Null form estimates for second order hyperbolic operators with rough coefficients

Daniel Tataru *
Department of Mathematics
Northwestern University

March 15, 2001

Abstract

It has been known for some time that in the $L^p_t L^q_x$ bilinear dispersive estimates for the wave equation the worst interaction occurs between spatially concentrated waves which travel in the same direction. Such interactions, however, are essentially canceled for a special class of bilinear expressions, called null forms. In this article we prove certain null form estimates for second order hyperbolic equations with rough coefficients. This extends some earlier L^p results of Wolff and Tao for the constant coefficient case, and also some L^2 null form estimates for operators with rough coefficients due to Smith and Sogge.

1 Introduction

We first describe the Strichartz estimates for the constant coefficient wave equation,

$$\Box u = 0 \quad in\ \mathbb{R}^n \times \mathbb{R}.$$

We denote by X the space of L^2 solutions to the homogeneous equation,

$$\|u\|_X^2 = \|u(0)\|_{L^2}^2 + \|u_t(0)\|_{\dot H^{-1}}^2.$$

One can think of this norm as an energy functional, since it is conserved in time.

We say that the function u is localized at frequency λ if the Cauchy data $(u(0), u_t(0))$ has Fourier transform supported in the region $\{|\xi| \in [\lambda, 4\lambda]\}$. We denote by X_λ the closed

*Research partially supported by NSF grant DMS-9622942 and by an Alfred P. Sloan fellowship

subspace of functions in X which are localized at frequency λ. Then one can express the Strichartz estimates in the form

$$\|u\|_{L^p_t L^q_x} \lesssim \lambda^s \|u\|_X, \qquad u \in X_\lambda.$$

We rewrite this as

$$X_\lambda \subset \lambda^s L^p(L^q) \tag{1}$$

and use this type of notation from here on. The exponent s is determined by scaling,

$$\frac{1}{p} + \frac{n}{q} = \frac{n}{2} - s,$$

and the exponents $2 \leq p, q \leq \infty$ satisfy the relation

$$\frac{2}{p} + \frac{n-1}{q} \leq \frac{n-1}{2}, \tag{2}$$

with the exception of the forbidden endpoint $(p, q, n) = (2, \infty, 3)$. Note that the interesting case is when the equality holds, the rest follows by Sobolev embeddings. Also one can recover the complete family of estimates from two endpoints[1], namely

$$\begin{array}{ccc} (\infty, 2) & (4, \infty) & n = 2 \\ (\infty, 2) & (2, \dfrac{2(n-1)}{n-3}) & n \geq 3. \end{array}$$

We refer the reader to the survey article Ginibre-Velo [3] and references therein, and also to Keel-Tao [4] for the endpoint estimates. If one considers instead second order hyperbolic operators with rough coefficients $g^{ij}\partial_i\partial_j$ then similar estimates hold locally in time if the coefficients are C^2 (see Smith [8], Tataru [12]) or more generally if $\nabla^2 g \in L^1 L^\infty$ (see Tataru [11]).

To see that the range of exponents (p, q) in (2) is sharp it suffices to look at Knapp's counterexample. This is a highly focused wave which at frequency λ and time-scale 1 is like a smooth bump function in a parallelepiped of size $\lambda^{-1} \times (\lambda^{-\frac{1}{2}})^{n-1} \times 1$. To understand its spatial orientation one can think of a $\lambda^{-1} \times (\lambda^{-\frac{1}{2}})^{n-1}$ parallelepiped at time 0, which travels in time with speed 1 in the "normal" direction. From here on we call such solutions more concisely *λ-wave packets.*

On the Fourier transform side, the initial data for λ-wave packets is concentrated in a dual region, namely a sector of the annulus $\{|\xi| \in [\lambda, 4\lambda]\}$ of size $\lambda \times (\lambda^{\frac{1}{2}})^{n-1}$. This corresponds to an angle of size $\lambda^{\frac{1}{2}}$.

[1] Strictly speaking this claim is false for $n = 3$ where the endpoint estimate is false; however, it is "almost" true

Consider now bilinear estimates of the same type. Then the situation becomes more complicated, because one needs to consider interactions between very different frequencies. There are two cases to consider:

a) The two factors have approximately the same frequency λ. Then the product is localized at frequency at most λ, therefore it is interesting to consider expressions of the form $S_\mu(X_\lambda \cdot X_\lambda)$ with $\mu \lesssim \lambda$. Here S_μ is a multiplier which selects the frequencies of size $O(\mu)$.

b) The two factors have very different frequencies μ and λ. Then the product is localized at frequency λ, therefore it suffices to look at expressions of the form $X_\lambda \cdot X_\mu$.

Corresponding to these two cases we now describe the bilinear estimates. However, instead of writing the complete family, it is easier to write the relevant endpoints, from which everything else follows by interpolation and Sobolev embeddings:

$$\begin{array}{llll} S_\mu(X_\lambda \cdot X_\lambda) \subset L^\infty L^1, & \mu\lambda^{\frac{1}{2}} L^2 L^\infty & & n=2, \\ S_\mu(X_\lambda \cdot X_\lambda) \subset L^\infty L^1, & \mu^{\frac{2}{n-1}}\lambda L^2 L^{\frac{n-1}{n-3}} & & n \geq 3. \end{array} \tag{3}$$

respectively

$$\begin{array}{llll} X_\lambda \cdot X_\mu \subset L^\infty L^1, & \mu^{\frac{3}{4}} L^4 L^2, & \mu^{\frac{3}{4}}\lambda^{\frac{3}{4}} L^2 L^\infty & n=2, \\ X_\lambda \cdot X_\mu \subset L^\infty L^1, & \mu^{\frac{n+1}{2(n-1)}} L^2 L^{\frac{n-1}{n-2}}, & \mu\lambda^{\frac{2}{n-1}} L^1 L^{\frac{n-1}{n-3}} & n=3,4, \\ X_\lambda \cdot X_\mu \subset L^\infty L^1, & \mu^{\frac{n+1}{2(n-1)}} L^2 L^{\frac{n-1}{n-2}}, & \mu^{\frac{n+3}{2(n-1)}}\lambda^{\frac{1}{2}} L^2 L^{\frac{n-1}{n-3}} & n \geq 5. \end{array} \tag{4}$$

Note that in these estimates it becomes important to have the correct balance of powers of λ and μ. In several cases this differs from the trivial one obtained simply by applying twice the linear estimate (1). This is the case with the second part of (3) which was proved in [6]. It is also the case with the third part of (4) for $n \geq 4$ is still a conjecture (see also [2]).

All these bilinear estimates are sharp in the case when both factors are λ-wave packets or μ-wave packets traveling in the same direction (or combinations thereof on different time scales). On the Fourier transform side, this means that the sectors of size $\lambda^{-\frac{1}{2}}$ or $\mu^{-\frac{1}{2}}$ where the Fourier transform of the corresponding initial data is concentrated are oriented in the same direction. However, it is best to look at this using a space-time Fourier transform. In space-time we denote by ξ the time Fourier variable and by τ the time Fourier variable. The symbol for $\Box$ is $\tau^2 - \xi^2$, which vanishes on the characteristic cone

$$K = \{\tau^2 - \xi^2 = 0\}.$$

The Fourier transform of a solution to the homogeneous wave equation is a distribution supported on K. If one multiplies two such solutions, this translates into a convolution in the Fourier space. Then it becomes interesting to look at sums of covectors on K. Observe that the sum of two such covectors is "close" to K if and only if the two covectors are "almost" colinear. This corresponds exactly to the case of the product of two wave packets which travel in the same direction.

It was observed some time ago in Klainerman-Machedon [5], Bourgain [1] that the range of exponents (p, q) for which bilinear estimates hold can be increased if one avoids the above mentioned type of interaction. This can be done either by replacing the product by a null-form, i.e. a bilinear form whose symbol vanishes when its two arguments are colinear and on the characteristic cone, or simply by assuming some angular separation between the Fourier transforms of the two factors.

Some of Klainerman-Machedon's L^2 null form estimates were later extended by Smith-Sogge [9] to second order hyperbolic operators with C^2 coefficients. However, obtaining L^p estimates for $p \neq 2$ turns out to be a considerably more difficult problem.

In a recent article Foschi and Klainerman [2] provide a tentative description of all possible estimates of the type

$$D^{\beta_0} D_+^{\beta_+} D_-^{\beta_-} (X_\lambda \cdot X_\mu) \subset \lambda^{\alpha_1} \mu^{\alpha_2} L^p L^q$$

where D, D_+ and D_- are multipliers with symbols $|\xi|$, $|\tau| + |\xi|$ respectively $|\tau| - |\xi|$. Note that D_- is the symbol which vanishes on vectors on the characteristic cone, thereby providing some cancellation for products of waves which travel in the same direction.

The conjectured range of p, q is[2]

$$\frac{2}{p} + \frac{n+1}{q} \leq n+1 \qquad 1 \leq p, q \leq \infty. \tag{5}$$

One should compare this range with the (doubled) range in (2), which is sharp for estimates where no cancellation occurs.

Inequalities of this type were studied recently by Wolff [13] and then Tao [10] in the special case when $p = q$. They consider products of two solutions to the homogeneous wave equation for which the supports Fourier transforms are separated by an $O(1)$ angle. If we denote such products by $X_\lambda \hat{\cdot} X_\mu$, then their result can be stated as

$$X_\lambda \hat{\cdot} X_\mu \subset \lambda^{\frac{n-1}{2(n+3)} + \epsilon} \mu^{\frac{n-1}{2(n+3)} - \epsilon} L^{\frac{n+3}{n+1}} \qquad \epsilon > 0.$$

Wolff's original result is an L^p estimate for $p > \frac{n+3}{n+1}$ and $\lambda = \mu$, while Tao obtains the endpoint and extends the estimate to the case $\lambda \neq \mu$. Combining this result with dyadic

[2]To this one should add the restriction $p > \frac{4}{3}$ in dimension $n = 2$

and angular decompositions and with Lorentz transforms, Tao [10] observes that this implies most of the Foschi-Klainerman conjecture in the case $p = q$.

Our goal here is twofold. On one hand we want to extend the results of Wolff and Tao to the case when $p \neq q$, while on the other hand we want to have the result applicable to second order hyperbolic equations with rough coefficients. To gain in generality in terms of the regularity of the coefficients we make some compromises with respect to the type of estimates we prove.

On one hand, we allow a loss of an ϵ in the power of μ in the estimates. In this regard our estimates are comparable to Wolff's non-endpoint result. Conceivably one could use some of Tao's arguments in [10] to gain the endpoints, but this would increase the complexity of the argument by orders of magnitude.

On the other hand, sharp angular separation is not easy to define in the rough coefficient case, nor is it easy to define the operator D_-. Consequently, we limit ourselves to bilinear null-form estimates. This is not a serious restriction since on one hand these are the type of estimates which are needed for applications, while on the other hand in the constant coefficient case one can combine the null form estimates with Lorentz transforms to recover the estimates in the form stated by Foschi-Klainerman.

Finally, we focus on L^pL^q estimates with p, q on the end line

$$\frac{2}{p} + \frac{n+1}{q} = n+1. \tag{6}$$

Of course, for larger p and/or q one can use Sobolev embeddings and/or interpolate with the bilinear Strichartz estimates (3), (4). However, there are still two end-points missing, one of which is the counterpart of Tao's missing endpoint in [10].

We consider a second order hyperbolic operator

$$\Box_g = g^{ij}\partial_i\partial_j$$

with variable coefficients g^{ij}. All our estimates are true when $\nabla^2 g \in L^1L^\infty$. However, instead of making such an assumption on the metric we postulate the existence of a parametrix, constructed as a superposition of wave packets. Then we show how the estimates follow from certain geometric properties of the wave packets. This has two advantages. On one hand it makes the connection between the properties of the Hamilton flow and the estimates very explicit. On the other hand the result is more robust and can be applied for certain metrics with less regularity which arise in the study of nonlinear hyperbolic equations (see Smith-Tataru [7]).

A null form is a bilinear quadratic form

$$B(u,v) = b^{ij}\partial_i u \partial_j v$$

whose symbol $b^{ij}\xi_i\eta_j$ vanishes when ξ and η are colinear and on the characteristic cone

$$K = \{g^{ij}\xi_i\xi_j = 0\}.$$

All null forms are linear combinations of two special types of null forms, namely

$$B_{ij}(u,v) = \partial_i u \partial_j v - \partial_j u \partial_i v, \quad B_0(u,v) = g^{ij}\partial_i u \partial_j v.$$

In our estimates we work with approximate solutions for the wave equation, therefore we should allow for in homogeneous terms in our equations. Hence we need an in homogeneous replacement for the X_λ spaces. Consequently, we set X^1 to be the space with norm

$$\|u\|_{X^1} = \|\nabla u\|_{L^\infty L^2} + \|\Box_g u\|_{L^1 L^2}.$$

The subspace of X^1 which contains functions with space-time Fourier transform supported in $\{\lambda \leq |\xi| \leq 2\lambda\}$ is denoted by X^1_λ. This is only a naive definition, it will be modified later to enlarge the space and make it more stable with respect to various operations.

