Low regularity solutions for the general quasilinear ultrahyperbolic Schrödinger equation

B. Pineau and M. Taylor

LOW REGULARITY SOLUTIONS FOR THE GENERAL QUASILINEAR ULTRAHYPERBOLIC SCHRÖDINGER EQUATION

BEN PINEAU AND MITCHELL A. TAYLOR

ABSTRACT. We present a novel method for establishing large data local well-posedness in low regularity Sobolev spaces for general quasilinear Schrödinger equations with non-degenerate and nontrapping metrics. Our result represents a definitive improvement over the landmark results of Kenig, Ponce, Rolvung and Vega [15, 16, 17, 18], as it weakens the regularity and decay assumptions to the same scale of spaces considered by Marzuola, Metcalfe and Tataru in [24], but removes the uniform ellipticity assumption on the metric from their result. Our method has the additional benefit of being relatively simple but also very robust. In particular, it only relies on the use of pseudodifferential calculus for classical symbols.

CONTENTS

1. INTRODUCTION

In this article, we consider the large data local well-posedness problem for general quasilinear ultrahyperbolic Schrödinger equations of the form

$$(1.1) \quad \begin{cases} i\partial_t u + g^{jk}(u, \overline{u}, \nabla u, \nabla \overline{u})\partial_j \partial_k u = F(u, \overline{u}, \nabla u, \nabla \overline{u}), & u : \mathbb{R} \times \mathbb{R}^d \to \mathbb{C}^m, \\ u(0, x) = u_0(x), \end{cases}$$

where g and F are assumed to be smooth functions of their arguments with g real, symmetric and uniformly non-degenerate and F vanishing at least quadratically at the origin.

In a recent series of articles [22, 23, 24], Marzuola, Metcalfe and Tataru have studied the well-posedness of the system (1.1) in low regularity Sobolev spaces. As a brief overview, the paper [23] considers the small data problem for cubic and higher nonlinearities in the Sobolev spaces $H^s(\mathbb{R}^d)$, $s > \frac{d+5}{2}$. The article [22],

on the other hand, permits quadratic terms in the nonlinearity, but assumes that the data comes from the smaller space $l^1 H^s(\mathbb{R}^d)$, $s > \frac{d}{2} + 3$. Here, $l^1 H^s$ is an appropriate translation invariant Sobolev type space, imposing similar regularity requirements as H^s, but slightly stronger decay. To see that some additional decay is needed, it is instructive to look at the leading part of the linearized flow, which can be written schematically as

$$(1.2) \qquad \begin{cases} i\partial_t v + \partial_j g^{jk} \partial_k v + b^j \partial_j v + \tilde{b}^j \partial_j \overline{v} = f, \\ v(0, x) = v_0(x). \end{cases}$$

Here, for the purposes of our heuristic discussion, we have written the principal operator in divergence form with $g^{jk} = g^{jk}(u, \overline{u})$ – we will elaborate further on this reduction later on. As is well-known, a necessary condition for L^2 well-posedness of a wide class of such linear systems is integrability of the first order coefficient $\mathrm{Re}(b^j)$ along the bicharacteristic (or Hamilton) flow of the principal differential operator $\partial_j g^{jk} \partial_k$. This is usually referred to as the Mizohata (or Takeuchi–Mizohata) condition. See, for instance, [9, 13, 21, 26, 27, 28, 30] for several manifestations of this ill-posedness mechanism. For cubic and higher nonlinearities, the integrability of $\mathrm{Re}(b^j)$ along the bicharacteristics is automatic for small H^s data, but for quadratic nonlinearities it is not. That being said, there are several natural ways to recover the above integrability condition. One common approach is to work in weighted Sobolev spaces. However, the alternative $l^1 H^s$ spaces also achieve this goal, but have the additional advantage of being translation invariant – they are also far less restrictive in terms of regularity and decay, as we will see below.

In contrast to the case of small data, the third paper in the series by Marzuola, Metcalfe and Tataru [24] considers the significantly more challenging large data problem. Here, the authors establish well-posedness in the same setting as their small data papers, but under two additional assumptions. The first assumption is that the initial metric $g(u_0)$ is nontrapping, meaning that all nontrivial bicharacteristics corresponding to the principal operator $\Delta_{g(u_0)}$ escape to spatial infinity at both ends. Such a condition is automatic in the small data regime (assuming sufficient regularity and asymptotic flatness of the metric) as in this setting the Hamilton trajectories are close to straight lines. For large data well-posedness, a nontrapping assumption is completely natural, in light of the Mizohata condition.

On the other hand, the methods in [24] also rely on the assumption of uniform ellipticity of the principal operator, i.e., the existence of a uniform constant $c > 0$ such that

$$(1.3) \qquad c^{-1} |\xi|^2 \le g^{jk}(x) \xi_j \xi_k \le c |\xi|^2.$$

This assumption is critically used in the above article to effectively diagonalize the linearized equation (1.2) and remove the complex conjugate first order term. Roughly speaking, this diagonalization proceeds by considering the new variable

$$Sv := v + \mathcal{R}\overline{v},$$

where $\mathcal{R}$ is a pseudodifferential operator of order -1 with symbol which is essentially of the form

$$r(x, \xi) = \frac{i\tilde{b}^l \xi_l}{g^{jk} \xi_j \xi_k},$$

when $|\xi| \ge 1$. It is not difficult to see that, to leading order, Sv formally satisfies an equation like (1.2), but without the complex conjugate first order term. This diagonalization procedure is then used as a key ingredient in the proofs of the requisite local smoothing and $L_T^\infty L_x^2$ estimates for the linearized flow. A similar diagonalization is heavily relied upon in [3] and [18].

The primary objective of the current article is to generalize the main result of [24] to the full class of ultrahyperbolic quasilinear Schrödinger flows, while keeping the regularity and function spaces identical. That is, we shall relax the uniform ellipticity assumption

$$c^{-1}|\xi|^2 \le g^{jk}\xi_j\xi_k \le c|\xi|^2$$

to the much weaker uniform non-degeneracy condition

$$c^{-1}|\xi| \le |g^{jk}\xi_k| \le c|\xi|.$$

The lack of an ellipticity assumption on the metric in (1.1) causes significant difficulties, and is what prompted the development of the new well-posedness scheme that we present in this article (we were also inspired by the scheme in [14]). On the other hand, there are several physical sources of motivation for studying the general ultrahyperbolic problem. Some well-known examples arise naturally in the study of water waves [5] and others arise in the theory of completely integrable models [11, 29]. More recently, the Hall and electron magnetohydrodynamic equations without resistivity have been shown to behave at leading order like degenerate quasilinear Schrödinger systems of ultrahyperbolic type [12]. This dispersive character of the equations was used to great effect in [12, 14], leading to well-posedness in certain regimes and ill-posedness in others.

Although [24] requires ellipticity of the metric in order to achieve their low regularity results, significant progress has been made towards removing the ellipticity assumptions from the well-posedness theory of (1.1) in the high regularity regime. This is best illustrated by the pioneering series [15, 16, 17, 18] of Kenig, Ponce, Rolvung and Vega, which culminates in a proof of large data well-posedness under the nontrapping assumption for systems of the form (1.1) in high regularity weighted Sobolev spaces of the form $H^s \cap L^2(\langle x \rangle^N dx)$, where s and N are suitably large, dimension dependent parameters. In this fundamental series of papers, [18] studies the well-posedness problem assuming ellipticity of the principal operator $\partial_j g^{jk} \partial_k$, while [15, 16, 17] consider symmetric, non-degenerate metrics, first in the constant coefficient case and then later for variable coefficients. As should be evident from these articles, the ellipticity assumption on the metric is not easy to remove, even in the high regularity regime. The main objective of the current paper is to give a much simpler proof of well-posedness for the general system (1.1) that is also robust enough to work in low regularity spaces. To the best of our knowledge, this is the first low regularity well-posedness result that applies to the full class of ultrahyperbolic quasilinear Schrödinger flows.

The rough strategy used in [15] to prove well-posedness of the ultrahyperbolic flow (1.1) in high regularity weighted spaces is to first establish an estimate for the local energy type norm

$$\|v\|_{LE} := \|\langle x \rangle^{-\frac{N}{2}} \langle \nabla \rangle^{\frac{1}{2}} v\|_{L^2_T L^2_x}, \qquad N = N(d) \in \mathbb{N},$$

for the linearized equation (1.2) (assuming suitably strong asymptotic decay of the coefficients b^j, $\tilde{b}^j$ and $\nabla_x g^{jk}$) of the form

$$(1.4) \qquad \qquad \|v\|_{LE} \lesssim \|v\|_{L^\infty_T L^2_x} + \|f\|_{LE^* + L^1_T L^2_x}.$$

Here, LE^* denotes the "dual" local energy space. The estimate (1.4) shows that the local energy norm of v remains under control, as long as v satisfies an a priori $L^\infty_T L^2_x$ bound. The preliminary estimate (1.4) follows, roughly speaking, from a suitable adaptation of Doi's construction in [6] to the ultrahyperbolic problem. The more significant technical obstruction in [15] is in establishing the a priori bound for the $L^\infty_T L^2_x$ norm. To

understand the difficulties, we first note that when the real part of the coefficient b^j vanishes, it is a relatively straightforward exercise (in view of (1.4)) to obtain the bound

$$\|v\|_{L_T^\infty L_x^2} \lesssim \|v_0\|_{L_x^2} + \|f\|_{LE^* + L_T^1 L_x^2}.$$

Indeed, this follows by a standard energy estimate, as one can integrate by parts to shift derivatives off of the first order terms and onto the coefficients b^j and $\tilde{b}^j$. Therefore, in the general case, one is motivated to try to conjugate away the badly behaved first order term $\operatorname{Re}(b^j)\partial_j v$. In [15], this conjugation is accomplished by constructing a (formally) zeroth order operator $\mathcal{O}$ which achieves the approximate cancellation

$$(1.5) \qquad\qquad [\mathcal{O}, \partial_j g^{jk}\partial_k] + \mathcal{O}\operatorname{Re}(b^j)\partial_j \approx 0.$$

The idea here is very loosely akin to the method of integrating factors from ODE. On a formal level, the symbol for the operator $\mathcal{O}$ achieving (1.5) is given by

$$(1.6) \qquad\qquad O(x,\xi) := \exp\left(-\int_{-\infty}^0 \operatorname{Re}(b(x^t))\cdot\xi^t dt\right),$$

where (x^t, ξ^t) denotes the bicharacteristic flow

$$(\dot{x}^t, \dot{\xi}^t) = (\nabla_\xi a(x^t, \xi^t), -\nabla_x a(x^t, \xi^t)), \qquad (x^0, \xi^0) = (x,\xi),$$

corresponding to the principal symbol $a(x,\xi) := -g^{jk}(x)\xi_j\xi_k$. Unfortunately, the symbol O does not belong to the standard symbol class S^0. Rather, (assuming that b has sufficient regularity and decay) it satisfies

$$(1.7) \qquad\qquad |\partial_\xi^\alpha \partial_x^\beta O(x,\xi)| \lesssim_{\alpha,\beta} \langle\xi\rangle^{-|\alpha|}\langle x\rangle^{|\alpha|}.$$

In the case when the metric is positive-definite (i.e. Δ_g is elliptic), the mapping properties of the pseudodifferential operators associated with this class of symbols were intensively studied in the paper [4] of Craig, Kappeler and Strauss. In the case of a merely non-degenerate metric, Kenig, Ponce, Rolvung and Vega in [15] execute a systematic study of this symbol class as well as a very careful analysis of the bicharacteristic flow for $-g^{jk}(x)\xi_j\xi_k$ to establish suitable mapping properties for $\mathcal{O}$. In contrast, in the current article, to obtain the $L_T^\infty L_x^2$ estimate for (1.2) we will instead use a spatially truncated version of the above renormalization operator which achieves a suitable cancellation of the form (1.5), at least within a large compact set. The key advantage of this truncation is that the corresponding renormalization operator will be a classical pseudodifferential operator of order 0, which will dramatically simplify the analysis (perhaps at the cost of estimating some extra error terms). Moreover, it will allow us to considerably lower the regularity and decay assumptions on the coefficients in (1.2) compared to [15] when estimating the $L_T^\infty L_x^2$ norm of v. Of course, this idea comes with some technical caveats of its own, which will be discussed later.

We remark that the idea of using the above spatial truncation to close the energy estimate for (1.2) is inspired by the article [14] of Jeong and Oh, where they consider the well-posedness problem for the electron MHD equations near non-zero, constant magnetic fields, and perform an analogous truncation in their setting. As we shall see below, such a construction turns out to be tied heavily to the direction of propagation of the bicharacteristics of the principal part of the corresponding linear flow. For the electron MHD equations, the bicharacteristics have a distinguished direction of propagation. However, the bicharacteristics for the Schrödinger equations that we consider in this article do not exhibit this feature. Therefore, one key novelty of the present paper is in dealing with the multi-directionality present in Schrödinger flows. Another important novelty is our ability to extend the truncation idea in order to give a new and very simple proof of the natural local smoothing type estimate for (1.2) in the local energy norms compatible with the translation

invariant function spaces used in this paper. The method that we present is very robust and requires only mild decay of the coefficients (e.g. uniform integrability along the Hamilton flow). A more detailed outline of the argument will be given in Section 3.

1.1. **Statements of the results.** We now state our main results more precisely. As in [24], our primary focus will be on the case of quadratic nonlinear interactions.

Let $d, m \geq 1$ and consider a system of equations of the form (1.1) where

$$(1.8) \qquad g : \mathbb{C}^m \times \mathbb{C}^m \times (\mathbb{C}^m)^d \times (\mathbb{C}^m)^d \to \mathbb{R}^{d \times d} \ \text{ and } \ F : \mathbb{C}^m \times \mathbb{C}^m \times (\mathbb{C}^m)^d \times (\mathbb{C}^m)^d \to \mathbb{C}^m$$

are smooth functions. We assume that F vanishes at least quadratically at the origin, so that

$$(1.9) \qquad |F(y, z)| \approx \mathcal{O}\left(|y|^2 + |z|^2\right) \ \text{ near } (y, z) = (0, 0).$$

In [24], the authors assume that the metric g is uniformly elliptic and coincides with the identity matrix at the origin. That is, they assume that $g(0) = I_{d \times d}$ and that there is a fixed constant $c > 0$ so that

$$c^{-1}|\xi|^2 \leq g^{jk}(y, z)\xi_j \xi_k \leq c|\xi|^2, \ \forall \xi \in \mathbb{R}^d, \ y, z \in \mathbb{C}^m \times (\mathbb{C}^m)^d.$$

In this article, we only assume that g is symmetric and (uniformly) non-degenerate, in the sense that

$$(1.10) \qquad c^{-1}|\xi| \leq |g(y, z)\xi| \leq c|\xi|, \ \forall \xi \in \mathbb{R}^d, \ y, z \in \mathbb{C}^m \times (\mathbb{C}^m)^d,$$

for some fixed constant $c > 0$.

As in [22, 23, 24], we also consider a second class of quasilinear Schrödinger equations of the form

$$(1.11) \qquad \begin{cases} i\partial_t u + \partial_j g^{jk}(u, \overline{u})\partial_k u = F(u, \overline{u}, \nabla u, \nabla \overline{u}), \quad u : \mathbb{R} \times \mathbb{R}^d \to \mathbb{C}^m, \\ u(0, x) = u_0(x), \end{cases}$$

where F is as in (1.9), but where the metric g depends on u but not on ∇u. Such an equation arises by formally differentiating the system (1.1). Indeed, if u solves (1.1) then $(u, \nabla u)$ solves an equation of the form (1.11) with a nonlinearity F which depends at most quadratically on ∇u.

Remark 1.1. Note that the second order operator in (1.11) has a divergence structure, which can be achieved by commuting the first derivative with g and viewing the commutator as an additional term on the right-hand side. In contrast, the second order operator in (1.1) cannot be written in divergence form without possibly changing the type of the equations.

To state our main well-posedness theorem, we must recall the function spaces used in [22, 23, 24]. For now, we limit ourselves to an expository summary, giving more precise definitions in Section 2.2.

Consider a standard spatial Littlewood-Paley decomposition

$$1 = \sum_{j \in \mathbb{N}_0} S_j,$$

where S_j, $j \geq 1$, selects frequencies of size $\approx 2^j$ and S_0 selects all frequencies of size $\lesssim 1$. Corresponding to each dyadic frequency scale $2^j \geq 1$, we consider an associated partition $\mathcal{Q}_j$ of $\mathbb{R}^d$ into cubes of side length 2^j and an associated smooth partition of unity

$$1 = \sum_{Q \in \mathcal{Q}_j} \chi_Q.$$

We define the $l_j^1 L^2$ norm by

$$\|u\|_{l_j^1 L^2} = \sum_{Q \in \mathcal{Q}_j} \|\chi_Q u\|_{L^2}, \tag{1.12}$$

and the space $l^1 H^s$ via the norm

$$\|u\|_{l^1 H^s}^2 = \sum_{j \geq 0} 2^{2sj} \|S_j u\|_{l_j^1 L^2}^2. \tag{1.13}$$

Note that if one replaces the ℓ^1 sum by an ℓ^2 sum in (1.12) and defines $l^2 H^s$ analogously to (1.13), then $H^s = l^2 H^s$ with equivalent norms. The extra summability in the definition of the $l^1 H^s$ norm yields the decay necessary to circumvent Mizohata's ill-posedness mechanism. However, unlike the high regularity weighted Sobolev spaces used in previous works, the function spaces $l^1 H^s$ admit translation invariant equivalent norms and contain functions exhibiting weaker regularity and decay.

As mentioned above, in the large data problem, one has to contend with trapping. This is an obvious obstruction to well-posedness, so we will need to impose a nontrapping assumption on the initial metric $g(u_0)$ to prevent this. Then, as part of our well-posedness theorem, we will show that the nontrapping assumption propagates on a time interval whose length depends on the data size and the profile of the initial metric. Our definition of nontrapping is the same as [24].

Definition 1.2. We say that the metric $g(u_0)$ is *nontrapping* if all nontrivial bicharacteristics for $\Delta_{g(u_0)}$ escape to spatial infinity at both ends.

The above qualitative definition of nontrapping suffices in order to state our main results. However, as we shall see, the proofs require us to introduce a parameter L which gives a quantitative description of nontrapping. The precise way in which we define L is slightly different than [24], so as to better handle the case when Δ_g is not elliptic.

With the above discussion in mind, we may state our main well-posedness theorem as follows.

Theorem 1.3. *Let $s > \frac{d}{2} + 3$ and suppose that the initial data $u_0 \in l^1 H^s$ makes $g(u_0)$ into a real, symmetric, uniformly non-degenerate, nontrapping metric. Then (1.1) with the quadratic nonlinearity (1.9) is locally well-posed in $l^1 H^s$. The same result holds if $s > \frac{d}{2} + 2$ for the equation (1.11).*

Remark 1.4. We will prove the latter result in Theorem 1.3 as it will imply the former by differentiating the equation.

Remark 1.5. As in [24], the regularity and decay assumptions in the above results can be weakened if the metric and nonlinearity satisfy the stronger vanishing conditions

$$g(y, z) = g(0) + \mathcal{O}\left(|y|^2 + |z|^2\right), \quad |F(y, z)| \approx \mathcal{O}\left(|y|^3 + |z|^3\right) \text{ near } (y, z) = (0, 0).$$

Namely, it can be shown that (1.1) is well-posed in the same sense as Theorem 1.3 when $u_0 \in H^s$ and $s > \frac{d+5}{2}$. An analogous result holds if $s > \frac{d+3}{2}$ for the equation (1.11). To prove this, one makes modifications to the quadratic case which are virtually identical to those made in [24]. In order to simplify our exposition, we omit the details for these relatively straightforward modifications and instead focus on the general case of quadratic nonlinearities.

Remark 1.6. In the above results, well-posedness is to be interpreted in the standard quasilinear fashion. More precisely, in the setting of Theorem 1.3 it includes the following key features.

- (Regular solutions). For large σ and nontrapping initial data $u_0 \in l^1 H^\sigma$ there is a unique solution $u \in C([0,T]; l^1 H^\sigma)$ which persists and remains nontrapping on some nontrivial maximal time interval $I = [0, T_*)$.
- (Rough solutions). For $s > \frac{d}{2} + 3$ and nontrapping data $u_0 \in l^1 H^s$ there is a unique solution $u \in C([0,T]; l^1 H^s) \cap l^1 X^s([0,T])$ which persists and remains nontrapping on some nontrivial maximal time interval $I = [0, T_*)$. Here, the auxiliary space $l^1 X^s$ is a natural analogue of the local energy space LE described earlier. A precise definition of this space will be given in Section 2.2.
- (Continuous dependence). The maximal time $T_*(u_0)$ is a lower semicontinuous function of u_0 with respect to the $l^1 H^s$ topology and for each $T < T_*(u_0)$ the data-to-solution map $v_0 \mapsto v$ is continuous near u_0 from $l^1 H^s$ into $C([0,T]; l^1 H^s) \cap l^1 X^s([0,T])$.

Remark 1.7. As in [24, Remark 1.3.2], the maximal existence time $T_*(u_0)$ a priori depends on the full profile of the initial data u_0 rather than just its size in $l^1 H^s$, due to the nontrapping condition on the metric.

Remark 1.8. As in [24], the arguments we use here are purely dispersive. This is in contrast to the viscosity methods used in earlier works, which are less tailored to the structure of the equations, and hence less suitable for low regularity analysis.

1.2. **Organization of the paper.** The paper is organized as follows. In Section 2 we recall the precise functional setting used in [24] as well as the standard Fourier-analytic, nonlinear, and pseudodifferential machinery that will be used throughout the article. In certain cases, we adapt this machinery in order to obtain refined estimates in the space-time function spaces where we aim to construct solutions to (1.1) and (1.11). In Section 3, we provide a detailed outline of the proof. Then, in Section 4, we analyze the bicharacteristic flow. The key objectives of this section are to quantify nontrapping, show that nontrapping is stable under small perturbations, and establish suitable asymptotic bounds for the bicharacteristics. In Section 5, we state our main well-posedness theorem for the linearized flow and reduce the main linear estimate to establishing a simplified bound for the corresponding inhomogeneous linear paradifferential flow in the $l^1 X^s$ spaces where we intend to construct solutions. Then, in Section 6, we aim to establish a suitable estimate for the $L_T^\infty L_x^2$ component of the $l^1 X^s$ norm by constructing a truncated version of the renormalization operator $\mathcal{O}$ in (1.6). Such an estimate will close on a short enough time interval, up to controlling a small factor of the local energy component of the $l^1 X^s$ norm. In Section 7, we control this remaining component of the $l^1 X^s$ norm for the linear paradifferential flow. Then, in Section 8, we deduce the full $l^1 X^s$ estimate for the paradifferential and linearized equations by combining the local energy estimate with the $L_T^\infty L_x^2$ estimate from Section 6. Finally, in Section 9 we use the linearized estimates from the previous sections along with a suitable paradifferential reduction of the full nonlinear equation to establish Theorem 1.3.

2. PRELIMINARIES

In this section, we recall some basic Fourier-analytic tools as well as the definitions and elementary properties of the function spaces that will be used in our analysis. We also recall some standard facts about pseudodifferential operators and establish some new estimates for these operators in our function spaces.

2.1. **Littlewood-Paley decomposition.** We begin by recalling the standard Littlewood-Paley decomposition. We let $\varphi : \mathbb{R}^d \to [0,1]$ be a smooth radial function supported in the ball of radius 2, $B_2 = B_2(0)$, which satisfies $\varphi = 1$ on B_1. We define Fourier multipliers S_0 and S_k by

$$\widehat{S}_k := \varphi(2^{-k}\xi) - \varphi(2^{-k+1}\xi), \qquad k \in \mathbb{N},$$

$$\widehat{S}_0 := \varphi(\xi).$$

We then define for each $k \in \mathbb{N}$,

$$S_{<k} := \sum_{0 \leq j < k} S_j, \qquad S_{\geq k} := \sum_{k \leq j < \infty} S_j.$$

With the above notation, we have the standard (inhomogeneous) Littlewood-Paley decomposition

$$1 = \sum_{k \geq 0} S_k.$$

In the sequel, we will often phrase our bilinear and nonlinear estimates in the language of paradifferential calculus. For a suitable pair of complex-valued functions f and g, we will write $T_g f$ to mean

$$(2.1) \qquad T_g f := \sum_{k \geq 0} S_{<k-4} g S_k f.$$

In other words, $T_g f$ selects the portion of the product fg where f is at high frequency compared to g. With this notation, we have the so-called Bony decomposition or Littlewood-Paley trichotomy,

$$fg = T_f g + T_g f + \Pi(f,g).$$

We refer the reader to [1] and [25] for some basic properties of these operators. To compactify the above notation, we will sometimes write $f_{<k}$ as shorthand for $S_{<k}f$ and $f_{\geq k}$ as shorthand for $S_{\geq k}f$.

2.2. **Function spaces.** Next, we recall the definitions and basic properties of the function spaces that will be used in our analysis. Much of the material here is recalled from [22] and the large data paper [24]. For each frequency scale 2^k, we consider a partition of $\mathbb{R}^d$ into a set Q_k of disjoint cubes of side length 2^k along with a smooth partition of unity in physical space,

$$1 = \sum_{Q \in Q_k} \chi_Q.$$

For a translation-invariant Sobolev type space U, we define the spaces $l_k^p U$ by

$$\|u\|_{l_k^p U} := \left(\sum_{Q \in Q_k} \|\chi_Q u\|_U^p \right)^{\frac{1}{p}}, \quad 1 \leq p < \infty, \qquad \|u\|_{l_k^\infty U} := \sup_{Q \in Q_k} \|\chi_Q u\|_U.$$

As noted in [22], these spaces have a translation invariant equivalent norm, obtained by replacing the sum over cubes with an integral. Moreover, up to norm equivalence, the smooth partition by compactly supported cutoffs can be replaced by a partition consisting of cutoffs which are all localized to frequency zero.

