L^2(\mathbf{Q}_R^{(0)})} \lesssim R^{\alpha-\sigma+C\varepsilon} (\sum_{T \in \mathbf{T}(\kappa)} E(\phi_T))^{1/2} (\sum_{T' \in \mathbf{T}'(\kappa')} E(\psi_{T'}))^{1/2} + R^{-50n} \tag{17}$$

for all close pairs of cubes κ, κ' of side-length ρ.

Fix κ, κ'. Observe that if two tubes T, T' in the above expression intersect, the angle between T and T' is comparable[10] to ρ/R. Since T, T' have width r and $\rho \geq \rho_0 \gg r$, we thus see by elementary geometry that $T \cap T'$ is contained in a cube of side-length $O(Rr/\rho) \ll R$. To exploit this we shall subdivide the cube $\mathbf{Q}_R^{(0)}$ into sub-cubes $\mathbf{Q}_{Rr/\rho}$ of side-length Rr/ρ. We can then split left-hand side of (17) as

$$(\sum_{\mathbf{Q}_{Rr/\rho}} \|Q(\phi_\kappa, \psi_{\kappa'})\|^2_{L^2(\mathbf{Q}_{Rr/\rho})})^{1/2}$$

(cf. (5)).

Consider the contribution of a single cube $\mathbf{Q}_{Rr/\rho}$. Intuitively we only expect the portion of ϕ, ψ which propagates along the backward light cone from $\mathbf{Q}_{Rr/\rho}$ to contribute to this portion. To make this intuition precise, we will now throw away the tubes T, T' which do not intersect $\mathbf{Q}_{Rr/\rho}$.

---

[10]This crucial fact, that separation in space implies separation in angle, relies on the choice of $\mathbf{Q}_R^{(0)}$ to be away from the initial spatial hypersurface $t = 0$.

Consider first the contribution of a single tube $T \in \mathbf{T}(\kappa)$ which does not intersect $\mathbf{Q}_{Rr/\rho}$. By Hölder, Bessel (10), and the decay (7) of ϕ_T outside of T, we have

$$\|Q(\phi_T, \psi_{\kappa'})\|_{L^2(\mathbf{Q}_{Rr/\rho})} \lesssim \left(\frac{Rr}{\rho}\right)^{\frac{1}{2}} \|\nabla_{x,t}\phi_T\|_{L_{x,t}^\infty} \|\nabla_{x,t}\psi_{\kappa'}\|_{L_t^\infty L_x^2}$$

$$\lesssim \left(\frac{Rr}{\rho}\right)^{\frac{1}{2}} R^{-100n} E(\psi_{\kappa'})$$

$$\lesssim R^{-99n}.$$

Summing this over all tubes T and all cubes $\mathbf{Q}_{Rr/\rho}$ we see that this contribution is acceptable. Thus we may replace ϕ_κ by $\sum_{T \in \mathbf{T}(\kappa):T \cap \mathbf{Q}_{Rr/\rho} \neq \emptyset} \phi_T$. Similarly for $\psi_{\kappa'}$ (but we continue to keep the low frequency in L_x^∞ and the high frequency in L_x^2).

We are now ready to apply the hypothesis $B(\alpha)$. Modulo acceptable errors, we thus have

$$\|Q(\phi_\kappa, \psi_{\kappa'})\|_{L^2(\mathbf{Q}_{Rr/\rho})} \lesssim \left(\frac{Rr}{\rho}\right)^\alpha E\left(\sum_{T \in \mathbf{T}(\kappa):T \cap \mathbf{Q}_{Rr/\rho} \neq \emptyset} \phi_T\right)^{1/2} E\left(\sum_{T' \in \mathbf{T}'(\kappa'):T' \cap \mathbf{Q}_{Rr/\rho} \neq \emptyset} \psi_{T'}\right)^{1/2}.$$

Applying (11) we obtain

$$\|Q(\phi_\kappa, \psi_{\kappa'})\|_{L^2(\mathbf{Q}_{Rr/\rho})} \lesssim \left(\frac{Rr}{\rho}\right)^\alpha R^{C\varepsilon}\left(\sum_{T \in \mathbf{T}(\kappa):T \cap \mathbf{Q}_{Rr/\rho} \neq \emptyset} \sum_{T' \in \mathbf{T}'(\kappa'):T' \cap \mathbf{Q}_{Rr/\rho} \neq \emptyset} E(\phi_T)E(\psi_{T'})\right)^{1/2}.$$

The expression inside the square root may be thought of as a measure of how much joint energy of ϕ and ψ can interact within the cube $\mathbf{Q}_{Rr/\rho}$. The point is that different cubes $\mathbf{Q}_{Rr/\rho}$ will require disjoint portions of the joint energy because of transversality, and this will allow us to sum. Indeed, if we insert the above bound into the left-hand side of (17), we obtain a total bound of

$$\left(\frac{Rr}{\rho}\right)^\alpha R^{C\varepsilon}\left(\sum_{\mathbf{Q}_{Rr/\rho}} \sum_{T \in \mathbf{T}(\kappa):T \cap \mathbf{Q}_{Rr/\rho} \neq \emptyset} \sum_{T' \in \mathbf{T}'(\kappa'):T' \cap \mathbf{Q}_{Rr/\rho} \neq \emptyset} E(\phi_T)E(\psi_{T'})\right)^{1/2}.$$

By transversality of κ and κ' and elementary geometry, we see that for each $T \in \mathbf{T}(\kappa)$, $T' \in \mathbf{T}'(\kappa)$ there are only $O(1)$ cubes $\mathbf{Q}_{Rr/\rho}$ for which the above summation is non-vacuous. Thus we can bound the previous by

$$(Rr/\rho)^\alpha R^{C\varepsilon}\left(\sum_{T \in \mathbf{T}(\kappa)} \sum_{T' \in \mathbf{T}'(\kappa')} E(\phi_T)E(\psi_{T'})\right)^{1/2}.$$

Since $\rho \geq \rho_0$, we see that this is acceptable if we set

$$\rho_0 := R^c r \tag{18}$$

for some small absolute constant $c > 0$ to be chosen later.

It remains to control the parallel interactions (16). By Bessel's inequality (10) and the same Cauchy-Schwarz trick used to reduce the transverse case to (17), we may reduce this portion of the estimate to

$$\|Q(\phi_\kappa, \psi_{\kappa'})\|_{L^2(\mathbf{Q}_R^{(0)})} \lesssim R^{(1-c)\alpha} R^{C\varepsilon}\left(\sum_{T \in \mathbf{T}(\kappa)} E(\phi_T)\right)^{1/2}\left(\sum_{T' \in \mathbf{T}'(\kappa')} E(\psi_{T'})\right)^{1/2} + R^{-10n} \tag{19}$$

for all cubes $\kappa \approx \kappa'$ of side-length $l(\kappa) = l(\kappa') = \rho_0$.

Fix κ, κ' and expand

$$\|Q(\phi_\kappa, \psi_{\kappa'})\|_{L^2(\mathbf{Q}_R^{(0)})} = \|\sum_{T \in \mathbf{T}(\kappa), T' \in \mathbf{T}'(\kappa')} Q(\phi_T, \psi_{T'})\|_{L^2(\mathbf{Q}_R^{(0)})}.$$

By arguments similar to those in the transverse case, we may use the decay estimate (7) to insert a characteristic function χ_T in front of the interaction $Q(\phi_T, \psi_{T'})$ since the error is acceptable. Similarly we may use (8) to insert a $\chi_{T'}$ in front of $\psi_{T'}$. Thus we can restrict to those pairs T, T' of tubes which intersect. But in the parallel interaction case we see from the definition (18) of ρ_0 that each T only intersects $O(R^{Cc})$ tubes T'. Also, the tubes T only overlap by a multiplicity of $O(R^{Cc})$. Thus we may estimate the left-hand side of (19) by

$$R^{Cc}(\|\sum_{T \in \mathbf{T}(\kappa), T' \in \mathbf{T}'(\kappa'): T \cap T' \neq \emptyset} Q(\phi_T, \psi_{T'})\|_{L^2(\mathbf{Q}_R^{(0)})}^2)^{1/2}.$$

We use the triangle inequality to pull the summations outside the norm. From another Cauchy-Schwartz and application of (11), (10), we see that it suffices to show that

$$R^{Cc}\|Q(\phi_T, \psi_{T'})\|_{L^2(\mathbf{Q}_R^{(0)})} \lesssim R^{(1-c)\alpha}E(\phi_T)^{1/2}E(\psi_{T'})^{1/2} + R^{-20n}$$

for all $T \in \mathbf{T}(\kappa)$ and $T' \in \mathbf{T}'(\kappa')$ such that $T \cap T' \neq \emptyset$. Indeed we shall show the stronger estimate

$$\|Q(\phi_T, \psi_{T'})\|_{L^2(\mathbf{Q}_R^{(0)})} \lesssim R^{-\beta}E(\phi_T)^{1/2}E(\psi_{T'})^{1/2} + R^{-30n} \qquad (20)$$

for some absolute constant β depending only on n, if c was sufficiently small and r was sufficiently close to $\sqrt{R}$. This shall be deferred to Section 5, after some preliminaries on vector fields.

4. A digression on vector fields

In this section we introduce our main tool for obtaining decay and localization estimates, namely the Killing and conformal Killing vector fields (see [13], [14] for a more comprehensive discussion). We begin with an informal discussion.

We are interested in those vector fields $X = X^\alpha \partial_\alpha$ which preserve the space of free waves. In addition to the constant co-efficient vector fields $(\partial_1, \dots, \partial_n, \partial_t)$, one also has the Poincare group of vector fields

$$\begin{aligned} S &:= t\partial_t + r\partial_r & \text{(Scaling)} \\ L_i &:= x_i\partial_t + t\partial_i \text{ for } 1 \leq i \leq n & \text{(Lorentz boosts)} \\ \Omega_{i,j} &:= x_i\partial_j - x_j\partial_i \text{. for } 1 \leq i < j \leq n & \text{(Angular rotations)} \end{aligned}$$

We use

$$\Gamma = \Gamma^{(0,0)} := (S, L_1, \dots, L_n, \Omega_{1,2}, \dots, \Omega_{n-1,n})$$

to denote the entire collection of Poincare vector fields, thus Γ is a vector-valued first order differentiation operator. We can then define Γ^k for any $k \geq 1$ as a k-vector-valued k^{th} order differentiation operator.