The estimates we prove in this article are local in time. For the reader's convenience we state them here in the same format as (3) and (4), including the trivial $L^\infty L^1$ bound coming from the energy estimates. Let $\epsilon > 0$. For $\mu \leq \lambda$ we get

$$\begin{array}{llll} S_\mu B(X^1_\lambda, X^1_\lambda) \subset \mu\lambda^{-1} L^\infty L^1, & \lambda^{\frac14}\mu^\epsilon L^{\frac43} L^2 & & n=2, \\ S_\mu B(X^1_\lambda, X^1_\lambda) \subset \mu\lambda^{-1} L^\infty L^1, & \lambda^{\frac{2}{n+1}}\mu^{\frac{n-3}{n+1}+\epsilon} L^1 L^{\frac{n+1}{n-1}} & & n \geq 3. \end{array} \tag{7}$$

while for $\mu \ll \lambda$ we have

$$\begin{array}{llll} B(X^1_\lambda, X^1_\mu) \subset L^\infty L^1, & \mu^{\frac16+\epsilon} L^2 L^{\frac32}, & \lambda^{\frac14}\mu^\epsilon L^{\frac43} L^2 & n=2, \\ B(X^1_\lambda, X^1_\mu) \subset L^\infty L^1, & \mu^{\frac{n-1}{2(n+1)}+\epsilon} L^2 L^{\frac{n+1}{n}} & \lambda^{\frac12}\mu^{\frac{n-3}{2(n+1)}+\epsilon} L^1 L^{\frac{n+1}{n-1}} & n \geq 3. \end{array} \tag{8}$$

Interpolating between these endpoints we obtain a set of sharp estimates (modulo the ϵ exponent) on the line (6). As mentioned before, there are two endpoint estimates away from this line but within the range conjectured by Foschi-Klainerman which do not follow from the above ones. One applies to the $(\lambda,\lambda) \to \mu$ case,

$$S_\mu B(X^1_\lambda, X^1_\lambda) \subset \mu\lambda^{-\frac14} L^{\frac43} L^4 \quad n=2,$$

$$S_\mu B(X^1_\lambda, X^1_\lambda) \subset \mu L^1 L^{\frac{n}{n-2}} \qquad n \geq 3$$

and the second to the $(\lambda,\mu) \to \lambda$ case:

$$B(X^1_\lambda, X^1_\mu) \subset \lambda^{\frac14}\mu^{\frac13} L^{\frac43} L^3 \quad n=2,$$

$$B(X^1_\lambda, X^1_\mu) \subset \lambda^{\frac12}\mu^{\frac{n^2-3}{2(n+1)}} L^1 L^{\frac{2(n+1)}{n-1}} \quad n \geq 3.$$

2 The wave equation parametrix

This section contains our assumptions on the metric g and on the corresponding hyperbolic equation. All the analysis we do is local in time. Hence, because of the finite speed of propagation, we can localize the estimates spatially as well. To fix the speed of propagation we assume that the coefficients g^{ij} are a small L^∞ perturbation of the coefficients of the d'Allembertian and that the speed of propagation is smaller than 1. We set

$$Q = \{|t| \leq 1, |x_i| < 1\}$$

to be the region where we want to obtain the bilinear estimates. For $C > 0$ we denote by CQ the cube with the same center but C times larger in all directions. By $C_x Q$ we denote the parallelepiped with the same center and the same time length, but whose spatial size is C times larger. In the sequel we assume that the coefficients are defined in a larger region $9_x Q$. By $X_0^1(Q)$ we denote the functions in $X^1(Q)$ whose Cauchy data vanishes on the time-like (i.e. lateral) boundary of Q. The notation $C_0^\infty(Q)$ has a similar meaning. On the metric g we impose a minimal regularity assumption, namely

(P1) *The metric g satisfies $\|\nabla g\|_{L^1 L^\infty(9_x Q)} \ll 1$.*

Here we assume smallness only in order to make this assumption consistent with the remaining assumptions. In general this can be always achieved by dividing the time interval into smaller subintervals.

Remark 1. *It is useful to observe that this assumption is scale invariant, i.e. for any cube $R \subset Q$ of size r the metric $g(rx, rt)$ uniformly satisfies the same condition in R.*

However, we make a stronger assumption concerning the existence of a wave packet type parametrix for the wave equation. The idea is quite simple, namely to decompose the frequency λ localized solutions u to $\Box_g u = 0$ into a sum of almost orthogonal wave packets. One can think of the wave packets as the building blocks for solutions to the wave equation.

There is more than one way of constructing wave packet parametrices for the wave equation. In the constant coefficient case this is essentially done by decomposing the initial data in both space and frequency on scales consistent with the uncertainty principle which remain coherent on the unit time scale, see e.g. Wolff [13]. In the variable coefficient case, a different wave packet parametrix was constructed in Smith [8] for operators with $C^{1,1}$ coefficients. His approach extents with little change to the case when $\partial^2 g \in L^1 L^\infty$ (see also Tataru [11]). Recently, a similar parametrix was constructed in Smith-Tataru [7] for certain operators whose coefficients have even less regularity, arising in the study of nonlinear second order hyperbolic equations.

We do not want to tie our present result to any single wave packet parametrix. Instead we assume that for each frequency λ we have a family $\mathcal{T}_\lambda$ of subsets of $8_x Q$ which we call λ-*tubes*. In addition, for each $T \in \mathcal{T}$ there is a family of functions in $8_x Q$ supported in T which we call *normalized* λ-*wave packets*. The properties of the wave packet parametrix for frequency λ localized solutions to the homogeneous equation $\Box_g u = 0$ are summarized as follows:

(P2) *Let* $\lambda > 1$. *Let* S_λ *be a multiplier whose symbol is supported in* $\{|\xi| \in [\frac{\lambda}{8}, 16\lambda]\}$. *Let* $\chi \in C_0^\infty(9_x Q)$. *Then*

(i) For any collection $\{u_T\}_{T\in\mathcal{T}_\lambda}$ *of normalized* λ*-wave packets we have*

$$\|\chi S_\lambda \sum_{T\in\mathcal{T}_\lambda} a_T u_T\|_{X^1}^2 \lesssim \sum_{T\in\mathcal{T}_\lambda} |a_T|^2.$$

(ii) For each solution $v \in X_0^1(6_x Q)$ *to the homogeneous problem* $\Box_g u = 0$ *there is a* λ*-wave packet superposition* u *which depends linearly on* u *of the form*

$$u = \sum_{T\in\mathcal{T}_\lambda} a_T u_T$$

so that

$$\sum_{T\in\mathcal{T}_\lambda} |a_T|^2 \lesssim \|u\|_{X^1(Q)}^2$$

and

$$\|\chi S_\lambda (u - v)\|_{X^1(Q)} \ll \|v\|_{X^1(Q)}.$$

Furthermore, this is assumed to hold uniformly at all scales less than 1 *(see Remark 1).*

Clearly, obtaining estimates for superpositions of wave packets must use some geometric information on the λ-tubes and λ-wave packets. The complete geometry is, however, not necessary for our arguments. Instead, we only need one geometric condition, namely a characteristic energy estimate. This would be easier to state in the absence of S_λ in condition (P2), namely

$$\|\sum_{T}^{P\in T} B(u_T, v)\|_{L^2(Q\setminus B(P,\delta))} \lesssim \delta^{-N} \lambda^{\frac{n-1}{4}} \|v\|_{X_0^1(Q)}. \tag{9}$$

However, the multiplier S_λ is needed because the functions u_T are sharply localized in the physical space so they cannot be simultaneously localized in frequency. To deal with this uncertainty principle problem we allow some more freedom in the estimate. Given a λ-tube T and a point P we define

$$\chi_T(P) = (1 + \lambda d(P,T))^{-N},$$

where N is a large number. Then the desired characteristic energy estimate is

(P3) *Let $\chi \in C_0^\infty(Q)$. For any point P in Q and any collection u_T of normalized λ-wave packets we have*

$$\sum_{T \in \mathcal{T}_\lambda} \chi_T^2(P) \|\chi_T^{-1} B(\chi S_\lambda u_T, v)\|^2_{L^2(Q \setminus B(P,\delta))} \lesssim \delta^{-N} \lambda^{\frac{n-1}{2}} \|v\|^2_{X_0^1(Q)}. \tag{10}$$

Furthermore, the above property is assumed to hold uniformly at all scales less than 1 *(see Remark 1).*

Remark 2. *The characteristic energy estimate (10) is true for Smith's parametrix, and it is relatively easy to obtain. This is discussed in the Appendix. Furthermore, although the matter clearly becomes more delicate, it appears that it also holds in the setup of Smith-Tataru [7].*

3 The main result

Before we state the main result we address the question of defining precisely the X_λ^1 spaces; one issue we must consider comes from the uncertainty principle, as we want to localize both in space and in frequency. Secondly, our proof uses a wave packet decomposition in which the range of frequencies is increased by a factor. This would not cause any problems in a direct argument, but is somewhat annoying in a proof as this which uses induction on scales. These factors motivate the set-up which follows.

The first step is to make precise choices for a spatial cutoff function χ in Q and the frequency cutoff multiplier S_λ. Let ϕ be a smooth nonnegative function with the following properties:

(i) $\phi = 1$ in $(-\infty, -1]$, $\phi > 0$ in $(-1, 0)$ and $\phi = 0$ in $(0, \infty)$.

(ii) $\phi(s) = e^{\frac{1}{s}}$ for $s \in [-\frac{1}{2}, 0)$.

Define the function ν supported in Q_4 which equals 1 in Q_3 by

$$\chi(x) = \prod_{i=1}^{n} \phi(x_i^2 - 1)$$

and the spatial multiplier S_λ whose symbol

$$s_\lambda(\xi) = \phi(\lambda^{-2}\xi^2 - 64)\phi(1 - 16\lambda^{-2}\xi^2)$$

is supported in $\{|\xi| \in [\lambda/4, 8\lambda]\}$ and equals 1 in $\{|\xi| \in [\lambda/2, 4\lambda]\}$. We also introduce the space-time multiplier T_λ with symbol

$$t_\lambda(\xi, \tau) = \phi(\lambda^{-2}(\xi^2 + \tau^2) - 64).$$

Definition 3. *$X^1_\lambda \subset X^1_0(Q)$ is the space of functions which can be represented as*

$$u = \chi S_\lambda v + r,$$

where $v \in X^1_0(6_x Q)$ and $\nabla r \in L^\infty C^N$, with norm

$$\|u\|_{X^1_\lambda} = \inf\{\|v\|_{X^1(6_x Q)} + \lambda^N \|\nabla r\|_{L^\infty(\dot{C}^N)};\ u = \nu A_\lambda u_1 + u_2\}$$

In other words, functions in X^1_λ are localized at frequency λ and in the interior of Q modulo $O(\lambda^{-N})$ errors. Here N is a large number. One can think of this definition as a more robust way of dealing with the notion of *margin* which is used by Wolff [13] and Tao [10].

Now we are ready to state our main result.

Theorem 1. *Assume the metric g satisfies the assumptions (P1-3). Let $\epsilon > 0$. Suppose B is a null form with respect to the metric g whose coefficients satisfy $\nabla b^{ij} \in L^1 L^\infty$.*

a) Let $1 \leq \mu \lesssim \lambda$. Then

$$T_\mu B(X^1_\lambda, X^1_\lambda) \subset \lambda^{\frac{1}{4}} \mu^\epsilon L^{\frac{4}{3}} L^2 \qquad n = 2, \tag{11}$$

$$T_\mu B(X^1_\lambda, X^1_\lambda) \subset \lambda^{\frac{2}{n+1}} \mu^{\frac{n-3}{n+1}+\epsilon} L^1 L^{\frac{n+1}{n-1}}, \qquad n \geq 3. \tag{12}$$

b) $1 \leq \mu \ll \lambda$. Then

$$B(X^1_\lambda, X^1_\mu) \subset \mu^{\frac{n-1}{2(n+1)}+\epsilon} L^2 L^{\frac{n+1}{n}} \qquad n \geq 2, \tag{13}$$

$$B(X^1_\lambda, X^1_\mu) \subset \lambda^{\frac{1}{4}} \mu^\epsilon L^{\frac{4}{3}} L^2 \qquad n = 2, \tag{14}$$

$$B(X^1_\lambda, X^1_\mu) \subset \lambda^{\frac{1}{2}} \mu^{\frac{n-3}{2(n+1)}+\epsilon} L^1 L^{\frac{n+1}{n-1}} \qquad n \geq 3. \tag{15}$$

To understand the applicability range of this result recall that[3] the assumptions (P1-3) are satisfied for $\partial^2 g \in L^1 L^\infty$. Hopefully, they can also be verified for the less regular coefficients arising in the context of the work in [7] on nonlinear hyperbolic equations.

