Title

Energy dispersed solutions for the (4 + 1)-dimensional Maxwell-Klein-Gordon equation

Journal
American Journal of Mathematics, 140(1)

ISSN
0002-9327

Authors
Oh, Sung-Jin
Tataru, Daniel

Publication Date
2018

DOI
10.1353/ajm.2018.0000

Peer reviewed

THE HYPERBOLIC YANG–MILLS EQUATION IN THE CALORIC GAUGE. LOCAL WELL-POSEDNESS AND CONTROL OF ENERGY DISPERSED SOLUTIONS

SUNG-JIN OH AND DANIEL TATARU

Abstract. This is the second part in a four-paper sequence, which establishes the Threshold Conjecture and the Soliton Bubbling vs. Scattering Dichotomy for the hyperbolic Yang–Mills equation in the $(4+1)$-dimensional space-time. This paper provides the key gauge-dependent analysis of the hyperbolic Yang–Mills equation.

We consider topologically trivial solutions in the caloric gauge, which was defined in the first paper [18] using the Yang–Mills heat flow. In this gauge, we establish a strong form of local well-posedness, where the time of existence is bounded from below by the energy concentration scale. Moreover, we show that regularity and dispersive properties of the solution persists as long as energy dispersion is small. We also observe that fixed-time regularity (but not dispersive) properties in the caloric gauge may be transferred to the temporal gauge without any loss, proving as a consequence small data global well-posedness in the temporal gauge.

The results in this paper are used in the subsequent papers [19, 20] to prove the sharp Threshold Theorem in caloric gauge in the trivial topological class, and the Dichotomy Theorem in arbitrary topological classes.

Contents

Date: March 19, 2021.

1. Introduction

In this paper, along with the companion papers [18], [19] and [20], we consider the hyperbolic Yang–Mills equation in the $(4+1)$-dimensional Minkowski space with a compact semi-simple structure group.

In [18], we defined the notion of *caloric gauge* with the help of the Yang–Mills heat flow on $\mathbb{R}^4$, and showed that every subthreshold connection admits a caloric gauge representative (see Section 1.2 below for a review). The first main result of the present paper (Theorem 1.13) is a strong form of local well-posedness of the hyperbolic Yang–Mills equation in the manifold of caloric gauge connections, where the time of existence is estimated from below by the scale of energy concentration. The second main result (Theorem 1.16) asserts that regularity and dispersive behaviors persist as long as a certain quantity called *energy dispersion*, which measures a certain type of non-dispersive concentration, remains small.

While the caloric gauge reveals the fine cancellation structure of the Yang–Mills equation, and is thus suitable for dispersive analysis at low regularity, it has the drawback that causality is lost. As a remedy, we also show that regularity (but not dispersive) properties in the caloric gauge may be transferred to the temporal gauge. As a corollary, we also obtain small data global well-posedness of the hyperbolic Yang–Mills equation in the temporal gauge (Theorem 1.18).

In the subsequent papers in the sequence [19], [20], we use the results proved in this paper to establish the Threshold Theorem (i.e., global well-posedness and scattering for subthreshold data) in the caloric gauge, as well as the Soliton Bubbling vs. Scattering Dichotomy Theorem for general finite energy solutions, formulated in more gauge-covariant fashion. An overview of the entire series is provided in [21].

1.1. **Hyperbolic Yang–Mills equation on $\mathbb{R}^{1+4}$.** Our set-up is as follows. Let $\mathbf{G}$ be a compact noncommutative Lie group and $\mathfrak{g}$ its associated Lie algebra. We denote by $Ad(O)X = OXO^{-1}$ the adjoint (or conjugation) action of $\mathbf{G}$ on $\mathfrak{g}$ and by $ad(X)Y = [X, Y]$ the Lie bracket on $\mathfrak{g}$. We use the notation $\langle X, Y\rangle$ for a bi-invariant inner product on $\mathfrak{g}$,

$$\langle [X, Y], Z\rangle = \langle X, [Y, Z]\rangle, \qquad X, Y, Z \in \mathfrak{g},$$

or equivalently

$$\langle X, Y\rangle = \langle Ad(O)X, Ad(O)Y\rangle, \qquad X, Y \in \mathfrak{g}, \quad O \in \mathbf{G}.$$

If $\mathbf{G}$ is semisimple then one can take $\langle X, Y\rangle = -\mathrm{tr}(ad(X)ad(Y))$ i.e. negative of the Killing form on $\mathfrak{g}$, which is then positive definite, However, a bi-invariant inner product on $\mathfrak{g}$ exists for any compact Lie group $\mathbf{G}$.