The Poincare vector fields $\Gamma = \Gamma^{(0,0)}$ are centered at the origin, but we may also translate them in spacetime to be centered at an arbitrary point (x_0, t_0), creating a new set of conformal Killing vector fields $\Gamma^{(x_0, t_0)}$. This will turn out to be important for tube localization.

If ϕ is a free wave and D is a collection of conformal Killing vector fields (e.g. $D = \Gamma$), then $D^k \phi$ is also a free wave, and in particular we have the energy identity

$$\|\nabla_{x,t} D^k \phi(t)\|_2 = \|\nabla_{x,t} D^k \phi(0)\|_2.$$

To exploit this we seek to use vector fields D which are "large" at time t and "small" at time zero. In practice, this means that we try to ensure $\phi(0)$ is localized in physical space (in order to make the weights associated with the X_j small at time zero), and at time t we interest ourselves with the regions where the vector fields D are large and span the tangent space. In such regions we will obtain extremely good decay estimates[11]. We observe the convenient fact that if ϕ has frequency μ and D is any collection of constant co-efficient or Poincare vector fields, then $D^k \phi$ also has frequency μ.

To localize a wave to a cone centered at the origin $(0,0)$ it is enough to use the Poincare vector fields $\Gamma^{(0,0)}$ (in fact one can use only the scaling vector field S to do this). Similarly for cones centered around other points (x_0, t_0) in spacetime. To localize instead to a tube, say one with velocity e_1, one shall need some additional (constant coefficient) vector fields, notably $\partial_2, \ldots, \partial_n$ and $\partial_t + \partial_1$.

We now turn to the details. Let ψ be a free wave with some frequency μ. Then $\Gamma^k \psi$ is also a free wave with frequency μ, and the energy identity yields

$$\|\nabla_{x,t} \Gamma^k \psi(t)\|_2 \lesssim \|\nabla_{x,t} \Gamma^k \psi(0)\|_2$$

for all $k \geq 0$ (here and in the sequel all implicit constants may depend on k). On the other hand, if we expand Γ^k out using the Leibnitz product rule, and use the equation $\Box \psi = 0$ to convert any repeated time derivative to a repeated space derivative, we obtain the pointwise estimate

$$|\nabla_{x,t} \Gamma^k \psi(x,0)| \lesssim \sum_{j=0}^{k} |x|^j |\nabla_x^j \nabla_{x,t} \psi(x,0)|.$$

Combining these bounds with (2) we obtain the weighted energy estimate

$$\|\nabla_{x,t} \Gamma^k \psi(t)\|_2 \lesssim \|(1 + \mu|x|)^k \nabla_{x,t} \psi(x,0)\|_{L_x^2}. \tag{21}$$

---

[11] By taking k arbitrarily large, we will in fact obtain arbitrarily fast decay; we will be able to justify taking these many derivatives because we will always be working in the frequency-localized setting. To some extent this is still true in such apparently rough situations as quasilinear wave equations, because of the frequency localization. In this paper we shall differentiate many times (e.g. $100n$ times) but one can get away with far less differentiation by being more careful with the error terms, etc.

As a first attempt to exploit this differentiability in Γ we observe the identities

$$\partial_t = \frac{t}{t^2 - r^2} S - \sum_{j=1}^n \frac{x_j}{t^2 - r^2} L_j$$

$$\partial_i = \frac{1}{t} L_i - \frac{x_i}{t} \partial_t$$

$$= \frac{1}{t} L_i - \frac{x_i}{t^2 - r^2} S + \sum_{j=1}^n \frac{x_i x_j}{t(t^2 - r^2)} L_j$$

where we use polar co-ordinates $x = r\omega$ As a caricature, we thus have something like

$$\nabla_{x,t} `` \sim `` \frac{x^2 + t^2}{t(t^2 - r^2)} \Gamma `` \lesssim `` \frac{1}{||t| - |x||} \Gamma.$$

If we combine this with repeated applications of the Leibnitz product rule, we conclude the pointwise estimate

$$|\nabla_{x,t}^k \psi(x,t)| \lesssim \frac{1}{||t| - |x||^k} \sum_{j=0}^k |\Gamma^j \psi(x,t)|. \tag{22}$$

This is estimate is good away from the light cone, but is not completely satisfactory in the region $|x|/2 < |t| < 2|x|$. In this case we introduce the null frame

$$\begin{array}{lll} E_+ & := \operatorname{sgn}(t)\partial_t + \partial_r & \text{good null direction} \\ E_- & := \operatorname{sgn}(t)\partial_t - \partial_r & \text{bad null direction} \\ A_i & := \partial_i - \omega_i \partial_r & \text{angular derivatives.} \end{array}$$

We write $\slashed{\nabla} := (A_1, \ldots, A_n)$ for the angular derivatives, and $\slashed{D} := (E_+, \slashed{\nabla})$ for the good derivatives.

These vector fields are not quite conformal Killing, but we can write them in terms of the Poincare vector fields:

$$(t + r)E_+ = S + \sum_{i=1}^n \omega_i L_i$$

$$(t - r)E_- = S - \sum_{i=1}^n \omega_i L_i$$

$$t A_i = L_i - \frac{x_i}{r} \sum_{j=1}^n \omega_i L_j.$$

As a caricature, we thus have

$$\slashed{D} `` \lesssim `` \frac{1}{|t|} \Gamma; \quad E_- `` \lesssim `` \frac{1}{||t| - |x||} \Gamma.$$

By repeated applications of the Leibnitz rule, we can thus improve (22) to

$$|\slashed{D}^j \nabla_{x,t}^k \psi(t,x)| \lesssim \frac{1}{|t|^j ||t| - |x||^k} \sum_{m=0}^{j+k} |\Gamma^m \psi(t,x)| \tag{23}$$

near the light cone for all $j, k \geq 0$. Informally, this means that if the Γ vector fields are under control, then good derivatives gain a $|t|$ while bad derivatives only gain a $||t| - |x||$.

We now specialize the above estimates. Fix $R \gtrsim 1$, $\mu \gtrsim 1$, and suppose that ψ is a wave of frequency μ such that $E(\psi) \lesssim 1$. We also assume that ψ is localized around the origin at time zero. More precisely, we assume there is a radius $1 \lesssim \Lambda \ll R$ such that we have the estimate

$$(\int_{|x| \gg \Lambda} |x|^{200n} |\nabla_{x,t} \psi(x,0)|^2 \, dx)^{1/2} \lesssim 1. \tag{24}$$

From (21) we thus obtain

$$\|\nabla_{x,t} \Gamma^k \psi(0)\|_2 \lesssim \Lambda^k \mu^k E(\psi)^{1/2} + \Lambda^{-100n+k} \mu^k$$

for all $0 \leq k \leq 100n$. Since $\Gamma^k \psi$ has frequency μ, we can use the reproducing formula (1) to then obtain

$$\|\Gamma^k \psi(t)\|_2 \lesssim \Lambda^k \mu^{k-1} E(\psi)^{1/2} + \Lambda^{-100n+k} \mu^{k-1}. \tag{25}$$

We now specialize to the case when ψ has frequency $\mu = 1$, and time is restricted to the region $R \leq t \leq 2R$. To emphasize the frequency 1 we shall write ϕ instead of ψ.

First suppose that we are in the region $r \leq |t|/2$ or $r \geq 2|t|$; in other words, we are away from the cone. By (22), (25) we see that

$$|\nabla_{x,t}^k \phi(t,x)| \lesssim \Lambda^k R^{-k} E(\phi)^{1/2} + \Lambda^{-100n}. \tag{26}$$

Now suppose that we are in the region $|t|/2 < r < 2|t|$, so that $r \sim R$. From (23) we have

$$|\not\!\nabla^j \nabla_{x,t}^k \phi(t,x)| \lesssim R^{-j} \sum_{m=0}^{j} |\Gamma^m \nabla_{x,t}^k \phi(t,x)|. \tag{27}$$

We now use the angular Sobolev inequality

$$\|u\|_{L^\infty(S^{n-1})} \lesssim \|u\|_{L^1(S^{n-1})} + \|\not\!\nabla^{n-1} u\|_{L^1(S^{n-1})},$$

which is easily verified by $n-1$ applications of the fundamental theorem of Calculus. Applying this inequality to $u := |\nabla_{x,t}^k \phi(t,x)|^2$, we obtain

$$|\nabla_{x,t}^k \phi(t,x)|^2 \lesssim \int_{S^{n-1}} \sum_{i+j \leq n-1} |\not\!\nabla^i \nabla_{x,t}^k \phi(t,r\omega')| |\not\!\nabla^j \nabla_{x,t}^k \phi(t,r\omega')| \, d\omega'.$$

By (27) we thus have

$$|\nabla_{x,t}^k \phi(t,x)|^2 \lesssim (\frac{\Lambda}{R})^{n-1} \int_{S^{n-1}} r^{n-1} (\sum_{0 \leq i \leq n-1} \Lambda^{-i} |\Gamma^i \nabla_{x,t}^k \psi(t,r\omega')|)^2 \, d\omega'.$$

This is a square of the L^2 norm on a sphere of radius r. We then use (2) to average out this sphere into what is essentially an annulus of thickness $O(1)$ around the sphere of radius r. This implies that

$$|\nabla_{x,t}^k \phi(t,x)|^2 \lesssim (\frac{\Lambda}{R})^{n-1} \sum_{0 \leq i \leq n-1} \Lambda^{-i} E(\Gamma^i \nabla_{x,t}^{k-1} \phi)^{1/2}.$$

Applying (25) we thus obtain the pointwise estimate

$$|\nabla_{x,t}^k \phi(t,x)| \lesssim (\frac{\Lambda}{R})^{\frac{n-1}{2}} E(\phi)^{1/2} + \Lambda^{-90n}.$$

If we combine this with (26) we obtain the decay estimate

$$\|\nabla_{x,t}^{k}\phi(t)\|_{\infty} \lesssim (\frac{\Lambda}{R})^{\frac{n-1}{2}} E(\phi)^{1/2} + \Lambda^{-90n} \tag{28}$$

for $k = 100n$ (say). But since ψ has frequency 1, this estimate in fact holds for all k by the reproducing formula (1). Finally, we remark that we may replace ϕ by $\Gamma\phi$ on both sides without affecting the estimate. In particular we have

$$\|\Gamma\phi(t)\|_{\infty} \lesssim (\frac{\Lambda}{R})^{\frac{n-1}{2}} E(\Gamma\phi)^{1/2} + \Lambda^{-90n} \lesssim (\frac{\Lambda}{R})^{\frac{n-1}{2}} \Lambda E(\phi)^{1/2} + \Lambda^{-90n}. \tag{29}$$

5. Parallel interactions

Fix T, T'. To complete the proof of Theorem 1.1 begin in Section 3, we need to prove (20). (We also need to prove Lemma 2.3, but this will be done in later sections).