[3] See Remark 2.

The remainder of the paper is devoted to the proof of this result. The main idea is the same as in the work of Wolff, namely to combine the wave packet decomposition with an inductive argument on scales. The parametrix we use here is different from the one Wolff [13] and Tao [10] use; the reason is that our parametrix seems to be more robust for low regularity coefficients (see again [7]).

Ideally, one would like to have a direct argument which only uses only the size of the wave packets, without taking advantage of any cancellations due to oscillations. This idea, however, does not work even in the constant coefficient case, because without oscillations one cannot localize the Fourier transform of wave packets away from the origin. Thus we get a low frequency overlapping of the wave packets which suffices in order to blow up all the estimates.

To avoid this difficulty, one tries to split the quadratic form into two parts. One contains relatively few wave packets interactions, and can be estimated simply based on the size of the wave packets combined with energy estimates. The second part contains most of the wave packets interactions, but instead has the redeeming feature that it is localized on a smaller spatial scale. After rescaling back to size 1, the estimate for the second part is identical to the original one, but at a lower frequency. This leads to the idea of induction with respect to scales. This idea works in this form for the proof of all estimates in the Theorem except (15), for which one needs a more sophisticated variation on the same theme based on an idea in Tao [10].

The outline of the paper is as follows. In the following section we show how to represent frequency localized functions in X^1 as truncated superpositions of wave packets; this essentially follows from a fixed point argument combined with the variation of parameters formula. Next we explore some simple properties of the X^1_λ spaces, including estimates for spatial localizations at different scales. Then we present the key bilinear decomposition lemma. This is followed in Section 7 by the inductive argument on scales. Finally, the Appendix contains a discussion of Smith's parametrix and of the characteristic energy estimate in (P3). Note that the null condition is only used in order to derive (P3) and improve the energy estimates in the $(\lambda, \lambda) \to \mu$ case.

4 Exact solutions vs the parametrix

The wave packet parametrix provides us with approximate solutions to the homogeneous wave equation, while the estimates we want to prove apply to exact solutions to the inhomogeneous wave equation. To bridge this gap we need the following representation of the exact solutions to the inhomogeneous wave equation as superpositions of truncated wave-packet combinations:

Proposition 4. *Let $\lambda > 1$. Let $u \in X_0^1(6_x Q)$. Then there is a representation*

$$\chi S_\lambda u = \chi S_\lambda \left(u_{-1} + \int_{-1}^{1} 1_{\{t \geq s\}} u_s(t,x) dt \right) + r, \tag{16}$$

with

$$\chi S_\lambda u_s = \chi S_\lambda \sum_{j=1}^{\infty} u_s^j + r_s, \qquad u_s^j = \sum_{T \in T_\lambda} a_{s,T}^j u_{s,T}^j \qquad s \in [-1,1],$$

where

(i) the functions u_s satisfy $u_s(s,x) = 0$ for $s \in (-1,1]$.

(ii) The remainders r, r_s are smooth and small, i.e. for any fixed large N we have

$$\|\nabla r\|_{L^\infty(C^N)} + \|\nabla r_{-1}\|_{L^\infty(C^N)} + \int_{-1}^{1} \|\nabla r_s\|_{L^\infty(C^N)} ds \leq \lambda^{-N} \|u\|_{X_0^1(6_x Q)}.$$

(iii) $u_s^{j,T}$ are normalized λ-wave packets.

(iv) The coefficients $a_{s,T}^j$ satisfy

$$\int_{-1}^{1} \sum_{j=1}^{\infty} \left(\sum_{T \in \mathcal{T}} |a_{s,T}^j|^2 \right)^{\frac{1}{2}} ds \lesssim \|u\|_{X_0^1(6_x Q)}.$$

Proof. We observe that it suffices to replace (16) with

$$\chi S_\lambda (u_1 - w) = \chi S_\lambda \left(u_{-1} + \int_{-1}^{1} 1_{\{t \geq s\}} u_s(t,x) dt \right) + r \tag{17}$$

for some $w \in X_0^1(6_x Q)$ which satisfies

$$\|w\|_{X_0^1(6_x Q)} \ll \|u\|_{X_0^1(6_x Q)}.$$

Indeed, if we know this then the Proposition follows using an iterative argument.

We represent $u = \chi S_\lambda v + r$ as in Definition 3. Then $v \in X_0^1(6_x Q)$. Since the speed of propagation is at most 1 we can represent it in terms of Pv using the variation of parameters formula,

$$v = u_{-1} + \int_{-1}^{1} 1_{t \geq s} u_s(t,x) ds \qquad \text{in } 1.5_x Q$$

where the functions $u_s \in X_0^1(6_x Q)$ solve

$$\Box_g u_s = 0, \qquad u_s(s) = 0, \quad \partial_t u_s(s) = (1_{4_x Q} P u)(s) \quad s \in [0,1],$$

$$\Box_g u_{-1} = 0, \qquad u_{-1}(-1) = \tilde{\chi} v(-1), \quad \partial_t u_{-1}(-1) = \tilde{\chi} \partial_t v(-1).$$

Here $\tilde{\chi} \in C_0^\infty(4_x Q)$ is a smooth cutoff which equals 1 in $3.5_x Q$. Then

$$\|u_{-1}\|_{X^1} + \int_{-1}^{1} \|u_s(t,x) ds\|_{X^1} \lesssim \|v\|_{X^1(6_x Q)}.$$

Since $\chi \in C_0^\infty(Q)$ we can easily bound the error due to the truncation of initial data for u_s,

$$\|\nabla \chi S_\lambda (v - u_{-1} - \int_{-1}^{1} 1_{t \geq s} u_s(t,x) ds)\|_{L^\infty C^N} \lesssim \lambda^{-N} \|u\|_{X^1(6_x Q)}.$$

To conclude the proof it remains to find an appropriate representation for the functions u_s. This follows from the following lemma with $f = 0$ and $g = (1_{4_x Q} P u_1)(s)$ for $s \in (-1, 1]$, respectively $f = \tilde{\chi} v(-1)$ and $g = \tilde{\chi} \partial_t v(-1)$ for $s = -1$.

Lemma 5. *Let $s \in [-1,1]$ and u_s be the solution to*

$$\Box_g u_s = 0, \qquad u_s(s) = f \in H^1, \quad \partial_t u_s(s) = g \in L^2$$

with f, g supported in $4_x Q$. Then there are functions v_s, w_s, r_s so that

$$\chi S_\lambda (u_s - w_s) = \chi S_\lambda v_s + r_s, \quad v_s = \sum_j v_s^j, \quad v_s^j = \sum_{T \in T_\lambda} a_{s,T}^j u_{s,T}^j$$

where

(i) the function w_s is in $X_0^1(6_x Q)$, satisfies $w_s(s) = 0$, $\partial_t w_s(s) = 0$ and

$$\|w_s\|_{X^1(6_x Q)} \ll \|f\|_{H^1} + \|g\|_{L^2}. \tag{18}$$

(ii) The remainder r_s is smooth and small, i.e. for any fixed large N we have

$$\|\nabla r\|_{L^\infty(C^N)} \lesssim \lambda^{-N} (\|f\|_{H^1} + \|g\|_{L^2}).$$

(iii) $u_s^{j,T}$ are normalized λ-wave packets.

(iv) The coefficients $a_{s,T}^j$ satisfy

$$\sum_j \left(\sum_{T \in \mathcal{T}} |a_{s,T}^j|^2 \right)^{\frac{1}{2}} ds \lesssim \|f\|_{H^1} + \|g\|_{L^2}.$$

To prove the Lemma we apply (P2) to the function u_s. We redenote the function u in (P2) by v_s and set

$$w_s = \tilde{\chi}(u_s - v_s).$$

Then by (P2) it follows that (18), (iii) and (iv) are satisfied. The remainder r_s is given by

$$r_s = \chi S_\lambda (1 - \tilde{\chi}) w_s.$$

But χ and $(1 - \tilde{\chi})$ have disjoint supports so (ii) holds as well.

The problem is, the condition $w_s(s) = \partial_t w_s(s) = 0$ is not satisfied. Instead, we only get the bound

$$\|w_s(s)\|_{H^1} + \|\partial_t w_s(s)\|_{L^2} \lesssim \|w\|_{X^1} \ll \|f\|_{H^1} + \|g\|_{L^2}.$$

However, if we have this then we can repeat the same argument with (f, g) replaced by $(w_s(s), \partial_t w_s(s))$ and iteratively arrive at the case when $w_s(s) = \partial_t w_s(s) = 0$. $\square$

5 Cubes and spaces

We begin with some simple observations about the X^1_λ space. Since both S_λ and multiplication by χ map X^1 into itself, it follows that $X^1_\lambda \subset X^1_0$. Next, note that without any restriction in generality we can assume that the function v in Definition 3 is supported inside $4_x Q$. Indeed, if we replace v by $\tilde{\chi} v$ then u is modified by $\chi S_\lambda (1 - \tilde{\chi}) v$. Since the supports of χ and $1 - \tilde{\chi}$ are separated, this is an $O(\lambda^{-N})$ term which can be included in the smooth remainder r.

The function v is not necessarily localized at frequency λ, but, once we have v supported inside $4_x Q$, this can almost be achieved by replacing v by $\tilde{S}_\lambda v$, where $\tilde{S}_\lambda$ is another multiplier localized at frequency λ and whose symbol is 1 within the support of the symbol of S_λ. This is not completely correct since we loose the support information on v. However, the part of $\tilde{S}_\lambda v$ which is outside $5_x Q$ is of size $O(\lambda^{-N})$ and can be harmlessly cut off. Hence we can replace v by $\tilde{S}_\lambda v + O(\lambda^{-N})$. This gives us better leverage on v, for instance by using estimates of the following type:

$$\|\tilde{S}_\lambda v\|_{L^\infty(L^2)} \lesssim \lambda^{-1} \|\nabla v\|_{L^\infty(L^2)}.$$

We summarize the above discussion as follows:

Lemma 6. *Any function $u \in X^1_\lambda$ can be represented as $u = \chi S_\lambda v + r$ where $v \in X^1_0(6_x Q)$ and*

$$\|v\|_{X^1} + \lambda \|v\|_{L^\infty L^2} + \lambda^N \|\nabla r\|_{L^\infty(\dot{C}^N)} \lesssim \|u\|_{X^1_\lambda}.$$

We next observe that X^1_λ is invariant with respect to multiplication by smooth functions.