We next recall the definition of the local energy type space X, which is defined for each $T > 0$ by

$$\|u\|_X := \sup_{l \in \mathbb{N}_0} \sup_{Q \in Q_l} 2^{-\frac{l}{2}} \|u\|_{L_T^2 L_x^2([0,T] \times Q)}.$$

Associated to X is the "dual" local energy type space $Y \subset L^2([0,T] \times \mathbb{R}^d)$ which satisfies the relation $X = Y^*$. See [22] for more details on the properties and construction of this space. For each non-negative integer k, we define

$$X_k := 2^{-\frac{k}{2}} X \cap L_T^\infty L_x^2, \quad \|u\|_{X_k} := 2^{\frac{k}{2}} \|u\|_X + \|u\|_{L_T^\infty L_x^2}$$

and

$$Y_k := 2^{\frac{k}{2}} Y + L_T^1 L_x^2, \quad \|u\|_{Y_k} := \inf\{2^{-\frac{k}{2}} \|u_1\|_Y + \|u_2\|_{L_T^1 L_x^2} : u = u_1 + u_2\}.$$

Loosely speaking, we will use X_k to measure solutions to the Schrödinger equation localized at frequency 2^k whereas Y_k will be used to measure inhomogeneous source terms localized at this frequency. Next, we define for each $s \in \mathbb{R}$,

$$\|u\|_{l^p X^s} := \left(\sum_{k \geq 0} 2^{2ks} \|S_k u\|_{l_k^p X_k}^2 \right)^{\frac{1}{2}}, \quad \|u\|_{l^p Y^s} := \left(\sum_{k \geq 0} 2^{2ks} \|S_k u\|_{l_k^p Y_k}^2 \right)^{\frac{1}{2}},$$

for $1 \leq p < \infty$ (with the natural modification for $p = \infty$). We will also work with the corresponding spaces without the ℓ^p summability,

$$\|u\|_{X^s} := \left(\sum_{k \geq 0} 2^{2ks} \|S_k u\|_{X_k}^2 \right)^{\frac{1}{2}}, \quad \|u\|_{Y^s} := \left(\sum_{k \geq 0} 2^{2ks} \|S_k u\|_{Y_k}^2 \right)^{\frac{1}{2}}.$$

As already mentioned, throughout the article we will frequently make use of the standard tools of paradifferential calculus to estimate various multilinear and nonlinear expressions. A very nice bookkeeping device for efficiently tracking the frequency distribution of such terms is the language of frequency envelopes introduced by Tao in [31]. To define these, suppose that we are given a translation-invariant Sobolev type space U with the orthogonality relation,

$$\|u\|_U \approx \left(\sum_{k \geq 0} \|S_k u\|_U^2 \right)^{\frac{1}{2}}.$$

An admissible frequency envelope for $u \in U$ is a positive sequence $(c_k) \subset \mathbb{N}_0$ such that for each $k \in \mathbb{N}_0$, we have

(i) (Boundedness and size).

$$\|S_k u\|_U \lesssim c_k \|u\|_U, \quad \|c_k\|_{l_k^2} \approx 1.$$

(ii) (Left-slowly varying).

$$c_j \geq 2^{\delta(j-k)} c_k, \quad j < k,$$

for some fixed parameter $\delta > 0$.

(iii) (Right-uniformly varying).

$$c_j \geq 2^{\sigma(k-j)} c_k, \quad j > k,$$

for some fixed parameter $\sigma > 0$.

For nonzero u, such a frequency envelope always exists. For instance, we may define

$$c_j = \|u\|_U^{-1} \left(\max_{k \geq j} 2^{-\delta|j-k|} \|S_k u\|_U + \max_{k \leq j} 2^{-\sigma|j-k|} \|S_k u\|_U \right).$$

In this article, the primary purpose of the above frequency envelopes will be to facilitate the proof of the continuity of the data-to-solution map for the quasilinear Schrödinger systems we consider.

2.3. **Pseudodifferential calculus.** Our objective in this subsection is to recall some basic properties of pseudodifferential operators and then establish some refined estimates for these operators in the local energy and "dual" local energy spaces defined above.

For $m \in \mathbb{R}$, we recall that the standard symbol class $S^m := S_{1,0}^m$ is defined by

$$S^m := \{a \in C^\infty(\mathbb{R}^{2d}) : |a|_{S^m}^{(j)} < \infty, \ j \in \mathbb{N}_0\},$$

where the corresponding seminorms $|a|_{S^m}^{(j)}$ are given by

$$|a|_{S^m}^{(j)} := \sup\{\|\langle\xi\rangle^{|\alpha|-m}\partial_x^\beta\partial_\xi^\alpha a(x,\xi)\|_{L^\infty(\mathbb{R}^{2d})} : \ |\alpha+\beta| \leq j\}.$$

To each symbol $a \in S^m$ we can associate the pseudodifferential operator $Op(a) \in OPS^m$, defined for $f \in \mathcal{S}(\mathbb{R}^d)$ by the quantization

$$Op(a)f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} a(x,\xi)e^{ix\cdot\xi}\widehat{f}(\xi)d\xi.$$

We now list some basic properties of pseudodifferential operators; proofs can be found in the standard reference [33]. We begin with an elementary result on Sobolev boundedness.

Proposition 2.1 (Sobolev boundedness). *Let $s, m \in \mathbb{R}$ and let $a \in S^m$. Then $Op(a)$ extends to a bounded linear operator from H^{s+m} to H^s and there exists j depending only on s, m and the dimension such that*

$$\|Op(a)\|_{H^{s+m}\to H^s} \lesssim |a|_{S^m}^{(j)}.$$

We next recall the sharp Gårding inequality for symbols $a \in S^1$.

Proposition 2.2 (Sharp Gårding inequality). *Let $a \in S^1$ and let $R > 0$ be such that $\mathrm{Re}(a) \geq 0$ for $|\xi| \geq R$. Then $Op(a)$ is semi-positive. That is, there exists j depending on d such that for $f \in \mathcal{S}(\mathbb{R}^d)$, we have*

$$\mathrm{Re}\langle Op(a)f, f\rangle \gtrsim_R -|a|_{S^1}^{(j)}\|f\|_{L^2}^2,$$

where $\langle\cdot,\cdot\rangle$ denotes the usual L^2 inner product.

Proof. See, e.g., [8]. □

Remark 2.3. As shown in [8, 20], a variant of Proposition 2.2 also holds for $N \times N$ matrix-valued symbols. More specifically, if $a \in S^1$ is an $N \times N$ symbol satisfying $\mathrm{Re}(a) \geq 0$ then the associated pseudodifferential operator $Op(a)$ is semi-positive in the sense that $\mathrm{Re}\langle Op(a)f, f\rangle \geq -c\|f\|_{L^2}^2$ for all f in the Schwartz class.

Next, we recall a (weak) version of the Calderon-Vaillancourt theorem [2].

Proposition 2.4 (Calderon-Vaillancourt theorem). *Let $a \in S^0$. There exists an integer $j > 0$ depending on the dimension such that*

$$\|Op(a)\|_{L^2\to L^2} \lesssim \sup_{|\alpha+\beta|\leq j} \|\partial_\xi^\alpha\partial_x^\beta a\|_{L^\infty(\mathbb{R}^{2d})},$$

where the implicit constant is universal.

Finally, we recall some elementary symbolic calculus which will allow us to perform basic manipulations with pseudodifferential operators.

Proposition 2.5 (Algebraic properties of pseudodifferential operators)**.** *Let* $m_1, m_2 \in \mathbb{R}$ *and let* $a_1 \in S^{m_1}$ *and* $a_2 \in S^{m_2}$. *The following properties hold.*

(i) *(Composition property). There is* $a \in S^{m_1+m_2-1}$ *such that*
$$Op(a_1)Op(a_2) = Op(a_1 a_2) + Op(a)$$
and for every $j \in \mathbb{N}_0$, $|a|_{S^{m_1+m_2-1}}^{(j)}$ *is controlled by* $|a_1|_{S^{m_1}}^{(k)} |a_2|_{S^{m_2}}^{(k)}$ *for some* k *depending on* j *and* d.

(ii) *(Adjoint). There is* $a \in S^{m_1-1}$ *such that*
$$Op(a_1)^* = Op(\overline{a}_1) + Op(a)$$
and for every $j \in \mathbb{N}_0$, $|a|_{S^{m_1-1}}^{(j)}$ *is controlled by* $|a_1|_{S^{m_1}}^{(k)}$ *for some* k *depending on* j *and* d.

(iii) *(Commutator). There is* $a \in S^{m_1+m_2-2}$ *such that*
$$Op(a_1)Op(a_2) - Op(a_2)Op(a_1) = Op(-i\{a_1, a_2\}) + Op(a)$$
where $\{\cdot, \cdot\}$ *denotes the Poisson bracket, which is defined by*
$$\{a_1, a_2\} = \nabla_\xi a_1 \cdot \nabla_x a_2 - \nabla_\xi a_2 \cdot \nabla_x a_1.$$
Moreover, for every $j \in \mathbb{N}_0$, $|a|_{S^{m_1+m_2-2}}^{(j)}$ *is controlled by* $|a_1|_{S^{m_1}}^{(k)} |a_2|_{S^{m_2}}^{(k)}$ *for some* k *depending on* j *and* d.

Proof. See, e.g., [15, Theorem 2.1.2] for a precise statement and [19, 33] for proofs. □

In our construction, we will need the following refinement of the Calderon-Vaillancourt theorem for symbols $a \in S^0$, which ensures that the $L^2 \to L^2$ operator bound for $Op(a)$ depends only on the L^∞ norm of a when applied to functions localized at sufficiently high frequency. This refinement will be important later when we attempt to spatially localize the renormalization operator mentioned in the introduction. We remark that Proposition 2.6 is also used in the paper [14] to achieve a similar purpose. We include the simple proof for completeness.

Proposition 2.6 (Calderon-Vaillancourt theorem at high frequency)**.** *Let* $a \in S^0$. *There is* k_0 *depending on* a *such that for* $k \geq k_0$, $Op(a)$ *satisfies the* $L^2 \to L^2$ *bound,*
$$\|Op(a)S_{\geq k}\|_{L^2 \to L^2} \lesssim \|a\|_{L^\infty}.$$
That is, the $L^2 \to L^2$ *bound for* $Op(a)$ *depends only on the* L^∞ *norm of the symbol* a *when applied to functions at sufficiently high frequency.*

Proof. The proof is a simple scaling argument. The symbol for $S_{>k}$ is of the form $\psi_k(\xi) := 1 - \varphi(2^{-k}\xi)$, where φ is a smooth bump function equal to one on the unit ball and supported in $B_2(0)$. Define the symbol $a_k := a\psi_k$. Let $\lambda > 0$ be some constant to be chosen, and define $a_{k,\lambda}(x,\xi) := a_k(\lambda^{-1}x, \lambda\xi)$, $v_\lambda(x) := v(\lambda x)$. We clearly have
$$Op(a_k)v = (2\pi)^{-d} \int_{\mathbb{R}^d} a_k(x, \lambda\xi) e^{i\lambda x \cdot \xi} \widehat{v_{\lambda^{-1}}}(\xi) d\xi.$$
Hence,
$$\|Op(a_k)v\|_{L^2} = \lambda^{-\frac{d}{2}} \|Op(a_{k,\lambda})v_{\lambda^{-1}}\|_{L^2}.$$

By Proposition 2.4, we have

$$\lambda^{-\frac{d}{2}}\|Op(a_{k,\lambda})v_{\lambda^{-1}}\|_{L^2} \lesssim \lambda^{-\frac{d}{2}} \sup_{|\alpha|,|\beta|\leq j(d)} \|\partial_x^\beta \partial_\xi^\alpha a_{k,\lambda}\|_{L^\infty}\|v_{\lambda^{-1}}\|_{L^2}$$

$$= \sup_{|\alpha|,|\beta|\leq j(d)} \|\partial_x^\beta \partial_\xi^\alpha a_{k,\lambda}\|_{L^\infty}\|v\|_{L^2},$$

where $j(d)$ depends only on the dimension. To conclude, we therefore only need to show that for a suitable choice of λ, we have

$$\sup_{|\alpha|,|\beta|\leq j(d)} \|\partial_x^\beta \partial_\xi^\alpha a_{k,\lambda}\|_{L^\infty} \lesssim \|a\|_{L^\infty}.$$

Taking $\lambda = 2^{\frac{k}{2}}$ and using that $a \in S^0$, we find

$$|\partial_x^\beta \partial_\xi^\alpha a_{k,\lambda}| \lesssim |a|_{S^0}^{(|\alpha|+|\beta|)} 2^{-|\alpha|k}\lambda^{|\alpha|-|\beta|} \lesssim |a|_{S^0}^{(|\alpha|+|\beta|)} 2^{-(|\alpha|+|\beta|)\frac{k}{2}}.$$

The proof is concluded by taking k sufficiently large (depending only on the symbol bounds for a). $\qquad\square$

Next, we extend the above bounds to the X^0 and Y^0 spaces.

Proposition 2.7 (Operator bounds for X^0 and Y^0). *Let $a \in S^0$ be time-independent and let $T \lesssim 1$. Then there is $j = j(d)$ such that we have the operator bounds*

$$(2.2) \qquad\qquad \|Op(a)\|_{X^0 \to X^0} + \|Op(a)\|_{Y^0 \to Y^0} \lesssim 1 + |a|_{S^0}^{(j)}.$$

Moreover, there is $k_0 > 0$ depending only on a such that if $k \geq k_0$, we also have

$$(2.3) \qquad\qquad \|Op(a)S_{\geq k}\|_{X^0 \to X^0} + \|Op(a)S_{\geq k}\|_{Y^0 \to Y^0} \lesssim 1 + \|a\|_{L^\infty}.$$

Remark 2.8. The inequality (2.3) can be thought of as the analogue of Proposition 2.6 for the X^0 and Y^0 spaces.

Proof. We prove (2.2) and remark on the very minor modifications required to prove (2.3) where necessary. For notational convenience, we let K_a denote the term on the right-hand side of (2.2). We begin with the $X^0 \to X^0$ bound. By definition, we have

$$(2.4) \qquad \begin{aligned} \|Op(a)f\|_{X^0}^2 &= \sum_{k\geq 0} \|S_k Op(a)f\|_{X_k}^2 \\ &\lesssim \sum_{k\geq 0} \|S_k[Op(a),\tilde{S}_k]f\|_{X_k}^2 + \sum_{k\geq 0}\|S_k Op(a)\tilde{S}_k f\|_{X_k}^2, \end{aligned}$$

for some fattened Littlewood-Paley projection $\tilde{S}_k$. For the first term, we can crudely estimate using Hölder in T and dyadic summation,

$$\left(\sum_{k\geq 0}\|S_k[Op(a),\tilde{S}_k]f\|_{X_k}^2\right)^{\frac{1}{2}} \lesssim \left(\sum_{k\geq 0} 2^k \|S_k[Op(a),\tilde{S}_k]f\|_{L_T^\infty L_x^2}^2\right)^{\frac{1}{2}}$$

$$\lesssim \sup_{k\geq 0}\|[Op(a),\tilde{S}_k]f\|_{L_T^\infty H_x^{\frac{1}{2}+\epsilon}}$$

$$\lesssim_\epsilon K_a\|f\|_{L_T^\infty H_x^{-\frac{1}{2}+\epsilon}}$$

$$\lesssim K_a\|f\|_{X^0},$$

where in the second to third line, we used Proposition 2.1 and that $[Op(a),\tilde{S}_k] \in OPS^{-1}$ (which has symbol bounds uniform in k, thanks to Proposition 2.5).

Remark 2.9. We remark briefly on one change needed here for the proof of (2.3). If f is replaced by $S_{>k}f$ for some sufficiently large k, then in the third line above, we can estimate

$$K_a\|S_{>k}f\|_{L_T^\infty H_x^{-\frac{1}{2}+\epsilon}} \lesssim K_a 2^{-k(\frac{1}{2}-\epsilon)}\|f\|_{X^0},$$

and so, the factor of K_a can be replaced by 1 in the above estimate by taking k large enough.

Now, we turn to the second term in (2.4). By square summing and Proposition 2.4 (or Proposition 2.6 when proving (2.3)), it suffices to estimate the $2^{-\frac{k}{2}}X$ component of the X_k norm. For this, we have

$$\|Op(a)\tilde{S}_k f\|_{2^{-\frac{k}{2}}X} = \sup_{l\in\mathbb{N}_0} \sup_{Q\in Q_l} 2^{\frac{k-l}{2}}\|\chi_Q Op(a)\tilde{S}_k f\|_{L_T^2 L_x^2}.$$

Using the $L^2 \to L^2$ bound for $Op(a)$ from Proposition 2.4 and that $[Op(a), \chi_Q] \in OPS^{-1}$ with bounds independent of l, we obtain for each $Q \in Q_l$,

(2.5)
$$2^{\frac{k-l}{2}}\|\chi_Q Op(a)\tilde{S}_k f\|_{L_T^2 L_x^2} \lesssim K_a 2^{\frac{k-l}{2}}\|\chi_Q \tilde{S}_k f\|_{L_T^2 L_x^2} + 2^{\frac{k}{2}}K_a\|\tilde{S}_k f\|_{L_T^2 H_x^{-1}}$$
$$\lesssim K_a\|\tilde{S}_k f\|_{X_k} + 2^{-\frac{k}{2}}K_a\|\tilde{S}_k f\|_{L_T^\infty L_x^2}.$$

Therefore,

$$\left(\sum_{k\geq 0}\|Op(a)\tilde{S}_k f\|_{X_k}^2\right)^{\frac{1}{2}} \lesssim K_a\|f\|_{X^0},$$

which establishes the $X^0 \to X^0$ bound. The high-frequency variant (2.3) is proved by using instead Proposition 2.6 in place of Proposition 2.4 above and using the frequency gain in the latter term in the second line of (2.5) to absorb the factor of K_a.

Next, we turn to the $Y^0 \to Y^0$ bound. Again, by definition, we have

$$\|Op(a)f\|_{Y^0}^2 = \sum_{k\geq 0}\|S_k Op(a)f\|_{Y_k}^2$$
$$\lesssim \sum_{k\geq 0}\|S_k[Op(a),\tilde{S}_k]f\|_{Y_k}^2 + \sum_{k\geq 0}\|S_k Op(a)\tilde{S}_k f\|_{Y_k}^2.$$

For the first term, we estimate using the embedding $L_T^1 L_x^2 \subset Y_k$ and that $[Op(a),\tilde{S}_k] \in OPS^{-1}$ to obtain

$$\left(\sum_{k\geq 0}\|S_k[Op(a),\tilde{S}_k]f\|_{Y_k}^2\right)^{\frac{1}{2}} \lesssim K_a\|f\|_{L_T^1 H_x^{-1+\epsilon}} \lesssim K_a\|f\|_{Y^0},$$

where the last inequality follows from the fact that $Y^0 \subset L_T^1 H_x^{-\frac{1}{2}-\epsilon}$. Similarly to before, for the bound (2.3) when f is replaced by $S_{>k}f$, we have

$$K_a\|S_{>k}f\|_{L_T^1 H_x^{-1+\epsilon}} \lesssim K_a 2^{-k(\frac{1}{2}-2\epsilon)}\|f\|_{Y^0},$$

and so, the factor of K_a can be replaced by 1 if k is large enough. For the second term, we use duality. Let $g \in X_k$ with $\|g\|_{X_k} \leq 1$. We have by Proposition 2.5 and similar embeddings as above,

$$|\langle S_k Op(a)\tilde{S}_k f, g\rangle| \lesssim \|\tilde{S}_k f\|_{Y_k}\|\tilde{S}_k(Op(\overline{a}))S_k g\|_{X_k} + K_a\|\tilde{S}_k f\|_{L_T^1 H_x^{-1}}\|S_k g\|_{L_T^\infty L_x^2}$$
$$\lesssim \|\tilde{S}_k f\|_{Y_k}\|\tilde{S}_k(Op(\overline{a}))S_k g\|_{X_k} + 2^{-k(\frac{1}{2}-\epsilon)}K_a\|\tilde{S}_k f\|_{Y_k}.$$

Again, if k is large enough, the K_a factor in the latter term can be discarded. Using the $X^0 \to X^0$ bound already established above, we also have

$$\|\tilde{S}_k(Op(\overline{a}))S_k g\|_{X_k} \lesssim \|Op(\overline{a})S_k g\|_{X^0} \lesssim K_a,$$

where K_a can be replaced by $1 + \|a\|_{L^\infty}$ if k is large enough. The proof of (2.2) is then concluded by dyadic summation. $\qquad\square$

In our analysis later, we will sometimes need to estimate commutators of pseudodifferential and paradifferential operators. For this purpose, we recall the following Coifman-Meyer type estimate from (3.6.4) and (3.6.5) of [33].

Proposition 2.10 (Coifman-Meyer type bound). *For every $m, \sigma \in \mathbb{R}$ and $P \in OPS^m$, we have*

$$(2.6) \qquad\qquad \|[P, T_g]f\|_{H^\sigma} \le C\|g\|_{W^{1,\infty}}\|f\|_{H^{\sigma+m-1}},$$

where $C > 0$ is a constant depending on P and σ.

2.4. **Multilinear and Moser estimates.** Here we recall several of the multilinear and Moser-type estimates for the local energy and dual local energy spaces defined above.

Proposition 2.11 (Proposition 3.1 in [22]). *Let $s > \frac{d}{2}$. Then for $u, v \in l^1 X^s$ we have the algebra property*

$$\|uv\|_{l^1 X^s} \lesssim \|u\|_{l^1 X^s}\|v\|_{l^1 X^s}.$$

We also have the Moser-type estimate,

$$\|F(u)\|_{l^1 X^s} \lesssim \|u\|_{l^1 X^s}(1 + \|u\|_{l^1 X^s})c(\|u\|_{L^\infty}),$$

for $s > \frac{d}{2}$ and any smooth function F with $F(0) = 0$.

We next recall some elementary bilinear estimates for the $l_k^1 Y_k$ spaces.

Proposition 2.12 (Bilinear estimates). *The following bilinear estimates hold for $l_k^1 Y_k$ spaces.*
  (i) (High-low interactions). If $j < k - 4$,

$$\|S_j u S_k v\|_{l_k^1 Y_k} \lesssim 2^{j(\frac{d}{2}+1)}2^{-k}\|S_k v\|_{2^{-\frac{k}{2}}X}\|S_j u\|_{l_j^1 L_T^\infty L_x^2}.$$

  (ii) (Balanced interactions). If $|i - j| \le 4$ and $i, j \ge k - 4$,

$$\|S_k(S_i u S_j v)\|_{l_k^1 Y_k} \lesssim 2^{\frac{jd}{2}}\|S_i u\|_{l_i^1 L_T^2 L_x^2}\|S_j v\|_{L_T^\infty L_x^2}.$$

Proof. This is a slight refinement of Lemma 4.3 in [24]. The proof is almost identical, so we omit the details. $\qquad\square$

By dyadic summation, the following is a consequence of Proposition 2.12.

Proposition 2.13 (Paradifferential bilinear estimates). *Let $s_0 > \frac{d}{2} + 2$. Then for every $\sigma \ge 0$, we have*

$$\|T_u v\|_{l^1 Y^\sigma} + \|(T_v - v)u\|_{l^1 Y^\sigma} \lesssim \|u\|_{l^1 X^{s_0-1}}\|v\|_{X^{\sigma-1}},$$

$$\|T_u v\|_{l^1 Y^\sigma} + \|(T_v - v)u\|_{l^1 Y^\sigma} \lesssim \|u\|_{l^1 X^{s_0-2}}\|v\|_{X^\sigma}.$$

If $0 \le \sigma \le s_0$, we also have

$$(2.7) \qquad \begin{aligned} \|(T_v - v)u\|_{l^1 Y^\sigma} &\lesssim \|u\|_{l^1 X^{\sigma-1}}\|v\|_{X^{s_0-1}}, \\ \|(T_v - v)u\|_{l^1 Y^\sigma} &\lesssim \|u\|_{l^1 X^{\sigma-2}}\|v\|_{X^{s_0}}. \end{aligned}$$

We next state a closely related commutator estimate, which is a slight refinement of the version in [22].

Proposition 2.14. *Let $s > \frac{d}{2} + 2$ and let $A \in S^0$ be a Fourier multiplier. Then we have*

$$\|\nabla[S_{<k-4}g, A(D)]\nabla S_k u\|_{l^1 Y^0} \lesssim_A \|g - g_\infty\|_{l^1 X^s}\|S_k u\|_{X^0},$$

where g_∞ is any constant matrix.

Proof. The proof of this is essentially identical to the proof of Proposition 3.2 in [22]. We omit the details. □

Remark 2.15. We note that if $A \in S^m$ is a Fourier multiplier for some real number $m \geq 0$ then we can write the commutator in the above proposition as

$$[S_{<k-4}g, A(D)]\nabla S_k u = 2^{mk}\tilde{S}_k[S_{<k-4}g, 2^{-mk}A(D)\tilde{S}_k]\nabla S_k u,$$

for some fattened projection $\tilde{S}_k$. Since $2^{-mk}A(D)\tilde{S}_k \in OPS^0$ with symbol bounds uniform in k, we have from Proposition 2.14,

$$\|[S_{<k-4}g, A(D)]\nabla S_k u\|_{l^1 Y^0} \lesssim 2^{(m-1)k}\|g - g_\infty\|_{l^1 X^s}\|S_k u\|_{X^0}.$$

The next bound will allow us to precisely estimate certain error terms in the dual local energy space Y^0 which involve commutators of pseudodifferential and paradifferential operators. This is essentially a variant of Proposition 2.10 but for the X and Y spaces.

Proposition 2.16 ($X^{-2} \to Y^0$ commutator estimate). *Let $T \lesssim 1$, $\mathcal{O} \in OPS^0$ be time-independent with symbol $O \in S^0$ and let $s_0 > \frac{d}{2} + 2$. Moreover, let g be a function such that $g - g_\infty \in l^1 X^{s_0}$ for some constant g_∞. Then we have the estimate*

$$\|[\mathcal{O}, T_g]f\|_{Y^0} \leq C\|g - g_\infty\|_{l^1 X^{s_0}}\|f\|_{X^{-2}},$$

where C depends only on $\mathcal{O}$.

Proof. Clearly, it suffices to prove the claim with $g_\infty = 0$. Moreover, it suffices to work with the principal part of the commutator since the remainder is bounded from $H_x^{-2} \to L_x^2$ uniformly in T with norm $\lesssim_{\mathcal{O}} \|g\|_{L_T^\infty C^{2,\epsilon}}$ for some sufficiently small $\epsilon > 0$, which by Sobolev embedding can be controlled by $\|g\|_{l^1 X^{s_0}}$. The principal symbol p for $[T_g, \mathcal{O}]$ is

$$p(x,\xi) := -i\sum_{k \geq 0}\{S_{<k-4}g(x)\widehat{S}_k(\xi), O\}$$
$$= -i\sum_{k \geq 0}S_{<k-4}g(x)\nabla_\xi\widehat{S}_k(\xi) \cdot \nabla_x O + i\sum_{k \geq 0}S_{<k-4}\nabla_x g(x)\widehat{S}_k(\xi) \cdot \nabla_\xi O =: p_1 + p_2.$$

First, we consider bounds for $P_1 := Op(p_1)$. Modulo an operator which is bounded from $H_x^{-2} \to L_x^2$ with norm $\lesssim_{\mathcal{O}} \|g\|_{L_T^\infty C^{2,\epsilon}}$, we can write

$$P_1 = -i\sum_{k \geq 0}S_{<k-4}g(x)(\nabla_\xi\widehat{S}_k)(D) \cdot Op(\nabla_x O)\tilde{S}_k + \mathcal{O}_{L_T^\infty H_x^{-2} \to L_T^\infty L_x^2}(1),$$

where $\tilde{S}_k$ is a slightly fattened Littlewood-Paley projection. As $\nabla_\xi\widehat{S}_k$ is localized at frequency $\approx 2^k$, we can use Proposition 2.12 and dyadic summation to estimate

$$\|P_1 f\|_{Y^0} \lesssim \|g - g_\infty\|_{l^1 X^{s_0-1}}\|f\|_{X^{-2}}.$$

A similar argument for $P_2 := Op(p_2)$ gives

$$\|P_2 f\|_{Y^0} \lesssim \|\nabla_x g\|_{l^1 X^{s_0-1}}\|f\|_{X^{-2}},$$

which concludes the proof. □

Finally, we state versions of some of the above bilinear and Moser estimates which are phrased in terms of frequency envelopes. This will be convenient for establishing the finer properties of the solution map later on, such as the continuous dependence of the solution on the initial data. From Proposition 3.2 of [22], we have the following estimates.