In earlier work such as [26], [29], [23] the proof of (20) was accomplished by using Lorentz transforms to reduce to the $R = 1$ case. However we do not wish to do this here as this method seems to rely heavily on the special properties of Minkowski space. Instead we shall use commutation with vector fields.

In proving (20) we will actually prove a somewhat stronger result:

Proposition 5.1. Let $\sqrt{R} \ll \Lambda \ll R$, and ϕ, ψ be free waves of frequency 1 and μ respectively, where $\mu \gtrsim 1$. Suppose we have the energy normalization

$$E(\phi), E(\psi) \lesssim 1$$

and the localization properties

$$(\int_{|x-x_0|\gg\Lambda} |x - x_0|^{200n} |\nabla_{x,t}\phi(x,t_0)|^2 \ dx)^{1/2} \lesssim 1$$

and

$$(\int_{|x-x_0|\gg\Lambda} |x - x_0|^{200n} |\nabla_{x,t}\psi(x,t_0)|^2 \ dx)^{1/2} \lesssim 1$$

for some $x_0 \in \mathbf{R}^n$ and $0 \le t_0 \le R/2$. Then we have the fixed-time estimate

$$\|Q(\phi,\psi)(t)\|_{L^2(\mathbf{R}^n)} \lesssim (\frac{\Lambda}{R})^{(n+1)/2} E(\phi)^{1/2} E(\psi)^{1/2} + R^{-40n} \tag{30}$$

for all $R \le t \le 2R$.

Proof By translation invariance we may set $x_0 = 0$. Also by a time translation and some redefinition of R if necessary we may set $t_0 = 0$.

Fix ϕ, ψ, R, μ. In the region away from the light cone, i.e. $|x| \le t/2$ or $|x| \ge 2t$, we can control ϕ using (26) and ψ using energy estimates to easily obtain (30). Now suppose we are near the light cone, so that $|x| \sim t \sim R$.

We shall write the null form Q in terms of the null frame A_i, E_+, E_- developed in the previous section. More precisely, we claim the pointwise estimate

$$|Q(\phi, \psi)| \lesssim |\not{D}\phi||\nabla_{x,t}\psi| + |\nabla_{x,t}\phi||\not{D}\psi|.$$

To see this, we substitute in the identities $\partial_i = A_i + \frac{x_i}{2r}E_+ - \frac{x_i}{2r}E_-$ and $\partial_t = \frac{1}{2}E_+ + \frac{1}{2}E_-$, and observe that in each of the three null forms Q_0, Q_i, Q_{ij} the $E_-\phi E_-\psi$ term (which is the only potentially bad term) always vanishes[12].

From the previous, (23) and Hölder we have

$$\|Q(\phi, \psi)(t)\|_2 \lesssim \frac{1}{R}(\|\Gamma\phi(t)\|_\infty \|\nabla_{x,t}\psi(t)\|_2 + \|\nabla_{x,t}\phi\|_\infty \|\Gamma\psi(t)\|_2).$$

The claim now follows from estimating ϕ using decay estimates (28), (29), and ψ using the standard energy estimate and (25). ∎

We now apply this Proposition with ϕ, ψ replaced by ϕ_T and $\psi_{T'}$ respectively, with t_0 set to $R/2$, and x_0 set to the center of κ. Recall that $\rho_0 = rR^c$ is the size of κ. By the Proposition and an L^2 norm in time we have

$$\|Q(\phi_T, \psi_{T'})\|_{L^2(\mathbf{R}^n \times [R, 2R]))} \lesssim R^{1/2}R^{Cc}(\frac{r}{R})^{(n+1)/2}E(\phi_T)^{1/2}E(\psi_{T'})^{1/2} + R^{-40n}.$$

If c is sufficiently small, and r is sufficiently close to $\sqrt{R}$, the claim follows. (Indeed from the argument we see that we just need r to be substantially less than $R^{n/(n+1)}$).

6. First proof of Lemma 2.3: Phase space preparation of the initial data.

Fix ϕ, μ, R, r. It suffices to prove the claim under the assumption on the initial data that $\phi_t(0) = \pm i\sqrt{-\Delta}\phi(0)$ for some sign $\pm$; this is because general initial data can be split into two pieces of the above form via the formula

$$\phi[0] = \sum_{\pm}(\frac{\phi(0) \pm i\sqrt{-\Delta}^{-1}\phi_t(0)}{2}, \pm\sqrt{-\Delta}\frac{\phi(0) \pm i\sqrt{-\Delta}^{-1}\phi_t(0)}{2})$$

Note that this decomposition is orthogonal with respect to the energy inner product $\langle \cdot, \cdot \rangle_e$, and does not affect the property of ϕ having frequency μ.

We shall assume that $\phi_t(0) = +i\sqrt{-\Delta}\phi(0)$; the argument for the other sign is similar and is left to the reader. It shall be convenient to define the length scales $s := R^{-\varepsilon/2}r$ and $\Lambda := R^{-\varepsilon/4}r$.

Since the ϕ_T are free waves, to construct ϕ_T it will suffice to construct their initial data $\phi_T[0]$. By standard arguments we may partition $1 = \sum_{x_0 \in \Sigma} \chi_{x_0}$ on $\mathbf{R}^n$, where $\Sigma \subset \mathbf{R}^n$ is a lattice of spacing $\sim s$, and χ_{x_0} is a bounded function with Fourier transform supported in the ball $|\xi| \ll \mu$ and which satisfies the estimates

$$|\chi_{x_0}(x)| \lesssim R^{-M}\mu^{-M} \text{ whenever } |x - x_0| \gg s \tag{31}$$

---

[12]In principle the Q_0 null form should be somewhat better than Q_i and Q_{ij}, as can be seen by inspection of the bilinear symbol (see e.g. [8]), but this cannot be detected by this simple null frame analysis.

for any $M > 0$; the implicit constants may depend on M. For instance, one can set $\chi_{x_0} := P_{\mu/C}\chi_{\mathbf{Q}(x_0)}$, where the $\mathbf{Q}(x_0)$ form a partition of $\mathbf{R}^n$ into cubes of side-length $\sim s$.

We also partition $1 = \sum_{\omega \in \Omega} \eta_\omega$ on S^{n-1}, where Ω is a s/R-separated set of directions and η_ω is a bump function adapted to the disk of radius $O(s/R)$ centered at ω.

We define the sector multipliers $P_{\mu,\omega}$ via the Fourier transform as

$$\widehat{P_{\mu,\omega}f}(\xi) := \eta(\frac{\xi}{|\xi|})a(|\xi|)\hat{f}(\xi)$$

where a is a bump function which equals one on a suitable annulus $|\xi| \sim \mu$ and is adapted to a dilate of this annulus. Thus $P_{\mu,\omega}$ is a Fourier projection to a tubular region of length μ and width $s\mu/R$, oriented in the direction ω. Observe from construction that $s \cdot \mu$ and $s\mu/R \cdot s$ are both $\gtrsim R^\varepsilon\mu$. By the uncertainty principle we thus see that the convolution kernel K_ω of $P_{\mu,\omega}$ is integrable and obeys the estimates

$$|\nabla K_\omega(x)| \lesssim R^{-M}\mu^{-M} \text{ whenever } |x| \gtrsim s \tag{32}$$

for any M.

Our collection of tubes $\mathbf{T}$ will be defined by

$$\mathbf{T} := \{T(R, r, x_0, \omega) : x_0 \in \Sigma; \omega \in \Omega\};$$

note that the claimed cardinality bound on the intersection of $\mathbf{T}$ with any R-cube is clear.

Next, our ϕ_T will be defined via their initial data as

$$\phi_{T(R,r,x_0,\omega)}(0) := P_{\mu,\omega}(\chi_{x_0}\phi(0))$$

and

$$\partial_t\phi_T(0) := i\sqrt{-\Delta}\phi_T(0).$$

We will not need an error term for this argument and will set $\phi_{error} := 0$.

By construction it is clear that the ϕ_T are free waves with frequency $\sim \mu$, and that (9) holds.

Observe that the physical space multipliers χ_{x_0} are almost orthogonal, as are the Fourier space multipliers $P_{\mu,\omega}$. Thus (10) holds. Now we show (11). Since the time derivative of ϕ_T is controlled by the spatial derivatives, it suffices to show that

$$\|\sum_{(\omega,x_0)\in X}\nabla_x P_{\mu,\omega}(\chi_{x_0}\phi(0))\|_2 \lesssim R^{C\varepsilon}(\sum_{(\omega,x_0)\in X}\|\nabla_x P_{\mu,\omega}(\chi_{x_0}\phi(0))\|_2^2)^{1/2} + R^{-100n}$$

for all subsets X of $\Omega \times \Sigma$.

Observe from Plancherel's theorem that the $\nabla_x P_{\mu,\omega}$ multipliers are essentially orthogonal as ω varies. Thus it suffices to show that

$$\| \sum_{x_0 \in Y} \nabla_x P_{\mu,\omega}(\chi_{x_0}\phi(0))\|_2 \lesssim R^{C\varepsilon}(\sum_{x_0 \in Y} \|\nabla_x P_{\mu,\omega}(\chi_{x_0}\phi(0))\|_2^2)^{1/2} + R^{-150n}$$

for all $\omega \in \Omega$ and $Y \subset \Sigma$.