Let $\mathbb{R}^{1+4}$ be the (4+1)-dimensional Minkowski space equipped with the Minkowski metric, which takes the form $\mathrm{diag}(-1, +1, \dots, +1)$ in the rectangular coordinates $(x^0, x^1, \dots, x^4)$. The coordinate x^0 serves the role of time, and we will often write $x^0 = t$. Throughout this paper, we will use the standard convention for raising or lowering indices using the Minkowski metric, and summing up repeated upper and lower indices.

Our objects of study are connection 1-forms A on $\mathbb{R}^{1+4}$ taking values in the Lie algebra $\mathfrak{g}$. They define covariant differentiation operators $\mathbf{D}_\mu = \mathbf{D}^{(A)}_\mu = \partial_\mu + A_\mu$ (in coordinates) acting on sections of any vector bundle with structure group $\mathbf{G}$. The commutator $\mathbf{D}_\mu \mathbf{D}_\nu - \mathbf{D}_\nu \mathbf{D}_\mu$ yields the curvature 2-form $F_{\mu\nu} = F[A]_{\mu\nu}$, which is given in terms of A_μ by the formula

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu].$$

Given a $\mathbf{G}$-valued function O on $\mathbb{R}^{1+4}$, we introduce the notation

$$O_{;\mu} = \partial_\mu O O^{-1}.$$

The pointwise action of O on the vector bundle induces a gauge transformation for A and F, namely

$$A_\mu \mapsto OA_\mu O^{-1} - \partial_\mu O O^{-1} = Ad(O)A_\mu - O_{;\mu}, \qquad F_{\mu\nu} \mapsto OF_{\mu\nu}O^{-1} = Ad(O)F_{\mu\nu}.$$

In view of this transformation property, F may be viewed as a 2-form taking values in the $\mathbf{G}$-vector bundle with fiber $\mathfrak{g}$, where $\mathbf{G}$ acts on $\mathfrak{g}$ by the adjoint action (geometrically, the adjoint vector bundle). Thus the covariant derivative $\mathbf{D}_\mu$ acts on F by

$$\mathbf{D}_\mu F_{\alpha\beta} = (\partial_\mu + ad(A_\mu))F_{\alpha\beta} = \partial_\mu F_{\alpha\beta} + [A_\mu, F_{\alpha\beta}].$$

The *hyperbolic Yang–Mills equation* on $\mathbb{R}^{1+4}$ is the Euler–Lagrange equation associated with the formal Lagrangian action functional

$$\mathcal{L}(A) = \frac{1}{2} \int_{\mathbb{R}^{1+4}} \langle F_{\alpha\beta}, F^{\alpha\beta} \rangle \, dxdt,$$

which takes the form

$$\mathbf{D}^\alpha F_{\alpha\beta} = 0. \tag{1.1}$$

Clearly, (1.1) is invariant under gauge transformations. This equation possesses a conserved energy, given by

$$\mathcal{E}_{\{t\}\times\mathbb{R}^4}(A) = \int_{\{t\}\times\mathbb{R}^4} \sum_{\alpha<\beta} |F_{\alpha\beta}|^2 \, dx. \tag{1.2}$$

Furthermore, both the equation (1.1) and the energy (1.2) are invariant under the scaling

$$A(t,x) \mapsto \lambda A(\lambda t, \lambda x) \qquad (\lambda > 0).$$

Hence, the hyperbolic Yang–Mills equation is energy critical in dimension (4+1), which is the reason why we focus on this dimension in the present series of papers.