Proposition 2.17 (Frequency localized estimates I). *Let $s > \frac{d}{2}$ and let $u, v \in l^1 X^s$ with $l^1 X^s$ frequency envelopes given by a_k and b_k, respectively. Then for each $k \in \mathbb{N}_0$, we have*

$$\|S_k(uv)\|_{l^1 X^s} \lesssim (a_k + b_k)\|u\|_{l^1 X^s}\|v\|_{l^1 X^s}.$$

Moreover, if F is a smooth function with $F(0) = 0$, then we have

$$\|S_k(F(u))\|_{l^1 X^s} \lesssim a_k \|u\|_{l^1 X^s}(1 + \|u\|_{l^1 X^s})c(\|u\|_{L^\infty}).$$

Proposition 2.18 (Frequency localized estimates II). *Let $s > \frac{d}{2} + 2$. The following estimates hold for $k \in \mathbb{N}_0$.*

(i) *Let $0 \le \sigma \le s$ and let $u \in l^1 X^{\sigma-1}$ and $v \in l^1 X^{s-1}$ with corresponding frequency envelopes a_k and b_k, respectively. We have*

$$\|S_k(uv)\|_{l^1 Y^\sigma} \lesssim (a_k + b_k)\|u\|_{l^1 X^{\sigma-1}}\|v\|_{l^1 X^{s-1}}.$$

(ii) *Let $0 \le \sigma \le s - 1$ and let $u \in l^1 X^\sigma$ and $v \in l^1 X^{s-2}$ with corresponding frequency envelopes a_k and b_k, respectively. We have*

$$\|S_k(uv)\|_{l^1 Y^\sigma} \lesssim (a_k + b_k)\|u\|_{l^1 X^\sigma}\|v\|_{l^1 X^{s-2}}.$$

(iii) *Let $0 \le \sigma \le s$ and let $u \in l^1 X^\sigma$ and $v \in l^1 X^{s-2}$ with corresponding frequency envelopes a_k and b_k, respectively. We have*

$$\|S_k(vS_{\ge k-4}u)\|_{l^1 Y^\sigma} \lesssim (a_k + b_k)\|u\|_{l^1 X^\sigma}\|v\|_{l^1 X^{s-2}}.$$

3. OVERVIEW OF THE PROOF

In this section, we give an overview of the key ideas that go into the proof of Theorem 1.3. We recall that our essential aim is to establish local well-posedness for the system

$$(3.1) \qquad \begin{cases} i\partial_t u + \partial_j g^{jk}(u, \overline{u})\partial_k u = F(u, \overline{u}, \nabla u, \nabla \overline{u}), & u : \mathbb{R} \times \mathbb{R}^d \to \mathbb{C}^m, \\ u(0, x) = u_0(x), \end{cases}$$

in the $l^1 H^s$ scale, for $s > s_0 > \frac{d}{2} + 2$. As we shall see, our scheme builds on and complements the ideas from [14, 15, 16, 17, 18, 24], but also has several important novelties.

3.1. **The linear and paradifferential ultrahyperbolic flows.** The main component of our argument involves a careful analysis of the linear ultrahyperbolic flow

$$(3.2) \qquad \begin{cases} i\partial_t v + \partial_j g^{jk}\partial_k v + b^j \partial_j v + \tilde{b}^j \partial_j \overline{v} = f, \\ v(0, x) = v_0(x), \end{cases}$$

which is naturally associated with the linearization of (3.1). Here, the metric g^{jk} is real, nontrapping, symmetric and non-degenerate, and the coefficients g, b, and $\tilde{b}$ satisfy the asymptotic flatness conditions

$$(3.3) \qquad \|g - g_\infty\|_{l^1 X^{s_0}} + \|\partial_t g\|_{l^1 X^{s_0-2}} + \|(b, \tilde{b})\|_{l^1 X^{s_0-1}} + \|\partial_t(b, \tilde{b})\|_{l^1 X^{s_0-3}} \le M,$$

where $M > 0$ is a fixed constant and g_∞ is a constant, non-degenerate, symmetric matrix. Note that in the special case where v corresponds to the linearization around a solution u to (3.1), the coefficients and inhomogeneous source term in (3.2) take the form

$$(3.4) \quad \begin{cases} b^j := \nabla_u g^{jk} \partial_k u - \nabla_{(\nabla u)_j} F, \quad \tilde{b}^j := \nabla_{\overline{u}} g^{jk} \partial_k u - \nabla_{(\nabla \overline{u})_j} F, \\ f := (\nabla_{\overline{u}} F - \partial_j (\nabla_{\overline{u}} g^{jk}) \partial_k u) \overline{v} + (\nabla_u F - \partial_j (\nabla_u g^{jk}) \partial_k u) v, \end{cases}$$

where we have suppressed the dependence on u in the coefficients for simplicity of notation. In our general analysis of (3.2), we will not require that b, $\tilde{b}$ and f arise from solutions to (3.1) via linearization.

An equation that is closely related to (3.2) is the associated linear paradifferential flow

$$(3.5) \quad \begin{cases} i\partial_t v + \partial_j T_{g^{jk}} \partial_k v + T_{b^j} \partial_j v + T_{\tilde{b}^j} \partial_j \overline{v} = f, \\ v(0, x) = v_0(x), \end{cases}$$

which extracts the leading part of the linear flow. Here, the paradifferential operator T_g is defined as in (2.1). Again, when v arises from the linearization around a solution u to (3.1), b^j and $\tilde{b}^j$ remain as in (3.4), but now

$$f := (\nabla_{\overline{u}} F - \partial_j (\nabla_{\overline{u}} g^{jk}) \partial_k u) \overline{v} + (\nabla_u F - \partial_j (\nabla_u g^{jk}) \partial_k u) v + (\partial_j T_{g^{jk}} \partial_k - \partial_j g^{jk} \partial_k) v$$
$$+ (T_{b^j} - b^j) \partial_j v + (T_{\tilde{b}^j} - \tilde{b}^j) \partial_j \overline{v}.$$

In this case, f can be thought of as being comprised of perturbative error terms when measured in the dual local energy space $l^1 Y^0$. These terms either have a suitable algebraic balance of derivatives between the coefficients and v or have coefficient functions that are at high or comparable frequency relative to v.

3.2. Quantitative nontrapping and the bicharacteristic flow.

As in [24], to adequately study the linear (and ultimately nonlinear) problem, we will need a quantitative measure of nontrapping. For our purposes, we will only need to define nontrapping for time-independent metrics g with regularity and decay given by

$$(3.6) \quad \|g - g_\infty\|_{l^1 H^{s_0}} \leq M, \qquad \frac{d}{2} + 2 < s_0 < s,$$

where $M > 0$ is a fixed constant and g_∞ is a constant, non-degenerate, symmetric matrix. Note that the condition (3.6) guarantees that $g \in C^{2,\delta}$, which in particular ensures that the corresponding Hamilton flow,

$$(\dot{x}^t, \dot{\xi}^t) = (\nabla_\xi a(x^t, \xi^t), -\nabla_x a(x^t, \xi^t)), \qquad a(x, \xi) = -g^{ij}(x) \xi_i \xi_j, \qquad (x^0, \xi^0) = (x, \xi),$$

is locally well-posed. The first preliminary objective of Section 4 is to show that under the nontrapping assumption on g and the asymptotic flatness condition (3.6), the Hamilton flow is in fact globally defined. This is not automatic when Δ_g is not elliptic. Indeed, although $g^{ij} \xi_i \xi_j$ is conserved by the Hamilton flow, unlike in the elliptic case, it does not necessarily control the size of $|\xi^t|$ in our setting.

The second objective of Section 4 is to provide a quantitative measure of nontrapping. For this, we define a function $L : [0, \infty) \to [0, \infty)$ where $L(R)$ measures (roughly speaking) the maximal amount of time any initially unit speed bicharacteristic can intersect the ball $B_R(0)$. Our definition differs slightly from the definition in [24], as they define L in terms of the Hamilton flow projected onto the co-sphere bundle $|\xi^t| = 1$. This latter definition is natural in the elliptic case, in light of the conservation of $g^{ij} \xi_i \xi_j$, but is not quite suitable for our problem. Analogously to [24], we show that our nontrapping parameter L is stable under small perturbations of the metric, which will be important later on when we analyze the linear and nonlinear Schrödinger flows.

3.3. **The main linear estimate.** The crux of our argument centers around establishing the following key estimate for the linear paradifferential flow (3.5):

$$\|v\|_{l^1 X^\sigma} \leq C(M, L)(\|v_0\|_{l^1 H^\sigma} + \|f\|_{l^1 Y^\sigma}), \quad 0 \leq \sigma, \tag{3.7}$$

which as a simple consequence yields the following estimate for the linear flow (3.2):

$$\|v\|_{l^1 X^\sigma} \leq C(M, L)(\|v_0\|_{l^1 H^\sigma} + \|f\|_{l^1 Y^\sigma}), \quad 0 \leq \sigma \leq s_0.$$

Here, $C(M, L)$ is a constant depending on the coefficient size M in (3.3) and on the nontrapping parameter L for g within a fixed compact set whose size depends on the profile of the metric g and the rate of decay of the coefficients b^j and $\tilde{b}^j$. In Section 5, we reduce establishing the above two estimates to establishing the following simpler bound for the linear paradifferential flow:

$$\|v\|_{X^\sigma} \leq C(M, L)(\|v_0\|_{H^\sigma} + \|f\|_{Y^\sigma}), \quad \sigma \geq 0, \tag{3.8}$$

in the setting where $\widehat{v}$ is supported at frequencies $\gtrsim 2^{k_1}$, where k_1 is some sufficiently large parameter. This latter reduction follows in a straightforward manner as low-frequency errors can be controlled by taking T small enough depending on k_1. The reason we perform this reduction is so that we can make use of the more precise pseudodifferential mapping properties in Propositions 2.6 and 2.7, which provide high frequency operator bounds for pseudodifferential operators that depend only on the L^∞ norm of their symbols (as long as the symbol itself does not depend on k_1). This will be of critical importance in Sections 6 and 7, as we will explain below.

3.4. L^2 **bounds for the linear flow.** We next give an outline of Section 6, where we prove the first of the two main components of the bound (3.8). The main aim of Section 6 is to establish control of the $L_T^\infty H_x^\sigma$ norm of v. Throughout the discussion, we assume that $\widehat{v}$ is supported at frequencies larger than 2^{k_1}. Given $\epsilon > 0$, our aim is to prove an estimate of the form

$$\|v\|_{L_T^\infty H_x^\sigma} \leq C(M, L)(\|v_0\|_{H^\sigma} + \|f\|_{Y^\sigma}) + \epsilon \|v\|_{X^\sigma}, \quad \sigma \geq 0, \tag{3.9}$$

on a time interval $[0, T]$ whose length depends on M, L and ϵ. In order to clearly outline the main techniques, we focus on the case $\sigma = 0$. The key objective in this part of the proof is to construct a spatially truncated version of the renormalization operator in (1.6) which conjugates away the "main" portion of the term $\mathrm{Re}(b^j)\partial_j v$. As noted in the introduction, obtaining an $L_T^\infty L_x^2$ bound for (3.5) is straightforward in the absence of such a term. Unlike the symbol in (1.6), however, we want the symbol $O(x, \xi)$ of our renormalization operator $\mathcal{O}$ to be time-independent and to belong to S^0. In view of the first goal (and also to ensure that our symbol is smooth) we truncate in frequency and time, rewriting the paradifferential linear flow as

$$\begin{cases} i\partial_t v + \partial_j T_{g^{ij}} \partial_i v + b_{<k_0}^j(0)\partial_j v + \tilde{b}_{<k_0}^j(0)\partial_j \overline{v} = f + \mathcal{R}^1, \\ v(0) = v_0. \end{cases} \tag{3.10}$$

We then prove that if the frequency truncation parameter k_0 is large enough and T is small enough, the resulting error term $\mathcal{R}^1$ satisfies the perturbative bound

$$\|\mathcal{R}^1\|_{Y^0} \leq \epsilon \|v\|_{X^0}.$$

In order to guarantee smoothness of our symbol O, we will only work with the Hamilton flow (x^t, ξ^t) for the truncated symbol $a := -g_{<k_0}^{ij}(0)\xi_i \xi_j$. By our stability results, the truncated metric $g_{<k_0}^{ij}(0)$ will be nontrapping with comparable parameters to g^{ij} if k_0 is sufficiently large and T is sufficiently small. The

downside of working exclusively with these truncated quantities, however, is that we will need to obtain an estimate of the form

$$\|[\mathcal{O}, \partial_i(T_{g^{ij}} - g^{ij}_{<k_0}(0))\partial_j]\|_{X^0 \to Y^0} \leq \epsilon,$$

when we commute the equation with $\mathcal{O}$. Establishing such a bound is not completely trivial, but can be handled expeditiously with the tools developed in Section 2. Knowing this, our construction of O proceeds as follows: First, we fix a large parameter R such that the coefficients in the equation are small outside of $B_R(0)$. That is,

$$(3.11) \qquad \|(g - g_\infty)\chi_{>R}\|_{l^1 X^{s_0}} + \|(b, \tilde{b})\chi_{>R}\|_{l^1 X^{s_0-1}} \ll \epsilon.$$

We then make the ansatz $O := e^{\psi_1 + \psi_2}$, where $\psi_1, \psi_2 \in S^0$. The purpose of the symbol ψ_1 is to arrange for the leading order cancellation

$$\{a, \psi_1\} + \mathrm{Re}(b^j_{<k_0}(0))\xi_j \geq 0,$$

within the region $B_R(0)$ where the coefficient b^j is potentially large. Roughly speaking (but not exactly), we will take

$$\psi_1(x, \xi) := -\chi_{<2R}(x) \int_{-\infty}^0 \mathrm{Re}((\chi_{<4R}b_{<k_0}(0))(x^t)) \cdot \xi^t dt.$$

The symbol ψ_2 will then be chosen to correct the error terms in the transition region $|x| \approx R$ which appear when derivatives are applied to the localization $\chi_{<2R}$. The resulting symbol O will turn out to be a classical time-independent S^0 symbol, allowing us to avoid the more exotic symbol class (1.7) used in [15].

We crucially note that the spatial localization in O comes with one significant caveat. Namely, it only conjugates away the bad first order term within the region $B_R(0)$. Therefore, we still have to estimate the residual error term $\chi_{>R} \mathrm{Re}(b^j_{<k_0}(0))\partial_j \mathcal{O}v$ in Y^0. Ideally, such an estimate would follow easily from the smallness (3.11) of b^j outside of $B_R(0)$. However, the symbol bounds for O grow in the parameter R. Therefore, we have to somehow ensure that the $X^0 \to X^0$ bounds for $\mathcal{O}$ do not counteract the smallness coming from b^j. This is accomplished by using the observation that, unlike the higher order symbol bounds, the L^∞ norm of the symbol O is independent of the parameter R (as $R \to \infty$). Therefore, since $\widehat{v}$ is supported at frequencies $\gtrsim 2^{k_1}$, we can make use of the bounds in Proposition 2.6 and Proposition 2.7 to ensure that we have an estimate essentially of the form

$$\|\mathcal{O}v\|_{X^0} \lesssim \|O\|_{L^\infty}\|v\|_{X^0}.$$

This is what will ultimately allow us to break any potential circularity in our analysis. As mentioned earlier, an analogous construction was used to establish well-posedness for the electron MHD equations in [14].

3.5. **Local energy bounds for the linear flow.** In Section 7, we turn to the second of the two main components of the bound (3.8). Again, for simplicity of discussion, we take $\sigma = 0$. Here, the aim is to establish control of the local energy component of the X^0 norm of v in terms of the dual norm of f and the $L^\infty_T L^2_x$ norm of v. More precisely, we aim to prove an estimate of the form

$$(3.12) \qquad \|v\|_{X^0} \leq C(M, L)(\|v\|_{L^\infty_T L^2_x} + \|f\|_{Y^0}).$$

Combining this bound with the $L^\infty_T L^2_x$ bound (3.9) (with ϵ sufficiently small), it is relatively straightforward to obtain the main bound (3.8). To obtain (3.12), we implement a novel approach based on the truncation idea used in the $L^\infty_T L^2_x$ bound. As before, we begin by fixing $R > 0$ so that we have the smallness (3.11) outside of $B_R(0)$.

Our first observation is that we can use the small data result from [22] (which holds for the general ul-trahyperbolic Schrödinger flows that we consider here) to reduce having to control the entire local energy component of v to having to obtain the corresponding estimate within the compact set $B_R(0)$. Precisely, we can reduce matters to establishing the bound

$$(3.13) \qquad \|\chi_{<R} v\|_{L_T^2 H_x^{\frac{1}{2}}} \le C(M, L)(\|v\|_{L_T^\infty L_x^2} + \|f\|_{Y^0}) + \epsilon \|v\|_{X^0}.$$

Note that in (3.13) we work with the stronger (but simpler) $L_T^2 H_x^{\frac{1}{2}}$ norm within the compact set $B_R(0)$. The starting point in the proof of this estimate is to rewrite (3.5) as a system for $\mathbf{v} := (v, \overline{v})$:

$$\partial_t \mathbf{v} + \mathbf{P} \mathbf{v} + \mathbf{B}_{k_0}^0 \mathbf{v} = \mathbf{R},$$

where $\mathbf{P}$ is the corresponding principal operator. As in the $L_T^\infty L_x^2$ estimate, the operator $\mathbf{B}_{k_0}^0$ is a suitable time and frequency truncated version of the first order differential operator in the paralinearized Schrödinger equation and $\mathbf{R}$ is a source term which can be controlled by the right-hand side of (3.13) in Y^0. We write $\mathbf{P}_{k_0}^0$ as a shorthand for the associated time and frequency truncated principal operator.

The estimate (3.13) proceeds via a positive commutator argument. Our implementation can be thought of as a spatially truncated version of Doi's argument in [6]. Precisely, we aim to construct a real symbol $q \in S^0$ and a corresponding pseudodifferential operator $\mathbf{Q}$ such that the principal symbol for the commutator $[\mathbf{Q}, \mathbf{P}_{k_0}^0]$ is elliptic within $B_R(0)$ and controls the first order term $\mathbf{B}_{k_0}^0$ up to a small error. Like before, we work with the bicharacteristic flow (x^t, ξ^t) for the time and frequency truncated metric $g_{<k_0}^{ij}(0)$ to ensure that the symbol we construct is smooth and time-independent. To construct q, we first fix a secondary parameter $R' \gg R$ to be chosen. Similarly to before, we make the ansatz

$$q := e^{C(M)(p_1 + p_2 + p_3)},$$

where $C(M)$ is a suitably large constant and p_1, p_2, p_3 are S^0 symbols to be chosen. The choice of p_1 will simply ensure the ellipticity of $[\mathbf{Q}, \mathbf{P}_{k_0}^0]$ in $B_R(0)$. We can take

$$p_1(x, \xi) := -\chi_{<R'} \int_0^\infty \chi_{<R}(x^t, \xi^t) |\xi^t| dt.$$

A natural next step would be to correct this symbol in the transition region $|x| \approx R'$ and use the smallness of the coefficients $(b, \tilde{b})$ outside of $B_R(0)$ as in the $L_T^\infty L_x^2$ estimate. However, this will not work because the L^∞ bound for p_1 will not be uniform in R. Instead, we consider a second symbol p_2 whose purpose will be to ensure that the commutator $[\mathbf{Q}, \mathbf{P}_{k_0}^0]$ controls the first order term $\mathbf{B}_{k_0}^0$ within the much larger compact set $B_{R'}(0)$ but with an L^∞ bound which does not depend on the larger parameter R'. Roughly speaking, we will take p_2 to be

$$p_2(x, \xi) := -\chi_{<R'} \int_0^\infty \chi_{<R'}(x^t) \sqrt{|(b_{<k_0}(0))(x^t)|^2 + |(\tilde{b}_{<k_0}(0))(x^t)|^2 + L(R')^{-2}} \langle \xi^t \rangle dt,$$

which turns out to be a S^0 symbol with the desired properties. The symbol p_3 will then be chosen to correct the error in the transition region $|x| \approx R'$ similarly to the $L_T^\infty L_x^2$ bound. If $\mathbf{v}$ is localized at high enough frequency, the multiplier $\mathbf{Q}$ will then achieve the following key outcomes.

- $[\mathbf{Q}, \mathbf{P}_{k_0}^0]$ will have an essentially positive-definite principal symbol which is elliptic of order 1 within $B_R(0)$. This will permit the use of Gårding's inequality to control $\chi_{<R} \mathbf{v}$ in $L_T^2 H_x^{\frac{1}{2}}$.

- $[\mathbf{Q}, \mathbf{P}_{k_0}^0] \mathbf{v}$ will control the first order term $\chi_{<R'} \mathbf{B}_{k_0}^0 \mathbf{Q} \mathbf{v}$.

- If $\mathbf{v}$ is at high enough frequency 2^{k_1}, $\|\mathbf{Q}S_{>k_1-4}\|_{X^0 \to X^0}$ will be independent of R'. This will allow us to control $\chi_{\geq R'}\mathbf{B}^0_{k_0}\mathbf{v}$ in Y^0 by a small factor of $\|v\|_{X^0}$ by taking R' sufficiently large and using the smallness of b^j and $\tilde{b}^j$ outside of $B_{R'}(0)$.

The above scheme turns out to be sufficient for closing the estimate (3.12). We remark that this method is very robust and works under extremely mild decay assumptions on the coefficients – we essentially only require integrability along the bicharacteristic flow (x^t, ξ^t). Such integrability is guaranteed by the asymptotic flatness condition (3.3) and the fact that the metric is nontrapping.

3.6. **Well-posedness for the nonlinear equation.** Finally, in Section 9 we will make use of the estimate (3.7) for both the linear and paradifferential flows as well as its various corollaries to establish well-posedness for the nonlinear flow. Having established our key linear estimate, the scheme for establishing well-posedness is virtually identical to the one implemented in Section 7 of [24]. We therefore only outline the minor changes, and refer to [24] for additional details. The interested reader may also consult [10] for an expository presentation of the overarching well-posedness scheme.

4. THE BICHARACTERISTIC FLOW

In this section, we define our quantitative measure of nontrapping and establish basic properties of the bicharacteristic flow corresponding to the symbol $a(x, \xi) := -g^{ij}(x)\xi_i\xi_j$. We begin by fixing $s_0 > \frac{d}{2} + 2$ and letting g be a time-independent metric satisfying

$$(4.1) \qquad \|g - g_\infty\|_{l^1 H^{s_0}} \leq M,$$

for some constant non-degenerate symmetric matrix g_∞. We moreover assume the non-degeneracy condition

$$(4.2) \qquad c^{-1}|\xi| \leq |g^{ij}\xi_j| \leq c|\xi|, \qquad \forall \xi \in \mathbb{R}^d,$$

for some constant $c > 0$. By Sobolev embedding, we have for some $\delta > 0$,

$$\|g\|_{C^{2,\delta}} \lesssim_{g_\infty} 1 + M.$$

As a consequence, for each $(x, \xi) \in \mathbb{R}^{2d}$, the bicharacteristic flow $(x^t, \xi^t) := (x^t_{(x,\xi)}, \xi^t_{(x,\xi)})$ given by

$$(4.3) \qquad (\dot{x}^t, \dot{\xi}^t) = (\nabla_\xi a(x^t, \xi^t), -\nabla_x a(x^t, \xi^t)), \qquad (x^0, \xi^0) = (x, \xi),$$

is well-defined in a neighborhood of $t = 0$ (whose size a priori depends on (x, ξ)).

In addition to the above decay and non-degeneracy assumptions, we will also impose the condition that the metric g be nontrapping. The meaning of this is given in a qualitative form by the following definition.

Definition 4.1 (Nontrapping metric). A non-degenerate metric g is said to be *nontrapping* if for every $(x, \xi) \in \mathbb{R}^d \times (\mathbb{R}^d - \{0\})$ and every compact set $K \subset \mathbb{R}^d$, the bicharacteristic x^t intersects K on a compact time interval.

As in [24], we will need a more quantitative description of the above definition. The quantitative parameter $L = L(R)$ we introduce should measure, in some sense, how long a given bicharacteristic can intersect $B_R(0)$. However, since the bicharacteristic flow satisfies the homogeneity law

$$(4.4) \qquad \xi \mapsto \lambda\xi, \qquad t \mapsto \lambda t,$$

such a parameter will not be uniform in the size of ξ. To deal with this, it is natural to restrict to data $\xi \in \mathbb{S}^{d-1}$. From the non-degeneracy of the metric, this restricts the initial speed of a given bicharacteristic to approximately unit size.

At this point, we fix a non-degenerate, nontrapping metric g satisfying (4.1) and (4.2). By a compactness argument, the function

$$L : [0, \infty) \to [0, \infty)$$

given formally by

(4.5) $$L(R) := \inf\{s \geq 0 : |x^t| > R, \ \forall \ |t| \geq s, \ \forall (x, \xi) \in B_R(0) \times \mathbb{S}^{d-1}\}$$

is well-defined. We will use $L := L(R)$ as a quantitative measure of nontrapping.