Fix ω, Y. From (31), (32) and the assumption $E(\phi) \lesssim 1$ we see that

$$|\nabla_x P_{\mu,\omega}(\chi_{x_0}\phi(0))(x)| \lesssim (R + |x - x_0|)^{-M}\mu^{-M} \text{ whenever } |x - x_0| \gtrsim \Lambda \tag{33}$$

for any M. Thus one can restrict $\nabla_x P_{\mu,\omega}(\chi_{x_0}\phi(0))(x)$ to a ball of radius $O(\Lambda)$ around x_0 by paying a total price of $O(R^{-150n})$. Since these balls only overlap by about $O(R^{C\varepsilon})$ or so, the claim follows. This proves (11).

It remains to show that each ϕ_T is localized to T. Fix T; without loss of generality we may take $x_0 = 0$ and $\omega = e_1$. We have to show (7), (8). We only show the L^2 bound (8); the L^∞ bound (7) essentially follows from Sobolev (replacing r by $r/2$ if necessary). We thus need to show

$$(\int_{|x-te_1|\geq r} |\nabla_{x,t}\phi_T(x,t)|^2 \, dx)^{1/2} \lesssim (R + |x - te_1|)^{-100n} \tag{34}$$

whenever $|t| \sim R$.

We shall achieve this using vector fields by means of the following lemma, which will also be used in our second proof in the next section.

Lemma 6.1. *Let $R^{1/2+\varepsilon} \lesssim r \ll R$, and let $\Lambda = R^{-\varepsilon/4}r$. Let ϕ_T be a free wave of some frequency $\mu \gtrsim 1$ such that*

$$|\nabla_{x,t}^k\phi_T(x,0)| \lesssim (R + |x|)^{-M}\mu^{-M} \text{ whenever } |x| \gtrsim \Lambda \tag{35}$$

for any $k \geq 1$ and any M, with the implicit constant depending on k and M of course. Also, if Σ denotes the vector fields

$$\Sigma := \{\partial_t + \partial_1, \partial_2, \dots, \partial_n\}$$

then we assume the estimates

$$E(\Sigma^k \phi_T) \lesssim R^{100n}(\frac{\Lambda\mu}{R})^k \tag{36}$$

for all $k \geq 0$. Then (34) holds for all $|t| \sim R$.

Proof Since ϕ_T has frequency μ, it will suffice to show that

$$(\int_{|x-te_1|\geq r} |\nabla_{x,t}^{k+1}\phi_T(x,t)|^2 \, dx)^{1/2} \lesssim (R + |x - te_1|)^{-100n}\mu^k \tag{37}$$

for some suitable k which we can choose at will.

Fix t. From (35) and (36) we have the regularity estimate

$$E(\Gamma^k \Sigma^j \phi_T)^{1/2} \lesssim R^{100n}(\Lambda\mu)^k (\frac{\Lambda\mu}{R})^j + R^{-M}\mu^{-M} \tag{38}$$

for any $k, j \geq 0$.

We now divide the region $|x - te_1| \geq r$ into several sub-regions and prove the localization estimate (37) separately for each sub-region.

Case 1. $||t| - |x|| \gtrsim r$

In this case we are away from the light cone and one can use the fact that the vector fields Γ span.

By (23) we have
$$|\nabla_{x,t}^{k+1} \phi_T(x,t)| \lesssim r^{-k} |\Gamma^k \nabla_{x,t} \phi_T(x,t)|$$
and hence by the regularity estimate (38)
$$(\int_{x:||t|-|x||\gtrsim r} |\nabla_{x,t}^{k+1} \phi_T(x,t)|^2 \, dx)^{1/2} \lesssim R^{100n} (\Lambda\mu/r)^k + R^{-M}\mu^{-M}.$$

Since $r = R^{\varepsilon/4}\Lambda$, we thus obtain (37) for this region by setting k sufficiently large.

Case 2. $||t| - |x|| \ll r$, **and** $x_1 > |t|/2$.

Without loss of generality we may take $|x_2| \geq |x_3|, \ldots, |x_n|$, thus $|x_2| \gtrsim r$. In particular $|\omega_2| \gtrsim r/R$. The idea is now to use the fact that Γ together with ∂_2 spans.

We now use the identity
$$E_- = E_+ + \frac{2}{\omega_2} A_2 - \frac{2}{\omega_2} \partial_2;$$
since E_- and $\not{D}$ form a null frame, we thus informally have
$$\nabla_{x,t} \text{``} \lesssim \text{''} (1 + \frac{1}{\omega_2})(\not{D}, \Sigma).$$

Since $|\omega_2| \gtrsim r/R$, we thus obtain after repeated applications of the Leibnitz rule
$$|\nabla_{x,t}^{k+1} \phi_T(x,t)| \lesssim \sum_{i+j\leq k} R^{i+j} r^{-k} |\nabla_{x,t} \not{D}^i \Sigma^j \phi_T(x,t)|$$

for any k. Applying the regularity estimate (38) and substituting $\Lambda = R^{-\varepsilon/4}r$, we thus obtain
$$(\int_{|x_2|\gtrsim r} |\nabla_{x,t}^{k+1} \phi_T(x,t)|^2 \, dx)^{1/2} \lesssim R^{100n}\mu^k R^{-\varepsilon k/2} E(\phi_T)^{1/2} + R^{-M}\mu^{-M}.$$

The claim follows by setting k sufficiently large.

Case 3. $||t| - |x|| \ll r$, **and** $x_1 \leq |t|/2$.

In this case we use the fact that $\not{D}$ together with $\partial_t + \partial_1$ span. Specifically, we use the identity
$$E_- = \frac{-1-\omega_1}{1-\omega_1} E_+ + \frac{2}{1-\omega_1}(\partial_t + \partial_1) - \frac{2}{1-\omega_1} A_1.$$
Since $|1 - \omega_1| \sim 1$ by hypothesis, we thus informally have
$$\nabla_{x,t} \text{``} \lesssim \text{''} (\not{D}, \Sigma).$$

One then argues as in Case 2.   ∎

Fix t. We now apply the Lemma just proven to obtain the localization estimate (34). By repeating the derivation of the initial data localization estimate (33) we can obtain the pointwise estimates (35). We now verify the regularity estimate (36).

For the derivatives $\partial_2, \ldots, \partial_n$ this is clear from the Fourier support of ϕ_T. Now consider the powers of $\partial_t + \partial_1$. By repeated application of the equation $\Box \phi_T = 0$ we see that

$$\partial_t^k \phi_T = (i\sqrt{-\Delta})^k \phi_T,$$

and hence

$$(\partial_t + \partial_1)^k \phi_T = (i\sqrt{-\Delta} - \partial_1)^k \phi_T.$$

On the other hand, the Fourier transform of $\nabla_{x,t}^k \phi_T$ is supported in the region $\xi_1 \sim \mu$, $\xi = \xi_1 e_1 + O(\mu s/R)$. In particular we have the crude estimate $i|\xi| - i\xi_1 = O(\mu s/R)$, and the claim follows. This concludes the proof of the localization estimate (34), and Lemma 2.3 follows.

7. Second proof of Lemma 2.3: Spatial truncation at two distinct times.

In the previous argument, tube localization was achieved by localizing $\phi[0]$ in physical space and then applying a Fourier projection. Although this method works, it is somewhat unsatisfactory as it (implicitly) uses the fact that the Fourier support of a free wave is preserved by the time evolution. This type of fact does not seem very stable in rough metrics, and so we would like to use a tube decomposition method which does not require as much usage of Fourier space.

The idea is to characterize a tube by its intersection with two time slices $t = 0$ and $t = 3R$ (cf. the strategy used to attack the Kakeya problem in [5]). We shall write τ for $3R$.

In this section it will be convenient to take $s := R^{1/2+\varepsilon/2}$ and $\Lambda := R^{1/2+3\varepsilon/4}$. With this value of s, we may partition $1 = \sum_{x_0 \in \Sigma} \chi_{x_0}$ on $\mathbf{R}^n$ as in the previous section.

Let $U(t)$ denote the free evolution operator on free waves, thus $U(t)\phi[0] := \phi[t]$. Then we may use linearity to split

$$\phi[t] = \sum_{x_0, x_\tau \in \Sigma} U(t)(\chi_{x_0} U(-\tau)(\chi_{x_\tau} \phi[\tau])).$$

In other words, for each x_0, x_τ, we localize ϕ to an s-neighbourhood of x_τ at time τ, evolve back by the free wave equation, localize to an s-neighbourhood of x_0 at time 0, and use this as initial data. As we shall see (using the fact that $s > \sqrt{R}$), the second localization will not significantly disturb the localization of the first.

We divide the summation into two terms, depending on whether we have a null ray (in the sense that $|x_\tau - x_0| = \tau + O(\Lambda)$) or not.

First consider the contribution of the case $||x_\tau| - |x_0|| \neq \tau + O(\Lambda)$, so that the two points do not form a null ray. This case is small and will be absorbed into the error term ϕ_{error}. To prepare for this, we fix x_τ, and define $\psi = \psi_{x_\tau}$ by $\psi[t] := U(t - \tau)(\chi_{x_\tau}\phi[\tau])$. Thus ψ is a free wave of frequency μ which is essentially localized to the disk $\{(x, \tau) : x = x_\tau + O(s)\}$ at time τ. To exploit this we recall that $\Gamma^{(x_\tau,\tau)}$ denotes the Poincare vector fields centered at the point (x_τ, τ) in spacetime. From the reproducing formula (1) and several applications of the Leibnitz rule we have

$$E((\Gamma^{(x_\tau,\tau)})^k \psi) \lesssim (s\mu)^{2k} \int (1 + \frac{|x - x_\tau|}{\Lambda})^{-100n} |\nabla_{x,t}\phi(x,\tau)|^2 \; dx$$

for any k, where of course the implicit constant depends on k. By the pointwise estimates (23) (adapted to $\Gamma^{(x_\tau,\tau)}$ rather than $\Gamma^{(0,0)}$) we can thus bound

$$(\int_{||x|-\tau|\gtrsim\Lambda} |\nabla_{x,t}^{k+1}\psi(x,0)|^2 \; dx)^{1/2} \lesssim \Lambda^{-k}(s\mu)^k(\int (1 + \frac{|x - x_\tau|}{\Lambda})^{-100n} |\nabla_{x,t}\phi(x,\tau)|^2 \; dx)^{1/2}.$$

Since ψ has frequency μ, we thus see from the pointwise bounds (2) that

$$(\int_{||x|-\tau|\gtrsim\Lambda} |\nabla_{x,t}\psi(x,0)|^2 \; dx)^{1/2} \lesssim (s/\Lambda)^k(\int (1 + \frac{|x - x_\tau|}{\Lambda})^{-100n} |\nabla_{x,t}\phi(x,\tau)|^2 \; dx)^{1/2}.$$

By making k sufficiently large we may set $(s/\Lambda)^k \lesssim R^{-100n}$. From this estimate it is easy to show that

$$E(\sum_{x_0,x_\tau\in\Sigma:||x_\tau|-|x_0||\neq\tau+O(\Lambda)} \chi_{x_0}U(-\tau)(\chi_{x_\tau}\phi[\tau])) \lesssim R^{-50n}$$

which is acceptable.