We are interested in the initial value problem for (1.1). For this purpose, we first formulate a gauge-covariant notion of an initial data set. We say that a pair (a, e) of a connection 1-form a and a $\mathfrak{g}$-valued 1-form e on $\mathbb{R}^4$ is an initial data set for a solution A to (1.1) if

$$(A_j, F_{0j}) \restriction_{\{t=0\}} = (a_j, e_j).$$

Here and throughout this paper, roman letter indices stand for the spatial coordinates $x^1, \ldots, x^4$. Note that (1.1) with $\beta = 0$ imposes the condition that

$$\mathbf{D}^j e_j = \partial^j e_j + [a^j, e_j] = 0. \tag{1.3}$$

This equation is the *Gauss* (or the *constraint*) *equation* for (1.1).

It turns out that (1.3) characterizes precisely those pairs (a, e) which can arise as an initial data set. Thus we make the following definition:

Definition 1.1. (1) A *regular initial data set* for the hyperbolic Yang–Mills equation is a pair $(a, e) \in H^N_{loc} \times H^{N-1}$ ($N \geq 2$), which has finite energy (i.e., $F[a] \in L^2$) and satisfies the constraint equation (1.3).

(2) A *finite energy initial data set* is a pair $(a, e) \in \dot{H}^1_{loc} \times L^2$ which has finite energy (i.e., $F[a] \in L^2$) and satisfies the constraint equation (1.3).

In this paper, we make an additional assumption that a decays suitably at infinity:

$$a \in \dot{H}^1. \tag{1.4}$$

This assumption turns out to be equivalent to the requirement that a is topologically trivial [19]. As this property is conserved under any continuous evolution in time, this is the natural setting for scattering and thus for the Threshold Conjecture for (1.1), which is one main subject of the final paper [20] of the series.

The hyperbolic Yang–Mills equation (1.1), when naively viewed as an evolution equation for A, fails to be locally well-posed; to restore (at least formally) well-posedness, we need to fix the gauge invariance.

There are several classical interesting gauge choices which can be made here, for instance the Coulomb gauge $\partial^j A_j = 0$, the temporal gauge $A_0 = 0$ and the Lorenz gauge $\partial^\alpha A_\alpha = 0$. For a more detailed discussion and comparison of these gauges we refer the reader to our first article [18].

However, the main gauge choice we use in this paper is the so-called *caloric gauge*, which was defined in the first paper of the series [18] with the help of a parabolic analogue of (1.1), namely the *Yang–Mills heat flow*. This is the subject of our next discussion.

1.2. **Yang–Mills heat flow and the caloric gauge.** Let a be a connection 1-form on $\mathbb{R}^4$ (in short, a spatial connection). We say that a connection $A = A(x, s)$ on $\mathbb{R}^4 \times J$ (where J is a subinterval of $[0, \infty)$) is a (covariant) *Yang–Mills heat flow* development of a if it solves

$$F_{sj} = \mathbf{D}^\ell F_{\ell j}, \quad A(s = 0) = a. \tag{1.5}$$

This equation is invariant under gauge transformations on $\mathbb{R}^4 \times J$. Under the *local caloric gauge* condition

$$A_s = 0, \tag{1.6}$$

the forward-in-s initial value problem for (1.5) is locally well-posed [18, Theorem 2.7] in $\dot{H}^1$. We remark that the evolution (1.5) under the gauge (1.6) is precisely the gradient flow for the (spatial) energy

$$\mathcal{E}_e(a) = \frac{1}{2} \int_{\mathbb{R}^4} \langle F_{jk}[a], F^{jk}[a] \rangle \, dx = \int_{\mathbb{R}^4} \sum_{j<k} |F_{jk}[a]|^2 \, dx.$$

The key controlling norm for the Yang–Mills heat flow in the local caloric gauge is $\|F\|_{L^3_s(J;L^3)}$, which is both scale- and gauge-invariant.