Lemma 7. *Let ϕ be a smooth function. Then*

$$\|\phi u\|_{X^1_\lambda} \lesssim \|u\|_{X^1_\lambda}.$$

Proof. As u is supported in Q, without any restriction in generality we assume that ϕ is supported in $4_x Q$. Start with $u = \chi S_\lambda v + r$ where

$$\|v\|_{X^1} + \lambda^N \|\nabla r\|_{L^\infty \dot{C}^N} \leq 2\|u\|_{X^\lambda_1}.$$

Then we want a similar decomposition for ϕu, $\phi u = \chi S_\lambda \tilde{v} + \tilde{r}$. The term ϕr is easily placed in $\tilde{r}$, so it remains to look at $\phi \chi S_\lambda v$. Naively we would like to conjugate ϕ by S_λ modulo small errors which we can put into $\tilde{r}$. The difficulty is that S_λ is not invertible, but this can be overcome by a careful truncation argument. We choose a spherically symmetric cutoff function $a(\xi)$ which equals 1 in $\{s_\lambda > 2\lambda^{-10N}\}$ and 0 in $\{s_\lambda < \lambda^{-10N}\}$. Given the definition of s_λ, it follows that a must vary from 0 to 1 on the $\frac{\lambda}{\ln \lambda}$ scale, i.e.

$$|\partial_\xi^\beta a| \leq c_\beta \left(\frac{\lambda}{\ln \lambda}\right)^{-|\beta|}.$$

We use the formal series for the product ϕS_λ up to the order $10N$ to write

$$\phi S_\lambda = \sum_{|\alpha|<10N} \frac{1}{i^{|\alpha|}\alpha!} (a\partial_\xi^\alpha s_\lambda)(D)(\partial_x^\alpha \psi) + R(x, D).$$

By the standard pseudodifferential calculus, the symbol of the remainder R satisfies

$$|\partial_x^\alpha \partial_\xi^\beta r(x,\xi)| \leq c_{\alpha\beta} \lambda^{-10N} \left(\frac{\lambda}{\ln \lambda}\right)^{-|\beta|}$$

and decays rapidly beyond frequency λ. This implies that the contribution of R to ϕu is negligible and can be placed in $\tilde{r}$. On the other hand we write

$$\chi \sum_{|\alpha|<10N} \frac{1}{i^{|\alpha|}\alpha!} (a\partial_\xi^\alpha s_\lambda)(D)(\partial_x^\alpha \phi) v = \chi S_\lambda \left(\sum_{|\alpha|<10N} \frac{1}{i^{|\alpha|}\alpha!} \left(\frac{a\partial_\xi^\alpha s_\lambda}{s_\lambda}\right)(D)(\partial_x^\alpha \phi)\right) v,$$

and claim that we can take

$$\tilde{v} = \left(\sum_{|\alpha|<10N} \frac{1}{i^{|\alpha|}\alpha!} \left(\frac{a\partial_\xi^\alpha s_\lambda}{s_\lambda} \right) (D)(\partial_x^\alpha \phi) \right) v.$$

For this we need to prove that

$$\|(\frac{a\partial_\xi^\alpha s_\lambda}{s_\lambda})(D)(\partial_x^\alpha \phi) v\|_{X^1} \lesssim \|v\|_{X^1}.$$

The case $\alpha = 0$ is trivial and the multiplication by $\partial_x^\alpha \phi$ is harmless, the difficulty may come from the multiplier with symbol

$$m(\xi) = \frac{a\partial_\xi^\alpha s_\lambda}{s_\lambda}, \qquad |\alpha| \geq 1.$$

But from the definition of b and a_λ we get

$$|\partial_\xi^\beta m(\xi)| \leq c_\beta \left(\frac{\lambda}{\ln \lambda} \right)^{-|\beta|} \left(\frac{\lambda}{(\ln \lambda)^2} \right)^{-|\alpha|}.$$

Hence this is in effect a nice multiplier so the desired estimate

$$\|m(D)u\|_{X^1} \lesssim \|u\|_{X^1}.$$

follows easily. □

Given a cube $R \subset Q$ and $0 < \gamma < 1$ we define a smaller parallelepiped $R_\gamma^- \subset R$ as follows. If the lateral boundary of R does not intersect the lateral boundary of Q then we set $R_\gamma^- = (1-\gamma)_x R$. Otherwise, we define R_γ^- to be the parallelepiped obtained by extending $(1-\gamma)_x R$ up to the common lateral boundary of R and Q.

Given $\delta > 0$ we cover the unit cube Q with a regular array $\mathcal{Q}^\delta$ of sub-cubes of size δ so that

$$Q = \bigcup_{R \in \mathcal{Q}^\delta} R = \bigcup_{R \in \mathcal{Q}^\delta} R_{\frac{1}{2}}^-.$$

Correspondingly we choose a smooth partition of unity in Q,

$$1_Q = \sum_{R \in \mathcal{Q}_\delta} \phi_R, \qquad \text{supp } \phi_R \subset R,$$

with $\phi_R \in C^\infty(R)$, supported in $R_{\frac{1}{4}}^-$, and with uniform bounds on the δ scale,

$$|\partial^\alpha \phi_R| \leq c_\alpha \delta^{-\alpha}, \qquad R \in \mathcal{Q}_\delta. \tag{19}$$

Note that the functions ϕ_R are compactly supported in R unless the boundary of R intersects the boundary of Q. However, we must allow ϕ_R to be nonzero in a compact subset of the common boundary of R and Q.

Let $\mathcal{Q}_\delta^t$ be the set of δ cubes in $\mathcal{Q}_\delta$ which intersect the t time slice. By J_δ we denote a set of equidistant times in $[-1,1]$ with separation δ.

Given $R \in \mathcal{Q}^\delta$ we denote by $X^1_\lambda(R)$ the space of functions which is obtained by rescaling R to scale 1, using Definition 3 with λ replaced by $\delta\lambda$, and then scaling back with the scaling factor which leaves the X^1 norm unchanged. Then the norm on $X^1_\lambda(R)$ is

$$\|u\|_{X^1_\lambda(R)} = \inf\{\|v\|_{X^1(6_x R)} + \delta^{2N+\frac{n}{2}}\lambda^N \|\nabla r\|_{L^\infty \dot{C}^N};\ u = \chi_R S_\lambda v + r\}.$$

where χ_R is the rescaled counterpart of χ.

Lemma 8. *Let $\delta > \lambda^{-1+\epsilon}$ and $t \in [-1,1]$. For $R \in \mathcal{Q}_\delta^t$ we consider smooth functions $\tilde{\phi}_R \in C^\infty(R)$ supported in $R^-_{\frac{1}{8}}$, with uniform bounds on the R scale as in* (19).

Then for $u \in X^1_\lambda$ we have

$$\sum_{R\in\mathcal{Q}_\delta^t} \|\tilde{\phi}_R u\|^2_{X^1_\lambda(R)} \lesssim \|u\|^2_{X^1_\lambda}. \tag{20}$$

Proof. For $R \in \mathcal{Q}_\delta^t$ we choose a cutoff function $\psi_R \in C_0^\infty(6_x R)$ which is 1 in $4_x R$. We represent $u = \chi S_\lambda v + r$ with v, r as in Lemma 6. Then we write

$$\tilde{\phi}_R u = b_R \chi_R S_\lambda v_R + r_R.$$

where

$$v_R = \psi_R v, \quad r_R = \psi_R r + \tilde{\phi}_R \chi S_\lambda (1-\psi_R) v, \quad b_R = \tilde{\phi}_R \frac{\chi}{\chi_R}.$$

We first note that the the functions b_R are uniformly smooth on the δ scale. This is trivial if R is not near the boundary of Q, when $\tilde{\phi}_R$ is supported away from the lateral boundary of R. Otherwise, we need to take a closer look at what happens near the common lateral boundary of R and Q. There $\tilde{\phi}_R$ needs not be zero and χ_R^{-1} is singular; however, we can take advantage of the fact that χ decays rapidly near the boundary of Q. More precisely, near a common face of Q and R, say $x_j = 1$, due to the definition of χ, χ_R we get factors of the form

$$e^{-(x_j^2-1)^{-1}} e^{[\delta^{-2}(x_j-1+\delta)^2-1]^{-1}},$$

which are smooth on the δ scale in a δ neighborhood $\{1-\delta \leq x_j \leq 1\}$ of the common boundary.

By Lemma 7 applied in the rescaled setting of the cubes in $\mathcal{Q}_\delta^t$, it remains to prove that

$$\sum_{R\in\mathcal{Q}_\delta^t} \left(\|v_R\|^2_{X^1(6_x R)} + \delta^{4N+n}\lambda^N \|\nabla r_R\|^2_{L^\infty(\mathbb{C}^N)} \right) \lesssim \|v\|^2_{X^1(6_x Q)} + \lambda^2 \|v\|^2_{L^\infty L^2} + \lambda^N \|\nabla r\|^2_{L^\infty \dot{C}^N}.$$

The bound for v_R follows easily from energy estimates combined with the finite speed of propagation. The first term in r_R can be estimated by

$$\|\nabla \psi_R r\|_{L^2(\dot{C}^N)} \lesssim \delta^{-N-1} \|\nabla r\|_{L^2(\dot{C}^N)},$$

which gives a better bound than needed, namely

$$\sum_{R\in\mathcal{Q}_\delta^t} \|\nabla \psi_R r\|^2_{L^2(\dot{C}^N)} \lesssim \delta^{-2N-n-1} \|\nabla r\|^2_{L^2(\dot{C}^N)}.$$

The second term in r_R is negligible since the kernel of S_λ decays rapidly beyond the λ^{-1} scale while the supports of $\tilde{\phi}_R$ and $(1-\psi_R)$ are separated by $\delta \geq \lambda^{-1+\epsilon}$. $\square$

6 The bilinear decomposition

The main step in the inductive argument is the following bilinear decomposition lemma.

Lemma 9. *a) Let $1 \leq \mu$, $\mu^{-1+\epsilon} \leq \delta \leq 1$. Let $u \in X^1_\mu$ be a superposition of normalized μ-wave packets,*

$$u = \chi S_\mu \sum_{T\in\mathcal{T}_\mu} a_T u_T,$$

and $v \in X^1_0(Q)$. Then there are functions $u_R \in X^1_\mu(R)$ for $R \in \mathcal{Q}_\delta$ so that the following properties are satisfied:

(i) (square summability)

$$\sum_{R\in\mathcal{Q}_\delta} \|u_R\|^2_{X^1_\mu(R)} \lesssim \sum_{T\in\mathcal{T}_\mu} |a_T|^2. \tag{21}$$

(ii) (remainder estimate)

$$\|B(u,v) - \sum_{R\in\mathcal{Q}_\delta} \phi_R B(u_R, v)\|^2_{L^2} \lesssim \delta^{-N} \mu^{\frac{n-1}{4}} (\sum_{T\in\mathcal{T}_\mu} |a_T|^2) \|v\|^2_{X^1}. \tag{22}$$

b) Assume that $1 \le \lambda$, $r < 1$, $\lambda^{-1+\epsilon} \le \delta \le 1$. Let $u_L \in X^1_0(L)$ for $L \in \mathcal{Q}_r$. Let $v \in X^1_\lambda$ be a superposition of normalized μ-wave packets,

$$v = \chi S_\lambda \sum_{S \in \mathcal{T}_\mu} b_S v_S,$$

Then there are functions $v_R \in X^1_\lambda(R)$ for $R \in \mathcal{Q}_\delta$ so that the following properties are satisfied:

(i) (square summability)

$$\sum_{R \in \mathcal{Q}_\delta} \|v_R\|^2_{X^1_\lambda(R)} \lesssim \sum_{S \in \mathcal{T}_\mu} |b_S|^2. \tag{23}$$

(ii) (remainder estimate)

$$\| \sum_{L \in \mathcal{Q}_r} \phi_L (B(u_L, v) - \sum_{R \in \mathcal{Q}_\delta} \phi_R B(u_L, v_R))\|^2_{L^2} \lesssim \delta^{-N} \lambda^{\frac{n-1}{4}} (\sum_{L \in \mathcal{Q}_r} \|u_L\|^2_{X^1_0(L)})(\sum_{S \in \mathcal{T}_\mu} |b_S|^2). \tag{24}$$

Note that part (a) is a special case of part (b). We state and prove it separately, however, since the ideas involved are more transparent in this case.