It thus remains to consider the terms when $||x_\tau| - |x_0|| = \tau + O(\Lambda)$. In this case we can find a tube $T = T(x_0, x_\tau)$ which is a distance $O(\Lambda)$ from $(x_0, 0)$ and (x_τ, τ), and we define $\phi_T(t) = U(t)(\chi_{x_0}U(-\tau)(\chi_{x_\tau}\phi[\tau]))$ accordingly. If we let $\mathbf{T}$ denote the collection of all such tubes (which we can of course make disjoint) then (9) holds. To show (10), we observe from spatial orthogonality and the energy identity that

$$\sum_{T\in\mathbf{T}} E(\phi_T) \lesssim \sum_{x_0,x_T\in\Sigma} E(U(t)(\chi_{x_0}U(-\tau)(\chi_{x_\tau}\phi[\tau])))$$

$$= \sum_{x_0,x_\tau\in\Sigma} E(\chi_{x_0}U(-\tau)(\chi_{x_\tau}\phi[\tau]))$$

$$\lesssim \sum_{x_\tau\in\Sigma} E(U(-\tau)(\chi_{x_\tau}\phi[\tau]))$$

$$= \sum_{x_\tau\in\Sigma} E(\chi_{x_\tau}\phi[\tau])$$

$$\lesssim E(\phi[\tau]) = E(\phi)$$

as desired.

Now we show the L^2 localization estimate (8); the L^∞ localization estimate (7) is similar and follows from (8) (with r replaced by $r/2$) and the pointwise estimate (2).

Without loss of generality we shall assume that $x_0 = 0$ and $x_\tau = \tau e_1 + O(\Lambda)$. In this case we need to prove the localization estimate (34). By Lemma 6.1 it suffices to verify the initial data localization estimate (35) and the regularity estimate (36).

The claim (35) follows easily from the decay of cutoff χ_{x_0} and crude energy estimates on ϕ_T (the Bessel inequality(10) will do). Now we verify (36). Once again we set $\psi = \psi_{x_\tau}$ by $\psi[t] := U(t - \tau)(\chi_{x_\tau}\phi[\tau])$. If derivatives Σ land on the cutoff χ_{x_0} then one gains a power of s for each derivative (note that we can use the equation $\Box\phi_T = 0$ to convert any repeated time derivative to a repeated space derivative). Thus it suffices to show that

$$E(\chi_{x_0}\Sigma^k\psi[0]) \lesssim R^{100n}(\frac{\Lambda\mu}{R})^k. \tag{39}$$

We may replace χ_{x_0} with the characteristic function of the disk $D := \{(x,0) : |x - x_0| \lesssim \Lambda\}$ since the rapid decay of χ_{x_0} ensures that the contribution outside this disk is acceptable.

This disk is close to the light cone centered at (x_τ, τ). We will thus use the good derivatives $\displaystyle{\not{D}}^{(x_\tau,\tau)}$, being the vector fields $\displaystyle{\not{D}}$ translated to be centered at (x_τ, τ) and adapted to the backward time flow.

We observe the identities

$$\partial_t + \partial_1 = A_1^{(x_\tau,\tau)} - \frac{1}{2}(1 - \omega_1^{(x_\tau,\tau)})E_+^{(x_\tau,\tau)} - \frac{1}{2}(1 + \omega_1^{(x_\tau,\tau)})E_-^{(x_\tau,\tau)}$$

$$\partial_i = A_i + \frac{1}{2}\omega_i^{(x_\tau,\tau)}E_+^{(x_\tau,\tau)} + \frac{1}{2}\omega_i^{(x_\tau,\tau)}E_-^{(x_\tau,\tau)}$$

where $\omega^{(x_\tau,\tau)} := (x - x_\tau)/|x - x_\tau|$ is the analogue of ω for the origin (x_τ, τ). On the disk D we have $\omega^{(x_\tau,\tau)} = -e_1 + O(\Lambda/R)$, and thus we informally have

$$\Sigma \text{ ``} \lesssim \text{''} {\not{D}}^{(x_\tau,\tau)} + \frac{\Lambda}{R}\nabla_{x,t}.$$

By many applications of the Leibnitz rule one thus obtains the rigorous estimates

$$|\nabla_{x,t}\Sigma^k\psi(0)| \lesssim \sum_{i+j\leq k}(\Lambda/R)^{k-i}|\nabla_{x,t}^{j+1}\tilde{X}^i\psi(0)|$$

on the disk D with $\tilde{X} = {\not{D}}^{(x_\tau,\tau)}$. To obtain the regularity estimate (39), it thus suffices to show that

$$\sum_{i+j\leq k}(\Lambda/R)^{k-i}E(\nabla_{x,t}^j\tilde{X}^i\psi[0]) \lesssim R^{100n}(\frac{\Lambda\mu}{R})^k.$$

Since ψ has frequency μ, it suffices by the reproducing formula (1) to show that

$$E(\tilde{X}^i\psi[0]) \lesssim R^{100n}(\frac{\Lambda\mu}{R})^i.$$

Observe that the "good" derivatives $\tilde{X} \lesssim \frac{1}{R}\Gamma^{(x_\tau, \tau)}$. Thus

$$E(\tilde{X}^i \psi[0]) \lesssim (\frac{\Lambda}{R})^i \sum_{m \leq i} \Lambda^{-m} E((\Gamma^{(x_\tau, \tau)})^m \psi[0])$$

$$= (\frac{\Lambda}{R})^i \sum_{m \leq i} \Lambda^{-m} E((\Gamma^{(x_\tau, \tau)})^m \psi[\tau])$$

$$= (\frac{\Lambda}{R})^i \sum_{m \leq i} \Lambda^{-m} E((\Gamma^{(x_\tau, \tau)})^m (\chi_{x_\tau} \phi[\tau]))$$

The claim now follows from the Leibnitz rule, energy estimates for ϕ, and the frequency μ hypothesis. This proves the localization estimate (8).

Finally, we show the orthogonality estimate (11). By the Cotlar-Knapp-Stein lemma (see e.g. [20]) it suffices to show that

$$|\langle \phi_T, \phi_{T'} \rangle_e| \lesssim (R + |x_0 - x_0'| + |x_\tau - x_T'|)^{-100n}$$

whenever $|x_0 - x_0'| + |x_\tau - x_\tau'| \gg r$ and $T = T(x_0, x_\tau)$, $T' = T(x_0', x_\tau')$ are in **T**.

Fix $x_0, x_0', x_\tau, x_\tau'$. Define $x_t := x_0 + \frac{t}{\tau}(x_\tau - x_0)$ and $x_t' := x_0' + \frac{t}{\tau}(x_\tau' - x_0')$ for each $0 \leq t \leq \tau$. Clearly we may find a time $R \leq t \leq 2R$ such that $|x_t - x_t'| \gg r$, which means that on this time slice the tubes T and T' are separated by $\gg r$. Since $\langle \phi_T, \phi_{T'} \rangle_e = \langle \phi_T[t], \phi_{T'}[t] \rangle_e$, the claim follows from the localization estimate (8). This completes the second proof of Lemma 2.3.

8. ADDITIONAL REMARKS

• The induction on scales argument in Section 3 lost a power of R^ε on a cube of length R. This R^ε loss can be removed in a number of ways. One possibility is to tighten the induction so that, if one inductively assumes a bilinear estimate with a constant of A (instead of something like R^ε), then one can deduce the same estimate with a constant of $A/2 + O(1)$. By standard bootstrap arguments this gives the bound with $A = O(1)$. This can be done provided one performs extremely careful tube decompositions in which one does not lose any constant at all (not even a factor of 2!) in the main term of the orthogonality inequality (11), as well as the Bessel identity (10). This approach was carried out in [23], however that argument is heavily dependent on Plancherel's identity (among other things) and will probably not be extendable to rough metrics.

• Another option is to tighten the L^2 estimate to an L^p estimate for some $p < 2$ while still losing an R^ε, in the spirit of [29], [28]; this can be accomplished (for $p > (n + 3)/(n + 1)$) by using further geometric facts about the tubes[13]. Standard "epsilon-removal" arguments (see [4], [26]; related arguments are in [3], [21]) exploiting some decay of the fundamental solution in a manner reminiscent of

---

[13]In our current argument, the only geometric fact used is that any two transverse tubes only intersect in a localized region of spacetime. A further fact, used to good effect in [29] (see also [23]) is that a tube which is transverse to an entire *light cone* (not just another tube) will still only intersect in a localized region. An alternative way to phrase this is that one can recover good energy estimates on light cones and similar null surfaces if one makes a transversality assumption.

the Tomas-Stein or Strichartz arguments, can then be used to trade in the R^ε loss at the expense of L^p integrability.