Theorem 1.2 ([18])**.** *Consider a Yang–Mills heat flow $A \in C_s(J; \dot{H}^1)$ in the local caloric gauge satisfying*

$$\|F\|_{L^3_s(J;L^3)} \leq \mathcal{Q} < \infty. \tag{1.7}$$

When $J = [0, s_0)$ for $s_0 < \infty$, A can be extended past s_0 as a (well-posed) Yang–Mills heat flow. When $J = [0, \infty)$, the solution has the property that the limit

$$\lim_{s \to \infty} A(s) = a_\infty$$

exists in $\dot{H}^1$. The limiting connection is flat ($F[a_\infty] = 0$) and the map $a \mapsto a_\infty$ is locally Lipschitz in $\dot{H}^1$, H^N ($N \geq 1$) and $\dot{H}^1 \cap \dot{H}^N$ ($N \geq 2$). Denoting by $O(a)$ a gauge transformation satisfying $O^{-1}\partial_j O = a_\infty$, the map $a \mapsto O(a)$ is continuous from $\dot{H}^1$ to $\dot{H}^2$ up to constant conjugations.

In the case when the Yang–Mills heat flow with initial data a admits a global solution with finite L^3 norm for the curvature as in (1.7), we define the *caloric size* $\mathcal{Q}(a)$ of a as

$$\mathcal{Q}(a) = \|F\|^3_{L^3_s(\mathbb{R}^+;L^3)} \tag{1.8}$$

We note that this is a gauge invariant quantity.

Remark 1.3. Here we need to clarify the topology on the (nonlinear) space of gauge transformations. We will say that a sequence $O^{(n)}$ converges to O if there exists a sequence $\tilde{O}^{(n)}$ of gauge transformations so that $\tilde{O}^{(n)}(O^{(n)})^{-1}$ are constant and so that we have

- Pointwise convergence[1]:

$$d(\tilde{O}^{(n)}, O) \to 0 \qquad \text{in } L^2_{loc}$$

- Convergence of derivatives

$$\tilde{O}^{(n)}_{;x} \to O_{;x} \qquad \text{in } \dot{H}^1$$

A simple but important case in which (1.7) holds with $J = [0, \infty)$ is when the initial energy $\mathcal{E}_e(a)$ is sufficiently small. The same conclusion holds as long as $\mathcal{E}_e(a)$ is below any nontrivial connection $a \in \dot{H}^1$ satisfying the harmonic Yang–Mills equation

$$\mathbf{D}^\ell F_{\ell j} = 0. \tag{1.9}$$

The above assertion is closely related to the topological class of connections. Relaxing the requirement $a \in \dot{H}^1$ to $a \in H^1_{loc}$ allows also topologically nontrivial initial data sets, in which case the ground state energy

$$E_{GS} = \inf\{\mathcal{E}_e(a) : a \in H^1_{loc} \text{ is nontrivial and solves (1.9)}\} \tag{1.10}$$

is nonzero, and the minimum is attained for a special class of solutions called instantons. However, within the trivial topological class we have

$$2E_{GS} \leq \inf\{\mathcal{E}_e(a) : a \in \dot{H}^1 \text{ is nontrivial and solves (1.9)}\}. \tag{1.11}$$

We further remark that in order for a connection a to have $\mathcal{Q}(a)$ finite, it must be topologically trivial. Because of this, the present paper is limited to topologically trivial connections, which are simply defined by the requirement that $a \in \dot{H}^1$ in a suitable gauge. For an extended discussion and further references we refer the reader to our next article in the series [19].

In view of this discussion, the following result is natural:

Theorem 1.4 (Threshold theorem for the Yang–Mills heat flow on $\mathbb{R}^4$ [18])**.** *Assume that a is topologically trivial and that*

$$\mathcal{E}_e(a) < 2E_{GS}.$$

Then the solution to (1.5) *exists globally on* $[0, \infty)$*. Moreover, there exists a non-decreasing function* $\mathcal{Q}(\cdot) : [0, 2E_{GS}) \to [0, \infty)$ *such that*

$$\mathcal{Q}(a) \leq \mathcal{Q}(\mathcal{E}_e(a)).$$

[1]The functions $O^{(n)}$ are uniformly bounded in BMO so this property essentially provides the additional information that in some sense the local averages converge as well.