Proof. a) We associate each μ-tube T to δ cubes which roughly carry most of the interaction of u_T with v. More precisely, we say that $T \sim R$ if

$$\|\chi_T^{-1} B(\chi S_\mu u_T, v)\|_{L^2(Q \setminus 8R)} \le \frac{1}{2} \|\chi_T^{-1} B(\chi S_\mu u_T, v)\|_{L^2(Q)}.$$

We choose cutoff functions $\tilde\phi_R \in C^\infty(R)$, supported in $R^-_{\frac{1}{8}}$ which equal 1 in the support of ϕ_R. Define

$$u_R = \tilde\phi_R \sum_{T \sim R} a_T \chi S_\mu u_T.$$

It is clear that each tube $T \in \mathcal{T}_\mu$ is associated to at most finitely many cubes. Then we use (P2)(i) and Lemma 8 separately for each δ cube R to obtain the condition (i). It remains to prove (22). Denote

$$E(u, v) = B(u, v) - \sum_{R \in \mathcal{Q}_\delta} \phi_R B(u_R, v) = \sum_{R \in \mathcal{Q}_\delta} \phi_R \sum_{T \not\sim R} a_T B(\chi S_\mu u_T, v).$$

For $P \in Q$ and $T \in \mathcal{T}_\mu$ we say that $P \sim T$ if $P \in R$ for some δ-cube $R \in \mathcal{Q}_\delta$ so that $T \sim R$. Then

$$|E(u, v)(P)| \le \sum_{T \in \mathcal{T}_\mu}^{P \not\sim T} |a_T||B(\chi S_\mu u_T, v)(P)|.$$

Using the condition $P \not\sim T$ we further get

$$|E(u,v)(P)| \leq \sum_{T\in\mathcal{T}_\mu}^{P\not\sim T} \frac{|a_T|\chi_T^{-1}(P)|B(\chi S_\mu u_T,v)(P)|}{\|\chi_T^{-1}B(\chi S_\mu u_T,v)\|_{L^2(Q)}}\chi_T(P)\|\chi_T^{-1}B(\chi S_\mu u_T,v)\|_{L^2(Q\setminus B(P,4\delta))}.$$

Applying the Cauchy-Schwartz inequality with respect to T in the previous estimate we obtain

$$|E(u,v)(P)|^2 \leq \left(\sum_{T\in\mathcal{T}_\mu} \frac{a_T^2|\chi_T^{-1}B(\chi S_\mu u_T,v)(P)|^2}{\|\chi_T^{-1}B(\chi S_\mu u_T,v)\|^2_{L^2(Q)}}\right)\left(\sum_{T\in\mathcal{T}_\mu}\chi_T^2(P)\|\chi_T^{-1}B(\chi S_\mu u_T,v)\|^2_{L^2(Q\setminus(B(P,4\delta))}\right).$$

According to the characteristic energy estimates (10), the second factor is bounded by $\delta^{-N}\mu^{\frac{n-1}{2}}\|v\|^2_{X^1}$ so we have

$$|E(u,v)(P)|^2 \lesssim \delta^{-N}\mu^{\frac{n-1}{2}}\left(\sum_{T\in\mathcal{T}_\mu} \frac{a_T^2|\chi_T^{-1}B(\chi S_\mu u_T,v)(P)|^2}{\|\chi_T^{-1}B(\chi S_\mu u_T,v)\|^2_{L^2(Q)}}\right)\|v\|^2_{X^1}.$$

Integrating this inequality we obtain

$$\|E(u,v)\|^2_{L^2} \lesssim \delta^{-N}\mu^{\frac{n-1}{2}}(\sum_{T\in\mathcal{T}_\mu} a_T^2)\|v\|^2_{X^1}.$$

and conclude the proof.

b) As before, for $S \in \mathcal{T}_\lambda$ and $R \in \mathcal{Q}_\delta$ we say that $S \sim R$ if

$$\sum_{L\in\mathcal{Q}_r} \|\chi_S^{-1}\phi_L B(u_L,\chi S_\lambda v_S)\|^2_{L^2(Q\setminus 8R)} \leq \frac{1}{4}\sum_{L\in\mathcal{Q}_r}\|\chi_S^{-1}\phi_L B(u_L,\chi S_\lambda v_S)\|^2_{L^2(Q)}.$$

For $R \in \mathcal{Q}_\delta$ set

$$v_R = \tilde{\phi}_R \sum_{S\sim R} a_S \chi S_\lambda v_S.$$

Again each λ-tube S is associated to at most finitely many δ cubes, which by (P2)(i) and Lemma 8 implies the square summability relation (i). It remains to estimate the error term

$$E(u,v) = \sum_{L\in\mathcal{Q}_r}\phi_L(B(u_L,v) - \sum_{R\in\mathcal{Q}_\delta}\phi_R B(u_L,v_R)) = \sum_{L\in\mathcal{Q}_r}\phi_L\sum_{R\in\mathcal{Q}_\delta}\phi_R\sum_{S\in\mathcal{T}_\lambda}^{S\not\sim R} b_S B(u_L,\chi S_\lambda v_S).$$

For each point $P \in Q$ there are finitely many indices L (depending only on the dimension) for which $\phi_L(P) \neq 0$. Then

$$\begin{aligned}|E(u,v)(P)| &\leq \sum_{S\in\mathcal{T}_\lambda}^{P\not\sim S}|b_S||\sum_{L\in\mathcal{Q}_r}\phi_L B(u_L,\chi S_\lambda v_S)(P)| \\ &\lesssim \sum_{S\in\mathcal{T}_\lambda}^{P\not\sim S}|b_S|(\sum_{L\in\mathcal{Q}_r}|\phi_L B(u_L,\chi S_\lambda v_S)(P)|)^{\frac{1}{2}}.\end{aligned}$$

From here on the argument proceeds exactly as in case (a). The modified characteristic energy estimates in (10) are now applied separately for each function u_L.

$\square$

7 The inductive argument

Proof of (11), (12): We denote by $M(\lambda, \mu)$ the best constant in (11) or (12) with $\epsilon = 0$ with respect to all scales less or equal than 1 and with respect to all null forms B whose coefficients are in a bounded set,

$$\|b^{ij}\|_{L^\infty} + \|\nabla b^{ij}\|_{L^1 L^\infty} \leq 1.$$

Note carefully that λ and μ stand for the frequencies after rescaling into an unit cube, and not before. Then we seek a recursive relation for $M(\lambda, \mu)$. For this proof it is sufficient to have the function M well defined up to a fixed multiplicative constant. This is unlike in the similar argument in [10] for the endpoint estimate. We claim that

$$M(\lambda, \mu) \leq C(\delta^{-N} + M(\delta\lambda, \delta\mu)) \qquad \mu^{-1+\epsilon} \leq \delta \leq 1. \tag{25}$$

We postpone for now the proof of (25) and instead show that it implies (11), (12) via an iterative argument.

As a starting point for the iteration we can take $100 = \mu \lesssim \lambda$. In this case (11), (12) follow trivially from the energy estimates therefore $M(\lambda, 100) \lesssim 1$. We successively apply (25) for an exponentially increasing sequence $\{\delta_k\}_{k=1,m}$ of values for δ,

$$\delta_k = AC^{\frac{k}{N}}, \qquad \prod_{k=1}^{m} \delta_k = \mu^{-1}.$$

This gives $A \approx \mu^{-\frac{1}{m}} C^{-\frac{m}{N}}$. On the other hand for $M(\lambda, \mu)$ we get the iterated bound

$$M(\lambda, \mu) \lesssim \sum_{k=1}^{m} C^k \delta_k^{-N} \approx mA^{-N} \approx m\mu^{\frac{N}{m}} C^m.$$

We optimize with respect to m and set $m \approx N^{\frac{1}{2}} (\log \mu)^{\frac{1}{2}}$. This gives

$$M(\lambda, \mu) \lesssim e^{DN^{\frac{1}{2}} (\log \mu)^{\frac{1}{2}}} \ll \mu^{\epsilon}.$$

It remains to prove (25). It suffices to do it for (12), as the case of (11) is similar to the 3-d case for (12). We need to prove that for $u, v \in X^1_\lambda$ we have

$$\|T_\mu B(u,v)\|_{L^1 L^{\frac{n+1}{n-1}}} \lesssim \lambda^{\frac{2}{n+1}} \mu^{\frac{n-3}{n+1}} (\delta^{-N} + M(\delta\lambda, \delta\mu)) \|u\|_{X^1_\lambda} \|v\|_{X^1_\lambda}, \qquad n \geq 3.$$

We use the representation in Lemma 4 for u and then for v. The bound for the remainder term r is trivial. The size of u_s is integrable in s, therefore it suffices to prove the Lemma with u replaced by $\chi S_\lambda 1_{\{t\geq s\}} u_s$. Furthermore, $u_s(s) = 0$ therefore $B(\chi S_\lambda 1_{\{t\geq s\}} u_s, v) = 1_{\{t\geq s\}} B(\chi S_\lambda u_s, v)$. Hence we may as well replace u by $\chi S_\lambda u_s$. The bounds for the remainder terms r_s are again trivial. Hence it suffices to consider the the case when both u and v are normalized λ-wave packet superpositions,

$$u = \chi S_\lambda \sum_{T\in\mathcal{T}_\lambda} a_T u_T, \qquad v = \chi S_\lambda \sum_{S\in\mathcal{T}_\lambda} b_S u_S$$

and show that

$$\|T_\mu B(u,v)\|_{L^1 L^{\frac{n+1}{n-1}}} \lesssim \lambda^{\frac{2}{n+1}} \mu^{\frac{n-3}{n+1}} (\delta^{-N} + M(\delta\lambda, \delta\mu)) (\sum a_T^2)^{\frac{1}{2}} (\sum b_S^2)^{\frac{1}{2}} \qquad n \geq 3. \tag{26}$$

Now we use the decomposition in Lemma 9, first for u and then for v at the same scale δ. We obtain functions $u_R, v_R \in X^1_\lambda(R)$ for $R \in \mathcal{Q}_\delta$, which satisfy (21), respectively (23). Combining (22) with (24) we get the L^2 estimate

$$\|B(u,v) - \sum_{L\in\mathcal{Q}_\delta} \sum_{R\in\mathcal{Q}_\delta} \phi_L \phi_R B(u_L, v_R)\|_{L^2} \lesssim \delta^{-N} \lambda^{\frac{n-1}{4}} (\sum a_T^2)^{\frac{1}{2}} (\sum b_S^2)^{\frac{1}{2}}. \tag{27}$$

In order to prove (26) it suffices to show that

$$\|T_\mu \sum_{L\in\mathcal{Q}_\delta} \sum_{R\in\mathcal{Q}_\delta} \phi_L \phi_R B(u_L, v_R)]\|_{L^1 L^{\frac{n+1}{n-1}}} \lesssim M(\delta\lambda, \delta\mu) \lambda^{\frac{2}{n+1}} \mu^{\frac{n-3}{n+1}} (\sum a_T^2)^{\frac{1}{2}} (\sum b_S^2)^{\frac{1}{2}}. \tag{28}$$

$$\|T_\mu [B(u,v) - \sum_{L\in\mathcal{Q}_\delta} \sum_{R\in\mathcal{Q}_\delta} \phi_L \phi_R B(u_L, v_R)]\|_{L^1 L^{\frac{n+1}{n-1}}} \lesssim \delta^{-N} \lambda^{\frac{2}{n+1}} \mu^{\frac{n-3}{n+1}} (\sum a_T^2)^{\frac{1}{2}} (\sum b_S^2)^{\frac{1}{2}}. \tag{29}$$

The plan is on one hand to prove (28) using (12) at a lower scale, and on the other hand, to obtain (29) either directly from (24) (for $n \leq 3$) or by interpolating it with energy estimates (for $n \geq 4$).

Since the cubes R and L have the same size, it follows that in the above summations each cube R interacts with a small number of cubes L, which depends only on the dimension n. Furthermore, we can write $\phi_L \phi_R B(u_L, v_R) = \phi_L \phi_R B(\tilde{\phi}_R \tilde{\phi}_L u_L, v_R)$ and it is easy to see that if $u_L \in X^1(L)$ then $\tilde{\phi}_R \tilde{\phi}_L u_L \in X^1_\lambda(R)$. Hence without any restriction in generality we can set $R = L$ and replace $\phi_R \phi_L$ by ϕ_R. Taking (21) and (23) into account, it follows that (28) is a consequence of the Cauchy-Schwartz inequality and the simpler bound

$$\|T_\mu \phi_R B(u_R, v_R)\|_{L^1 L^{\frac{n+1}{n-1}}(R)} \lesssim M(\delta\lambda, \delta\mu) \lambda^{\frac{2}{n+1}} \mu^{\frac{n-3}{n+1}} \|u_R\|_{X^1_\lambda(R)} \|v_R\|_{X^1_\lambda(R)}. \tag{30}$$

If we rescale R to size 1 then the rescaled functions $u_R(\delta^{-1}t, \delta^{-1}x)$, $v_R(\delta^{-1}t, \delta^{-1}x)$ are localized at frequency $\delta\lambda$ and S_μ is replaced by $S_{\delta\mu}$. Also ϕ_R is replaced by a function ϕ which is smooth on the unit scale. Hence, without any restriction in generality, we can include ϕ in the coefficients of B. We apply (12) corresponding to frequencies $\delta\lambda, \delta\mu$ and then rescale back to scale δ. One can easily verify that (12) with $\epsilon = 0$ is dimensionally correct, i.e. the constant in the estimate does not depend on the scale. Hence we get (30) and conclude the proof of (28).