Remark 4.2. In the case where Δ_g is elliptic, it is automatic that the bicharacteristic flow is globally well-defined because the quantity

(4.6) $$g^{ij}\xi_i\xi_j$$

is preserved by the flow, which, by ellipticity, implies that $|\xi^t|$ remains bounded uniformly in t by $|\xi|$. The same is not immediate when the symbol $g^{ij}\xi_i\xi_j$ is not elliptic and therefore it still needs to be proved that the bicharacteristic flow is globally well-defined. We remark that our definition of the nontrapping parameter L is slightly different than the one used in [24]. In their article, they define L in terms of the maximal time any bicharacteristic for the projected flow onto $\{|\xi^t| = 1\}$ can intersect $B_R(0)$. In light of the above discussion, this is a natural definition in the case of an elliptic symbol, but is not so natural for our purposes because the bicharacteristic flow should not in general preserve any normalization of $|\xi|$ (even though (4.6) is still preserved by the flow). We therefore only restrict the initial ξ to the unit sphere in our quantitative measure of nontrapping.

Our next proposition addresses the problem of global existence and asymptotic bounds for the bicharacteristic flow when the metric is nontrapping and satisfies the decay condition $g - g_\infty \in l^1 H^{s_0}$.

Proposition 4.3. *Let $s_0 > \frac{d}{2} + 2$ and let g be a non-degenerate, nontrapping metric satisfying (4.1). Then*

(i) *For each $(x, \xi) \in \mathbb{R}^d \times (\mathbb{R}^d - \{0\})$, the bicharacteristic flow for $a(x, \xi) := -g^{ij}\xi_i\xi_j$ is globally defined.*

(ii) *For every $\epsilon_0 > 0$ sufficiently small, there exists $R_0 > 0$ such that for any initially outgoing bicharacteristic (i.e. $\dot{x}^t(0) \cdot x \geq 0$) with data $(x, \xi) \in (\mathbb{R}^d - B_{R_0}(0)) \times (\mathbb{R}^d - \{0\})$, x^t is defined for $t \geq 0$ and is close to the flat flow in the sense that for all $t \geq 0$, we have*

(4.7) $$|x^t - x + 2tg_\infty^{ij}\xi_j| \leq t\epsilon_0|\xi|, \quad |\xi^t - \xi| \leq \epsilon_0|\xi|.$$

Proof. The proof of this is very similar to Lemma 5.1 in [24]. We include the short argument for completeness. We begin by choosing R_0 large enough so that g is sufficiently close to the flat metric g_∞ in $l^1 H^{s_0}$ outside of $B_{\frac{R_0}{2}}(0)$. That is,

(4.8) $$\|\chi_{>\frac{R_0}{2}}(g - g_\infty)\|_{l^1 H^{s_0}} < \epsilon,$$

where $0 < \epsilon \ll \epsilon_0 \ll 1$ is some sufficiently small constant relative to ϵ_0. We let (x^t, ξ^t) be any initially outgoing bicharacteristic with data $(x, \xi) \in (\mathbb{R}^d - B_{R_0}(0)) \times (\mathbb{R}^d - \{0\})$ and make the bootstrap assumption that the bicharacteristic (x^t, ξ^t) satisfies (4.7) on a time interval $t \in [0, T]$. Our goal will be to show that when $\epsilon > 0$ is small enough, the factor of ϵ_0 in the bootstrap hypothesis can be improved to $\frac{\epsilon_0}{2}$. Thanks to the nontrapping assumption on g, this will clearly suffice for establishing both (i) and (ii).

To close the bootstrap, we note that on $[0, T]$, thanks to (4.7) and the fact that x^t is initially outgoing, the bicharacteristic x^t remains outside $B_{\frac{3}{4}R_0}(0)$. Using this and the bootstrap hypothesis, we aim to prove the following simple lemma.

Lemma 4.4. *The following estimate holds for every $t \in [0, T]$:*

$$\int_0^t |\nabla_x g(x^s)| ds \lesssim \epsilon |\xi|^{-1}.$$

Proof. We estimate

$$\int_0^t |\nabla_x g(x^s)| ds \lesssim \sum_{k \geq 0} \sum_{Q \in Q_k} \int_0^t |\chi_Q(x^s)(S_k(\chi_{> \frac{R_0}{2}} \nabla_x g))(x^s)| ds$$

$$\lesssim |\xi|^{-1} \sum_{k \geq 0} \sum_{Q \in Q_k} 2^k \|\chi_Q S_k(\chi_{> \frac{R_0}{2}} \nabla_x g)\|_{L^\infty}$$

$$\lesssim |\xi|^{-1} \|\chi_{> \frac{R_0}{2}} \nabla_x g\|_{l^1 H^{s_0 - 1}}$$

$$\leq |\xi|^{-1} \epsilon,$$

where in the second line we used (4.7) and the non-degeneracy of g_∞, which ensures that the bicharacteristic x^s intersects a cube of size 2^k for time at most $\lesssim 2^k |\xi|^{-1}$. In the third line, we used Bernstein's inequality and in the fourth line, we used (4.8). This concludes the proof. $\square$

To close the bootstrap, we note that by (4.7) we have $|\xi^t| \leq (1 + \epsilon_0)|\xi|$. Therefore, by using Lemma 4.4 and integrating in time the equation

$$\frac{d}{dt}(\xi^t - \xi) = \nabla_x g^{ij}(x^t)\xi_i^t \xi_j^t$$

we obtain

$$|\xi^t - \xi| \lesssim \epsilon |\xi|.$$

Using this bound, integrating in time the equation

$$\frac{d}{dt}(x^t - x + 2tg_\infty^{ij}\xi_j) = 2(g_\infty^{ij} - g^{ij})(x^t)\xi_j^t - 2g_\infty^{ij}(\xi_j^t - \xi_j)$$

and using that $|(g - g_\infty)(x^t)| \lesssim \epsilon$, we also obtain

$$|x^t - x + 2tg_\infty^{ij}\xi_j| \lesssim t\epsilon |\xi|,$$

which improves the bootstrap (4.7) if ϵ is small enough relative to ϵ_0. This concludes the proof of the proposition. $\square$

The next proposition shows that the size of the nontrapping function L as well as the bicharacteristic bounds are stable under small perturbations of the metric.

Proposition 4.5. *Let g_0 be a non-degenerate nontrapping metric satisfying (4.1). For every sufficiently small $\epsilon_0 > 0$, there is a radius $R_0(\epsilon_0) > 0$ and a constant $C_0 > 0$ depending only on M and the profile of g_0 such that if g_1 is another non-degenerate metric satisfying*

$$(4.9) \qquad\qquad \|g_0 - g_1\|_{l^1 H^{s_0}} < e^{-C_0 L(R_0)}$$

then the bicharacteristics corresponding to g_1 satisfy (ii) in Proposition 4.3 with comparable parameters R_0 and ϵ_0 and, moreover, g_1 is also nontrapping with comparable parameters L_1 and data size M_1.

Proof. Choosing $e^{-C_0 L(R_0)}$ so small that

$$\|g_0 - g_1\|_{l^1 H^{s_0}} \ll \epsilon_0$$

ensures that the data size M_1 is comparable to M and also that the proof of part (ii) of Proposition 4.3 works equally well for the metric g_1. It therefore suffices to show that L_1 is comparable to L for $R \leq R_0$. To do

this, we fix $(x, \xi) \in B_{R_0}(0) \times \mathbb{S}^{d-1}$. The desired conclusion will follow if we can show that the bicharacteristic flows corresponding to g_0 and g_1 are close within $B_{R_0}(0)$ in the sense that

$$(4.10) \qquad |x_0^t - x_1^t|_{L_t^\infty} + |\xi_0^t - \xi_1^t|_{L_t^\infty} \lesssim e^{-C(M)L(R_0)}$$

for times in which x_0^t intersects $B_{R_0}(0)$. The proof of this is similar to the proof of Proposition 5.2 in [24] but since our nontrapping parameter L is slightly different than theirs, we include the short proof.

We implement a simple bootstrap. First, we can restrict to a time interval J such that $|J| \leq L(R_0)$. We will then assume the bound (4.10) on some smaller interval $I \subset J$ and establish the same bound with an improved constant. We begin by writing the equation for $\delta x^t := x_0^t - x_1^t$ and $\delta \xi^t := \xi_0^t - \xi_1^t$. Dropping the i, j indices, we obtain

$$\begin{cases} \frac{d}{dt}\delta x^t = 2(g_1 - g_0)(x_1^t)\xi_1^t + 2(g_0(x_1^t) - g_0(x_0^t))\xi_0^t - 2g_0(x_1^t)\delta\xi^t, \\ \frac{d}{dt}\delta\xi^t = -\xi_1^t \nabla(g_1 - g_0)(x_1^t)\xi_1^t - \xi_1^t(\nabla g_0(x_1^t) - \nabla g_0(x_0^t))\xi_1^t + (\xi_0^t \nabla g_0(x_0^t)\xi_0^t - \xi_1^t \nabla g_0(x_0^t)\xi_1^t), \\ (\delta x^0, \delta\xi^0) = (0, 0). \end{cases}$$

By definition, we have $|I| \leq L(R_0)$. Moreover, by a compactness argument, there is a constant $K_0 > 1$ depending on the profile of g_0 (but not on (x, ξ)) such that

$$|\xi_0^t| \lesssim K_0$$

for every $t \in J$. By the bootstrap hypothesis, this implies the same bound for ξ_1^t on I. From this, (4.9), the bootstrap hypothesis and the fact that $g_0 \in C^2$, we obtain the bound

$$\frac{d}{dt}[(\delta x^t)^2 + (\delta\xi^t)^2] \lesssim e^{-2C_0 L(R_0)} + C(K_0)(1 + M)[(\delta x^t)^2 + (\delta\xi^t)^2].$$

By Grönwall's inequality and the bound $|I| \leq L(R_0)$, we obtain

$$(\delta x^t)^2 + (\delta\xi^t)^2 \lesssim e^{-2C_0 L(R_0)} e^{C(K_0)(1+M)L(R_0)}$$

on I. Choosing C_0 large enough improves the bootstrap hypothesis and concludes the proof. $\qquad \square$

By combining Proposition 4.3 with Proposition 4.5, we have the following immediate corollary which gives a precise quantitative bound for ξ^t.

Corollary 4.6. *Let g_0 be as in Proposition 4.3. Then the corresponding bicharacteristic ξ_0^t (which is defined for all t) satisfies the bound*

$$|\xi_0^t| \lesssim C_0|\xi|$$

for all $(x, \xi) \in \mathbb{R}^{2d}$ and some constant $C_0 > 1$ depending only on M and the profile of g_0. Moreover, if g_1 is any other metric satisfying the conditions of Proposition 4.5, then the corresponding bicharacteristic ξ_1^t is globally defined and satisfies the same bound with a similar constant.

Proof. For $|\xi| = 1$, this follows immediately from Proposition 4.3, Proposition 4.5, and the nontrapping assumption. The general case follows from this case and the homogeneity law (4.4). $\qquad \square$

We next note the following bounds for the x and ξ derivatives of x^t and ξ^t.

Proposition 4.7 (Higher regularity bounds). *Let $(x, \xi) \in B_R(0) \times \mathbb{S}^{d-1}$. Let k be a positive integer. Assume that the metric satisfies $g \in C^{k+1}$ and write $M_k := \|g\|_{C^{k+1}}$. Then if $|x^t| \leq R$, there holds*

$$|\partial_\xi^\alpha \partial_x^\beta x_{(x,\xi)}^t| + |\partial_\xi^\alpha \partial_x^\beta \xi_{(x,\xi)}^t| \leq e^{C(M_k)L(R)}, \qquad |\alpha + \beta| \leq k.$$

Proof. The proof follows by differentiating (4.3) in the x and ξ variables, which leads to a differential inequality for

$$\frac{d}{dt}\left(|\partial_\xi^\alpha \partial_x^\beta x^t|^2 + |\partial_\xi^\alpha \partial_x^\beta \xi^t|^2\right).$$

One then concludes by inductively applying Grönwall's inequality. We omit the details which are straightforward. $\square$

The final result of this section shows that functions in $l^1 H^s$ with $s > \frac{d}{2} + 1$ are uniformly integrable along the bicharacteristic flow. This is what will allow us to recover the Mizohata condition.

Proposition 4.8. *Let g be as in Proposition 4.3 and let $s > \frac{d}{2} + 1$. Let $v \in l^1 H^s$. Then v is integrable along the bicharacteristic flow and satisfies the bound*

$$\sup_{(x,\xi)\in\mathbb{R}^d\times\mathbb{S}^{d-1}} \|v(x^t_{(x,\xi)})\|_{L^1_t(\mathbb{R})} \lesssim (1 + L(R_0))\|v\|_{l^1 H^s},$$

where R_0 is as in Proposition 4.3.

Proof. We abbreviate $x^t_{(x,\xi)}$ by x^t. Without loss of generality, we may assume that x^t intersects $B_{R_0}(0)$ only if $|t| < L(R_0)$. We then have

$$\int_{-L(R_0)}^{L(R_0)} |v(x^t)|dt \leq 2L(R_0)\|v\|_{L^\infty} \lesssim L(R_0)\|v\|_{l^1 H^s}.$$

By homogeneity of the flow, it therefore suffices to show that

$$\int_{L(R_0)}^{\infty} |v(x^t)|dt \lesssim \|v\|_{l^1 H^s}.$$

Without loss of generality, we may assume that $x^t(L(R_0))$ is outgoing. Using Proposition 4.3, we see that if $t \geq L(R_0)$, then for every cube $Q \subset \mathbb{R}^d$, x^t intersects the cube on a time interval I of size at most $|I| \lesssim |Q|^{\frac{1}{d}}$. Therefore, we have

$$\int_{L(R_0)}^{\infty} |v(x^t)|dt \lesssim \sum_{k\geq 0}\sum_{Q\in Q_k} 2^k \|\chi_Q S_k v\|_{L^\infty} \lesssim \|v\|_{l^1 H^s},$$

where in the last step we used Bernstein's inequality and dyadic summation. Here, the strict inequality $s > \frac{d}{2} + 1$ was what allowed us to retain summability in k. This completes the proof. $\square$

5. The linear ultrahyperbolic flow

Let $s_0 > \frac{d}{2} + 2$ and let $0 \leq \sigma \leq s_0$. Here we consider the $l^1 H^\sigma$ well-posedness of the linear ultrahyperbolic flow,

$$(5.1) \qquad \begin{cases} i\partial_t v + \partial_j g^{jk}\partial_k v + b^j \partial_j v + \tilde{b}^j \partial_j \overline{v} = f, \\ v(0,x) = v_0(x), \end{cases}$$

as well as the corresponding linear paradifferential flow,

$$(5.2) \qquad \begin{cases} i\partial_t v + \partial_j T_{g^{jk}}\partial_k v + T_{b^j}\partial_j v + T_{\tilde{b}^j}\partial_j \overline{v} = f, \\ v(0,x) = v_0(x). \end{cases}$$

We make the following basic assumptions on the metric g^{jk} and the coefficients b^j in the above equations:

(i) (Non-degeneracy). The metric g^{jk} is real, symmetric and non-degenerate. That is, there is $c > 0$ such that for all $\xi \in \mathbb{R}^d$ we have,

$$c^{-1}|\xi| \leq |g^{jk}\xi_k| \leq c|\xi|.$$

(ii) (Asymptotic flatness and size). There is a constant, symmetric, non-degenerate matrix g_∞ and a constant $M > 0$ such that

(5.3) $$\|g - g_\infty\|_{l^1 X^{s_0}} + \|\partial_t g\|_{l^1 X^{s_0-2}} + \|(b, \tilde{b})\|_{l^1 X^{s_0-1}} + \|\partial_t (b, \tilde{b})\|_{l^1 X^{s_0-3}} \leq M.$$

(iii) (Asymptotic smallness). For every $\epsilon_0 > 0$, there is $R_0 > 0$ such that

(5.4) $$\|(g^{jk} - g_\infty^{jk})\chi_{>R_0}\|_{l^1 X^{s_0}} + \|(b^j, \tilde{b}^j)\chi_{>R_0}\|_{l^1 X^{s_0-1}} \leq \epsilon_0,$$

where $0 \leq \chi_{>R_0} \leq 1$ is a smooth cutoff which vanishes on $B_{R_0}(0)$ and is equal to 1 outside of $B_{2R_0}(0)$.

(iv) (Nontrapping). The metric is nontrapping with parameter L as defined in (4.5).

Note that condition (iii) follows from the asymptotic flatness condition (ii). However, we prefer to make statement (iii) explicit, as it will play a prominent role in the analysis.

In the sequel, we will write $C(L)$ to denote a constant which depends on the parameter L within some fixed compact set whose size depends on the profile of the metric g. The main result we aim to prove is the following.

Theorem 5.1. *Let $s_0 > \frac{d}{2} + 2$ and $0 \leq \sigma \leq s_0$. Moreover, assume that $g^{jk}, b^j, \tilde{b}^j$ satisfy the above assumptions with parameters M and L. Then for every $f \in l^1 Y^\sigma$, the equation (5.1) is well-posed in $l^1 H^\sigma$. Furthermore, there is $T_0 > 0$ depending on the size of L within a compact set and on the data size M such that for every $0 \leq T \leq T_0$, we have*

(5.5) $$\|v\|_{l^1 X^\sigma} \leq C(M, L)(\|v_0\|_{l^1 H^\sigma} + \|f\|_{l^1 Y^\sigma}).$$

The same result holds for the paradifferential flow (5.2) for every $\sigma \geq 0$.

As the above result holds for the paradifferential flow for all $\sigma \geq 0$, it is a straightforward consequence to deduce the following frequency envelope variant using similar reasoning to Section 5 of [22] (see also [10]).

Corollary 5.2. *Let $\sigma \geq 0$ and assume the other properties in the statement of Theorem 5.1. Let a_k be an admissible $l^1 H^\sigma$ frequency envelope for v_0 and let b_k be an admissible $l^1 Y^\sigma$ frequency envelope for f. Then the solution v to the paradifferential equation (5.2) satisfies the bound*

$$\|S_k v\|_{l^1 X^\sigma} \leq C(M, L)\,(a_k\|v_0\|_{l^1 H^\sigma} + b_k\|f\|_{l^1 Y^\sigma})$$

on a time interval $[0, T]$ whose length depends on the size of L within a compact set and on the data size M.

The main component of the proof of well-posedness for the equations (5.1) and (5.2) is the energy estimate (5.5). This is because the adjoint equation, which has essentially the same form, will also satisfy a similar energy estimate. Well-posedness then follows by a standard duality argument. Therefore, we focus our attention mainly on the bound (5.5).

5.1. **Some simplifying reductions.** We begin our analysis by making some straightforward but useful reductions which will allow us to simplify some of the steps in the proof of Theorem 5.1. Our first reduction shows that by restricting the time interval to be small enough, we may assume that $\hat{v}$ is supported at high frequency. More precisely, we have the following lemma.

Lemma 5.3 (High frequency reduction). *Let $\epsilon > 0$. Under the assumptions of Theorem 5.1, for every $k_1 > 0$ there is $T_0 > 0$ depending on k_1, ϵ, M and σ such that for $0 < T \leq T_0$, $v_{>k_1} := S_{>k_1}v$ satisfies the equation*

$$\begin{cases} i\partial_t v_{>k_1} + \partial_j g^{ij}\partial_i v_{>k_1} + b^j\partial_j v_{>k_1} + \tilde{b}^j\partial_j \overline{v_{>k_1}} = h, \\ v_{>k_1}(0) := S_{>k_1}v_0, \end{cases}$$

where h and $S_{\leq k_1}v$ satisfy the estimate

$$\|S_{\leq k_1}v\|_{l^1 X^\sigma} + \|h\|_{l^1 Y^\sigma} \leq C(M, k_1, \sigma)(\|v_0\|_{l^1 H^\sigma} + \|f\|_{l^1 Y^\sigma}) + \epsilon\|v\|_{l^1 X^\sigma}$$

for $0 \leq \sigma \leq s_0$. The analogous result holds for the paradifferential equation (5.2) for $\sigma \geq 0$.

Proof. We show the proof for the full linear equation. The proof for the paradifferential flow is similar. Using the notation of the lemma, we easily compute that

$$h = S_{>k_1}f - (\partial_j g^{ij}\partial_i S_{\leq k_1}v + b^j\partial_j S_{\leq k_1}v + \tilde{b}^j\partial_j S_{\leq k_1}\overline{v})$$
$$+ S_{\leq k_1}(\partial_j g^{ij}\partial_i v + b^j\partial_j v + \tilde{b}^j\partial_j\overline{v}).$$

We clearly have

$$\|S_{>k_1}f\|_{l^1 Y^\sigma} \lesssim \|f\|_{l^1 Y^\sigma}.$$

For the remaining source terms, if $0 \leq \sigma \leq s_0 - 1$, we can estimate in $l^1 L_T^1 H_x^\sigma \subset l^1 Y^\sigma$ in a naïve fashion using the frequency projection $S_{\leq k_1}$ to obtain

$$\|h\|_{l^1 Y^\sigma} \lesssim \|f\|_{l^1 Y^\sigma} + \epsilon\|v\|_{l^1 L_T^\infty H_x^\sigma},$$

by applying Hölder's inequality in T and taking T small enough (depending on k_1). On the other hand, for $s_0 - 1 < \sigma \leq s_0$, we can instead use the bilinear estimates in Proposition 2.13 to obtain

$$\|h\|_{l^1 Y^\sigma} \lesssim \|f\|_{l^1 Y^\sigma} + C(M, k_1, \sigma)\|v\|_{l^1 L_T^\infty L_x^2}.$$

We can estimate the latter term on the right using the crude energy inequality

$$\|v\|_{l^1 L_T^\infty L_x^2} \lesssim_M \|v_0\|_{l^1 L_x^2} + \|v\|_{l^1 L_T^1 H_x^1} + \|f\|_{l^1 Y^\sigma},$$

which follows from a direct energy estimate for (5.1) where the first order terms are estimated directly in $L_T^1 L_x^2$. Since $\sigma > 1$, we may conclude by applying Hölder in T and taking $T \ll \epsilon$ to control the second term on the right by $\epsilon\|v\|_{l^1 X^\sigma}$. It remains to estimate $\|S_{\leq k_1}v\|_{l^1 X^\sigma}$. Using that $S_{\leq k_1}v$ is frequency localized, we easily have

$$\|S_{\leq k_1}v\|_{l^1 X^\sigma} \lesssim 2^{k_1(\sigma+\frac{1}{2})}\|S_{\leq k_1}v\|_{l^1 L_T^\infty L_x^2}.$$

We then note the naïve energy type estimate

$$\|S_{\leq k_1}v\|_{l^1 L_T^\infty L_x^2} \lesssim_{M,k_1} \|v_0\|_{l^1 L_x^2} + \|v\|_{l^1 L_T^1 L_x^2} + \|f\|_{l^1 Y^0},$$

which follows from inspecting the equation for $S_{\leq k_1}v$ and using the fact that the first and second-order terms in the resulting equation are localized to frequencies $\lesssim k_1$. Then using Hölder in T and taking T small enough (depending on M, k_1 and ϵ) we can again control the second term on the right by $\epsilon\|v\|_{l^1 X^\sigma}$. This

concludes the proof of the lemma for (5.1). A very similar argument works for the paradifferential analogue. We omit the details. $\qquad\square$

5.2. **Reduction to the paradifferential flow.** As a second reduction, we reduce proving Theorem 5.1 to proving the corresponding estimate for the paradifferential equation. We begin by writing (5.1) in the paradifferential form

$$\begin{cases} i\partial_t v + \partial_j T_{g^{ij}} \partial_i v + T_{b^j} \partial_j v + T_{\tilde{b}^j} \partial_j \overline{v} = f + \mathcal{R}, \\ v(0) = v_0, \end{cases}$$

where $\mathcal{R}$ is a remainder term given by

$$(5.6) \qquad \mathcal{R} = (T_{b^j} \partial_j v - b^j \partial_j v) + (T_{\tilde{b}^j} \partial_j \overline{v} - \tilde{b}^j \partial_j \overline{v}) + \partial_j (T_{g^{ij}} \partial_i v - g^{ij} \partial_i v).$$

Thanks to Lemma 5.3, we may harmlessly assume that v is localized to frequencies $\gtrsim 2^{k_1}$. Our next lemma shows that the error term $\mathcal{R}$ can be treated perturbatively if k_1 is large enough.

Lemma 5.4 (Paradifferential source terms). *Assume that the estimate in Theorem 5.1 holds for the paradifferential flow for each $\sigma \geq 0$. Let $\epsilon > 0$ and assume that $\widehat{v}$ is supported at frequencies $|\xi| \gtrsim 2^{k_1}$. Then for k_1 large enough and T small enough depending on ϵ and k_1, the remainder term $\mathcal{R}$ satisfies the estimate*

$$\|\mathcal{R}\|_{l^1 Y^\sigma} \leq C(M, k_1, \sigma)(\|v_0\|_{l^1 H^\sigma} + \|f\|_{l^1 Y^\sigma}) + \epsilon \|v\|_{l^1 X^\sigma}.$$

Proof. We show the details for the first term as the estimates for the other two are similar. We split the analysis into two cases. First, assume that $\sigma \leq s_0 - \delta$ where $\delta > 0$ is such that $s_0 - 2\delta > \frac{d}{2} + 2$. Then since v is localized to frequencies $\gtrsim 2^{k_1}$, we may replace the coefficient b^j in $(T_{b^j} \partial_j v - b^j \partial_j v)$ with $S_{\geq k_1 - 5} b^j$. Therefore, by (2.7) in Proposition 2.13 and Bernstein's inequality, we have

$$\|(T_{b^j} \partial_j v - b^j \partial_j v)\|_{l^1 Y^\sigma} \lesssim \|S_{\geq k_1 - 5} b^j\|_{l^1 X^{s_0 - 1 - \delta}} \|v\|_{l^1 X^\sigma}$$
$$\lesssim_M 2^{-\delta k_1} \|v\|_{l^1 X^\sigma}.$$

Taking k_1 large enough, we therefore have

$$\|(T_{b^j} \partial_j v - b^j \partial_j v)\|_{l^1 Y^\sigma} \leq \epsilon \|v\|_{l^1 X^\sigma}.$$

The other terms in (5.6) can be estimated similarly to obtain

$$\|\mathcal{R}\|_{l^1 Y^\sigma} \leq \epsilon \|v\|_{l^1 X^\sigma}.$$

In the case $s_0 \geq \sigma \geq s_0 - \delta > \frac{d}{2} + 2$, we use instead the first estimate in Proposition 2.13 to obtain

$$\|(T_{b^j} \partial_j v - b^j \partial_j v)\|_{l^1 Y^\sigma} \lesssim \|b^j\|_{l^1 X^{\sigma-1}} \|v\|_{l^1 X^{s_0 - 2\delta}}$$
$$\lesssim_M 2^{-k_1 \delta} \|v\|_{l^1 X^\sigma},$$

where we used the fact that v is localized to frequencies greater than 2^{k_1}. Estimating the other terms in (5.6) in a similar fashion, and again taking k_1 large enough, we obtain

$$\|\mathcal{R}\|_{l^1 Y^\sigma} \leq \epsilon \|v\|_{l^1 X^\sigma}.$$

This concludes the proof. $\qquad\square$

5.3. **Reduction to the X^σ estimate.** To summarize what we have so far, it now suffices to establish (5.5) for the paradifferential flow under the assumption that v is localized to high frequency. As one final simplification, we reduce the proof of this estimate for the paradifferential flow to the corresponding X^σ estimate without the l^1 summability. For this, we will need the small data result from [22].