We now return to the discussion of an arbitrary (not necessarily subthreshold) spatial connection a, whose Yang–Mills heat flow development satisfies (1.7) with $J = [0, \infty)$. Since the limiting connection a_∞ is flat, it must be gauge equivalent to the zero connection. This motivates the following definition of the *caloric gauge*:

Definition 1.5 (Caloric gauge)**.** We say that a connection $a_j \in \dot{H}^1$ is *caloric* if $J = [0, \infty)$ and a_∞ in Theorem 1.2 is equal to zero. We denote the set of all such connections by $\mathcal{C}$. More quantitatively, we denote by $\mathcal{C}_\mathcal{Q}$ the set of all caloric connections whose Yang–Mills heat flow development satisfies

$$\mathcal{Q}(a) \leq \mathcal{Q}. \tag{1.12}$$

Given a connection $a \in \dot{H}^1$ satisfying (1.7) with $J = [0, \infty)$, note that

$$Cal(a)_j = Ad(O(a))a_j - O(a)_{;j}$$

is its caloric representative, which is unique up to constant conjugations.

To solve the Yang–Mills equation in the caloric gauge, we need to view the family $\mathcal{C}$ of the caloric gauge connections as an infinite dimensional manifold. Here the $\dot{H}^1$ topology is no longer sufficient, so we introduce the slightly stronger topology

$$\mathbf{H} = \{a \in \dot{H}^1 : \|a\|_{\mathbf{H}} < \infty\}, \text{ where } \|a\|_{\mathbf{H}} := \|a\|_{\dot{H}^1} + \sum_j \|P_j(\partial^\ell a_\ell)\|_{L^2}.$$

Here, $\{P_j\}$ refer to the standard Littlewood–Paley projections to dyadic frequency annuli on $\mathbb{R}^4$. It turns out that every caloric connection belongs to $\mathbf{H}$, which reflects the fact, to be discussed in Section 3 in greater detail, that caloric connections satisfy a nonlinear form of the Coulomb gauge condition. Moreover, the following theorem holds.

Theorem 1.6. *(1) For a connection $a \in \mathcal{C}$ with energy $\mathcal{E}$ and caloric size $\mathcal{Q}$ we have*

$$\|a\|_{\mathbf{H}} \lesssim_{\mathcal{E},\mathcal{Q}} 1.$$

(2) Consider a connection $a \in \mathbf{H}$ (not necessarily caloric) satisfying (1.12)*. Then $O(a)$ in Theorem 1.2 may be uniquely fixed by imposing $\lim_{|x|\to\infty} O(a) = I$. Such a map $a \mapsto O(a)$ is locally C^1 from $\mathbf{H}$ to $\dot{H}^2 \cap C^0$, and also from H^N to $\dot{H}^2 \cap \dot{H}^{N+1}$* $(N \geq 2)$.

Essentially as a corollary, we have:

Theorem 1.7. *The set $\mathcal{C}$ is an infinite dimensional C^1 submanifold of $\mathbf{H}$.*

The spatial components of a finite energy Yang–Mills waves will be continuous functions of time which take values into $\mathcal{C}$. They are however not C^1 in time; instead their time derivative will merely belong to L^2. Because of this, we need to take the closure of its tangent space $T\mathcal{C}$ (which a-priori is a closed subspace of $\mathbf{H}$) in L^2. This is denoted by $T_a^{L^2}\mathcal{C}$. It is also convenient to have a direct way of characterizing this space; that is naturally done via the linearization of (1.5):

Definition 1.8. For a caloric gauge connection $a \in \mathcal{C}$, we say that $L^2 \ni b \in T_a^{L^2}\mathcal{C}$ iff the solution to the linearized local caloric gauge Yang–Mills heat flow equation

$$\partial_s B_k = [B^j, F_{kj}] + \mathbf{D}^j(\mathbf{D}_j B_k - \mathbf{D}_k B_j), \qquad B_k(s=0) = b_k, \tag{1.13}$$

(where $\mathbf{D} = \mathbf{D}^{(a)}$) satisfies

$$\lim_{s\to\infty} B(s) = 0.$$

We say that $(a,b) \in T^{L^2}\mathcal{C}_{\mathcal{Q}}$ (resp. $T^{L^2}\mathcal{C}$) if $a \in \mathcal{C}_{\mathcal{Q}}$ (resp. $\mathcal{C}$) and $b \in T^{L^2}_a\mathcal{C}$.