Since this argument does not appear to be well-known in the PDE literature we briefly sketch it as follows. Suppose for instance that we have obtained an estimate of the form

$$\|\phi\psi\|_{L^p(\mathbf{Q}_R)} \lesssim R^\varepsilon E(\phi)^{1/2} E(\psi)^{1/2}$$

for some $p < 2$ and all $\varepsilon > 0$, $R \gg 1$, and all free waves ϕ, ψ of frequency 1. We sketch how this estimate implies the global estimate

$$\|\phi\psi\|_{L^2_{x,t}} \lesssim E(\phi)^{1/2} E(\psi)^{1/2}. \tag{40}$$

For simplicity we shall reduce to the case when $\phi_t = \pm i\sqrt{-\Delta}\phi$, and similarly for ψ. In this case the energy inner product is equivalent to the usual inner product. We shall now cease to distinguish between $\phi(t)$ and $\phi[t]$ to simplify the notation.

Firstly, we use finite speed of propagation to replace $\mathbf{Q}_R$ by a spacetime slab $\mathbf{R}^n \times I$, where I is any interval of length R. By Sobolev embedding we may then replace the L^p_x norm with an L^2_x norm:

$$\|\phi\psi\|_{L^p_t L^2_x(\mathbf{R}^n \times I)} \lesssim R^\varepsilon E(\phi)^{1/2} E(\psi)^{1/2}. \tag{41}$$

Now we prove (40). It suffices to prove the weak-type estimate

$$\|\phi\psi\|_{L^{q,\infty}_t L^2_x} \lesssim E(\phi)^{1/2} E(\psi)^{1/2}$$

for $q < p$. By duality it suffices to show that

$$\int_\Omega \|\phi(t)\psi(t)\|_{L^2_x} \, dt \lesssim |\Omega|^{1/q} E(\phi)^{1/2} E(\psi)^{1/2}. \tag{42}$$

This estimate is clear for $|\Omega| \lesssim 1$ just by crudely bounding ϕ in L^2 and ψ in L^∞ (for instance), so we may assume $|\Omega| \gg 1$.

To compare with (42), observe that the estimate (41) yields

$$\int_{\Omega \cap I} \|\phi(t)\psi(t)\|_{L^2_x} \, dt \lesssim |\Omega|^{1/p} |I|^\varepsilon E(\phi)^{1/2} E(\psi)^{1/2} \tag{43}$$

for any interval I. This is superior if Ω is concentrated in an interval I of size $O(|\Omega|^C)$, but we still must consider the case when Ω is widely dispersed along the time axis. To resolve this we use the decay of the fundamental solution as well as a TT^* argument.

To see this, we fix Ω and let $R := |\Omega|^C$ for some large C to be chosen later. We then choose ε so that $|\Omega|^{1/q} \lesssim |\Omega|^{1/p} R^\varepsilon$. We then partition the time axis $\mathbf{R}$ into disjoint intervals I of length R, and observe from (43), the hypothesis $q < p$, and Cauchy-Schwarz that we have the "weakly global" inequality

$$\sum_I \int_{\Omega \cap I} \|\phi_I(t)\psi_I(t)\|_{L^2_x} \, dt \lesssim |\Omega|^{1/q} \Big(\sum_I E(\phi_I)\Big)^{1/2} \Big(\sum_I E(\psi_I)\Big)^{1/2} \tag{44}$$

for all free waves ϕ_I, ψ_I. Our next objective is to bootstrap this inequality to a "semi-weakly global" inequality

$$\sum_I \int_{\Omega \cap I} \|\phi(t)\psi_I(t)\|_{L_x^2} \, dt \lesssim |\Omega|^{1/q} E(\phi)^{1/2} (\sum_I E(\psi_I))^{1/2}. \qquad (45)$$

A similar bootstrap will then obtain (42).

To obtain (45), we fix the ψ_I and write $\phi[t] = U(t)P_1\phi[0]$ (using the reproducing formula (1)) to rewrite the estimate as

$$|\sum_I \int_{\Omega \cap I} \int U(t)P_1\phi(0)\psi_I(t)F(x,t) \, dt| \lesssim |\Omega|^{1/q} E(\phi)^{1/2}(\sum_I E(\psi_I))^{1/2}\|F\|_{L_t^\infty L_x^2}$$

By duality it suffices to show that

$$E(\sum_I \int_{\Omega \cap I} P_1 U(-t)(\psi_I(t)F(t)) \, dt)^{1/2} \lesssim |\Omega|^{1/q}(\sum_I E(\psi_I))^{1/2}\|F\|_{L_t^\infty L_x^2}.$$

We square this, and expand this as

$$|\sum_{I,J} \int_{\Omega \cap I} \int_{\Omega \cap J} \langle P_1 U(-t)(\psi_I(t)F(t)), P_1 U(-s)(\psi_I(s)F(s)) \rangle \, dsdt| \lesssim (|\Omega|^{1/q}(\sum_I E(\psi_I))^{1/2}\|F\|_{L_t^\infty L_x^2})^2. \qquad (46)$$

On the other hand, if we start with (44) and apply similar considerations we obtain

$$|\sum_I \int_{\Omega \cap I} \int_{\Omega \cap I} \langle P_1 U(-t)(\psi_I(t)F(t)), P_1 U(-s)(\psi_I(s)F(s)) \rangle \, dsdt| \lesssim (|\Omega|^{1/q}(\sum_I E(\psi_I))^{1/2}\|F\|_{L_t^\infty L_x^2})^2;$$

in other words, we already control the diagonal contribution $I = J$ of (46). By Cauchy-Schwarz we also control the near-diagonal contribution when I and J are adjacent. It remains to consider the contribution when I and J are not adjacent. In this case $|t - s| \gtrsim R$. To exploit this we rewrite this contribution as

$$|\sum_{I,J} \int_{\Omega \cap I} \int_{\Omega \cap J} \langle P_1^2 U(s-t)(\psi_I(t)F(t)), \psi_J(s)F(s) \rangle \, dsdt|$$

The decay estimate shows that $P_1^2 U(s-t)$ maps L^1 to L^∞ with a norm of $O(R^{-(n-1)/2})$, thus we can estimate this by

$$R^{-(n-1)/2} \sum_{I,J} \int_{\Omega \cap I} \int_{\Omega \cap J} E(\psi_I)\|F\|_{L_t^\infty L_x^2} E(\psi_J)\|F\|_{L_t^\infty L_x^2} \, dsdt$$

which we can easily bound by

$$R^{-(n-1)/2}|\Omega|^2((\sum_I E(\psi_I))^{1/2}\|F\|_{L_t^\infty L_x^2})^2,$$

which is acceptable if we choose $R = |\Omega|^C$ sufficiently large. (Note that we do not need the optimal decay of $(n-1)/2$ to run this argument). This proves (45); the extension to (42) is done similarly and is omitted. This completes the derivation of (40).

• In what follows we sketch a slightly different proof of Theorem 1.1, in which the parallel case is treated somewhat differently. We start, as before, with the tube

decomposition

$$\phi \approx \sum_{T \in \mathbf{T}} \phi_T; \quad \psi \approx \sum_{T' \in \mathbf{T}'} \psi_{T'}.$$

We observe that the wave packets ϕ_T, ψ_T verify the properties (7) and (8) of definition 2.2, not only for $R/2 \leq t \leq 5R/2$ but also at $t = 0$. This can easily be checked from either of the two construction of the wave packets. For simplicity we shall assume that in fact, at $t = 0$, both ϕ_T, ψ_T are localized to balls of radius $r = R^{\frac{1}{2}+\varepsilon}$. As before we write:

$$Q(\phi, \psi) \approx \sum_{T \in \mathbf{T}} \sum_{T' \in \mathbf{T}'} Q(\phi_T, \psi_{T'}) \approx Q_{||} + Q_{\times}$$

with the transversal interaction component

$$Q_{\times} = \sum_{\kappa,\kappa' \in \mathbf{K}:\kappa \sim \kappa'} Q(\phi_\kappa, \psi_{\kappa'})$$

and the parallel interaction component

$$Q_{||} = \sum_{\kappa,\kappa' \in \mathbf{K}:\kappa \approx \kappa', l(\kappa)=l(\kappa')=\rho_0} Q(\phi_\kappa, \psi_{\kappa'}).$$

Recall that,

$$\phi_\kappa := \sum_{T \in \mathbf{T}(\kappa)} \phi_T \text{ and } \psi_\kappa := \sum_{T' \in \mathbf{T}'(\kappa)} \psi_{T'}.$$

Due to the additional localization properties of the wave packets at $t = 0$ we observe that $\phi_k(0), \psi_k(0)$ are essentially localized in the cube k. We start by estimating $Q_{||}$. Recalling, the estimate

$$\|Q(\phi_k, \psi_{k'})(t)\|_2 \lesssim \frac{1}{R}(\|\Gamma\phi_k(t)\|_\infty\|\nabla_{x,t}\psi_{k'}(t)\|_2 + \|\nabla_{x,t}\phi_k\|_\infty\|\Gamma\psi_{k'}(t)\|_2),$$

we write

$$\begin{aligned}
\|Q_{||}\|_{L^2(\mathbf{Q}_R^{(0)})} &\lesssim \sum_{\kappa,\kappa' \in \mathbf{K}:\kappa \approx \kappa', l(\kappa)=l(\kappa')=\rho_0} \|Q(\phi_\kappa, \psi_{\kappa'})\|_{L^2(\mathbf{Q}_R^{(0)})} \\
&\lesssim R^{\frac{1}{2}}(\frac{\rho_0}{R})^{\frac{n+1}{2}} \sum_{\kappa,\kappa' \in \mathbf{K}:\kappa \approx \kappa', l(\kappa)=l(\kappa')=\rho_0} E(\phi_k(0))^{\frac{1}{2}} E(\psi_{k'}(0))^{\frac{1}{2}} \\
&\lesssim R^{\frac{1}{2}}(\frac{\rho_0}{R})^{\frac{n+1}{2}} E(\phi)^{\frac{1}{2}} E(\psi)^{\frac{1}{2}}
\end{aligned}$$

We can therefore chose $\rho_0 = R^{\frac{n}{n+1}}$ rather than $\rho_0 \approx R^{\frac{1}{2}+\varepsilon}$ as used earlier. The proof of the estimate for $Q_{\times}$ proceeds exactly as before.