Theorem 5.5 (Small data well-posedness). *Let b^j, $\tilde{b}^j$, g^{ij} and M, σ be as above. Let $0 < T \leq 1$. For every $\sigma \geq 0$, there is $\delta > 0$ such that if $M \leq \delta$ then (5.2) is well-posed in both H^σ and $l^1 H^\sigma$ with the uniform bounds*

$$\|v\|_{X^\sigma} \lesssim \|v_0\|_{H^\sigma} + \|f\|_{Y^\sigma},$$
$$\|v\|_{l^1 X^\sigma} \lesssim \|v_0\|_{l^1 H^\sigma} + \|f\|_{l^1 Y^\sigma}.$$

Remark 5.6. Strictly speaking, the small data result above is only explicitly stated in the case when g_∞ is the identity, but as remarked on page 1154 of [22], the result is also true when g_∞ is of the form we consider here, and the estimates above follow almost verbatim from the proof of Proposition 4.1 in their paper.

We may now phrase our final reduction as follows.

Lemma 5.7. *Let b^j, $\tilde{b}^j$, g^{ij} and M, σ be as in Theorem 5.1. Assume that the paradifferential flow (5.2) admits the estimate*

$$\|v\|_{X^\sigma} \leq C(M, L)(\|v_0\|_{H^\sigma} + \|f\|_{Y^\sigma}) \tag{5.7}$$

for each $\sigma \geq 0$. Then the corresponding estimate in Theorem 5.1 in $l^1 X^\sigma$ also holds for (5.2) for each $\sigma \geq 0$.

Proof. We can again harmlessly assume that v is localized to frequencies $\gtrsim 2^{k_1}$. Now let $\epsilon > 0$ and let $R(\epsilon)$ be such that (5.4) holds. Using Proposition 2.13 and Theorem 5.5, our first aim will be to reduce to estimating v in a compact set. More precisely, we aim to prove the estimate

$$\|\chi_{>2R} v\|_{l^1 X^\sigma} \lesssim \|v_0\|_{l^1 H^\sigma} + \|f\|_{l^1 Y^\sigma} + \|\chi_{<4R} v\|_{l^1 X^\sigma}. \tag{5.8}$$

This is a straightforward computation which follows by inspecting the equation for $v_{ext} := \chi_{>2R} v$. Indeed, if we define $g_{ext} := \chi_{>R} g + \chi_{\leq R} g_\infty$, $b_{ext} := \chi_{>R} b$ and $\tilde{b}_{ext} := \chi_{>R} \tilde{b}$, we obtain

$$\begin{cases} i\partial_t v_{ext} + \partial_i T_{g_{ext}^{ij}} \partial_j v_{ext} + T_{b_{ext}^j} \partial_j v_{ext} + T_{\tilde{b}_{ext}^j} \partial_j \overline{v}_{ext} = f_{ext}, \\ v_{ext}(0) = \chi_{>2R} v(0), \end{cases}$$

where

$$f_{ext} := \chi_{>2R} f + [\partial_i T_{g^{ij}} \partial_j + T_{b^j} \partial_j + T_{\tilde{b}^j} \partial_j, \chi_{>2R}] v + (\partial_i T_{g_{ext}^{ij}} \partial_j - \partial_i T_{g^{ij}} \partial_j) v_{ext}$$
$$+ (T_{b_{ext}^j} \partial_j - T_{b^j} \partial_j) v_{ext} + (T_{\tilde{b}_{ext}^j} \partial_j - T_{\tilde{b}^j} \partial_j) \overline{v}_{ext}.$$

Making use of Proposition 2.13 and paradifferential calculus, we can easily estimate

$$\|[\partial_i T_{g^{ij}} \partial_j + T_{b^j} \partial_j + T_{\tilde{b}^j} \partial_j, \chi_{>2R}] v\|_{l^1 Y^\sigma} \leq C(M, R)(\|\chi_{<4R} v\|_{l^1 X^\sigma} + \|v\|_{l^1 L_T^1 H_x^\sigma})$$
$$\leq C(M, R)\|\chi_{<4R} v\|_{l^1 X^\sigma} + \delta\|v\|_{l^1 X^\sigma}$$

for some small $\delta > 0$. We note that in the last inequality, we used Hölder's inequality in T and took T sufficiently small depending on R and M. Using the disjointness of the supports of $g_{ext} - g$ and v_{ext}, we obtain from the embedding $l^1 L_T^1 H_x^\sigma \subset l^1 Y^\sigma$ and paradifferential calculus,

$$\|(\partial_i T_{g_{ext}^{ij}} \partial_j - \partial_i T_{g^{ij}} \partial_j) v_{ext}\|_{l^1 Y^\sigma} \lesssim_M \|v\|_{l^1 L_T^1 H_x^\sigma} \lesssim \delta\|v\|_{l^1 X^\sigma}.$$

We can similarly estimate the last two terms in the definition of f_{ext}. In light of this and the small data result Theorem 5.5 which applies to the equation for v_{ext}, we obtain (5.8). We have therefore reduced the estimate for v in $l^1 X^\sigma$ to obtaining the bound

$$\|\chi_{<4R} v\|_{l^1 X^\sigma} \leq C(M, L)(\|v_0\|_{l^1 H^\sigma} + \|f\|_{l^1 Y^\sigma}).$$

However, this simply follows from (5.7) and the fact that the $l^1 X^\sigma$ and X^σ norms are equivalent within the set $B_{4R}(0)$ (with equivalence constant depending on R). $\qquad\square$

6. THE L^2 ESTIMATE FOR THE LINEAR FLOW

We begin our analysis by showing that we can close an estimate for the $L_T^\infty H_x^\sigma$ norm of a solution to the paradifferential linear equation (5.2) up to a small error term in X^σ as long as the time interval is small enough. Thanks to Lemma 5.3, we may from here on harmlessly assume that

$$\text{supp}(\widehat{v}) \subset \{|\xi| > 2^{k_1}\}$$

for some large parameter k_1 to be chosen. We will make this assumption for the rest of the section. The main estimate we wish to prove is the following.

Proposition 6.1 (L^2 estimate for the paradifferential linear flow). *Let s_0, g^{ij}, b^j and $\tilde{b}^j$ be as in Theorem 5.1 with parameters M and L. Let $\epsilon > 0$. There is $T_0 = T_0(\epsilon) > 0$ such that for $0 \leq T \leq T_0$, we have the a priori bound for v satisfying (5.2),*

$$\|v\|_{L_T^\infty H^\sigma} \leq C(M, L)(\|v_0\|_{H^\sigma} + \|f\|_{Y^\sigma}) + \epsilon\|v\|_{X^\sigma},$$

for every $\sigma \geq 0$.

As noted earlier, by $C(M, L)$ we mean a constant which depends on M and the trapping parameter L within some fixed compact set (which is allowed to depend on ϵ). The main obstruction to establishing Proposition 6.1 is essentially the presence of the real part of the first order term $T_{\text{Re}(b^j)} \partial_j v$. This is characterized somewhat by the following basic estimate for a truncated version of the linear flow in which the coefficient b^j is purely imaginary.

Lemma 6.2 (Basic energy estimate). *Let g^{ij} be smooth, real and symmetric and let b^j and $\tilde{b}^j$ be smooth functions. Assume that we have the size condition (5.3). Moreover, let $A(x, D) \in OPS^1$ be a time-independent pseudodifferential operator with symbol satisfying $\text{Re}(A) \geq 0$ and assume that v solves the equation*

$$(6.1) \qquad\qquad i\partial_t v + \partial_i T_{g^{ij}} \partial_j v + i\,\text{Im}(b^j)\partial_j v + \tilde{b}^j \partial_j \overline{v} + iA(x, D)v = f.$$

Then for every $0 < \delta \ll 1$ there is $T_0 > 0$ depending on M, δ and A such that for $0 < T \leq T_0$, we have the L^2 estimate,

$$\|v\|_{L_T^\infty L_x^2}^2 \lesssim \|v_0\|_{L_x^2}^2 + \|v\|_{X^0}\|f\|_{Y^0} + \delta\|v\|_{X^0}^2.$$

In the above lemma, we allow for the extra first order term $iA(x, D)v$. This will afford us some flexibility when dealing with commutations of the principal operator $\partial_j T_{g^{ij}} \partial_i$ with various zeroth order Fourier multipliers and pseudodifferential operators later on when we deal with the full linear paradifferential flow.

Proof. We start with the basic energy identity:

$$\|v(t)\|_{L_x^2}^2 + 2\,\text{Re}\langle A(x, D)v, v\rangle = \|v_0\|_{L_x^2}^2 + 2\,\text{Re}\langle i\partial_i T_{g^{ij}}\partial_j v, v\rangle - 2\,\text{Re}\langle \text{Im}(b^j)\partial_j v, v\rangle + 2\,\text{Re}\langle i\tilde{b}^j \partial_j \overline{v}, v\rangle$$
$$- 2\,\text{Re}\langle if, v\rangle,$$

which holds for each $0 \le t \le T$. Here $\langle \cdot, \cdot \rangle$ denotes the inner product on $L^2_t L^2_x$. Unlike the operator $\partial_i g^{ij} \partial_j$, the paradifferential operator $\partial_i T_{g^{ij}} \partial_j$ is not quite self-adjoint. However, we do have the relation

$$\mathrm{Re} \langle i \partial_i T_{g^{ij}} \partial_j v, v \rangle = \mathrm{Re} \langle i \partial_i (T_{g^{ij}} - g^{ij}) \partial_j v, v \rangle.$$

By standard paradifferential calculus and the fact that $\|g^{ij}\|_{L^\infty_T C^{2,\alpha}} \le C(M)$ for some $\alpha > 0$, we have

$$\|\partial_i (T_{g^{ij}} - g^{ij}) \partial_j v\|_{L^2_x} \lesssim_M \|v\|_{L^2_x}.$$

Hence, by Hölder in T and taking T sufficiently small, we have

$$2 \mathrm{Re} \langle i \partial_i T_{g^{ij}} \partial_j v, v \rangle \lesssim_M T \|v\|^2_{L^\infty_T L^2_x} \le \delta \|v\|^2_{X^0}.$$

Now, we turn to the other terms in the energy estimate. Integrating by parts and making use of Sobolev embeddings, we obtain the bound

$$-2 \mathrm{Re} \langle \mathrm{Im}(b^j) \partial_j v, v \rangle + 2 \mathrm{Re} \langle i \tilde{b}^j \partial_j \overline{v}, v \rangle \lesssim MT \|v\|^2_{L^\infty_T L^2_x} \le \delta \|v\|^2_{X^0},$$

if T is small enough. Moreover, by the $Y^* = X$ duality, we have

$$-2 \mathrm{Re} \langle if, v \rangle \lesssim \|v\|_{X^0} \|f\|_{Y^0}.$$

Therefore, if T is small enough, we arrive at the bound

$$\|v\|^2_{L^\infty_T L^2_x} + \mathrm{Re} \langle A(x,D) v, v \rangle \lesssim \|v_0\|^2_{L^2_x} + \|v\|_{X^0} \|f\|_{Y^0} + \delta \|v\|^2_{X^0}.$$

Finally, by the sharp Gårding inequality Proposition 2.2 and Hölder in time, we have

$$\mathrm{Re} \langle A(x,D) v, v \rangle \gtrsim_A -T \|v\|^2_{L^\infty_T L^2_x}.$$

Taking T sufficiently small concludes the proof. $\qquad \square$

The remainder of this section will be essentially devoted to transforming the equation (5.2) into an equation of the ideal form (6.1). Our primary means of doing this will be to construct a time-independent pseudodifferential renormalization operator $\mathcal{O} = Op(O) \in OPS^0$ which upon commuting $\mathcal{O}$ with the equation achieves this transformation within a compact ball $B_R(0)$. The hope is then to use the asymptotic smallness (5.4) to control the residual error terms outside $B_R(0)$. Quite a bit of care is needed here to avoid a circular argument because the higher order symbol bounds for O will grow in the parameter R, and so, at first glance, the operator bounds for $\mathcal{O}$ could counteract any smallness coming from the remaining error terms. Therefore, we will need to carefully track the dependence of the operator bounds for $\mathcal{O}$ on the parameters R and L. In our construction, it will turn out that the L^∞ norm of the symbol O will have a R independent bound (as $R \to \infty$). Therefore, for large enough k_1, the operator $\mathcal{O} S_{\ge k_1}$ will have R independent $L^2 \to L^2$, $X^0 \to X^0$ and $Y^0 \to Y^0$ bounds thanks to Proposition 2.6 and Proposition 2.7, respectively. This is how we will break the potential circularity.

6.1. **First order truncations.** Since we want the symbol for $\mathcal{O}$ to be time-independent and smooth, our first aim will be to show that the first order paradifferential coefficients in (5.2) can be replaced by smooth time-independent coefficients localized at a suitable frequency scale. To achieve this, let us fix another large parameter k_0 with $0 \ll k_0 \ll k_1$ to be chosen. We can rearrange the paradifferential equation as

$$(6.2) \qquad \begin{cases} i \partial_t v + \partial_j T_{g^{ij}} \partial_i v + b^j_{<k_0}(0) \partial_j v + \tilde{b}^j_{<k_0}(0) \partial_j \overline{v} = f + \mathcal{R}^1, \\ v(0) = v_0, \end{cases}$$

where

$$(6.3) \qquad \mathcal{R}^1 = (b^j_{<k_0}(0)\partial_j v - T_{b^j}\partial_j v) + (\tilde{b}^j_{<k_0}(0)\partial_j \overline{v} - T_{\tilde{b}^j}\partial_j \overline{v}).$$

We have the following short lemma which shows that for large enough k_0, k_1 and small enough T, the error term $\mathcal{R}_1$ can be treated perturbatively.

Lemma 6.3. *For k_0 and k_1 sufficiently large and T sufficiently small, we have*

$$\|\mathcal{R}^1\|_{Y^\sigma} \le \epsilon \|v\|_{X^\sigma}.$$

Proof. We estimate the first term in (6.3) as the other term is essentially identical. By Bernstein's inequality, averaging in T and the assumption (5.3), we have

$$\|b^j_{<k_0} - b^j_{<k_0}(0)\|_{l^1 X^{s_0-1}} \lesssim_M 2^{2k_0}T.$$

Therefore, by the assumption $k_1 \gg k_0$, Proposition 2.13 and taking T small enough (depending on k_0 and M), we have

$$\|(b^j_{<k_0} - b^j_{<k_0}(0))\partial_j v\|_{Y^\sigma} = \|T_{(b^j_{<k_0} - b^j_{<k_0}(0))}\partial_j v\|_{Y^\sigma} \le \epsilon \|v\|_{X^\sigma}.$$

Next, using $k_1 \gg k_0$, we can write

$$T_{b^j}\partial_j v - b^j_{<k_0}\partial_j v = T_{S_{\ge k_0}b^j}\partial_j v.$$

So, from Proposition 2.13, there is $\delta > 0$ depending only on s_0 such that

$$\|T_{b^j}\partial_j v - b^j_{<k_0}\partial_j v\|_{Y^\sigma} \lesssim \|S_{\ge k_0}b^j\|_{l^1 X^{s_0-1-\delta}}\|v\|_{X^\sigma}$$
$$\lesssim_M 2^{-k_0\delta}\|v\|_{X^\sigma}.$$

The above term can be controlled by $\epsilon\|v\|_{X^\sigma}$ by taking k_0 large enough. This completes the proof. $\qquad \square$

6.2. **Commuting with derivatives.** The next step is to commute (5.2) with $\langle\nabla\rangle^\sigma$. This will essentially reduce matters to proving an L^2 estimate for the paradifferential flow and get us one step closer to a situation in which we can apply Lemma 6.2. This would typically be a completely straightforward matter since the equation is already in paradifferential form; however, the commutation of the principal operator $\mathcal{P}$ with $\langle\nabla\rangle^\sigma$ will generate a further first order term which cannot be treated perturbatively in the large data regime.

To proceed, we define $u := \langle\nabla\rangle^\sigma v$. We also compactify the notation for the principal and new first order terms by defining

$$\mathcal{P} := \partial_j T_{g^{ij}}\partial_i,$$
$$\mathcal{B} := b^j_{<k_0}(0)\partial_j - [\mathcal{P}, \langle\nabla\rangle^\sigma]\langle\nabla\rangle^{-\sigma},$$
$$\tilde{\mathcal{B}} := \tilde{b}^j_{<k_0}(0)\partial_j.$$

By commuting (6.2) with $\langle\nabla\rangle^\sigma$, we obtain

$$i\partial_t u + \mathcal{P}u + \mathcal{B}u + \tilde{\mathcal{B}}\overline{u} = \langle\nabla\rangle^\sigma f + \mathcal{R}^1_\sigma + \mathcal{R}^2_\sigma,$$

where $\mathcal{R}^1_\sigma := \langle\nabla\rangle^\sigma \mathcal{R}^1$ and

$$\mathcal{R}^2_\sigma := -[\langle\nabla\rangle^\sigma, b^j_{<k_0}(0)]\partial_j v - [\langle\nabla\rangle^\sigma, \tilde{b}^j_{<k_0}(0)]\partial_j \overline{v}.$$

Thanks to Lemma 6.3, we have a suitable estimate for $\mathcal{R}^1_\sigma$ in Y^0 which allows us to treat this term perturbatively. The following lemma shows that $\mathcal{R}^2_\sigma$ can be estimated naïvely in $L^1_T L^2_x \subset Y^0$.

Lemma 6.4. *For T small enough, the source term $\mathcal{R}_\sigma^2$ satisfies the bound*

(6.4)
$$\|\mathcal{R}_\sigma^2\|_{L_T^1 L_x^2} \leq \epsilon \|v\|_{X^\sigma}.$$

Proof. Since $k_0 \ll k_1$ and $\widehat{v}$ is supported at frequencies $\gtrsim 2^{k_1}$, we can write

$$[\langle\nabla\rangle^\sigma, b_{<k_0}^j(0)]\partial_j v = [\langle\nabla\rangle^\sigma, T_{b_{<k_0}^j(0)}]\partial_j v.$$

Hence, by Proposition 2.10, Sobolev embedding and the regularity assumptions on b^j, we have

$$\|[\langle\nabla\rangle^\sigma, b_{<k_0}^j(0)]\partial_j v\|_{L_T^1 L_x^2} \lesssim_{M,k_0} \|v\|_{L_T^1 H_x^\sigma} \lesssim_M T\|v\|_{L_T^\infty H_x^\sigma}.$$

The other term in $\mathcal{R}_\sigma^2$ can be estimated similarly. Hence, by taking T small enough, we obtain (6.4), as desired. $\square$

Next, we further frequency and time truncate the commutator in the term $\mathcal{B}$. As we will see later, such truncations will ensure that our renormalization operator $\mathcal{O}$ belongs to OPS^0. Note that while we cannot directly truncate the principal operator $\mathcal{P}$ because it is second order, it is reasonable to expect that we can do this (as long as the truncation is sharp enough) for commutators involving $\mathcal{P}$, which are first order. We therefore define time and frequency truncated variants of $\mathcal{P}$, $\mathcal{B}$ and $\tilde{\mathcal{B}}$ (technically, this last term is unchanged) via

$$\mathcal{P}_{k_0}^0 := \partial_j g_{<k_0}^{ij}(0)\partial_i,$$
$$\mathcal{B}_{k_0}^0 := b_{<k_0}^j(0)\partial_j - [\mathcal{P}_{k_0}^0, \langle\nabla\rangle^\sigma]\langle\nabla\rangle^{-\sigma},$$
$$\tilde{\mathcal{B}}_{k_0}^0 := \tilde{b}_{<k_0}^j(0)\partial_j,$$

and obtain the equation

(6.5)
$$i\partial_t u + \mathcal{P}u + \mathcal{B}_{k_0}^0 u + \tilde{\mathcal{B}}_{k_0}^0 \overline{u} = \langle\nabla\rangle^\sigma f + \mathcal{R}_\sigma^1 + \mathcal{R}_\sigma^2 + \mathcal{R}_\sigma^3,$$

where

$$\mathcal{R}_\sigma^3 := (\mathcal{B}_{k_0}^0 - \mathcal{B})u = -[\mathcal{P}_{k_0}^0 - \mathcal{P}, \langle\nabla\rangle^\sigma]v.$$

The next lemma treats the new source term $\mathcal{R}_\sigma^3$.

Lemma 6.5. *For k_0 and k_1 large enough and T small enough, we have*

(6.6)
$$\|\mathcal{R}_\sigma^3\|_{Y^0} \leq \epsilon \|v\|_{X^\sigma}.$$

Proof. We begin by writing

$$\mathcal{P}_{k_0}^0 - \mathcal{P} = (\partial_i g_{<k_0}^{ij}(0) - T_{\partial_i g^{ij}})\partial_j + (g_{<k_0}^{ij}(0) - T_{g^{ij}})\partial_i\partial_j.$$

As with the estimate for $\mathcal{R}_\sigma^2$, we have

$$\|[\langle\nabla\rangle^\sigma, (\partial_i g_{<k_0}^{ij}(0) - T_{\partial_i g^{ij}})]\partial_j v\|_{L_T^1 L_x^2} \leq \epsilon \|v\|_{X^\sigma},$$

by taking T small enough. The term $[\langle\nabla\rangle^\sigma, (T_{g^{ij}} - g_{<k_0}^{ij}(0))]\partial_i\partial_j v$ is more difficult to deal with since it is like an operator of order $\sigma + 1$ applied to v, and therefore cannot be estimated in $L_T^1 L_x^2$ without losing derivatives. Consequently, we must estimate it in the weaker space Y^0. Since $k_1 \gg k_0$, we have the identity

$$[\langle\nabla\rangle^\sigma, (T_{g^{ij}} - g_{<k_0}^{ij}(0))]\partial_i\partial_j v = \sum_{k \geq 0} \tilde{S}_k[\langle\nabla\rangle^\sigma, S_{<k-4}(g^{ij} - g_{<k_0}^{ij}(0))]\partial_i\partial_j S_k v,$$

where $\tilde{S}_k$ is a fattened Littlewood-Paley projection. Therefore, by almost orthogonality, Proposition 2.14 and Remark 2.15 we have

$$\|[\langle\nabla\rangle^\sigma, (T_{g^{ij}} - g^{ij}_{<k_0}(0))]\partial_i\partial_j v\|_{Y^0} \lesssim \|g^{ij} - g^{ij}_{<k_0}(0)\|_{l^1 X^{s_0-\delta}} \left(\sum_{k\geq 0} 2^{2k(\sigma-1)}\|S_k\nabla v\|^2_{X^0}\right)^{\frac{1}{2}}$$

$$\lesssim \|g^{ij} - g^{ij}_{<k_0}(0)\|_{l^1 X^{s_0-\delta}}\|v\|_{X^\sigma}$$

for some $\delta > 0$. By taking k_0 large enough and then T small enough, we can estimate using Bernstein type inequalities and the fundamental theorem of calculus,

$$\|g^{ij} - g^{ij}_{<k_0}(0)\|_{l^1 X^{s_0-\delta}} \leq \epsilon.$$

Combining this with the above estimates concludes the proof of (6.6), as desired. $\qquad\square$

To summarize what we have so far, $u := \langle\nabla\rangle^\sigma v$ solves the equation

$$(6.7) \qquad\qquad i\partial_t u + \mathcal{P}u + \mathcal{B}^0_{k_0}u + \tilde{\mathcal{B}}^0_{k_0}\overline{u} = \mathcal{R},$$

where the source term $\mathcal{R}$ can be estimated in Y^0 by

$$\|\mathcal{R}\|_{Y^0} \leq C\|f\|_{Y^\sigma} + \epsilon\|v\|_{X^\sigma},$$

for some universal constant C.

6.3. **Renormalization construction.** Now we are ready to construct the renormalization operator $\mathcal{O}$ whose role will be to transform (6.7) into an equation essentially of the form (6.1). As alluded to earlier, the main enemy we have to deal with is the first order term $\mathrm{Re}(\mathcal{B}^0_{k_0})u$. The strategy will be to construct an operator with symbol in S^0 which conjugates away the "worst part" of this term. As noted in [15], conjugating the entire term away would give a symbol that does not belong to S^0. We opt therefore to conjugate away only a portion of the first-order term whose principal part is supported within some large compact set $B_R(0)$. The hope is that the remaining error term will contribute errors of size $\approx \epsilon\|v\|_{X^\sigma}$ due to the smallness of the coefficients in (5.4) outside of $B_R(0)$. As mentioned earlier, this does not come for free. The trade-off is that we will also need to control the $X^0 \to X^0$, $Y^0 \to Y^0$, and $L^2 \to L^2$ norms of our renormalization operator to ensure that the smallness is retained when applying this operator (as the ξ derivatives of its symbol will not have uniform in R bounds).

The details of this construction will be given below. To set the stage, let us fix a large constant $R \gg 1$ to be chosen. We also define for each $\rho > 0$, the function $\chi_{<\rho}(x) := \chi(\rho^{-1}x)$ where χ is a radial cutoff function equal to 1 on the unit ball and vanishing outside $|x| > 2$. As a first constraint, we demand for R to be such that (5.4) holds with some $R_0 < \frac{R}{8}$ and $\epsilon_0 \ll \epsilon$. The bulk of the renormalization construction is given by the following proposition.

Proposition 6.6. *Let u be as above. Let k_0 be large enough so that $g^{ij}_{<k_0}(0)$ is a nontrapping metric with comparable parameters to $g^{ij}(0)$ (the existence of which is guaranteed by Proposition 4.5). Define the truncated symbol $a(x,\xi) := -g^{ij}_{<k_0}(0)\xi_i\xi_j$, which is the principal symbol for $\mathcal{P}^0_{k_0}$. Write also $iB(x,\xi) :=$ $i\,\mathrm{Re}(b^j_{<k_0}(0))\xi_j + i\{a, \langle\xi\rangle^\sigma\}\langle\xi\rangle^{-\sigma}$ to denote the principal symbol of $\mathrm{Re}\,\mathcal{B}^0_{k_0}$ and*

$$H_a := \nabla_\xi a \cdot \nabla_x - \nabla_x a \cdot \nabla_\xi$$

to denote the Hamiltonian vector field for a. Let the parameters R, M and L be as above. Then there exists a smooth, non-negative, real-valued, time-independent symbol $O \in S^0$ with the following properties.