We now prove (29). If $n = 3$ this is a trivial consequence of (26). If $n \geq 4$ then (29) follows by interpolating (26) with an L^1 bound coming from an energy estimate, namely

$$\|T_\mu[B(u,v) - \sum_{L\in\mathcal{Q}_\delta}\sum_{R\in\mathcal{Q}_\delta} \phi_L\phi_R B(u_L, v_R)]\|_{L^1} \lesssim \delta^{-N}\lambda^{-1}\mu(\sum a_T^2)^{\frac{1}{2}}(\sum b_S^2)^{\frac{1}{2}}. \tag{31}$$

Note that the trivial bound in this case is

$$\|B(u,v)\|_{L^1} \leq \|B(u,v)\|_{L^\infty L^1} \lesssim \|u\|_{X^1}\|v\|_{X^1}.$$

and in order to improve it we need to use the null condition. We use the triangle inequality the left hand side. Arguing as before we can take $L = R$ and use the Cauchy-Schwartz inequality in order to take advantage of the square summability for the size of u_R and v_R with respect to R. Then we are left with two estimates to prove, namely

$$\|T_\mu B(u,v)\|_{L^1} \lesssim \mu\lambda^{-1}\|u\|_{X^1_\lambda}\|v\|_{X^1_\lambda}, \qquad \|T_\mu \phi_R B(u_R, v_R)\|_{L^1} \lesssim \mu\lambda^{-1}\|u_R\|_{X^1_\lambda}\|v_R\|_{X^1_\lambda}.$$

The first follows directly and the second after rescaling from

$$\|T_\mu \phi B(u,v)\|_{L^1} \lesssim \mu\lambda^{-1}\|u\|_{X^1_\lambda}\|v\|_{X^1_\lambda}.$$

Here ϕ is a smooth function in Q, which we insert into B. To prove this we consider the two possible forms of a null form B. If $B(u,v) = b^{0j}Q_{0j}(u,v)$ then we write

$$B(u,v) = \partial_0(b^{0j}u\partial_j v) - \partial_j(b^{0j}u\partial_0 v) - (\partial_0 b^{ij})u\partial_j v - (\partial_j b^{ij})u\partial_0 v + \delta_{t+1}b^{0j}u\partial_j v - \delta_{t-1}b^{0j}u\partial_j v.$$

The last two terms arise since u, v might not be 0 at the initial or the final time. For the first two terms we use

$$\|T_\mu \nabla(bu\nabla v)\|_{L^1} \lesssim \mu\|u\nabla v\|_{L^1} \lesssim \mu\lambda^{-1}\|u\|_{X^1_\lambda}\|v\|_{X^1},$$

while for the next two we have

$$\|T_\mu(\nabla b)u(\nabla v)\|_{L^1} \lesssim \|(\nabla b)u(\nabla v)\|_{L^1} \lesssim \lambda^{-1}\|\nabla g\|_{L^1L^\infty}\|u\|_{X^1_\lambda}\|v\|_{X^1}.$$

Finally, the last two terms can also be bounded by

$$\|T_\mu b^{0j} u\partial_j v \delta_{t\pm 1}\|_{L^1} \lesssim \lambda^{-1}\|u\|_{X^1_\lambda}\|v\|_{X^1}.$$

If $B(u,v) = b^{ij}Q_{ij}(u,v)$ with $i,j \neq 0$ then the problem is simpler as we no longer get the last two terms. Finally, if $B = b^0 Q_0$ then we write

$$B(u,v) = \partial^i(ub^0 g^{ij}\partial_j v) - ub^0 g^{ij}\partial_i\partial^j v - (\partial^i b^0 g^{ij})u\partial_j v - ub^0 g^{0j}\partial_j v\delta_{t+1} + ub^0 g^{0j}\partial_j v\delta_{t-1}.$$

All terms are treated as before except for the second one, for which we have the better estimate

$$\|b^0 u g^{ij}\partial^i\partial^j v\|_{L^1} \lesssim \|u\|_{L^\infty L^2}\|\partial^i g^{ij}\partial^j v\|_{L^1L^2} \lesssim \lambda^{-1}\|u\|_{X^1_\lambda}\|v\|_{X^1_\lambda}.$$

Proof of (14). The argument is similar to the previous one in the case $n = 2,3$, but simpler, because the multiplier T_μ is no longer present.

Proof of (13). Observe that we can get an estimate similar to (13) by interpolating the trivial $L^\infty L^1$ estimate with either (14) or (15). The problem with this reasoning is that we get the wrong balance of powers of λ and μ. To remedy this, we need to find a substitute for the arguments above which only uses the scale reduction argument for the lower frequency factor u i.e. the estimate (22) in Lemma 9.

As before we denote by $M(\lambda,\mu)$ the best constant in (13) with $\epsilon = 0$ with respect to all scales less than 1. We claim that $M(\lambda,\mu)$ satisfies the same recursive relation as before, namely (25). Since from energy estimates we trivially have $M(\lambda, 100) \lesssim 1$, the desired conclusion $M(\lambda,\mu) \lesssim \mu^\epsilon$ follows as in the previous case.

We now prove (25). Precisely, for $u \in X^1_\mu$ and $v \in X^1_\lambda$ with $\mu \leq \lambda$ we need to show that

$$\|B(u,v)\|_{L^2 L^{\frac{n+1}{n}}} \lesssim (\delta^{-N} + M(\delta\lambda,\delta\mu))\mu^{\frac{n-1}{2(n+1)}}\|u\|_{X^1_\mu}\|v\|_{X^1_\lambda}$$

Using the representation in Lemma 4 for u we reduce the problem to the case when u is a superposition of wave packets,

$$u = \chi S_\mu \sum_{T\in\mathcal{T}_\mu} a_T u_T,$$

where we need to prove that

$$\|B(u,v)\|_{L^2 L^{\frac{n+1}{n}}} \lesssim (\delta^{-N} + M(\delta\lambda,\delta\mu))\mu^{\frac{n-1}{2(n+1)}}(\sum_{T\in\mathcal{T}_\mu}|a_T|^2)^{\frac{1}{2}}\|v\|_{X^1_\lambda}. \tag{32}$$

We use part (a) of Lemma 9 to construct the functions $u_R \in X^1_\mu(R)$ for $R \in \mathcal{Q}_\delta$. If we apply (13) to $B(u_R, v)$ in R then we get

$$\|\phi_R B(u_R, v)\|_{L^2 L^{\frac{n+1}{n}}} \lesssim M(\delta\lambda, \delta\mu)\mu^{\frac{n-1}{2(n+1)}} \|u_R\|_{X^1_\mu(R)} \|\tilde{\phi}_R v\|_{X^1_\lambda(R)}.$$

Here we have replaced v by $\tilde{\phi}_R v$. Since $\tilde{\phi}_R = 1$ within the support of ϕ_R this does not change anything. Now we sum this up with respect to R. For $\|u_R\|_{X^1_\mu(R)}$ we have square summability, but for $\|\tilde{\phi}_R v\|_{X^1}$ we have to contend ourselves with the square summability on time slices given by (20). This yields

$$\|\sum_{R \in \mathcal{Q}^t_\delta} \phi_R B(u_R, v)\|^2_{L^2 L^{\frac{n+1}{n}}} \lesssim M(\delta\lambda, \delta\mu)\mu^{\frac{n-1}{n+1}} (\sum_{R \in \mathcal{Q}^t_\delta} \|u_R\|^2_{X^1_\mu(R)}) \|v\|^2_{X^1_\lambda}.$$

Finally, we sum this with respect to $t \in J_\delta$ and use (21). This gives

$$\|\sum_{R \in \mathcal{Q}_\delta} \phi_R B(u_R, v)\|_{L^2 L^{\frac{n+1}{n}}} \lesssim M(\delta\lambda, \delta\mu)\mu^{\frac{n-1}{2(n+1)}} (\sum_{T \in \mathcal{T}_\mu} |a_T|^2)^{\frac{1}{2}} \|v\|_{X^1_\lambda}. \tag{33}$$

On the other hand, we do a parallel computation for the energy estimates. Start with

$$\|\phi_R B(u_R, v)\|_{L^2 L^1} \lesssim \|\phi_R B(u_R, v)\|_{L^\infty L^1} \lesssim \|u_R\|_{X^1} \|\tilde{\phi}_R v\|_{X^1}.$$

and use the same summation procedure as above to get

$$\|\sum_{R \in \mathcal{Q}_\delta} \phi_R B(u_R, v)\|_{L^2 L^1} \lesssim (\sum_{T \in \mathcal{T}_\mu} |a_T|^2)^{\frac{1}{2}} \|v\|_{X^1_\lambda}.$$

Using also the energy estimate for $B(u, v)$ we obtain

$$\|B(u, v) - \sum_{R \in \mathcal{Q}_\delta} \phi_R B(u_R, v)\|_{L^2 L^1} \lesssim (\sum_{T \in \mathcal{T}_\mu} |a_T|^2)^{\frac{1}{2}} \|v\|_{X^1_\lambda}.$$

This we interpolate with (22) to get

$$\|B(u, v) - \sum_{R \in \mathcal{Q}_\delta} \phi_R B(u_R, v)\|_{L^2 L^{\frac{n+1}{n}}} \lesssim \delta^{-N} \mu^{\frac{n-1}{2(n+1)}} (\sum_{T \in \mathcal{T}_\mu} |a_T|^2)^{\frac{1}{2}} \|v\|^2_{X^1_\lambda}.$$

Adding (33) to this yields (32), and the proof is concluded.

Proof of (15) Observe first that the previous arguments fail[4] to provide a proof for (15). The difficulty is as follows. We want to estimate an L^1 norm in time, therefore it seems that

[4]Except for the case $n = 3$, when the proof is similar to the proof of (12).

we need to decompose both u and v to gain the summability with respect to the δ-cubes. However, when we use the decomposition argument for the frequency λ factor we get a too large power of λ. Hence it would appear that we should use the decomposition argument only for the frequency μ factor. This gives initially a better balance of powers of λ and μ than needed, but worsens as the scales decrease because we only get square summability in time.

The solution to this dilemma is to first use the decomposition argument only for the frequency μ factor for as long as possible, which turns out to be up to the spatial scale $\frac{\mu}{\lambda}$. It is at this point only where we start decomposing both factors at the same rate. The advantage of doing this is that the already decomposed frequency μ factor satisfies a better energy type estimate, which suffices in order to close the argument.

The first iterative procedure The intermediate step in the argument can be stated as follows:

Proposition 10. *Let $\lambda^{\frac{1}{2}} \leq \mu \ll \lambda$. If $\{u_R\}_{R\in\mathcal{Q}_{\frac{\mu}{\lambda}}} \in X^1_\mu$ and $v \in X^1_\lambda$ then*

$$\|\sum_{R\in\mathcal{Q}_{\frac{\mu}{\lambda}}} \phi_R B(u_R, v)\|_{L^1 L^{\frac{n+1}{n-1}}} \lesssim \lambda^{\frac{1}{2}} \mu^{\frac{n-3}{2(n+1)}+\epsilon} \left(\sum_{R\in\mathcal{Q}_{\frac{\mu}{\lambda}}} \|u_R\|^2_{X^1_\mu(R)} \right)^{\frac{1}{2}} \|v\|_{X^1_\lambda} \quad n \geq 3. \tag{34}$$

We postpone the proof of the Proposition and instead show that the estimate (15) reduces to (34). For this we use the decomposition of the frequency μ factor only, using the estimate (22). Given $\max\{\mu^{-1}, \mu\lambda^{-1}\} \leq r \leq 1$ we denote by $C(r)$ the best constant in the estimate

$$\|\sum_{R\in\mathcal{Q}_r} \phi_R B(u_R, v)\|_{L^1 L^{\frac{n+1}{n-1}}} \leq C(r,\mu)\lambda^{\frac{1}{2}} \mu^{\frac{n-3}{2(n+1)}} \left(\sum_{R\in\mathcal{Q}_r} \|u_R\|^2_{X^1_\mu(R)} \right)^{\frac{1}{2}} \|v\|_{X^1_\lambda}.$$

Then we can argue exactly as in the proof of (13) and get a recursive relation for $C(r)$, namely

$$C(r) \leq C(\delta^{-N}(\frac{\mu}{\lambda})^{\frac{1}{2}} r^{-\frac{1}{2}} + C(\delta r)), \quad \delta \geq (\mu r)^{1-\epsilon}. \tag{35}$$

The good factor $(\frac{\mu}{\lambda})^{\frac{1}{2}}$ appears because we only decompose the frequency μ function, so we get only a power of μ from Lemma 9. The bad factor $r^{-\frac{1}{2}}$ comes from the fact that we are estimating an L^1 norm in time but only have square summability in time on the r scale. To prove that $C(1) \lesssim \mu^\epsilon$ we iterate (35), but for this to work we need a good stopping point. We have to consider two cases.

a) Suppose $\mu \leq \sqrt{\lambda}$. Then we can stop the iterations at $r = 100\mu^{-1}$. Indeed, in this case we can use the trivial estimate

$$\|\nabla u_R\|_{L^\infty} \leq \mu^{\frac{n}{2}} \|u_R\|_{X^1_\mu}.$$

If J is an equidistant set of times with step μ^{-1} then

$$\begin{aligned}
\|\sum_{R\in\mathcal{Q}_r} \phi_R B(u_R, v)\|_{L^1 L^{\frac{n+1}{n-1}}} &\lesssim \mu^{\frac{1}{2}} \left(\sum_{t\in J} \| \sum_{R\in\mathcal{Q}_r^t} \phi_R B(u, v_R)\|^2_{L^1 L^{\frac{n+1}{n-1}}} \right)^{\frac{1}{2}} \\
&\lesssim \mu^{\frac{1}{2}} \mu^{-\frac{n}{2}+\frac{n-1}{n+1}} \left(\sum_{t\in J} (\sum_{R\in\mathcal{Q}_r^t} \|\nabla u_R\|_{L^\infty})^2 \|\phi_R \nabla v\|_{L^\infty L^2} \right)^{\frac{1}{2}} \\
&\lesssim \mu^{\frac{1}{2}+\frac{n-1}{n+1}} \left(\sum_{R\in\mathcal{Q}_r} \|u_R\|^2_{X^1_\mu(R)} \right)^{\frac{1}{2}} \sup_t \left(\sum_{R\in\mathcal{Q}_r^t} \|\phi_R \nabla v\|^2_{L^\infty L^2} \right)^{\frac{1}{2}} \\
&\lesssim \mu^{1+\frac{n-3}{2(n+1)}} \left(\sum_{R\in\mathcal{Q}_r} \|u_R\|^2_{X^1_\mu(R)} \right)^{\frac{1}{2}} \|v\|_{X^1_\lambda}.
\end{aligned}$$

This implies that

$$C(100\mu^{-1}) \lesssim \mu^{-1}\lambda^{\frac{1}{2}} \leq 1.$$

b) If $\mu \geq \lambda^{\frac{1}{2}}$ then we stop the iterations instead at $r = \frac{\mu}{\lambda}$, where we use (34).