9. Proof of Theorem 1.2

Fix ϕ, ψ, μ, λ, q, r. By a scaling argument we may take $\lambda = 1$. We may assume that $\mu \lesssim 1$ since the left-hand side vanishes otherwise. Our task is now to prove

$$\|P_\mu(\phi\psi)\|_{L_t^{q/2}L_x^{r/2}(\mathbf{R}^{n+1})} \lesssim \mu^{n-\frac{4}{q}-\frac{2n}{r}} E(\phi)^{1/2} E(\psi)^{1/2}. \tag{47}$$

If one simply discards the P_μ projection and uses Hölder and Strichartz then one does not obtain the decay in μ in (47). To recover this decay we will use the averaging property of P_μ at scale $1/\mu$ (given by (2)), interacted against the fact that the fundamental solution at frequency 1 is essentially supported on a set of "thickness" 1. (In this we are very much in the spirit of [22]).

We divide space $\mathbf{R}^n$ into cubes Q of side-length $1/\mu$. From (5) we have

$$\|P_\mu(\phi(t)\psi(t))\|_{L_x^{r/2}(\mathbf{R}^n)} = (\sum_Q \|P_\mu(\phi(t)\psi(t))\|_{L_x^{r/2}(Q)}^{r/2})^{2/r}.$$

By (2) we have

$$\|P_\mu(\phi(t)\psi(t))\|_{L_x^{r/2}(Q)} \lesssim \mu^{n(1-\frac{2}{r})} \sum_{Q'}(1 + \mu\mathrm{dist}(Q,Q'))^{-100n}\|\phi(t)\psi(t)\|_{L_x^1(Q)}.$$

Inserting this into the previous and using Young's inequality, we obtain

$$\|P_1(\phi(t)\psi(t))\|_{L_x^{r/2}(\mathbf{R}^n)} \lesssim \mu^{n(1-\frac{2}{r})}(\sum_Q \|\phi(t)\psi(t)\|_{L_x^1(Q)}^{r/2})^{2/r}.$$

To prove (47), it thus suffices by Hölder and Cauchy-Schwartz to show the "improved Strichartz" estimate

$$\|(\sum_Q \|\phi(t)\|_{L_x^2(Q)}^r)^{1/r}\|_{L_t^q} \lesssim \mu^{-\frac{2}{q}}E(\phi)^{1/2}. \tag{48}$$

together with the identical estimate for ψ.

It suffices to prove (48) for ϕ. We remark that this estimate has essentially appeared in [22], and we shall repeat the proof here.

Observe that the claim is trivial when $q = \infty$ (thanks to the embedding $l^r \subset l^2$), so we will assume that $q < \infty$. This forces r to be strictly greater than 2.

We shall use the abstract Strichartz estimate in [9] (Theorem 10.1). Let H be the Hilbert space of initial data $\phi[0]$ with frequency 1 with the energy inner product. We let $U(t) : H \to H$ be the map $U(t)\phi[0] := \phi[t]$; note that this map is unitary on H. For any $0 \le \theta \le 1$, let B_θ denote the Banach space given by the norm

$$\|\phi[t]\|_{B_\theta} := \mu^{-\theta\sigma}(\sum_Q \|\phi[t]\|_{L_x^2(Q)}^{r_\theta})^{1/r_\theta}$$

where $r_\theta := 2/(1+\theta)$ and $\sigma := \frac{2r}{q(r-2)}$. Since (q,r) is admissible we observe that $0 < \sigma \le (n-1)/2$. Observe that the dual space B_θ^* is given by

$$\|\phi[t]\|_{B_\theta^*} := \mu^{\theta\sigma}(\sum_Q \|\phi[t]\|_{L_x^2(Q)}^{r_\theta'})^{1/r_\theta'}.$$

Also, one has the real interpolation embedding $(B_0^*, B_1^*)_{\theta,2} \subseteq B_\theta^*$ (see e.g. [2]).

To prove (48) it will suffice to prove the estimate

$$\|U(t)\phi[0]\|_{L_t^q B_\theta^*} \lesssim \|\phi[0]\|_H$$

with $\theta := 1 - 2/r$ (so that $r_\theta' = r$).

Observe that $q = 2/\sigma\theta$. By the abstract Strichartz estimate in [9], Theorem 10.1, we see that it will suffice to prove the decay estimate

$$\|U(t)U(s)^*\phi[s]\|_{B_1^*} \lesssim |t - s|^{-\sigma}\|\phi[s]\|_{B_1}$$

for all s, t. From the group properties of the evolution we see that $U(t)U(s)^* = U(t - s)$, so by time translation invariance it suffices to show that

$$\|U(t)\phi[0]\|_{B_1^*} \lesssim |t|^{-\sigma}\|\phi[0]\|_{B_1}.$$

We expand out the norms to write this as

$$\mu^{\sigma} \sup_Q \|\phi[t]\|_{L_x^2(Q)} \lesssim \mu^{-\sigma}|t|^{-\sigma} \sum_Q \|\phi[0]\|_{L_x^2(Q)}.$$

It thus suffices to show that

$$\|\phi[t]\|_{L_x^2(Q)} \lesssim (1/\mu^2|t|)^{\sigma} \sum_{Q'} \|\phi[0]\|_{L_x^2(Q')}$$

for all cubes Q of side-length μ.

We may assume that $|t| \gg \mu^{-2}$ since the claim follows immediately from energy estimates otherwise.

We shall assume zero initial velocity $\phi_t(0) = 0$; the case of zero initial position is similar and is left to the reader. By the reproducing formula (1) we may write $\phi(0) = \phi(0) * \psi(0)$ for some Schwartz function $\psi(0)$ with frequency 1. We extend ψ to be a free wave by setting zero initial velocity $\psi(0) = 0$. By uniqueness we see that $\phi[t] = \phi(0) * \psi[t]$, thus $\psi[t]$ is a frequency-localized fundamental solution for the wave equation. By Young's inequality we thus see that it suffices to show the decay estimate

$$\|\psi[t]\|_{L_x^1(Q)} \lesssim (1/\mu^2|t|)^{\sigma}$$

for all cubes Q of side-length $1/\mu$.

Since $\psi[0]$ is a Schwartz function, we have the standard decay estimate

$$|\nabla_{x,t}^k \psi(x, t)| \lesssim |t|^{-(n-1)/2}(1 + ||t| - |x||)^{-k}$$

(see e.g. [19]; one can also use the machinery of Section 4 to achieve this bound). Since ψ has frequency 1, we thus have in particular that

$$|\psi[t](x)| \lesssim |t|^{-(n-1)/2}(1 + ||t| - |x||)^{-k}$$

for any k. A computation then shows that

$$\|\psi[t]\|_{L_x^1(Q)} \lesssim (1/\mu^2|t|)^{(n-1)/2}.$$

Since $\sigma \leq (n - 1)/2$, the claim follows. This completes the proof of Theorem 1.2.

10. Appendix: Proof of Corollaries 1.3, 1.4

We now prove Corollary 1.3. We first begin with the preliminary remark that in Theorem 1.1 the spacetime cube $\mathbf{Q}_R$ can be replaced with the slab $[0, R] \times \mathbf{R}^n$. To see this we divide into cases.

Case 1: $R \min(\lambda, \mu) \geq 1$.

In this case we use finite speed of propagation and smooth cutoff functions applied to the initial data to replace the energy $E(\phi)$ in (4) by the more localized quantity

$$\int (|\phi_t(x,0)|^2 + |\nabla\phi(x,0)|^2)(1 + \frac{\text{dist}\,((x,0),\mathbf{Q}_R)}{R})^{-100n} \, dx.$$

(One has to construct the cutoff functions so as not to destroy the property of having frequency λ, but since $R\lambda \geq 1$ this can be done without violating the uncertainty principle. See [23] for further discussion). Similarly for $E(\psi)$. The claim then follows by partitioning $[0, R] \times \mathbf{R}^n$ into cubes $\mathbf{Q}_R$, applying the above localized estimates, and then Cauchy-Schwarz.

Case 2: $R \min(\lambda, \mu) \leq 1$.

Without loss of generality we may assume that $\lambda \leq \mu$. From the definition of energy and Bernstein's inequality we see that

$$\|\nabla_{x,t}\phi\|_{L_t^\infty L_x^\infty} \lesssim \lambda^{n/2} E(\phi)^{1/2}$$

and

$$\|\nabla_{x,t}\psi\|_{L_t^\infty L_x^2} \lesssim E(\psi)^{1/2}.$$

The claim then follows from Hölder.

We now return to the proof of Corollary 1.3. Fix Q, ϕ, ψ. We use suitably chosen Littlewood-Paley projections to split

$$Q(\phi, \psi) = \sum_{\lambda, \mu} Q(P_\lambda \phi, P_\mu \psi).$$

We split into three cases: the high-low case $\lambda \gg \mu$, the high-high case $\lambda \sim \mu$, and the low-high case $\lambda \ll \mu$.

In the high-low case we see that $Q(P_\lambda \phi, P_\mu \psi)$ has Fourier transform supported in the region $|\xi| \sim \lambda$. By orthogonality we thus have

$$\| \sum_{\lambda \gg \mu} Q(P_\lambda \phi, P_\mu \psi)\|_{L_{t,x}^2([0,1] \times \mathbf{R}^n)} \lesssim (\sum_\lambda \| \sum_{\mu : \mu \ll \lambda} Q(P_\lambda \phi, P_\mu \psi)\|_{L_{t,x}^2([0,1] \times \mathbf{R}^n)}^2)^{1/2}.$$

First consider the contribution when $\mu \lesssim 1$. Applying Theorem 1.1 (with the previously mentioned modification) we may estimate this by

$$\lesssim (\sum_\lambda (\sum_{\mu : \mu \lesssim 1} \mu^{n/2} E(P_\lambda \phi)^{1/2} E(P_\mu \psi)^{1/2})^2)^{1/2}.$$

This easily sums to be bounded by $E(\phi)^{1/2} E(\psi)^{1/2}$, which is acceptable.