(i) *(Positive commutator with good error). There exists $r \in S^1$ such that if T is sufficiently small,*

$$H_a O + \chi_{<2R} B(x,\xi) O(x,\xi) + r(x,\xi) O(x,\xi) \geq 0, \qquad \|Op(r)\|_{X^0 \to Y^0} \lesssim_M \epsilon.$$

(ii) *(Uniform L^2 bound at high frequency). For k_0, k_1 large enough and T small enough depending on R, M and L, $\mathcal{O} := Op(O)$ satisfies the estimates*

(6.8)
$$\|\mathcal{O}u\|_{L^2} \approx \|u\|_{L^2}, \qquad \|\mathcal{O}u\|_{Y^0} \lesssim \|u\|_{Y^0}, \qquad \|\mathcal{O}u\|_{X^0} \lesssim \|u\|_{X^0},$$

with implicit constants depending only on M and on L within a fixed compact set whose size is independent of R.

(iii) *(Even in ξ within $B_{\frac{R}{8}}(0)$). The symbol $s := O(x,\xi) - O(x,-\xi)$ is supported in the region $|x| > \frac{R}{8}$ and for k_1 large enough, there holds*

$$\|Op(s)S_{\geq k_1}\|_{Y^0 \to Y^0} \lesssim 1,$$

with implicit constants depending only on M and on L within a fixed compact set whose size is independent of R.

The first property will allow us to transform (6.7) into an equation of the type (6.1) up to an error term supported outside $B_R(0)$ (plus an acceptable remainder). The second property ensures that the $L^2 \to L^2$, $Y^0 \to Y^0$ and $X^0 \to X^0$ operator bounds for $\mathcal{O}$ do not depend on R, at least at high frequency. The third property ensures that $\mathcal{O} := Op(O)$ commutes with complex conjugation to leading order (i.e. within $B_{\frac{R}{8}}(0)$ where the coefficient $\tilde{b}^j$ can be large). The second and third properties will be important for avoiding the circularity mentioned earlier when trying to estimate the error terms supported outside $B_{\frac{R}{8}}(0)$.

We also emphasize that a is the principal symbol for the truncated operator $\mathcal{P}^0_{k_0}$ and not $\mathcal{P}$. This is to ensure that O will be a classical (time-independent) S^0 symbol with bounds not depending on higher derivatives of g^{ij} (however, they will depend on the frequency truncation scale 2^{k_0}). The trade-off is that when commuting the equation for u with $\mathcal{O}$, we will need to estimate an additional first order error term of the form

$$[\mathcal{P} - \mathcal{P}^0_{k_0}, \mathcal{O}]u$$

in Y^0. It will turn out that this can be made small by taking k_0, k_1 large enough and T small enough. We will discuss how to estimate this term later. For now, we start by proving Proposition 6.6.

Proof. We make the ansatz $O(x,\xi) = e^{\psi(x,\xi)}$ where ψ is some smooth real-valued function to be chosen. We begin by trying to enforce condition (i). For this, we recall that the vector field H_a corresponds to differentiation along the Hamilton flow of a, which is given by (4.3). That is,

$$(H_a \psi)(x,\xi) = \frac{d}{dt}\psi(x^t, \xi^t)_{|t=0},$$

where (x^t, ξ^t) are the bicharacteristics for a with initial data (x,ξ). We will perform our construction in two stages. That is, we will define two symbols ψ_1 and ψ_2 in S^0. The symbol ψ_1 will be chosen so that $H_a \psi_1$ cancels the bulk of the term $\chi_{<2R} B(x,\xi)$ but possibly with an additional error term which isn't small but has the redeeming feature that it is supported in the transition region $|x| \approx R$ where $g^{ij}_{<k_0}(0)$ is close to the corresponding flat metric. The second symbol ψ_2 will be chosen to correct ψ_1 so that the error term can be made sufficiently small. The full symbol ψ will then be defined by $\psi := \psi_1 + \psi_2$. Inspired by the previous works [4, 6, 7, 16], our starting point is to consider the ideal "symbol"

$$\psi_{ideal}(x,\xi) := -\frac{1}{2}\chi_{>1}(|\xi|) \int_{-\infty}^0 B(x^t_{(x,\xi)}, \xi^t_{(x,\xi)}) + B(x^t_{(x,-\xi)}, \xi^t_{(x,-\xi)})dt,$$

where $\chi_{>1}(|\xi|)$ is an increasing Fourier multiplier selecting frequencies ≥ 1. We note that since $g^{ij}_{<k_0}(0)$ is nontrapping and $b^j, \nabla_x g^{ij} \in l^1 X^{s_0-1}$, the integral in ψ_{ideal} is well-defined. On a formal level, the commutator of the principal part of the equation with $Op(e^{\psi_{ideal}})$ conjugates away the leading part of the term $\mathrm{Re}\, \mathcal{B}^0_{k_0} u$, but as mentioned above, the symbol ψ_{ideal} is not a classical S^0 symbol, so it is not ideal to work with such a construction directly. In order to resolve this issue, we localize this symbol to the compact set $B_{2R}(0)$ by instead defining

$$\psi_1(x,\xi) := -\frac{1}{2}\chi_{>1}(|\xi|)\chi_{<2R}(x)\int_{-\infty}^0 (\chi_{<4R}B)(x^t_{(x,\xi)}, \xi^t_{(x,\xi)}) + (\chi_{<4R}B)(x^t_{(x,-\xi)}, \xi^t_{(x,-\xi)})dt.$$

The corresponding pseudodifferential operator $Op(e^{\psi_1})$ will conjugate away the leading part of the first order term $\mathrm{Re}\, \mathcal{B}^0_{k_0} u$ within the ball $B_{2R}(0)$, which is the region where the $X^0 \to Y^0$ operator bounds for $\mathcal{B}^0_{k_0}$ are expected to be large. The difficulty is then shifted to controlling the remaining errors in the exterior region, but now we have the benefit of ψ_1 being a genuine S^0 symbol (this fact will be confirmed below). We remark that since $B(x,\xi)$ is real, ψ_1 is as well. Moreover, ψ_1 is even in ξ.

Since B is odd in ξ, it is straightforward to verify that we have the leading order cancellation,

$$(6.9) \qquad H_a\psi_1 + \chi_{<2R}B(x,\xi) \geq -KR^{-1}|\chi'(\frac{1}{2}R^{-1}r)||\xi| - K\chi_{<2}(|\xi|),$$

where $K > 0$ is such that $K \gg_M \|\psi_{ideal}\|_{L^\infty}$. We remark that K is uniformly bounded in R because of Proposition 4.8. The term on the right-hand side of (6.9) is not quite suitable for defining a symbol r ensuring the bound in (i) (the corresponding operator need not have small $X^0 \to Y^0$ bound due to the insufficient spatial decay in the first term). For this reason, we seek to further correct ψ_1 by a symbol ψ_2 which is supported in the region $|x| \gtrsim R$. Precisely, our aim will be to construct ψ_2 so that

$$(6.10) \qquad H_a\psi_2 - KR^{-1}|\chi'(\frac{1}{2}R^{-1}r)||\xi| - K\chi_{<2}(|\xi|) + r(x,\xi) \geq 0$$

where $r \in S^1$ is a suitable remainder term satisfying the bound in (i). Before proceeding, to simplify the notation somewhat, for the remainder of the proof we will write $A := A(x)$ as a shorthand for $g^{ij}_{<k_0}(0)$ and A_∞ as a shorthand for g^{ij}_∞. We also define the functions $\theta(x,\xi) := \angle(x, A_\infty\xi)$, $\alpha(x,\xi) := \angle(x, A\xi)$, $\beta(\xi) := \angle(A\xi, A_\infty\xi)$ and $\gamma(x,\xi) := \frac{1}{2}(1 + \cos(\theta))$.

Now, to proceed, we begin by recalling that the assumption (5.4) ensures that we have the bounds

$$(6.11) \qquad |A - A_\infty| + |\nabla A| \ll \epsilon, \qquad |x| > \frac{R}{8}.$$

In particular, A is close to the flat metric in L^∞ when $|x| > \frac{R}{8}$. Now, let:

(i) ρ be a smooth, increasing function such that $\rho' \approx 1$ for $\frac{1}{7} \leq r \leq 2$, $\rho = 0$ for $r \leq \frac{1}{8}$ and $\rho = 1$ for $r \geq 3$. Define $\rho_R(x) = \rho(R^{-1}r)$ and $\rho_\theta(x,\xi) = \rho_R(x\gamma)$.

(ii) For $c \in [-1, 1]$ and some fixed positive $\delta_0 \ll 1$, let $\varphi_{<c}$ be a decreasing smooth function which vanishes for $x > c + \delta_0$ and is identically one for $x \leq c$. Define also $\varphi_{>c} := 1 - \varphi_{<c}$.

We then define the symbol ψ_2 by

$$(6.12) \qquad \psi_2(x,\xi) := K'\chi_{>1}(|\xi|)\left(\rho_R\varphi_{<-\frac{1}{2}}(\cos(\theta)) - \rho_\theta\varphi_{>-\frac{1}{2}}(\cos(\theta))\right)$$

where $K' \gg K$ is a constant to be chosen. We note that the weight ρ_R is increasing in the direction of the bicharacteristics in the regions of phase space where they are outgoing with respect to the flat metric. In such regions, this will give a good bound from below for the bulk of $H_a\psi_2$. The purpose of ρ_θ will be to accomplish the same task in the incoming region as well as the regions of phase space where $A_\infty\xi$ is nearly

orthogonal to x. In such regions, a purely radially increasing cutoff (such as ρ_R) would be insufficient. The reason we use the average $\frac{1}{2}(1+\cos(\theta))$ in the definition of ρ_θ is to ensure that ρ_θ still vanishes for a suitable range of r on the support of $\varphi_{>-\frac{1}{2}}(\cos(\theta))$ ($r < \frac{R}{8}$, say). This, in particular, ensures that the pointwise error between A and A_∞ is small on the support of $\rho_\theta \varphi_{>-\frac{1}{2}}(\cos(\theta))$. To verify that ψ_2 has the required properties, we first make note of the following simple algebraic computation.

Lemma 6.7. *For $r > \frac{R}{8}$, we have*

$$A\xi \cdot \nabla_x \cos(\theta) = |A\xi| \left(\frac{\sin^2(\theta)}{r} + \delta(x,\xi) \right),$$

where $\delta(x,\xi)$ is an error term with $|\delta(x,\xi)| \ll \frac{1}{r}$.

Proof. This is a simple computation. We have

$$(6.13) \qquad \begin{aligned} A\xi \cdot \nabla_x \cos(\theta) &= \frac{|A\xi|}{r}(\cos(\beta) - \cos(\alpha)\cos(\theta)) \\ &= \frac{|A\xi|}{r}\sin^2(\theta) + \frac{|A\xi|}{r}((\cos(\beta) - 1) + \cos(\theta)(\cos(\theta) - \cos(\alpha))). \end{aligned}$$

By non-degeneracy of A and A_∞ and (6.11), we have

$$|\cos(\alpha) - \cos(\theta)| + |\cos(\beta) - 1| \ll 1, \quad r \geq \frac{R}{8}.$$

Taking δ to be the coefficient of $|A\xi|$ in the second term in the second line of (6.13) concludes the proof. $\square$

Now, we compute the Hamilton vector field applied to ψ_2. We define the remainder symbol $r \in S^1$ by

$$(6.14) \qquad r(x,\xi) := -\xi_i \xi_j \nabla_\xi \psi_2 \cdot \nabla_x A^{ij} + K'' \chi_{<2}(|\xi|),$$

where $K'' \gg K'$ is some sufficiently large constant. We note that r essentially consists of the part of $H_a \psi_2$ in which ψ_2 is differentiated in ξ. This is expected to contribute a small $X^0 \to Y^0$ operator norm because its principal part includes a factor of $\nabla_x A$ which is small in $l^1 X^{s_0-1}$ when $|x| > \frac{R}{8}$. The subprincipal terms will contribute small $L_T^1 L_x^2 \to L_T^1 L_x^2$ operator bounds by taking T to be sufficiently small. We then have

$$(6.15) \qquad H_a \psi_2 + r(x,\xi) \geq -2A\xi \cdot \nabla_x \psi_2 + K'' \chi_{<2}(|\xi|).$$

We now expand the first term on the right-hand side of (6.15) to obtain

$$\begin{aligned} -A\xi \cdot \nabla_x \psi_2 &= -\frac{K'}{R}\chi_{>1}(|\xi|)|A\xi|\left(\cos(\alpha)\rho'(R^{-1}r)\varphi_{<-\frac{1}{2}}(\cos(\theta)) + \frac{R}{r}\left(\sin^2(\theta) + r\delta\right)\rho(R^{-1}r)\varphi'_{<-\frac{1}{2}}(\cos(\theta)) \right) \\ &\quad + \frac{K'}{R}\chi_{>1}(|\xi|)|A\xi|\left(\frac{1}{2}(\cos(\alpha) + \cos(\beta))\rho'(R^{-1}r\gamma)\varphi_{>-\frac{1}{2}}(\cos(\theta)) \right. \\ &\qquad\qquad \left. + \frac{R}{r}\left(\sin^2(\theta) + r\delta\right)\rho(R^{-1}r\gamma)\varphi'_{>-\frac{1}{2}}(\cos(\theta)) \right), \end{aligned}$$

where α and β are as in Lemma 6.7. If ϵ_0 is small enough in (5.4), we observe that on the support of $\rho'(R^{-1}r)\varphi_{<-\frac{1}{2}}(\cos(\theta))$, we have $\cos(\alpha) < -\frac{1}{3}$. Additionally, $(\sin^2(\theta) + r\delta)$ is non-negative on the support of $\rho(R^{-1}r)\varphi'_{<-\frac{1}{2}}(\cos(\theta))$ and $\rho(R^{-1}r\gamma)\varphi'_{>-\frac{1}{2}}(\cos(\theta))$. Moreover, on the support of $\rho'(R^{-1}r\gamma)\varphi_{>-\frac{1}{2}}(\cos(\theta))$, we have $(\cos(\alpha) + \cos(\beta)) > \frac{1}{3}$.

By non-degeneracy of A, we can choose K' depending only on g so that

$$K'|A\xi| \gg K|\xi|.$$

Combining the above, we can arrange for

$$-A\xi \cdot \nabla_x \psi_2 \geq K R^{-1} |\chi'(\frac{r}{2R})| |\xi| - \frac{K''}{2} \chi_{<2}(|\xi|),$$

where K'' is as in (6.14). We then define the full symbol ψ by

$$\psi := \psi_1 + \psi_2.$$

It is left to verify the properties (i), (ii) and (iii) in Proposition 6.6. The positive commutator bound

$$H_a O + \chi_{<2R} B(x,\xi) O(x,\xi) + r(x,\xi) O(x,\xi) \geq 0$$

follows easily from the chain rule and the above construction if K' is large enough. Next, we verify that $r \in S^1$ and that r has the operator bound

(6.16) $$\|Op(r)\|_{X^0 \to Y^0} \leq \epsilon.$$

The fact that $r \in S^1$ is clear so we turn our attention to (6.16). Using the definition of r, we can write

$$Op(r) = (\chi_{>\frac{R}{8}} \nabla_x A^{ij}) \cdot Op(\xi_i \xi_j \nabla_\xi \psi_2) + K'' \chi_{<2}(|D|).$$

Using the embedding $L_T^1 L_x^2 \subset Y^0$ and that $\xi_i \xi_j \nabla_\xi \psi_2 \in S^1$, we can estimate using simple paradifferential calculus and Proposition 2.5,

$$\|(\chi_{>\frac{R}{8}} \nabla_x A^{ij} - T_{\chi_{>\frac{R}{8}} \nabla_x A^{ij}}) \cdot Op(\xi_i \xi_j \nabla_\xi \psi_2)\|_{L_T^1 L_x^2 \to Y^0} \lesssim_{M,k_0} 1.$$

Therefore, by Hölder's inequality in T, we have for T small enough (depending on M and k_0),

$$\|(\chi_{>\frac{R}{8}} \nabla_x A^{ij} - T_{\chi_{>\frac{R}{8}} \nabla_x A^{ij}}) \cdot Op(\xi_i \xi_j \nabla_\xi \psi_2)\|_{X^0 \to Y^0} \leq \epsilon.$$

Hence, we now only need to show that the $X^0 \to Y^0$ norm for $T_{\chi_{>\frac{R}{8}} \nabla_x A^{ij}} \cdot Op(\xi_i \xi_j \nabla_\xi \psi_2)$ can be made small. For this, let us define $\tilde{r} \in S^0$ by

$$\tilde{r}(x,\xi) = \langle\xi\rangle^{-1} \xi_i \xi_j \nabla_\xi \psi_2.$$

By Proposition 2.5, one can verify that the operator $\langle\nabla\rangle Op(\tilde{r}) - Op(\xi_i \xi_j \nabla_\xi \psi_2)$ is bounded from $L_T^1 L_x^2 \to L_T^1 L_x^2$ with norm depending only on M and k_0. Therefore, by taking T small, we can make the $X^0 \to Y^0$ bound of this operator smaller than ϵ. From Proposition 2.13 and the smallness assumption (5.4), we then have

$$\|T_{\chi_{>\frac{R}{8}} \nabla_x A^{ij}} \cdot Op(\xi_i \xi_j \nabla_\xi \psi_2)\|_{X^0 \to Y^0} \lesssim \|\chi_{>\frac{R}{8}} \nabla_x A^{ij}\|_{l^1 X^{s_0-1}} \|\langle\nabla\rangle\|_{X^0 \to X^{-1}} \|Op(\tilde{r})\|_{X^0 \to X^0} + \epsilon$$

$$\lesssim \epsilon \|Op(\tilde{r})\|_{X^0 \to X^0} + \epsilon$$

$$\lesssim_M \epsilon.$$

Clearly, the $X^0 \to Y^0$ bound for the remaining subprincipal term $K'' \chi_{<2}(|D|)$ can be made small by taking T small. This concludes the proof of (i). Now, we turn to (ii) and (iii). For this, we need the following lemma involving symbol bounds for O.

Lemma 6.8. *The symbol O constructed above satisfies the following bounds.*

(i) *(R independent L^∞ bound). There is a constant C_0 depending only on the profile of $g(0)$ and on M but not on R or k_0 such that*

$$\|O(x,\xi)\|_{L_{x,\xi}^\infty} \lesssim C_0.$$

(ii) *(Higher order symbol bounds). For every $|\alpha + \beta| \geq 2$, there is a constant $C_{\alpha,\beta}$ depending on M, $L(R)$, R and k_0 such that*

$$\|\langle\xi\rangle^{|\alpha|}\partial_\xi^\alpha\partial_x^\beta O(x,\xi)\|_{L^\infty_{x,\xi}} \leq C_{\alpha,\beta}.$$

If $|\alpha + \beta| = 1$, the constant can be taken to be uniform in k_0.

The crucial thing to note here is that only the higher order symbol bounds for O depend on R and k_0 while the L^∞ bound does not.

Proof. Clearly, it suffices to show each of the above symbol bounds for ψ_1 and ψ_2. Given the requisite bounds for ψ_1, the bounds for ψ_2 are clear. Therefore, we focus on ψ_1. We begin with the L^∞ bound. By homogeneity of the bicharacteristic flow, it further suffices to show that

$$(6.17) \qquad \int_\mathbb{R} |B(x^t, \xi^t)|dt \lesssim_M C_0.$$

By homogeneity and a change of variables, we have

$$\int_\mathbb{R} |B(x^t, \xi^t)|dt = \int_\mathbb{R} |\xi|^{-1}|B(x^t_\omega, |\xi|\xi^t_\omega)|dt,$$

where $(x^t_\omega, \xi^t_\omega)$ denote the bicharacteristics with data $(x,\omega) := (x, \xi|\xi|^{-1})$. Then, we use Corollary 4.6 and the definition of the symbol B to obtain

$$|B(x^t_\omega, |\xi|\xi^t_\omega)| \leq C_0|\xi||(b^j_{<k_0}(0))(x^t_\omega)| + C_0|\xi||(\nabla_x A)(x^t_\omega)|.$$

The estimate (6.17) then follows (after possibly relabelling C_0) from Proposition 4.8, using the fact that $b^j_{<k_0}(0), \nabla_x A \in l^1 H^{s_0-1}$ with norm $\lesssim_M 1$. This yields the L^∞ bound for ψ_1. The higher order symbol bounds follow immediately from Proposition 4.7 and repeated applications of the chain rule. $\qquad\square$

Now, we return to the proof of (6.8). From the above lemma and Proposition 2.6, we have the L^2 bound,

$$(6.18) \qquad \|\mathcal{O}u\|_{L^2} \lesssim C_0\|u\|_{L^2}$$

for k_1 large enough, with universal implicit constant. We next aim to establish the bound

$$(6.19) \qquad \|u\|_{L^2} \lesssim C_0\|\mathcal{O}u\|_{L^2}.$$

Using Proposition 2.5, we see that $Op(e^{-\psi})$ is an approximate inverse for $Op(e^\psi)$ in the sense that we have

$$Op(e^{-\psi})Op(e^\psi) = 1 + Op(q),$$

where $q \in S^{-1}$ with symbol bounds depending only on the symbol bounds for ψ. Therefore, we have

$$u = S_{\geq k_1-4}u = Op(e^{-\psi})S_{\geq k_1-4}\mathcal{O}u + Op(\tilde{q})u,$$

where $\tilde{q} \in S^{-1}$ with uniform in k_1 symbol bounds. Hence, from Proposition 2.1 we obtain

$$\|u\|_{L^2_x} \lesssim C_0\|\mathcal{O}u\|_{L^2_x} + C_1\|u\|_{H^{-1}_x},$$

where C_0 depends only on M and $g(0)$ and C_1 depends on a finite collection of semi-norms $|O|^{(j)}_{S^0}$. Since $\|u\|_{H^{-1}_x} \lesssim 2^{-k_1}\|u\|_{L^2_x}$, we can take k_1 large enough so that

$$\|u\|_{L^2_x} \lesssim C_0\|\mathcal{O}u\|_{L^2}.$$

This gives (6.19). The $Y^0 \to Y^0$ and $X^0 \to X^0$ bounds for $\mathcal{O}$ follow from Proposition 2.7. This establishes property (ii) of Proposition 6.6. The proof of property (iii) follows almost identical reasoning to the proof of (ii), using the fact that ψ is even in ξ for $|x| < \frac{R}{8}$. This completes the proof of Proposition 6.6. $\qquad\square$

6.4. **Proof of Proposition 6.1.** Now, we complete the proof of Proposition 6.1. We will slightly abuse notation from here on and write $\lesssim_M$ to mean that the implicit constant in the corresponding estimate depends on M and C_0 as above (but not on R). Moreover, we let $\mathcal{R}$ generically denote an error term such that

(6.20) $$\|\mathcal{R}\|_{Y^0} \leq C(M, L)(\|f\|_{Y^\sigma} + \|v_0\|_{H^\sigma}) + \epsilon\|v\|_{X^\sigma}.$$

We apply $\mathcal{O} := Op(O)$ from Proposition 6.6 to equation (6.5). Writing $w := \mathcal{O}u$, we obtain

$$i\partial_t w + \mathcal{P}w + \mathcal{O}\mathcal{B}^0_{k_0}u + [\mathcal{O}, \mathcal{P}]u + \mathcal{O}\tilde{\mathcal{B}}^0_{k_0}\overline{u} = \mathcal{R},$$

where by the $Y^0 \to Y^0$ bound for $\mathcal{O}$ (see (ii) in Proposition 6.6), $\mathcal{R}$ still satisfies the estimate (6.20) as long as k_1 is large enough. Performing similar frequency and time truncations as before and commuting $\mathcal{O}$ with the first order terms, we obtain

$$i\partial_t w + \mathcal{P}w + i\operatorname{Im}(\mathcal{B}^0_{k_0})w + \tilde{\mathcal{B}}^0_{k_0}\overline{w} + [\mathcal{O}, \mathcal{P}^0_{k_0}]u + \chi_{<2R}\operatorname{Re}(\mathcal{B}^0_{k_0})\mathcal{O}u = \tilde{\mathcal{R}},$$

where

(6.21) $$\tilde{\mathcal{R}} = \mathcal{R} - [\mathcal{O}, \tilde{\mathcal{B}}^0_{k_0}]\overline{u} - [\mathcal{O}, \mathcal{B}^0_{k_0}]u - \chi_{\geq 2R}\operatorname{Re}(\mathcal{B}^0_{k_0})\mathcal{O}u + \tilde{\mathcal{B}}^0_{k_0}(\overline{\mathcal{O}u} - \mathcal{O}\overline{u}) + [\mathcal{O}, (\mathcal{P}^0_{k_0} - \mathcal{P})]u.$$

We next estimate $\tilde{\mathcal{R}}$. To begin, note that the second and third terms in (6.21) are zeroth order and can be estimated in $L^1_T L^2_x \subset Y^0$, so that

$$\|[\mathcal{O}, \tilde{\mathcal{B}}^0_{k_0}]\overline{u} + [\mathcal{O}, \mathcal{B}^0_{k_0}]u\|_{Y^0} \lesssim_{M,L} T\|v\|_{X^\sigma},$$

which by taking T small can be controlled by $\epsilon\|v\|_{X^\sigma}$. To get a suitable error estimate for the fourth term in (6.21), we first note that by property (ii) in Proposition 6.6, we have

$$\|\mathcal{O}u\|_{X^0} \lesssim C_0\|u\|_{X^0}$$

if k_1 is large enough. Here, we recall crucially that C_0 is a R independent constant. Therefore, it suffices to establish the bound

(6.22) $$\|\chi_{\geq 2R}\operatorname{Re}(\mathcal{B}^0_{k_0})\|_{X^0 \to Y^0} \leq \epsilon.$$

Clearly, it suffices to work with the principal part of $\chi_{\geq 2R}\operatorname{Re}(\mathcal{B}^0_{k_0})$ as the error term is bounded from $L^1_T L^2_x \to L^1_T L^2_x$. We can expand the principal part as

$$\chi_{\geq 2R}\operatorname{Re}(b_{<k_0}(0))m_1(D) + \chi_{\geq 2R}\nabla_x A m_2(D)$$

where $m_1, m_2 \in S^1$ are suitable (matrix-valued) Fourier multipliers with symbol bounds independent of M, L and R. We can replace the coefficients of m_1 and m_2 above with the paradifferential operators $T_{\chi_{\geq 2R}\operatorname{Re}(b_{<k_0}(0))}$ and $T_{\chi_{\geq 2R}\nabla A}$, as the error is an operator which maps $L^1_T L^2_x$ to $L^1_T L^2_x$ with norm depending only on M and k_0. Therefore, if T is small enough, such an error term can be discarded. Using Proposition 2.13 and the asymptotic smallness (5.4), the remaining term satisfies

$$\|T_{\chi_{\geq 2R}\operatorname{Re}(b_{<k_0}(0))}m_1(D) + T_{\chi_{\geq 2R}\nabla A}m_2(D)\|_{X^0 \to Y^0} \lesssim_M \epsilon(\|m_1(D)\|_{X^0 \to X^{-1}} + \|m_2(D)\|_{X^0 \to X^{-1}}) \lesssim \epsilon.$$

To deal with the fifth term in (6.21), we do a similar analysis. Using the definition of $\tilde{\mathcal{B}}^0_{k_0}$ and property (iii) in Proposition 6.6, we can write

$$\tilde{\mathcal{B}}^0_{k_0}(\overline{\mathcal{O}u} - \mathcal{O}\overline{u}) = \chi_{>\frac{R}{8}}\tilde{b}^j_{<k_0}(0)\partial_j(\overline{\mathcal{O}u} - \mathcal{O}\overline{u}).$$

If $\widehat{u}$ is supported at high enough frequency, we can estimate using (ii) in Proposition 6.6,

$$\|\partial_j(\overline{\mathcal{O}u} - \mathcal{O}\overline{u})\|_{X^{-1}} \lesssim_M \|u\|_{X^0}.$$

Combining this with the smallness

$$\|\chi_{> \frac{R}{8}} \tilde{b}_{<k_0}(0)\|_{l^1 X^{s_0-1}} \leq \epsilon,$$

we can argue as with the previous term to obtain

$$\|\tilde{B}_{k_0}^0(\overline{\mathcal{O}u} - \mathcal{O}\overline{u})\|_{Y^0} \lesssim_M \epsilon\|v\|_{X^\sigma}.$$

Now, we turn to the most tricky part, which is estimating the last term in (6.21). For this, we have the following lemma.