The second iterative procedure To prove Proposition 10 we seek a recursive relation for the best constant $M(\lambda, \mu)$ in (34) with $\epsilon = 0$. We claim that

$$M(\lambda, \mu) \leq C(\delta^{-N} + M(\delta\lambda, \delta\mu)), \qquad \delta \geq (\frac{\mu^2}{\lambda})^{-1+\epsilon}.$$

The conclusion $M(\lambda, \mu) \lesssim \mu^\epsilon$ follows by iterating this up to the point where $\mu = 100\lambda^{\frac{1}{2}}$, in which case the bound $M(\lambda, \mu) \lesssim 1$ was verified directly in the previous paragraph.

Given $u_R \in X^1_\mu(R)$ for $R \in \mathcal{Q}_{\frac{\mu}{\lambda}}$ and $v \in X^1_\lambda$ we want to prove that

$$\|\sum_{R\in\mathcal{Q}_r} \phi_R B(u_R, v)\|_{L^1 L^{\frac{n+1}{n-1}}} \lesssim (\delta^{-N} + M(\delta\lambda, \delta\mu)) \left(\sum_{R\in\mathcal{Q}_r} \|u_R\|^2_{X^1_\mu(R)} \right)^{\frac{1}{2}} \|v\|_{X^1_\lambda}.$$

We use the representation in Lemma 4 for both u_R and v to reduce the problem to the case when u_R, v are wave packet superpositions,

$$u_R = \chi_R S_\mu \sum_{T \in \mathcal{T}_\mu(R)} a_{R,T} u_{R,T}, \qquad v = \chi S_\lambda \sum_{S \in \mathcal{T}_\lambda} b_S u_S$$

In this case the estimate to prove is

$$\| \sum_{R \in \mathcal{Q}_r} \phi_R B(u_R, v) \|_{L^1 L^{\frac{n+1}{n-1}}} \lesssim (dd^{-N} + M(\delta\lambda, \delta\mu)) \left(\sum_{R \in \mathcal{Q}_\delta} \sum_{T \in \mathcal{T}_\mu(R)} |a_{R,T}|^2 \right)^{\frac{1}{2}} \left(\sum_{S \in \mathcal{T}_\lambda} |a_S|^2 \right)^{\frac{1}{2}}. \tag{36}$$

Next we use Lemma 9 twice. First we apply (a rescaling of) part (a) to the pair (u_R, v) relative to the cube R of size $\frac{\mu}{\lambda}$ to produce the functions u_L on the new scale $\delta\frac{\mu}{\lambda}$, and then we apply part (b) to the pair u_L, v to obtain the functions v_P on the δ scale. From (22) we get the L^2 estimate

$$\|\phi_R(B(u_R, v) - \sum_{L \in \mathcal{Q}_{\frac{\delta\mu}{\lambda}}(R)} \phi_L B(u_L, v))\|_{L^2} \lesssim \delta^{-N} \mu^{\frac{n-1}{4}} (\frac{\mu}{\lambda})^{-\frac{n-1}{4}} (\sum_{T \in \mathcal{T}_\mu(R)} |a_{R,T}|^2)^{\frac{1}{2}} \|v\|_{X^1_\lambda}$$

where the factor $(\frac{\mu}{\lambda})^{-\frac{n-1}{4}}$ comes from scaling. Without any restriction in generality we redenote $\phi_R \phi_L := \phi_L$, which is a nice cutoff function even if we think of L as an element of $\mathcal{Q}_{\frac{\delta\mu}{\lambda}}$ (instead of $\mathcal{Q}_{\frac{\delta\mu}{\lambda}}(R)$). After summation with respect to R this gives

$$\| \sum_{R \in \mathcal{Q}_{\frac{\mu}{\lambda}}} \phi_R(B(u_R, v) - \sum_{L \in \mathcal{Q}_{\frac{\delta\mu}{\lambda}}(R)} \phi_L B(u_L, v)\|_{L^2} \lesssim \delta^{-N} \lambda^{\frac{n-1}{4}} (\sum_{R \in \mathcal{Q}_{\frac{\mu}{\lambda}}} \sum_{T \in \mathcal{T}_\mu(R)} |a_{R,T}|^2)^{\frac{1}{2}} \|v\|_{X^1_\lambda}.$$

On the other hand, from (24) we get

$$\| \sum_{L \in \mathcal{Q}_{\frac{\delta\mu}{\lambda}}} \phi_L(B(u_L, v) - \sum_{P \in \mathcal{Q}_\delta} \phi_P B(u_L, v_P))\|_{L^2} \lesssim \delta^{-N} \lambda^{\frac{n-1}{4}} (\sum_{L \in \mathcal{Q}_{\frac{\delta\mu}{\lambda}}} \|u_L\|^2_{X^1_\mu(L)})^{\frac{1}{2}} (\sum_{S \in \mathcal{T}_\lambda} |a_S|^2)^{\frac{1}{2}}.$$

Adding up the last two relations yields

$$\|\sum_{R \in \mathcal{Q}_{\frac{\mu}{\lambda}}} \phi_R B(u_R, v) - \sum_{L \in \mathcal{Q}_{\frac{\delta\mu}{\lambda}}} \sum_{P \in \mathcal{Q}_\delta} \phi_L \phi_P B(u_L, v_P))\|_{L^2} \lesssim \delta^{-N} \lambda^{\frac{n-1}{4}} (\sum_{R \in \mathcal{Q}_{\frac{\mu}{\lambda}}} \sum_{T \in \mathcal{T}_\mu(R)} |a_{R,T}|^2)^{\frac{1}{2}} (\sum_{S \in \mathcal{T}_\lambda} |a_S|^2)^{\frac{1}{2}}.$$

For $n \geq 4$ we interpolate this with an L^1 estimate, namely

$$\|\sum_{R \in \mathcal{Q}_{\frac{\mu}{\lambda}}} \phi_R B(u_R, v) - \sum_{L \in \mathcal{R}_{\frac{\delta\mu}{\lambda}}} \sum_{P \in \mathcal{R}_\delta} \phi_L \phi_P B(u_L, v_P)\|_{L^1} \lesssim (\frac{\mu}{\lambda})^{\frac{1}{2}} (\sum_{R \in \mathcal{Q}_{\frac{\mu}{\lambda}}} \sum_{T \in \mathcal{T}_\mu(R)} |a_{R,T}|^2)^{\frac{1}{2}} (\sum_{S \in \mathcal{T}_\lambda} |a_S|^2)^{\frac{1}{2}}.$$

Although easy to prove[5], this is really the key improvement which makes the argument work. The crucial $(\frac{\mu}{\lambda})^{\frac{1}{2}}$ factor arises because the X^1_μ norms of u_R (and u_L) are square summable in time on the $\frac{\mu}{\lambda}$ scale (instead of just uniformly bounded). This is why we had to localize v first on the $\frac{\mu}{\lambda}$ scale. After interpolation we get

$$\| \sum_{R\in\mathcal{Q}_{\frac{\mu}{\lambda}}} \phi_R B(u_R, v) - \sum_{L\in\mathcal{R}_{\frac{\delta\mu}{\lambda}}} \sum_{P\in\mathcal{R}_\delta} \phi_L\phi_P B(u_L, v_P)\|_{L^{\frac{n+1}{n-1}}} \lesssim \tag{37}$$
$$\delta^{-N}\lambda^{\frac{1}{2}}\mu^{\frac{n-3}{2(n+1)}}(\sum_{R\in\mathcal{Q}_{\frac{\mu}{\lambda}}} \sum_{T\in\mathcal{T}_\mu(R)} |a_{R,T}|^2)^{\frac{1}{2}}(\sum_{S\in\mathcal{T}_\lambda} |a_S|^2)^{\frac{1}{2}}.$$

To complete the $L^1L^{\frac{n+1}{n-1}}$ estimate for $\sum_{R\in\mathcal{R}_{\frac{\lambda}{\mu}}} \phi_R B(u_R, v)$ it remains to use (34) at a lower scale in order to get a bound for $\sum_{L\in\mathcal{R}_{\frac{\delta\mu}{\lambda}}} \sum_{P\in\mathcal{R}_\delta} \phi_L\phi_P B(u_L, v_P)$. The size of the functions in this expression is summable with respect to δ-cubes. Hence we can rescale δ-cubes to size 1, use (34) with (λ,μ) replaced with $(\delta\lambda,\delta\mu)$, scale back and sum up with respect to δ cubes. Since (34) is dimensionally correct, all powers of δ cancel so in the end we get

$$\| \sum_{L\in\mathcal{R}_{\frac{\delta\mu}{\lambda}}} \sum_{P\in\mathcal{R}_\delta} \phi_L\phi_P B(u_L, v_P)\|_{L^1L^{\frac{n+1}{n-1}}} \lesssim \tag{38}$$
$$M(\delta\lambda,\delta\mu)\lambda^{\frac{1}{2}}\mu^{\frac{n-3}{2(n+1)}}(\sum_{R\in\mathcal{Q}_{\frac{\delta\mu}{\lambda}}} \|u_L\|^2_{X^1_\mu(L)})^{\frac{1}{2}}(\sum_{P\in\mathcal{Q}_\delta} \|u_P\|^2_{X^1_\lambda(P)})^{\frac{1}{2}}.$$

Adding (37) and (38) we get (36), which concludes the proof.

A A parametrix for the wave equation.

Our aim here is to outline Smith's [8] construction of a wave packet parametrix.

The λ-tubes. We begin with the constant coefficient case of the D'Allembertian. Then the λ-tubes are parallelepipeds of size $\lambda^{-1}\times(\lambda^{-\frac{1}{2}})^{n-1}\times 1$ which are produced as follows. Given a unit covector e in $\mathbb{R}^n$ we partition the space into a grid of parallelepipeds which have size $\lambda^{-\frac{1}{2}}$ in the directions which are normal to e, and thickness λ^{-1}. Then the λ-tubes are obtained by moving these $\lambda^{-1}\times(\lambda^{-\frac{1}{2}})^{n-1}$ parallelepipeds with speed ± 1 in the normal direction. The allowable conormal directions are spaced about $\lambda^{\frac{1}{2}}$ apart.

In the variable coefficient case considered by Smith [8] in the case of C^2 coefficients one starts with a similar partition of the initial data space $\mathbb{R}^n$, but now the $\lambda^{-1}\times(\lambda^{-\frac{1}{2}})^{n-1}$

[5] i.e. using the energy estimates for all factors involved

parallelepipeds are transported along the Hamilton flow with initial conormal direction e. Thus the λ-tubes become curved parallelepipeds of size $\lambda^{-1} \times (\lambda^{-\frac{1}{2}})^{n-1} \times 1$. It is important here that the Hamilton flow remains nondegenerate on the unit time scale, i.e. that flat initial surfaces do not fold when evolved according to the eikonal equation. If the coefficients are in C^2 then this is true, and it remains true as well if $\partial^2 g \in L^1 L^\infty$.