A key property of the tangent space $T^{L^2}_a\mathcal{C}$ is the following nonlinear div-curl type decomposition:

Theorem 1.9. *Let $a \in \mathcal{C}_{\mathcal{Q}}$ with energy $\mathcal{E}$. Then for each $\mathfrak{g}$-valued 1-form $e \in L^2$ there exists a unique decomposition*

$$e = b - \mathbf{D}^{(a)} a_0, \qquad b \in T^{L^2}_a\mathcal{C}, \qquad a_0 \in \dot{H}^1, \tag{1.14}$$

where b is a $\mathfrak{g}$-valued 1-form and a_0 is a $\mathfrak{g}$-valued function, with the corresponding bound

$$\|b\|_{L^2} + \|a_0\|_{\dot{H}^1} \lesssim_{\mathcal{E},\mathcal{Q}} \|e\|_{L^2}. \tag{1.15}$$

A hyperbolic Yang–Mill connection consists not only of spatial components (the sole subject of discussion so far), but also of a temporal component. As in the Coulomb gauge, we will consider the spatial components of the connection as the dynamic variables, which satisfy a system of wave equations. The temporal components, on the other hand, will be viewed as an auxiliary variable determined from the spatial components. This point of view motivates the following definition.

Definition 1.10 (Initial data in the caloric gauge)**.** An initial data for the Yang–Mills equation in the caloric gauge is a pair (a,b) where $(a,b) \in T^{L^2}\mathcal{C}$.

The notion of covariant Yang–Mills initial data (Definition 1.1) is connected to the preceding definition by the following result proved in [18] (which motivates the notation in Theorem 1.9):

Theorem 1.11. *(1) Given any Yang–Mills initial data pair $(a,e) \in \dot{H}^1 \times L^2$ such that the Yang–Mills heat flow development of a satisfies (1.12), there exists a caloric gauge Yang–Mills data $(\tilde{a}, b) \in T^{L^2}\mathcal{C}$ and $a_0 \in \dot{H}^1$, so that the initial data pair $(\tilde{a}, \tilde{e})$ is gauge equivalent to (a,e), where*

$$\tilde{e}_k = b_k - \mathbf{D}^{(\tilde{a})}_k a_0.$$

In addition, $(\tilde{a}, b)$ and a_0 are unique up to constant conjugations, and depend continuously on (a,e) in the corresponding quotient topology. Further, the map $(a,e) \mapsto (\tilde{a}, b)$ is locally C^1 in the stronger topology[2] $\mathbf{H} \times L^2 \to \mathbf{H} \times L^2$, as well as in more regular spaces $H^N \times H^{N-1} \to H^N \times H^{N-1}$ $(N \geq 2)$.

(2) Given any caloric gauge data $(a,b) \in T^{L^2}\mathcal{C}$, there exists an unique $a_0 \in \dot{H}^1$, with Lipschitz dependence on $(a,b) \in \dot{H}^1 \times L^2$, so that

$$e_k = b_k - \mathbf{D}^{(a)}_k a_0$$

satisfies the constraint equation (1.3). Further, the map $(a,b) \to a_0$ is also Lipschitz from $H^N \times H^{N-1}$ to H^N for $N \geq 3$.

Remark 1.12. The caloric gauge just described is a global version of a local caloric gauge previously introduced by the first author [13, 14], and is based on an idea by Tao [26] in his study of the energy critical wave maps into the hyperbolic space [27, 28, 29, 30, 31].

[2]Here we impose again the condition $\lim_{|x|\to\infty} O(a) = I$ in order to fix the choice of $O(a)$.