Lemma 6.9 (Commutator bound). *For k_0 large enough and T sufficiently small, there holds*

$$(6.23) \qquad \|[\mathcal{O}, (\mathcal{P}_{k_0}^0 - \mathcal{P})]u\|_{Y^0} \leq \epsilon\|v\|_{X^\sigma}.$$

Proof. Clearly, we can write

$$(6.24) \quad \begin{aligned} [\mathcal{O}, (\mathcal{P} - \mathcal{P}_{k_0}^0)] &= [\mathcal{O}, (T_{\partial_i g^{ij}} - \partial_i g_{<k_0}^{ij}(0))\partial_j] + (T_{g^{ij}} - g_{<k_0}^{ij}(0))[\mathcal{O}, \partial_i \partial_j] \\ &\quad + [\mathcal{O}, (T_{g^{ij}} - g_{<k_0}^{ij}(0))]\partial_i \partial_j. \end{aligned}$$

The first term is zeroth order and is bounded from $L_T^1 L_x^2 \to L_T^1 L_x^2$. Indeed, by taking T small enough, it is a straightforward consequence of Proposition 2.5 and Proposition 2.10 that

$$\|[\mathcal{O}, (T_{\partial_i g^{ij}} - \partial_i g_{<k_0}^{ij}(0))\partial_j]\|_{L_T^\infty L_x^2 \to Y^0} \leq \epsilon.$$

The second and third terms are first order, and, as usual, must be dealt with carefully to extract the necessary smallness. We start with the second term which is a bit easier. Since $(T_{g_{<k_0}^{ij}(0)} - g_{<k_0}^{ij}(0))[\mathcal{O}, \partial_i \partial_j]$ is bounded from $L_T^1 L_x^2 \to L_T^1 L_x^2$, we can replace $T_{g^{ij}} - g_{<k_0}^{ij}(0)$ with $T_{g^{ij} - g_{<k_0}^{ij}(0)}$. Then by Proposition 2.13 and Bernstein inequalities, we have

$$\|T_{g^{ij} - g_{<k_0}^{ij}(0)}[\mathcal{O}, \partial_i \partial_j]u\|_{Y^0} \lesssim (\|S_{\geq k_0} g^{ij}\|_{l^1 X^{s_0-1-\delta}} + \|g_{<k_0}^{ij} - g_{<k_0}^{ij}(0)\|_{l^1 X^{s_0-1-\delta}})\|\langle\nabla\rangle^{-1}[\mathcal{O}, \partial_i \partial_j]u\|_{X^0}$$

$$\lesssim_M 2^{-(1+\delta)k_0}\|\langle\nabla\rangle^{-1}[\mathcal{O}, \partial_i \partial_j]u\|_{X^0},$$

for some $\delta > 0$. As $\langle\nabla\rangle^{-1}[\mathcal{O}, \partial_i \partial_j] \in OPS^0$, it suffices to consider its principal part when estimating the last term. This is because the subprincipal part is (crudely) bounded from $L_T^\infty H_x^{-\frac{1}{2}}$ to $L_T^\infty H_x^{\frac{1}{2}} \subset X^0$, so we can control such terms by using the fact that u is localized to frequencies $\gtrsim 2^{k_1}$ to gain a smallness factor $2^{-\frac{k_1}{2}}$, and then take k_1 sufficiently large. To estimate the principal symbol c_p for $\langle\nabla\rangle^{-1}[\mathcal{O}, \partial_i \partial_j]$, we can use that $\nabla g^{ij}(0), b^j \in C^{1,\delta}$ and Proposition 4.7 to obtain the bound

$$\|c_p\|_{L^\infty} \lesssim \|\nabla_x \mathcal{O}\|_{L^\infty} \lesssim_{M,R,L} 1,$$

with implicit constant independent of k_0. Therefore, by taking k_0 and k_1 large enough and applying Proposition 2.7, we obtain

$$2^{-(1+\delta)k_0}\|Op(c_p)u\|_{X^0} \lesssim 2^{-(1+\delta)k_0}\|c_p\|_{L^\infty}\|u\|_{X^0} \leq \epsilon\|v\|_{X^\sigma}.$$

Consequently, the second term in (6.24) can be estimated by

$$\|(T_{g^{ij}} - g_{<k_0}^{ij}(0))[\mathcal{O}, \partial_i \partial_j]u\|_{Y^0} \leq \epsilon\|v\|_{X^\sigma}.$$

It remains to estimate the third term in (6.24) which is the most delicate because the commutator itself involves the metric g^{ij} at high frequencies. Our aim as above is to show that

$$\|[\mathcal{O},(T_{g^{ij}}-g^{ij}_{<k_0}(0))]\partial_i\partial_j u\|_{Y^0} \leq \epsilon\|v\|_{X^\sigma}.$$

Intuitively, this should be possible by taking k_0 large enough. There are, however, two complications in dealing with this. Firstly, the symbol bounds for O depend on k_0. Secondly, the coefficient in the paradifferential operator $T_{g^{ij}}$ has limited regularity, so the standard pseudodifferential calculus cannot be directly applied. Our strategy is to split this term into three parts to separate the issues. We write

(6.25)
$$\begin{aligned}
[\mathcal{O},(T_{g^{ij}}-g^{ij}_{<k_0}(0))]\partial_i\partial_j u &= [\mathcal{O},(g^{ij}_{<m}-g^{ij}_{<k_0})]\partial_i\partial_j u\\
&+ [\mathcal{O},(T_{g^{ij}}-g^{ij}_{<m})]\partial_i\partial_j u\\
&+ [\mathcal{O},(g^{ij}_{<k_0}-g^{ij}_{<k_0}(0))]\partial_i\partial_j u,
\end{aligned}$$

where m is some universal parameter with $k_0 \ll m \ll k_1$. For the first term, we do not need to worry about the presence of any functions of limited regularity, but we still need to worry about the dependence of O on k_0. For the second term, by taking m large enough, the k_0 dependence in O should be a non-issue, which puts us in a position to use Proposition 2.16. Control of the final term follows by taking $T \ll 2^{-2k_0}$ and averaging in T.

Let us begin by analyzing the first term. The principal symbol c_p for $[\mathcal{O},(g^{ij}_{<m}-g^{ij}_{<k_0})]\partial_i$ is given by

$$c_p = \{O,(g^{ij}_{<m}-g^{ij}_{<k_0})\}\xi_i.$$

Analogously to the principal part for the second term in (6.24), we have the bound

$$\|\langle\xi\rangle\nabla_\xi O\|_{L^\infty} \lesssim_{M,R,L} 1.$$

To estimate the full commutator, we then use Proposition 2.5 to write

$$[\mathcal{O},(g^{ij}_{<m}-g^{ij}_{<k_0})]\partial_i = \nabla_x(g^{ij}_{<m}-g^{ij}_{<k_0})\cdot Op(\nabla_\xi O\xi_i) + Op(r)$$

where $r \in S^0$ (with symbol bounds depending on m). Arguing as in the estimate for the second term in (6.24), it follows by using Proposition 2.13, then Proposition 2.7, then taking k_0 large enough and T small enough (depending on m, M, R and L) that

$$\|[\mathcal{O},(g^{ij}_{<m}-g^{ij}_{<k_0})]\partial_i\partial_j u\|_{Y^0} \leq \epsilon\|v\|_{X^\sigma}.$$

This takes care of the first term in (6.25). For the second term, it suffices to estimate $[\mathcal{O},T_{g^{ij}-g^{ij}_{<m}}]\partial_i\partial_j u$, as the error will be bounded from $L^1_T L^2_x \to L^1_T L^2_x$. To estimate this term, we simply use Proposition 2.16 to obtain

$$\|[\mathcal{O},T_{g^{ij}-g^{ij}_{<m}}]\partial_i\partial_j u\|_{Y^0} \lesssim_{M,L,R,k_0} \|g^{ij}_{<m}-g^{ij}\|_{l^1 X^{s_0-\delta}}\|u\|_{X^0}.$$

We then recall the smallness bound

$$\|g^{ij}_{<m}-g^{ij}\|_{l^1 X^{s_0-\delta}} \lesssim_M 2^{-\delta m},$$

which tells us that if m is large enough relative to k_0, R, L and M then we have the estimate

$$\|[\mathcal{O},T_{g^{ij}-g^{ij}_{<m}}]\partial_i\partial_j u\|_{Y^0} \leq \epsilon\|u\|_{X^0} \lesssim \epsilon\|v\|_{X^\sigma}.$$

Finally, by averaging in T and arguing similarly to the above, the last term in (6.25) can be controlled by $\epsilon\|v\|_{X^\sigma}$ by taking T small enough. This completes the proof of Lemma 6.9. $\qquad\square$

Using the above lemma and Proposition 6.6, we now arrive at the following equation for w:

$$i\partial_t w + \mathcal{P}w + i\,\text{Im}(\mathcal{B}^0_{k_0})w + \tilde{\mathcal{B}}^0_{k_0}\overline{w} + [\mathcal{O}, \mathcal{P}^0_{k_0}]u + \chi_{<2R}\,\text{Re}(\mathcal{B}^0_{k_0})\mathcal{O}u = \mathcal{R},$$

where $\mathcal{R}$ is as in (6.20). To conclude, we make one final reduction. From Proposition 2.5, $\mathcal{O}^{-1} := Op(e^{-\psi})$ is an approximate inverse for $\mathcal{O}$ in the sense that we have $Op(\mathcal{O})Op(\mathcal{O}^{-1}) = 1 + Op(q)$ for $q \in S^{-1}$. Therefore, by estimating the error term generated by $Op(q)$ in $L^1_T L^2_x$, we can write

$$i\partial_t w + \mathcal{P}w + i\,\text{Im}(\mathcal{B}^0_{k_0})w + \tilde{\mathcal{B}}^0_{k_0}\overline{w} + iA(x,D)\mathcal{O}^{-1}w = \mathcal{R},$$

where

$$A(x,D) := -i[\mathcal{O}, \mathcal{P}^0_{k_0}] - i\chi_{<2R}\,\text{Re}(\mathcal{B}^0_{k_0})\mathcal{O} + Op(r)\mathcal{O},$$

and $\mathcal{R}$ is again of the form (6.21) (as long as T is small enough). By construction, $A(x,D)\mathcal{O}^{-1}$ is a time-independent pseudodifferential operator of order 1 with non-negative principal symbol in S^1. Therefore, the above equation for w is now in the form (6.1) with a source term $\mathcal{R}$ satisfying (6.20). Hence, Proposition 6.1 easily follows by applying Lemma 6.2.

7. The local energy decay estimate

In this section, we complement the L^2 estimate in the previous section with an estimate for the local energy component of the norm $\|v\|_{X^\sigma}$ for a solution v to (5.2). For every $\sigma \geq 0$, we denote the local energy component of X^σ by

$$\|v\|_{\mathcal{X}^\sigma} = \left(\sum_{j\geq 0} 2^{2j(\sigma+\frac{1}{2})}\|S_j u\|^2_X\right)^{\frac{1}{2}}.$$

We remark that we have the obvious embedding $\|v\|_{\mathcal{X}^\sigma} \lesssim \|v\|_{L^2_T H^{\sigma+\frac{1}{2}}_x}$.

7.1. **The local energy estimate.** The local energy estimate we will need for (5.2) is given by the following proposition.

Proposition 7.1. *Let $\sigma \geq 0$ and let s_0, g^{ij}, b^j and $\tilde{b}^j$ be as in Theorem 5.1 with parameters M and L. Suppose that v solves (5.2) and let $\epsilon > 0$. There is $T_0 = T_0(\epsilon) > 0$ such that for $0 \leq T \leq T_0$, we have the local energy bound*

$$\|v\|_{\mathcal{X}^\sigma} \leq C(M,L)(\|v\|_{L^\infty_T H^\sigma_x} + \|f\|_{Y^\sigma}) + \epsilon\|v\|_{X^\sigma},$$

where $C(M,L)$ depends on M and on the parameter L within some fixed compact set depending on ϵ.

Fix $\delta > 0$ to be some small parameter to be chosen. From (6.7) in the previous section, we can choose k_0 sufficiently large and T sufficiently small so that $u := \langle\nabla\rangle^\sigma v$ solves the equation

$$i\partial_t u + \mathcal{P}u + \mathcal{B}^0_{k_0}u + \tilde{\mathcal{B}}^0_{k_0}\overline{u} = \mathcal{R},$$

with the remainder estimate

$$\|\mathcal{R}\|_{Y^0} \lesssim \|f\|_{Y^\sigma} + \delta\|v\|_{X^\sigma}.$$

Also, as in the previous section, we may assume that u is localized to frequencies $\gtrsim 2^{k_1}$, where k_1 is some sufficiently large parameter to be chosen. Unlike with the L^2 estimate, however, we will not need the added energy structure coming from the complex-conjugate first order term. It is therefore convenient to write the equation as a system in u and $\overline{u}$. In doing this, we obtain the following compact form of the paradifferential linear equation:

$$\partial_t \mathbf{u} + \mathbf{P}\mathbf{u} + \mathbf{B}^0_{k_0}\mathbf{u} = \mathbf{R},$$

where

$$\mathbf{P} := i \begin{pmatrix} -\mathcal{P} & 0 \\ 0 & \mathcal{P} \end{pmatrix}, \quad \mathbf{B}^0_{k_0} := i \begin{pmatrix} -\mathcal{B}^0_{k_0} & -\tilde{\mathcal{B}}^0_{k_0} \\ \tilde{\mathcal{B}}^0_{k_0} & \mathcal{B}^0_{k_0} \end{pmatrix}, \quad \mathbf{u} := \begin{pmatrix} u \\ \overline{u} \end{pmatrix},$$

and $\mathbf{R}$ is a source term satisfying the bound

$$(7.1) \qquad \qquad \|\mathbf{R}\|_{Y^0} \lesssim \|f\|_{Y^\sigma} + \delta\|v\|_{X^\sigma}.$$

We define analogously to before the truncated principal operator $\mathbf{P}^0_{k_0}$ by replacing the nonzero entries in $\mathbf{P}$ with $\mathcal{P}^0_{k_0}$ in the natural way.

By using Theorem 5.5 and arguing similarly to the proof of Lemma 5.7, for each $R > 0$ large enough and T small enough, there holds

$$\|\chi_{\geq R}\mathbf{u}\|_{\mathcal{X}^0} \leq C(M,R)(\|v\|_{L^\infty_T H^\sigma_x} + \|\mathbf{R}\|_{Y^0} + \|\chi_{<2R}\mathbf{u}\|_{L^2_T H^{\frac{1}{2}}_x}) + \epsilon\|v\|_{X^\sigma}.$$

Therefore, to prove Proposition 7.2, it suffices to establish the bound

$$(7.2) \qquad \qquad \|\chi_{<2R}\mathbf{u}\|_{L^2_T H^{\frac{1}{2}}_x} \leq C(M,L)(\|v\|_{L^\infty_T H^\sigma_x} + \|\mathbf{R}\|_{Y^0}) + \epsilon\|v\|_{X^\sigma}.$$

This latter estimate is where we will concentrate the bulk of our efforts in this section.

7.2. **Interior estimate.** Now we turn to establishing the required interior estimate (7.2). The main construction we will need is given by the following result, which can very loosely be thought of as a spatially truncated version of Doi's construction in [6]. Our method will work under far less stringent decay assumptions, however. For similar reasons to the previous section, we will again work with the principal symbol for the truncated operator $\mathbf{P}^0_{k_0}$ in our analysis rather than $\mathbf{P}$ directly (at the cost of estimating a term with the same flavor as (6.23)). We will also write $|\mathbf{B}^0_{k_0}|$ to denote the maximum of the absolute values of the entries of the principal symbol for $\mathbf{B}^0_{k_0}$.

Proposition 7.2. *Let $C(M) > 1$ be a constant depending on M to be chosen. Moreover, let k_0 be large enough so that $g^{ij}_{<k_0}(0)$ is nontrapping with comparable parameters to $g^{ij}(0)$ (which is possible by Proposition 4.5). Define $a := -g^{ij}_{<k_0}(0)\xi_i\xi_j$. Then for every $R' \gg R$ sufficiently large, there is a smooth, nonnegative, time-independent S^0 symbol $q \geq 1$ with the following properties:*

(i) *(Positive commutator in $B_{R'}(0)$ with small error). There exists $r \in S^1$ such that if R' and k_1 are large enough and T is sufficiently small relative to R' and k_1, then we have*

$$H_a q + C(M)rq \gtrsim C(M)\chi_{<R'}|\mathbf{B}^0_{k_0}|q, \quad \|Op(rq)S_{\geq k_1}\|_{X^0 \to Y^0} \lesssim \frac{\epsilon}{C(M)}.$$

(ii) *(Ellipticity in $B_{2R}(0)$).*

$$H_a q + C(M)rq \geq C(M)\chi_{<2R}|\xi|q,$$

where r is as in (i).

(iii) *(Zeroth order symbol bound). There is a constant $C_0(M,R)$ depending on M and R but not on R' such that*

$$|q| \leq C_0.$$

(iv) *(First order symbol bound). There is a constant $C_1(M,R,R')$ depending on M, R and R' such that*

$$|\xi||\nabla_\xi q| + |\nabla_x q| \leq C_1.$$

(v) *(Higher order symbol bounds). There is a constant $C_2(M, R, R', k_0)$ depending on M, R, R' and k_0 such that*

$$\langle \xi \rangle^{|\alpha|} |\partial_\xi^\alpha \partial_x^\beta q| \lesssim_{\alpha,\beta} C_2, \qquad |\alpha + \beta| \geq 2.$$

In the R and R' dependent constants above, we also allow for dependence on L within $B_R(0)$ and $B_{R'}(0)$, respectively. The first property will allow us to control the contribution of the first-order terms in the equation within the larger compact set $B_{R'}(0)$, up to a small error term, as long as $\mathbf{u}$ is localized at high enough frequency. The second property will give us the required control of $\chi_{<2R}\mathbf{u}$ in $L_T^2 H_x^{\frac{1}{2}}$ up to a suitable error term. We importantly remark that the zeroth order symbol bounds in (iii) for q depend only on M and R (more precisely, $L(R)$). This will ensure that the $Y^0 \to Y^0$ bound for $Op(q)$ depends only on M and R as long as $\mathbf{u}$ is at sufficiently high frequency, thanks to Proposition 2.7. As a consequence, we may argue similarly to the previous section and treat the first-order terms in the region outside of $B_{R'}(0)$ perturbatively as long as R' is large enough relative to R. We note that unlike in the construction in Proposition 6.6, this second parameter R' is needed because the uniform norm of the symbol q necessarily depends on the smaller radius R.

We emphasize that the first order symbol bounds in (iv) depend on M, R and R' but not on k_0. The purpose of this will be to control an error term that is similar to the commutator (6.23) from the previous section by taking k_0 large relative to M, R, and R'. The higher order symbol bounds in property (v) will ensure that q is a classical S^0 symbol and will allow us to estimate lower order error terms in $L_T^1 L_x^2$ by taking T small depending on M, R, R' and k_0, similarly to the previous section.

Proof. We begin by defining a smooth function that will be suitable for controlling the size of the first order coefficients within the larger compact set $B_{R'}(0)$. A reasonable choice is the following:

$$\eta_{R'} = \chi_{<2R'} \sqrt{|\tilde{b}_{<k_0}(0)|^2 + |b_{<k_0}(0)|^2 + |\nabla_x g_{<k_0}^{ij}(0)|^2 + L(2R')^{-2}}.$$

The term $L(2R')^{-2}$ is for technical convenience. It ensures that $\eta_{R'}$ is smooth and allows us to invoke Proposition 4.8 to obtain uniform integrability along the bicharacteristic flow for the truncated metric $g_{<k_0}^{ij}(0)$ with a bound independent of R'. Precisely, we have

$$(7.3) \qquad \int_{\mathbb{R}} \eta_{R'}(x^t, \xi^t) |\xi^t| dt \leq C_0$$

where C_0 is as above. Moreover, for $|x| \leq R'$, we clearly have $|\nabla_x g_{<k_0}^{ij}(0)||\xi| + |\mathbf{B}_{k_0}^0| \lesssim \eta_{R'}|\xi|$. Now, we move to constructing the symbol q. We start by defining a preliminary symbol p_1 via

$$p_1(x, \xi) := -\chi_{>1}(|\xi|)\chi_{<R'} \int_0^\infty (\chi_{<2R} + \eta_{R'})(x^t, \xi^t)|\xi^t| dt,$$

where similarly to the construction for O in Proposition 6.6, we localized the symbol in space to $B_{R'}(0)$ so that it will ultimately belong to S^0. As in Proposition 6.6, $H_a p_1$ will generate an error term coming from the localization $\chi_{<R'}$. To deal with this, we correct p_1 by another symbol p_2. To define p_2, we take our cue from the definition (6.12) in the previous section. Using the same notation as in (6.12) with the parameter R' replacing R in all instances, we define

$$p_2(x, \xi) := K' \chi_{>1}(|\xi|) \left(\rho_{R'} \varphi_{<-\frac{1}{2}}(\cos(\theta)) - \rho_\theta \varphi_{>-\frac{1}{2}}(\cos(\theta)) \right),$$

where $K' := K'(R, M)$ is a constant such that

$$(7.4) \qquad K' \gg \sup_{(x,\xi) \in \mathbb{R}^{2d}} \chi_{>1}(|\xi|) \int_0^\infty (\chi_{<2R} + \eta_{R'})(x^t, \xi^t)|\xi^t| dt.$$

We note that thanks to the nontrapping assumption and (7.3), K' can be chosen to depend only on R and M, but not on R'. We then define $p := p_1 + p_2$ and analogously to (6.14), we define the remainder symbol r by

$$r(x,\xi) := -\xi_i \xi_j \nabla_\xi p_2 \cdot \nabla_x g^{ij}_{<k_0}(0) + K'' \chi_{<2}(|\xi|),$$

where $K'' \gg K'$ is some sufficiently large constant. We then define the required symbol q by

$$q := e^{C(M)p}, \tag{7.5}$$

for some sufficiently large constant $C(M) > 0$. Now, we turn to establishing each property in Proposition 7.2. First, arguing similarly to the proof of the first property in Proposition 6.6, we compute directly that

$$H_a q + C(M) r q \geq C(M)(\chi_{<2R} + \chi_{<R'} \eta_{R'})|\xi|q. \tag{7.6}$$

From this, we immediately obtain the positive commutator bounds in (i) and (ii) in Proposition 7.2. The $X^0 \to Y^0$ estimate for $Op(rq)S_{\geq k_1}$ follows from properties (iii)-(v) (to be established below), Proposition 2.7 and the fact that $\|\chi_{>R'} \nabla_x g^{ij}_{<k_0}(0)\|_{L^\infty} \to 0$ as $R' \to \infty$. Next, we verify the symbol bounds (iii)-(v). It clearly suffices to establish the analogous symbol bounds for the symbol p_1. To do this, we split

$$p_1 := -\chi_{>1}(|\xi|)\chi_{<R'} \left(\int_0^\infty \eta_{R'}(x^t, \xi^t)|\xi^t|dt + \int_0^\infty \chi_{<2R}(x^t,\xi^t)|\xi^t|dt \right) =: \chi_{>1}(|\xi|)\chi_{<R'}(a_1 + a_2).$$

By a change of variables and homogeneity, we have for each $\xi \neq 0$,

$$a_i(x,\xi) = a_i\left(x, \frac{\xi}{|\xi|}\right), \quad i = 1, 2.$$

By (7.3) and Corollary 4.6, we then easily verify property (iii) for $\chi_{>1}(|\xi|)\chi_{<R'}a_1$. By the nontrapping assumption, one may verify (iii) for the symbol $\chi_{>1}(|\xi|)\chi_{<R'}a_2$ as well. Properties (iv) and (v) for both $\chi_{>1}(|\xi|)\chi_{<R'}a_1$ and $\chi_{>1}(|\xi|)\chi_{<R'}a_2$ are a straightforward consequence of Proposition 4.7. This completes the proof of Proposition 7.2. $\qquad\square$

Now, we turn to establishing the main estimate (7.2). We begin by defining the symbols $\mathbf{q}$ and $|\mathbf{q}|$:

$$\mathbf{q} := \begin{pmatrix} q & 0 \\ 0 & -q \end{pmatrix}, \quad |\mathbf{q}| := q\mathbf{I}_{2\times 2}.$$