The λ-wave packets In the constant coefficient case a λ-wave packet associated to a slab $T \in T_\lambda$ is a bump function which is smooth on the scale of T. One can think of it as a $\lambda^{-\frac{1}{2}}$ wide piece cut from a traveling wave. This is the smallest scale on which such solutions can remain localized on the unit time scale.

In the variable coefficient case, Smith's construction was to parallel transport the initial bump in a $\lambda^{-1} \times (\lambda^{-\frac{1}{2}})^{n-1} \times 1$ along its "central" null bicharacteristic. The size and orientation of the bump are adjusted as it moves in time, according to the linearization of the flow around the "central" null bicharacteristic.

The parametrix The parametrix for the wave equation is constructed as follows:

(i) Localize the frequency λ initial data in frequency with respect to angular regions of size $(\lambda^{\frac{1}{2}})^{n-1} \times \lambda$.

(ii) Each such piece is further localized spatially with respect to dual regions.

(iii) Transport the doubly localized initial data pieces along the (straight) Hamilton flow as described above. These are the λ-wave packets.

(iv) Show that the λ-wave packets are approximate solutions to the wave equation.

(v) Show that superpositions of λ-wave packets are approximate solutions to the wave equation. This requires orthogonality estimates for λ-wave packets associated to different λ-tubes.

In the variable coefficient case, this program was carried out in Smith [8] for operators with $C^{1,1}$ coefficients. His approach extents with little change to the case when $\partial^2 g \in L^1 L^\infty$ (see also Tataru [11]). Recently, a similar parametrix was constructed in Smith-Tataru [7] for certain operators whose coefficients have even less regularity, arising in the study of nonlinear second order hyperbolic equations.

B Characteristic energy estimates

In the following discussion we place ourselves in the case when g^{ij} are a small C^2 perturbation of the coefficients of the d'Allembertian. However, the geometric properties involved in the argument remain valid as well in the case when $\partial^2 g \in L^1 L^\infty$.

The model estimate (9). We begin with a proof of the simpler estimate (9) in the special case when $\delta = O(1)$. Thus we want to show that

$$\| \sum_{T \in \mathcal{T}_\lambda}^{P \in T} B(u_T, v) \|_{L^2(Q \setminus B(P, \frac{1}{2}))} \lesssim \lambda^{-1} \|v\|_{X_0^1(Q)}. \tag{39}$$

For $|h| \ll 1$ we construct the characteristic cones K_h starting at $P_h = (x_P, t_P + h)$, with equation

$$K_h = \{t = \phi(h, x)\}.$$

These can be obtained by solving the Hamilton flow starting at P_h, and are non-degenerate on the time-scale 1. In effect one can compute their regularity by looking at the linearization of the Hamilton flow, as in Tataru [11] and show that away from P, ϕ is a C^2 function. At each point on K_h we introduce a so-called null frame $l, \bar{l}, e_a$. This is a basis in the tangent space which is the pseudo-riemmanian replacement of an orthonormal basis. More precisely l is the conormal vector (which is also tangent to K_h because K_h is characteristic), l and e_a span the tangent space to K_h and

$$< l, e_a >= 0, \quad < \bar{l}, e_a >= 0, \quad < e_a, e_b >= \delta_{ab}, \quad < \bar{l}, \bar{l} >= 0, \quad < l, \bar{l} >= -2, \quad < l, l >= 0.$$

For the present argument the regularity of the null frame is irrelevant.

The geometric properties of the λ-tubes and λ-wave packets which are needed are summarized in the following

Lemma 11. *Let T be a λ-tube containing P and u_T a corresponding λ-wave packet. Then*

(i) $T \cap B(P, \frac{1}{2})^c$ is contained between $K_{-\lambda^{-1}}$ and $K_{\lambda^{-1}}$.

(ii) Within $T \cap B(P, \frac{1}{2})^c$ we have

$$|\nabla(u_T)| \lesssim \lambda^{\frac{n+1}{4}}, \quad |(l, e_a)(u_T)| \lesssim \lambda^{\frac{n-1}{4}}.$$

(iii) The intersections $T \cap B(P, \frac{1}{2})^c$, where T is a λ-tube containing P, are locally finite (independent of λ).

The property (ii) allows us to compute pointwise

$$|B(u_T, v)| \lesssim \lambda^{\frac{n-1}{4}} |\bar{l}v| + \lambda^{\frac{n+1}{4}} (|lv| + |e_a v|).$$

To see this it suffices to express the null form B in the null frame and observe that the expression $\bar{l}u_T \bar{l}v$ cannot occur. This is clear for the analogue of B_{ij}, while, on the other hand, B_0 has the form

$$B_0(u,v) = < e_a u, e_a v > - < lu, \bar{l}v > - < lv, \bar{l}u > .$$

Using this pointwise bound, the properties (i)(iii) allow us to write

$$\| \sum_{T \in \mathcal{T}_\lambda}^{P \in T} B(u_T, v) \|^2_{L^2(B(P,\frac{1}{2})^c)} \lesssim \int_D \lambda^{\frac{n-1}{2}} |\bar{l}v|^2 + \lambda^{\frac{n+1}{2}} |lv|^2 + |e_a v|^2.$$

where

$$D = \{ \phi(\lambda^{-1}, x) \le t \le \phi(\lambda^{-1}, x) \} \cap B(P, \frac{1}{2})^c.$$

For the term involving $\bar{l}v$ we have already gained λ^{-1} so we can neglect D and simply use the energy estimates in the unit cube. For the rest we use the foliation of D into cones to write

$$\int_D |lv|^2 + |e_a v|^2 = \int_{-\lambda^{-1}}^{\lambda^{-1}} (\int_{K_h} |lv|^2 + |e_a v|^2) dh.$$

Hence to get (39) it suffices to prove that

$$\int_{K_h} |lv|^2 + |e_a v|^2 \lesssim \|v\|_{X^1_0(Q)}.$$

But here both derivatives of v are tangential to K_h, so this is a characteristic energy estimate for v.

The estimate (10) For the reader's convenience we restate (10) here:

$$\sum_{T \in \mathcal{T}_\lambda} \phi_T^2(P) \| \phi_T^{-1} B(\chi S_\lambda u_T, v) \|^2_{L^2(B(P,\delta)^c)} \lesssim \delta^{-N} \lambda^{-1} \|v\|^2_{X^1}.$$

Since u_T is supported in T it follows that

$$|\phi_T^{-1} \nabla \chi S_\lambda u_T| \lesssim \lambda^{\frac{n+1}{4}}.$$

Hence we get the trivial estimate

$$\| \phi_T^{-1} B(\chi S_\lambda u_T, v) \|_{L^2} \lesssim \lambda^{\frac{n+1}{4}} \|v\|_{X^1}.$$

Since the number of λ tubes is a power of λ it follows that without any restriction in generality we can assume that $\delta > \lambda^{-\epsilon}$ and that we can also restrict the sum to those tubes T for which $d(T,P) \leq \lambda^{-1+\epsilon}$. All these tubes must intersect the line $h \to (x_P, t_P + h)$ at some $h_T \leq \lambda^{-1+\epsilon}$.

We split the sum over T with respect to the position oh h_T in increments of λ^{-1}. If $h_T \approx j\lambda^{-1}$ then $\phi_T(P) \approx (1+|j|)^{-N}$. Hence

$$\sum_{T\in\mathcal{T}_\lambda} \chi_T^2(P)\|\chi_T^{-1}B(\chi S_\lambda u_T, v)\|^2_{L^2(B(P,\delta)^c)} \approx \sum_{|j|\leq\lambda^\epsilon}(1+|j|)^{-2N} \sum_{P_{j\lambda^{-1}}\in T} \|\chi_T^{-1}B(\chi S_\lambda u_T, v)\|^2_{L^2(B(P,\delta)^c)}.$$

This converges rapidly in j, therefore it suffices to do the estimate for $j=0$, i.e. for the tubes which contain P. Hence we still need to show that

$$\sum_{T\in\mathcal{T}_\lambda}^{P\in T} \|\chi_T^{-1}B(\chi S_\lambda u_T, v)\|^2_{L^2(B(P,\delta)^c)} \lesssim \delta^{-N}\lambda^{\frac{n-1}{2}}\|v\|^2_{X^1}. \tag{40}$$

The replacement of Lemma 11 is now

Lemma 12. *Let T be a λ-tube containing P and u_T a corresponding λ-wave packet. Then*

(i) $T\cap B(P,\delta)^c$ is contained between $K_{-(\delta\lambda)^{-1}}$ and $K_{(\delta\lambda)^{-1}}$.

(ii) Within $T\cap B(P,\delta)^c$ we have

$$|\nabla(u_T)| \lesssim \lambda^{\frac{n+1}{4}}, \quad |(l,e_a)(u_T)| \lesssim \delta^{-1}\lambda^{\frac{n-1}{4}}.$$

(iii) The number of tubes S containing P which are at distance less than $\lambda^{-\frac{1}{2}}$ from $T\cap B(P,\frac{1}{2})^c$ is about $O(\delta^{n-1})$.

The main difference is that we are getting all these powers of δ, but these are nonessential in our argument. We can use the property (ii) to obtain similar bounds for $\chi S_\lambda u_T$:

$$\nabla\chi S_\lambda u_T \lesssim \lambda^{\frac{n+1}{4}}\chi_T^2, \quad |(l,e_a)(\chi S_\lambda u_T)| \lesssim \chi_T^2\delta^{-1}\lambda^{\frac{n-1}{4}}.$$

The second part is true because S_λ is a mollifier on the λ^{-1} scale, while the changes in the tangent plane to the cones K_h (i.e. the span of l, e_a) are only significant on the larger $\delta^{-1}\lambda^{-\frac{1}{2}}$ scale. On $K_h \cap B(P,\frac{1}{2})^c$ we get

$$\chi_T^{-1}|\nabla\chi S_\lambda u_T| \lesssim \lambda^{\frac{n+1}{4}}(1+\frac{h}{\delta\lambda})^{-N}, \quad \chi_T^{-1}|(l,e_a)(\chi S_\lambda u_T)| \lesssim \delta^{-1}\lambda^{\frac{n-1}{4}}(1+\frac{h}{\delta\lambda})^{-N}.$$

Using these bounds the proof of (40) can be completed as the proof of (39). Note that part (iii) of Lemma 12 shows that the overlapping between the functions $\chi S_\lambda u_T$ yields only an additional factor which is a power of δ.

References

[1] J. Bourgain. Estimates for cone multipliers. In *Geometric aspects of functional analysis (Israel, 1992–1994)*, pages 41–60. Birkhäuser, Basel, 1995.

[2] Damiano Foschi and Sergiu Klainerman. Bilinear space-time estimates for homogeneous wave equations. *Ann. Sci. École Norm. Sup. (4)*, 33(2):211–274, 2000.

[3] J. Ginibre and G. Velo. Generalized Strichartz inequalities for the wave equation. *J. Funct. Anal.*, 133(1):50–68, 1995.

[4] Markus Keel and Terence Tao. Endpoint Strichartz estimates. *Amer. J. Math.*, 120(5):955–980, 1998.

[5] S. Klainerman and M. Machedon. Space-time estimates for null forms and the local existence theorem. *Comm. Pure Appl. Math.*, 46(9):1221–1268, 1993.

[6] Sergiu Klainerman and Daniel Tataru. On the optimal local regularity for Yang-Mills equations in R^{4+1}. *J. Amer. Math. Soc.*, 12(1):93–116, 1999.

[7] H. Smith and D. Tataru. Sharp local well-posedness results for the nonlinear wave equation. preprint.

[8] Hart F. Smith. A parametrix construction for wave equations with $C^{1,1}$ coefficients. *Ann. Inst. Fourier (Grenoble)*, 48(3):797–835, 1998.

[9] Hart F. Smith and Christopher D. Sogge. Null form estimates for $(1/2, 1/2)$ symbols and local existence for a quasilinear Dirichlet-wave equation. *Ann. Sci. École Norm. Sup. (4)*, 33(4):485–506, 2000.

[10] Terence Tao. Endpoint bilinear restriction theorems for the cone, and some sharp null form estimates. *Math. Z.*, 238(2):215–268, 2001.

[11] Daniel Tataru. Strichartz estimates for operators with nonsmooth coefficients III. to appear, J. Amer. Math. Soc.

[12] Daniel Tataru. Strichartz estimates for second order hyperbolic operators with nonsmooth coefficients. II. *Amer. J. Math.*, 123(3):385–423, 2001.

[13] Thomas Wolff. A sharp bilinear cone restriction estimate. *Ann. of Math. (2)*, 153(3):661–698, 2001.