Now consider the contribution when $\mu \gtrsim 1$. Again using Theorem 1.1, we may bound this by

$$\lesssim (\sum_\lambda (\sum_{\mu : 1 \lesssim \mu \ll \lambda} \mu^{(n-1)/2+\varepsilon} E(P_\lambda \phi)^{1/2} E(P_\mu \psi)^{1/2})^2)^{1/2}.$$

This can be controlled by

$$\lesssim (\sum_\lambda (\lambda^{(n-1)/2+\varepsilon} E(P_\lambda \phi)^{1/2} E(\psi)^{1/2})^2)^{1/2} \lesssim E(\langle D \rangle^{(n-1)/2+\varepsilon} \phi)^{1/2} E(\psi)^{1/2}$$

as desired.

In the high-high case we use the triangle inequality followed by Theorem 1.1. In the $\mu \lesssim 1$ case we thus obtain a bound of

$$\sum_{\lambda \sim \mu \lesssim 1} \mu^{n/2} E(P_\lambda \phi)^{1/2} E(P_\mu \psi)^{1/2} \lesssim E(\phi)^{1/2} E(\psi)^{1/2}$$

which is acceptable. In the $\mu \gtrsim 1$ case we obtain a bound of

$$\sum_{\lambda \sim \mu \gtrsim 1} \lambda^{(n-1)/2+\varepsilon} E(P_\lambda \phi)^{1/2} E(P_\mu \psi)^{1/2}$$

which by Cauchy-Schwarz is bounded by $E(\langle D \rangle^{(n-1)/2+\varepsilon} \phi)^{1/2} E(\psi)^{1/2}$ as desired.

The low-high case is similar to the high-low case. We only show the treatment of the $\lambda \gtrsim 1$ contribution. By orthogonality and Theorem 1.1, this is bounded by

$$\lesssim (\sum_\mu (\sum_{\lambda:1 \lesssim \lambda \ll \mu} \lambda^{(n-1)/2+\varepsilon} E(P_\lambda \phi)^{1/2} E(P_\mu \psi)^{1/2})^2)^{1/2} \lesssim (\sum_\mu E(\langle D \rangle^{(n-1)/2+2\varepsilon} \phi) E(P_\mu \psi))^{1/2}$$

which is acceptable (replacing ε by $\varepsilon/2$ as necessary). This concludes the proof of Corollary 1.3.

We now prove Corollary 1.4. We use suitably chosen Littlewood-Paley projections and the triangle inequality to split

$$\||D|^{-\sigma}(\phi\psi)\|_{L_t^{q/2} L_x^{r/2}} \le \sum_{\lambda,\mu} \||D|^{-\sigma}(P_\lambda \phi P_\mu \psi)\|_{L_t^{q/2} L_x^{r/2}}$$

and consider the high-low, high-high, and low-low cases separately.

In the high-low case $\lambda \gg \mu$ the symbol $|D|^{-\sigma}$ can be replaced by $\lambda^{-\sigma}$. We then use Hölder and Strichartz' inequality (see e.g. [9]) to estimate

$$\||D|^{-\sigma}(P_\lambda \phi P_\mu \psi)\|_{L_t^{q/2} L_x^{r/2}} \lesssim \lambda^{-\sigma} \|P_\lambda \phi\|_{L_t^q L_x^r} \|P_\mu \psi\|_{L_t^q L_x^r}$$
$$\lesssim \lambda^{-\sigma} \lambda^{\frac{n}{2}-\frac{1}{q}-\frac{n}{r}} E(P_\lambda \phi)^{1/2} \mu^{\frac{n}{2}-\frac{1}{q}-\frac{n}{r}} E(P_\mu \psi)^{1/2}.$$

Thus we may estimate this contribution by

$$\sum_\lambda \lambda^{-\sigma} \lambda^{\frac{n}{2}-\frac{1}{q}-\frac{n}{r}} E(P_\lambda \phi)^{1/2} (\sum_{\mu \ll \lambda} \mu^{\frac{n}{2}-\frac{1}{q}-\frac{n}{r}} E(P_\mu \psi)^{1/2}).$$

The inner sum can be estimated by

$$\sum_{\mu \ll \lambda} \mu^{\frac{n}{2}-\frac{1}{q}-\frac{n}{r}} E(P_\mu \psi)^{1/2} \lesssim \lambda^{\frac{\sigma}{2}} E(|D|^{\frac{n-\sigma}{2}-\frac{1}{q}-\frac{n}{r}} \psi)^{1/2},$$

and the claim then follows by performing the λ summation.

The low-high case $\lambda \ll \mu$ is similar, so we turn to the high-high case $\lambda \sim \mu$. To simplify the notation we shall take $\lambda = \mu$, although this does not substantially affect the argument. By Cauchy-Schwarz it suffices to show that

$$\||D|^{-\sigma}(P_\lambda \phi P_\lambda \psi)\|_{L_t^{q/2} L_x^{r/2}} \lesssim E(|D|^{\frac{n-\sigma}{2}-\frac{1}{q}-\frac{n}{r}}P_\lambda \phi)^{1/2} E(|D|^{\frac{n-\sigma}{2}-\frac{1}{q}-\frac{n}{r}}P_\lambda \psi)^{1/2}.$$

From scale invariance we may take $\lambda \sim 1$, so that the derivatives $|D|$ on the right hand side may be discarded.

The function $P_\lambda \phi P_\lambda \psi$ has frequency $\lesssim 1$. By the Littlewood-Paley decomposition and the triangle inequality it thus suffices to show that

$$\|P_\nu |D|^{-\sigma}(P_\lambda \phi P_\lambda \psi)\|_{L_t^{q/2} L_x^{r/2}} \lesssim \nu^\varepsilon E(P_\lambda \phi)^{1/2} E(P_\lambda \psi)^{1/2}$$

for some $\varepsilon > 0$. But we may replace $|D|^{-\sigma}$ by $\nu^{-\sigma}$, and the claim then follows from Theorem 1.2 and the hypothesis $\sigma < n - \frac{4}{q} - \frac{2n}{r}$.

References

[1] H. Bahouri and J.Y. Chemin *Cubic quasilinear wave equation*, preprint
[2] J. Bergh, J. Löfström, *Interpolation Spaces: An Introduction*, Springer-Verlag, 1976.
[3] J. Bourgain, *Besicovitch-type maximal operators and applications to Fourier analysis*, Geom. and Funct. Anal. **22** (1991), 147–187.
[4] J. Bourgain, *Estimates for cone multipliers*, Operator Theory: Advances and Applications, **77** (1995), 41–60.
[5] J. Bourgain, *On the dimension of Kakeya sets and related maximal inequalities*, Geom. Funct. Anal. **9** (1999), no. 2, 256–282.
[6] A. Córdoba *The Kakeya maximal function and spherical summation multipliers.* Am. J. Math. **99**, (1977), 1-22.
[7] C. Fefferman *A note on spherical summation multipliers* Israel J. Math. **15** (1973), 44-52.
[8] D. Foschi, S. Klainerman, *Bilinear Space-time Estimates for Homogeneous Wave Equations*, Ann. Scient. ENS, 4^e serie, **33**(2000), 211-274.
[9] M. Keel, T. Tao, *Endpoint Strichartz Estimates*, Amer. Math. J. 120 (1998), 955–980.
[10] S. Klainerman, M. Machedon, *Space-time Estimates for Null Forms and the Local Existence Theorem*, Comm. Pure Appl. Math., **46** (1993), 1221–1268.
[11] S. Klainerman, M. Machedon, *Remark on Strichartz type inequalities*, with appendices by J. Bourgain and D. Tataru, I.M.R.N , 1996, No 5, 201-220.
[12] S. Klainerman, *A commuting vectorfield approach to Strichartz type inequalities and applications to quasilinear wave equations*, preprint.
[13] S. Klainerman, *Uniform decay estimates and the Lorentz invariance of of the classical wave equation* Comm.Pure Appl.Math. **35**(1985), 321-332.
[14] S. Klainerman, *The null condition and global existence to nonlinear wave equations*, Lect. Appl. Math. **23**(1986), 293-326.
[15] S.Klainerman, I. Rodnianski *Improved local well posedness for quasilinear wave equations in dimension three*, preprint
[16] S. Klainerman, S. Selberg, *Bilinear estimates and applications to nonlinear wave equations*, preprint.
[17] S. Klainerman, D. Tataru, *On the optimal regularity for Yang-Mills equations in $\mathbf{R}^{4+1}$*, J. Amer. Math. Soc. 12 (1999), 93–116.
[18] S. Selberg, *Multilinear space-time estimates and applications to local existence theory for non-linear wave equations*, Princeton University Thesis.
[19] C. D. Sogge, *Lectures on Nonlinear Wave Equations*, Monographs in Analysis II, International Press, 1995.
[20] E. M. Stein, *Harmonic Analysis*, Princeton University Press, 1993.
[21] T. Tao, *The Bochner-Riesz conjecture implies the Restriction conjecture*, Duke Math J. **96** (1999), 363-376.

[22] T. Tao, *Low regularity semi-linear wave equations*, Comm. PDE **24** (1999), 599–630.

[23] T. Tao, *Endpoint bilinear restriction theorems for the cone, and some sharp null form estimates*, to appear, Math Z.

[24] T. Tao, *Multilinear weighted convolution of L^2 functions, and applications to non-linear dispersive equations*, to appear, Amer. J. Math.

[25] T. Tao, A. Vargas, *A bilinear approach to cone multipliers I. Restriction theorems*, GAFA **10** (2000), 185–215.

[26] T. Tao, A. Vargas, *A bilinear approach to cone multipliers II. Applications*, GAFA **10** (2000), 216–258.

[27] T. Tao, A. Vargas, L. Vega, *A bilinear approach to the restriction and Kakeya conjectures*, J. Amer. Math. Soc. **11** (1998), pp. 967–1000.

[28] D. Tataru, *Null form estimates for second order operators with rough co-efficients*, preprint.

[29] T. H. Wolff, *A sharp bilinear cone restriction estimate*, to appear, Annals of Math.

[30] T. H. Wolff, *Local smoothing type estimates on L^p for large p*, preprint.

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON NJ 08544

E-mail address: seri@math.princeton.edu

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON NJ 08544

E-mail address: irod@math.princeton.edu

DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES CA 90095-1555

E-mail address: tao@math.ucla.edu