Define $\mathbf{Q} := \frac{1}{2}Op(\mathbf{q}) + \frac{1}{2}Op(\mathbf{q})^*$. Performing a similar calculation to Lemma 6.2, we note that $\mathbf{P}$ is skew-adjoint up to a $L^2_x \to L^2_x$ bounded error. Therefore, it is a straightforward algebraic manipulation to verify the inequality

$$\mathrm{Re}\langle \mathbf{QP}\mathbf{u}, \mathbf{u}\rangle \geq \frac{1}{2}\mathrm{Re}\langle [\mathbf{Q}, \mathbf{P}]\mathbf{u}, \mathbf{u}\rangle - C_2\|v\|^2_{L^\infty_T H^\sigma_x},$$

where C_2 is as in Proposition 7.2. We then obtain the basic preliminary energy estimate,

$$
\begin{aligned}
&\frac{1}{2}\mathrm{Re}\langle \mathbf{Qu}, \mathbf{u}\rangle(T) + \int_0^T \mathrm{Re}\langle (\frac{1}{2}[\mathbf{Q}, \mathbf{P}^0_{k_0}] + \mathbf{QB}^0_{k_0})\mathbf{u}, \mathbf{u}\rangle dt \\
&\leq \frac{1}{2}\mathrm{Re}\langle \mathbf{Qu}, \mathbf{u}\rangle(0) + \frac{1}{2}\int_0^T \mathrm{Re}\langle [\mathbf{Q}, (\mathbf{P}^0_{k_0} - \mathbf{P})]\mathbf{u}, \mathbf{u}\rangle dt + \int_0^T \mathrm{Re}\langle \mathbf{QR}, \mathbf{u}\rangle dt + C_2\|\mathbf{u}\|^2_{L^\infty_T L^2_x}.
\end{aligned}
\tag{7.7}
$$

Now, we estimate each term in (7.7). By Cauchy-Schwarz and Proposition 2.1, we have

$$|\mathrm{Re}\langle \mathbf{Qu}, \mathbf{u}\rangle(T)| + |\mathrm{Re}\langle \mathbf{Qu}, \mathbf{u}\rangle(0)| \leq C_2\|\mathbf{u}\|^2_{L^\infty_T L^2_x}.$$

Next, by Proposition 7.2, the principal symbol $\mathbf{c}(x,\xi)$ of $[\mathbf{Q}, \mathbf{P}^0_{k_0}]$ satisfies,

$$\mathbf{c}(x,\xi) + C(M)rq\mathbf{I}_{2\times 2} \geq \frac{1}{2}C(M)(\chi_{<2R}|\xi|q + \chi_{<R'}|\mathbf{B}^0_{k_0}|q)\mathbf{I}_{2\times 2}.$$

We can therefore choose $C(M)$ large enough so that

$$\mathbf{c}(x,\xi) + C(M)rq\mathbf{I}_{2\times 2} - \chi_{<2R}|\xi|q\mathbf{I}_{2\times 2} - \frac{1}{2}C(M)\chi_{<R'}q|\mathbf{B}^0_{k_0}|\mathbf{I}_{2\times 2} \geq \mathbf{0}.$$

Then, the classical Gårding inequality Proposition 2.2 along with its matrix version (see Remark 2.3) yields

$$(7.8) \quad \int_0^T \operatorname{Re}\langle (\frac{1}{2}[\mathbf{Q},\mathbf{P}^0_{k_0}] + \mathbf{Q}\mathbf{B}^0_{k_0})\mathbf{u},\mathbf{u}\rangle dt \gtrsim \|\chi_{<2R}\mathbf{u}\|^2_{L^2_T H^{\frac{1}{2}}_x} - C_2\|\mathbf{u}\|^2_{L^\infty_T L^2_x} - C(M)\|Op(rq)\mathbf{u}\|_{Y^0}\|\mathbf{u}\|_{X^0}$$
$$- \|\mathbf{Q}\chi_{>R'}\mathbf{B}^0_{k_0}\mathbf{u}\|_{Y^0}\|\mathbf{u}\|_{X^0},$$

where we applied Hölder's inequality in T to control the lower order error term in Remark 2.3 by the $L^\infty_T L^2_x$ norm of $\mathbf{u}$ and the $Y^* = X$ duality to control the remaining first order terms. To control the first Y^0 error term on the right, we use property (i) from Proposition 7.2 to estimate

$$C(M)\|Op(rq)\mathbf{u}\|_{Y^0}\|\mathbf{u}\|_{X^0} \lesssim_M \epsilon\|v\|^2_{X^\sigma},$$

which holds as long as $\mathbf{u}$ is localized at high enough frequency (i.e. k_1 is large enough). To control the latter Y^0 error term, we first note that by Proposition 2.5, the embedding $L^1_T L^2_x \subset Y^0$ and Hölder in T, we have

$$\|\mathbf{Q}\chi_{>R'}\mathbf{B}^0_{k_0}\mathbf{u}\|_{Y^0} \leq \|\chi_{>R'}\mathbf{B}^0_{k_0}\mathbf{Q}\mathbf{u}\|_{Y^0} + C_2\|\mathbf{u}\|_{L^\infty_T L^2_x}.$$

Then by using Proposition 2.13 and arguing as with the analogous terms in the previous section, we have

$$\|\chi_{>R'}\mathbf{B}^0_{k_0}\mathbf{Q}\mathbf{u}\|_{Y^0} \lesssim \|\chi_{>R'}(\tilde{b}_{<k_0}(0), b_{<k_0}(0), \nabla_x g^{ij}_{<k_0}(0))\|_{l^1 X^{s_0-1}}\|\mathbf{Q}S_{>k_1-4}\|_{X^0\to X^0}\|\mathbf{u}\|_{X^0} + C_2\|\mathbf{u}\|_{L^\infty_T L^2_x}.$$

Using the fact that the L^∞ norm of the symbol $\mathbf{q}$ depends only on R and M and not on R', we can take k_1 and R' large enough so that Proposition 2.7 and (5.4) ensure that

$$\|\chi_{>R'}(\tilde{b}_{<k_0}(0), b_{<k_0}(0), \nabla_x g^{ij}_{<k_0}(0))\|_{l^1 X^{s_0-1}}\|\mathbf{Q}S_{>k_1-4}\|_{X^0\to X^0} \leq \epsilon.$$

It then follows by Cauchy-Schwarz and (7.8) that we have

$$\int_0^T \operatorname{Re}\langle (\frac{1}{2}[\mathbf{Q},\mathbf{P}^0_{k_0}] + \mathbf{Q}\mathbf{B}^0_{k_0})\mathbf{u},\mathbf{u}\rangle dt \gtrsim \|\chi_{<2R}\mathbf{u}\|^2_{L^2_T H^{\frac{1}{2}}_x} - C_2\|\mathbf{u}\|^2_{L^\infty_T L^2_x} - \epsilon\|v\|^2_{X^\sigma}.$$

Next, we estimate the contribution of the second term in the second line of (7.7). The procedure here is essentially identical to the estimate in (6.23). Using the symbol bounds for q in Proposition 7.2 (specifically, that the derivatives of q up to first order have uniform in k_0 bounds), we can estimate by taking k_0 large enough and T small enough as in the proof of Lemma 6.9 to obtain

$$\|[\mathbf{Q},(\mathbf{P}^0_{k_0} - \mathbf{P}_k)]\mathbf{u}\|_{Y^0} \leq \epsilon\|v\|_{X^\sigma}.$$

To estimate the third term in the second line of (7.7), we use the $Y^* = X$ duality and Proposition 2.7 to obtain

$$\int_0^T \operatorname{Re}\langle \mathbf{Q}\mathbf{R},\mathbf{u}\rangle dt \leq C_0\|\mathbf{R}\|_{Y^0}\|v\|_{X^\sigma},$$

where the constant C_0 depends only on M and R if k_1 is large enough. Taking T small enough in (7.1) and using Cauchy-Schwarz, we have

$$\int_0^T \operatorname{Re}\langle \mathbf{Q}\mathbf{R},\mathbf{u}\rangle dt \leq C_0\|f\|^2_{Y^\sigma} + \epsilon^2\|v\|^2_{X^\sigma}.$$

Putting the above estimates together, we obtain

$$\|\chi_{<2R}\mathbf{u}\|_{L^2_T H^{\frac{1}{2}}_x} \leq C_2(\|v\|_{L^\infty_T H^\sigma_x} + \|f\|_{Y^\sigma}) + \epsilon\|v\|_{X^\sigma}.$$

This establishes (7.2), which completes the proof of Proposition 7.1.

8. Proof of the main linear estimate

In this short section, we complete the proof of Theorem 5.1 by combining Proposition 6.1 and Proposition 7.1. First, note that by Lemma 5.3, Lemma 5.4 and Lemma 5.7, it suffices to establish for small enough T, the bound

$$(8.1) \qquad \|v\|_{X^\sigma} \leq C(M, L)(\|v_0\|_{H^\sigma} + \|f\|_{Y^\sigma}), \qquad \sigma \geq 0,$$

when v is a solution to (5.2) with $\widehat{v}$ supported at frequencies $\gtrsim 2^{k_1}$ for some arbitrarily large (but fixed) parameter k_1. Let $\epsilon > 0$ be a small positive constant to be chosen. By Proposition 6.1 and Proposition 7.1, we have the initial estimate

$$(8.2) \qquad \|v\|_{L_T^\infty H_x^\sigma} + \|v\|_{\mathcal{X}^\sigma} \leq C(M, L)(\|v_0\|_{H^\sigma} + \|f\|_{Y^\sigma}) + \epsilon \|v\|_{X^\sigma}.$$

We would like to strengthen this bound by replacing the left-hand side of (8.2) with $\|v\|_{X^\sigma}$, which would suffice to complete the proof. For this, we require control of the slightly stronger (than the $L_T^\infty H_x^\sigma$) norm

$$\|v\|_{\mathcal{Z}^\sigma} := \left(\sum_{j \geq 0} 2^{2j\sigma} \|S_j v\|_{L_T^\infty L_x^2}^2 \right)^{\frac{1}{2}}.$$

Clearly, (8.1) will follow from (8.2) and the following lemma, for ϵ small enough (depending on M and L).

Lemma 8.1. *Under the above assumptions, v satisfies the following estimate in the space $\mathcal{Z}^\sigma$:*

$$\|v\|_{\mathcal{Z}^\sigma} \leq C(M, L)(\|v_0\|_{H^\sigma} + \|f\|_{Y^\sigma} + \|v\|_{\mathcal{X}^\sigma}) + \epsilon \|v\|_{X^\sigma}.$$

Proof. We begin by defining $v_k := S_k v$ for each $k \geq 0$. We see that v_k satisfies the equation

$$(8.3) \qquad \begin{cases} i\partial_t v_k + \partial_j T_{g^{ij}} \partial_i v_k + T_{b^j} \partial_j v_k + T_{\tilde{b}^j} \partial_j \overline{v}_k = S_k f + \mathcal{R}_k, \\ v_k(0) = S_k v_0, \end{cases}$$

where

$$\mathcal{R}_k := [T_{g^{ij}}, S_k] \partial_i \partial_j \tilde{S}_k v + [T_{\partial_j g^{ij}}, S_k] \partial_i \tilde{S}_k v + [T_{b^j}, S_k] \partial_j \tilde{S}_k v + [T_{\tilde{b}^j}, S_k] \partial_j \tilde{S}_k \overline{v}$$

and $\tilde{S}_k$ is a fattened version of the dyadic multiplier S_k. By dyadic summation and Proposition 6.1, the proof of the lemma will be concluded if we can show that

$$(8.4) \qquad \|\mathcal{R}_k\|_{Y^\sigma} \leq C(M) \|\tilde{S}_k v\|_{\mathcal{X}^\sigma} + \epsilon \|\tilde{S}_k v\|_{X^\sigma}$$

for some $\epsilon > 0$ sufficiently small. This is an easy consequence of Proposition 2.10 for the latter three terms as we can estimate these in $L_T^1 H_x^\sigma$ and take T small. To estimate the remaining term, we first observe that

$$[T_{g^{ij}}, S_k] \partial_i \partial_j \tilde{S}_k v = [T_{S_{<k} g^{ij}}, S_k] \partial_i \partial_j \tilde{S}_k v = [S_{<k} g^{ij}, S_k] \partial_i \partial_j \tilde{S}_k v + [T_{S_{<k} g^{ij}} - S_{<k} g^{ij}, S_k] \partial_i \partial_j \tilde{S}_k v.$$

The latter term above can be estimated easily in $L_T^1 H_x^\sigma$ by the right-hand side of (8.4) by using paradifferential calculus and then by taking T small. For the remaining term, we use that

$$[S_{<k} g^{ij}, S_k] \partial_i \partial_j \tilde{S}_k v = 2^{-k} L(S_{<k} \nabla_x g^{ij}, \partial_i \partial_j \tilde{S}_k v),$$

where L is a translation invariant operator of the form

$$L(\phi_1, \phi_2)(x) = \int \phi_1(x + y) \phi_2(x + z) K(y, z) dy dz, \qquad \|K\|_{L^1} \lesssim 1.$$

See, for instance, [32]. As the spaces $l^1 X^{s_0}$ and $\mathcal{X}^\sigma$ are translation invariant (in that they admit translation invariant equivalent norms), it follows from Proposition 2.12 that we have

$$\|[S_{<k}g^{ij}, S_k]\partial_i\partial_j \tilde{S}_k v\|_{Y^\sigma} \leq C(M)\|\tilde{S}_k v\|_{\mathcal{X}^\sigma}.$$

This completes the proof of the lemma. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

9. Well-posedness for the nonlinear flow

Now, we proceed with the proof of Theorem 1.3. By differentiating (1.1), we obtain an equation for $(u, \nabla u)$ of the form (1.11). Therefore, it suffices to prove the second part of the theorem for (1.11). Given the key estimate and well-posedness in Theorem 5.1, the scheme for proving this follows a very similar path to [24, Section 7]. We only outline the main results and procedure here for the convenience of the reader, and refer to the corresponding parts of [24] where relevant. A fully detailed exposition of a simplified version of the scheme that we employ below can be found in [10].

The starting point is to rewrite the equation

$$(9.1) \qquad \begin{cases} i\partial_t u + \partial_j g^{ij}(u, \overline{u})\partial_i u = F(u, \overline{u}, \nabla u, \nabla \overline{u}), \\ u(0, x) = u_0(x), \end{cases}$$

in the paradifferential form

$$\begin{cases} i\partial_t u + \partial_j T_{g^{ij}}\partial_i u + T_{b^j}\partial_j u + T_{\tilde{b}^j}\partial_j \overline{u} = G(u, \overline{u}, \nabla u, \nabla \overline{u}), \\ u(0, x) = u_0(x), \end{cases}$$

where

$$b := -\partial_{(\nabla u)}F, \qquad \tilde{b} := -\partial_{(\nabla \overline{u})}F$$

and

$$G(u, \overline{u}, \nabla u, \nabla \overline{u}) := (\partial_j T_{g^{ij}}\partial_i - \partial_j g^{ij}\partial_i)u + F(u, \overline{u}, \nabla u, \nabla \overline{u}) + T_{b^j}\partial_j u + T_{\tilde{b}^j}\partial_j \overline{u}.$$

9.1. Existence of $l^1 X^s$ solutions to the nonlinear equation.
Our first aim is to establish existence of $l^1 X^s$ solutions to the equation (1.11) for small time. This is given by the following proposition.

Proposition 9.1. *Let $s > \frac{d}{2}+2$ and let $u_0 \in l^1 H^s$ with $\|u_0\|_{l^1 H^s} = M$. Suppose that $g(u_0)$ is a nontrapping, non-degenerate metric with parameters R_0 and L. Then there is $T_0 > 0$ depending on M, $L(R_0)$ and R_0 such that for every $T \leq T_0$, there exists a solution $u \in l^1 X^s$ to (1.11) such that*

    (i) ($l^1 X^s$ bound).

$$\|u\|_{l^1 X^s} \leq C(M, L)\|u_0\|_{l^1 H^s}.$$

    (ii) (Smallness outside B_{R_0}).

$$\|\chi_{>R_0}u\|_{l^1 X^s} \leq 2\epsilon.$$

    (iii) (Comparable nontrapping parameter).

$$L(u) \leq 2L(u_0).$$

As in Section 7 of [24], for each $n \geq 0$ we consider the following iteration scheme for the paradifferential form of the nonlinear equation:

$$(9.2) \quad \begin{cases} i\partial_t u^{n+1} + \partial_i T_{g^{ij}(u^n)} \partial_j u^{n+1} + T_{b^j(u^n)} \partial_j u^{n+1} + T_{\tilde{b}^j(u^n)} \partial_j \overline{u}^{n+1} = G(u^n), \\ u^{n+1}(0,x) = u_0(x), \end{cases}$$

with initialization $u^0 = 0$. Here, we are suppressing the dependence on derivatives of u^n and its complex conjugate in b^j, $\tilde{b}^j$ and G. It is clear that Proposition 9.1 will follow from our next proposition, which addresses the convergence and bounds for the iteration scheme.

Proposition 9.2. *Let s, M, L, R_0, T_0 and u_0 be as in Proposition 9.1. Then there exists a constant $C(M, L)$ such that for every $n \geq 0$ there exists a solution u^n to (9.2) on $[0, T]$ such that*

    (i) ($l^1 X^s$ bound).
$$\|u^n\|_{l^1 X^s} \leq C(M, L)\|u_0\|_{l^1 H^s}.$$

    (ii) (Smallness outside B_{R_0}).
$$\|\chi_{>R_0} u^n\|_{l^1 X^s} \leq 2\epsilon.$$

    (iii) (Comparable nontrapping parameter).
$$L(u^n) \leq 2L(u_0).$$

Moreover, there is a function $u \in l^1 X^s$ satisfying the same bounds as above such that u^n converges strongly to u in $l^1 X^\sigma$ for every $0 \leq \sigma < s$.

Remark 9.3. For simplicity of presentation, we have omitted the parameter $\frac{d}{2} + 2 < s_0 < s$ used in [24, Section 7] from the statements of the results in this section. This parameter still needs to be taken into account in the (omitted) proofs to ensure that the bounds for the low-frequency coefficients g, b, and $\tilde{b}$ stay under control in each iteration.

The proof of the above proposition follows from a virtually identical line of reasoning as [24, Sections 7.1-7.3]. We simply use Proposition 4.5 and Theorem 5.1 in place of the analogues in their proof. We omit the details.

9.2. **Uniqueness and the weak Lipschitz bound.** In this subsection, we establish uniqueness of solutions in the class $l^1 X^s$ when $s > \frac{d}{2} + 2$. In fact, our uniqueness result follows as a corollary of a weak Lipschitz type bound as noted in the following proposition.

Proposition 9.4. *Let $s > \frac{d}{2} + 2$ and let $u_0^1 \in l^1 H^s$. Assume that $g(u_0^1)$ is a non-degenerate, nontrapping metric with parameters M, R_0 and L as above. Suppose that $u_0^2 \in l^1 H^s$ is another initial datum satisfying*

$$\|u_0^2\|_{l^1 H^s} \lesssim M,$$

and suppose that u_0^2 is close to u_0^1 in the $l^1 L^2$ topology in the sense that

$$\|u_0^1 - u_0^2\|_{l^1 L^2} \ll_M e^{-C(M)L(R_0)}.$$

Then the following statements hold:

    (i) $g(u_0^2)$ is nontrapping with comparable parameters to $g(u_0^1)$.
    (ii) The solutions u^1 and u^2 generated by u_0^1 and u_0^2 exist on a time interval $[0, T]$ whose length depends only on the parameters M, R_0 and $L(R_0)$.
    (iii) For $0 \leq \sigma < s_0 - 1$, we have the following weak Lipschitz type bound:
$$\|u^1 - u^2\|_{l^1 X^\sigma} \leq C(M, L)\|u_0^1 - u_0^2\|_{l^1 H^\sigma}.$$

Proof. The proof follows an identical line of reasoning as Section 7.4 in [24] except that we use Proposition 4.5 in place of Proposition 5.2 in [24] to prove (i). $\square$

9.3. **Frequency envelope bounds and continuous dependence.** In this final subsection, our main objective is to establish continuous dependence for (9.1). More precisely, for $s > \frac{d}{2} + 2$, we want to show that the data-to-solution map (given nontrapping data) $u_0 \mapsto u$ is continuous from $l^1 H^s$ to $l^1 X^s$. As in [24], the main ingredient is the following frequency envelope bound for the solution $u \in l^1 X^s$ in terms of the data.

Proposition 9.5. *Let $u \in l^1 X^s$ be a solution to* (9.1) *as in Proposition 9.1 with initial data $u_0 \in l^1 H^s$. Let a_k be an admissible frequency envelope for u_0 in $l^1 H^s$. Then the solution u satisfies the bound*

$$\|S_k u\|_{l^1 X^s} \le a_k C(M, L) \|u_0\|_{l^1 H^s}.$$

Proof. The proof follows identical reasoning as the proof of Proposition 7.5 in [24]. The only difference is that we use Corollary 5.2 in place of the analogous bound in their proof. $\square$

Armed with Proposition 9.5, the proof of the continuity of the data-to-solution map in Section 7.6 of [24] now applies verbatim to establish the same property in our setting.

References

[1] Jean-Michel Bony. Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires. *Ann. Sci. École Norm. Sup. (4)*, 14(2):209–246, 1981.

[2] Alberto-P. Calderón and Rémi Vaillancourt. A class of bounded pseudo-differential operators. *Proc. Nat. Acad. Sci. U.S.A.*, 69:1185–1187, 1972.

[3] Hiroyuki Chihara. Local existence for semilinear Schrödinger equations. *Math. Japon.*, 42(1):35–51, 1995.

[4] Walter Craig, Thomas Kappeler, and Walter Strauss. Microlocal dispersive smoothing for the Schrödinger equation. *Comm. Pure Appl. Math.*, 48(8):769–860, 1995.

[5] A. Davey and K. Stewartson. On three-dimensional packets of surface waves. *Proc. Roy. Soc. London Ser. A*, 338:101–110, 1974.

[6] Shin-ichi Doi. Remarks on the Cauchy problem for Schrödinger-type equations. *Comm. Partial Differential Equations*, 21(1-2):163–178, 1996.

[7] Nakao Hayashi and Tohru Ozawa. Remarks on nonlinear Schrödinger equations in one space dimension. *Differential Integral Equations*, 7(2):453–461, 1994.

[8] Lars Hörmander. Pseudo-differential operators and non-elliptic boundary problems. *Ann. of Math. (2)*, 83:129–209, 1966.

[9] Wataru Ichinose. On L^2 well posedness of the Cauchy problem for Schrödinger type equations on the Riemannian manifold and the Maslov theory. *Duke Math. J.*, 56(3):549–588, 1988.

[10] Mihaela Ifrim and Daniel Tataru. Local well-posedness for quasi-linear problems: A primer. *Bulletin of the American Mathematical Society*, 60(2):167–194, 2023.

[11] Yuji Ishimori. Multivortex solutions of a two-dimensional nonlinear wave equation. *Progr. Theoret. Phys.*, 72(1):33–37, 1984.

[12] In-Jee Jeong and Sung-Jin Oh. On the Cauchy problem for the Hall and electron magnetohydrodynamic equations without resistivity I: Illposedness near degenerate stationary solutions. *Ann. PDE*, 8(2):Paper No. 15, 106, 2022.

[13] In-Jee Jeong and Sung-Jin Oh. Illposedness for dispersive equations: Degenerate dispersion and Takeuchi–Mizohata condition. *arXiv preprint arXiv:2308.15408*, 2023.

[14] In-Jee Jeong and Sung-Jin Oh. Wellposedness of the electron MHD without resistivity for large perturbations of the uniform magnetic field. *preprint*, 2023.

[15] C. E. Kenig, G. Ponce, C. Rolvung, and L. Vega. Variable coefficient Schrödinger flows for ultrahyperbolic operators. *Adv. Math.*, 196(2):373–486, 2005.

[16] Carlos E. Kenig, Gustavo Ponce, Christian Rolvung, and Luis Vega. The general quasilinear ultrahyperbolic Schrödinger equation. *Adv. Math.*, 206(2):402–433, 2006.

[17] Carlos E. Kenig, Gustavo Ponce, and Luis Vega. Smoothing effects and local existence theory for the generalized nonlinear Schrödinger equations. *Invent. Math.*, 134(3):489–545, 1998.

[18] Carlos E. Kenig, Gustavo Ponce, and Luis Vega. The Cauchy problem for quasi-linear Schrödinger equations. *Invent. Math.*, 158(2):343–388, 2004.

[19] H. Kumano-Go. *Pseudo-Differential Operators*. MIT Press, Cambridge, 1981.

[20] P. D. Lax and L. Nirenberg. On stability for difference schemes: A sharp form of Gàrding's inequality. *Comm. Pure Appl. Math.*, 19:473–492, 1966.

[21] Jeremy Marzuola, Jason Metcalfe, and Daniel Tataru. Wave packet parametrices for evolutions governed by PDO's with rough symbols. *Proc. Amer. Math. Soc.*, 136(2):597–604, 2008.

[22] Jeremy L. Marzuola, Jason Metcalfe, and Daniel Tataru. Quasilinear Schrödinger equations I: Small data and quadratic interactions. *Adv. Math.*, 231(2):1151–1172, 2012.

[23] Jeremy L. Marzuola, Jason Metcalfe, and Daniel Tataru. Quasilinear Schrödinger equations, II: Small data and cubic nonlinearities. *Kyoto J. Math.*, 54(3):529–546, 2014.

[24] Jeremy L. Marzuola, Jason Metcalfe, and Daniel Tataru. Quasilinear Schrödinger equations III: Large data and short time. *Arch. Ration. Mech. Anal.*, 242(2):1119–1175, 2021.

[25] Guy Métivier. *Para-differential calculus and applications to the Cauchy problem for nonlinear systems*, volume 5 of *Centro di Ricerca Matematica Ennio De Giorgi (CRM) Series*. Edizioni della Normale, Pisa, 2008.

[26] Sigeru Mizohata. Some remarks on the Cauchy problem. *J. Math. Kyoto Univ.*, 1:109–127, 1961/62.

[27] Sigeru Mizohata. Sur quelques équations du type Schrödinger. *Journées équations aux dérivées partielles*, pages 1–5, 1981.

[28] Sigeru Mizohata. *On the Cauchy problem*, volume 3 of *Notes and Reports in Mathematics in Science and Engineering*. Academic Press, Inc., Orlando, FL; Science Press Beijing, Beijing, 1985.

[29] E. I. Schulman and V. E. Zakharov. Degenerative dispersion laws, motion invariants and kinetic equations. *Phys. D*, 1(2):192–202, 1980.

[30] Jiro Takeuchi. On the Cauchy problem for some non-Kowalewskian equations with distinct characteristic roots. *J. Math. Kyoto Univ.*, 20(1):105–124, 1980.

[31] Terence Tao. Global regularity of wave maps. I. Small critical Sobolev norm in high dimension. *Internat. Math. Res. Notices*, pages 299–328, 2001.

[32] Terence Tao. Global regularity of wave maps. II. Small energy in two dimensions. *Comm. Math. Phys.*, 224(2):443–544, 2001.

[33] Michael E. Taylor. *Pseudodifferential operators and nonlinear PDE*, volume 100 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1991.

Department of Mathematics, University of California at Berkeley
Email address: bpineau@berkeley.edu

Department of Mathematics, ETH Zürich
Email address: mitchell.taylor@math.ethz.ch