1.3. **The main results.** The first main result is a strong gauge-dependent local well-posedness theorem for the Yang–Mills equation as an evolution in the manifold of caloric connections. To state this result, we define the *energy concentration scale* r_c of a Yang–Mills initial data set (a, e) with threshold ϵ_* (or the ϵ_*-energy concentration scale) to be

$$r_c^{\epsilon_*} = r_c^{\epsilon_*}[a, e] = \sup\{r > 0 : \mathcal{E}_{B_r(x)}(a, e) \leq \epsilon_*^2 \text{ for all } x \in \mathbb{R}^4\}.$$

Theorem 1.13 (Local well-posedness in caloric gauge)**.** *There exists a non-increasing function $\epsilon_*(\mathcal{E}, \mathcal{Q}) > 0$ and a non-decreasing function $M_*(\mathcal{E}, \mathcal{Q}) > 0$ such that the Yang–Mills equation in the caloric gauge is locally well-posed on the time interval of length $r_c = r_c^{\epsilon_*}(\mathcal{E}, \mathcal{Q})$ for initial data (a, e) with energy $\leq \mathcal{E}$ and $a \in \mathcal{C}_{\mathcal{Q}}$. More precisely, the following statements hold.*

(1) (Regular data) Let (a, e) be a smooth initial data set with energy $\leq \mathcal{E}$, where $a \in \mathcal{C}_{\mathcal{Q}}$. Then there exists a unique smooth solution $A_{t,x}$ to the Yang–Mills equation in caloric gauge on $I = [-r_c, r_c]$ such that $(A_j, F_{0j}) \restriction_{\{t=0\}} = (a_j, e_j)$.

(2) (Rough data) The data-to-solution map admits a continuous extension

$$\mathcal{C} \times L^2 \ni (a, e) \mapsto (A_x, \partial_t A_x) \in C(I, T^{L^2}\mathcal{C})$$

in the class of initial data with energy $\leq \mathcal{E}$, $a \in \mathcal{C}_{\mathcal{Q}}$ and energy concentration scale $\geq r_c$.

(3) (A-priori bound) The solution defined as above obeys the a-priori bound

$$\|A_x\|_{S^1[I]} \leq M_*(\mathcal{E}, \mathcal{Q}).$$

(4) (Weak Lipschitz dependence) Let $(a', e') \in \mathcal{C} \times L^2$ be another initial data set with energy concentration scale $\geq r_c$. For $\sigma < 1$ close to 1, we have the global bound

$$\|A_x - A'_x\|_{S^\sigma[I]} \lesssim_{M_*(\mathcal{E},\mathcal{Q}),\sigma} \|(a, e) - (a', e')\|_{\dot{H}^\sigma \times \dot{H}^{\sigma-1}}.$$

The a-priori bound (3) is highly gauge-dependent and has strong consequences. The S^1-norm, which is essentially the same as in [10] and is recalled in Section 4.1 below, serves the role of a controlling (or scattering) norm for the Yang–Mills equation in the caloric gauge. As we will see in Section 5, finiteness of the S^1-norm implies fine properties of the solution itself, such as frequency envelope control, persistence of regularity, continuation and scattering towards endpoints of I, and also for those nearby, such as weak Lipschitz dependence and local-in-time continuous dependence.

Theorem 1.13 implies small energy global well-posedness in the caloric gauge, analogous to the similar Coulomb gauge result in [11]:

Corollary 1.14. *If the energy of the initial data set is smaller than $\epsilon_*^2 := \min\{1, \epsilon_*^2(1, \mathcal{Q}(1))\}$, then the corresponding solution $A_{t,x}$ in the caloric gauge exists globally and obeys*

$$\|A_x\|_{S^1[(-\infty,\infty)]} \leq M_*(\mathcal{E}).$$

Moreover, if the initial data set (a, e) has subthreshold energy, then by Theorem 1.4 we have $a \in \mathcal{C}_{\mathcal{Q}}$ with $\mathcal{Q} \leq \mathcal{Q}(\mathcal{E})$. Therefore, we immediately obtain:

Corollary 1.15. *For initial data with subthreshold energy, the conclusions of Theorem 1.13 hold with ϵ_*, M_* and r_c depending only on the energy $\mathcal{E}$.*

The local well-posedness result (Theorem 1.13) provides a basic framework for considering dynamics of the Yang–Mills equation in the manifold of caloric connections $\mathcal{C}$. The second main result, which we now state, is a continuation/scattering criterion for this equation in terms of smallness of a quantity called *energy dispersion* (denoted by $ED[I]$ below).

Theorem 1.16 (Regularity and scattering of energy dispersed YM solutions)**.** *There exists a non-increasing function* $\epsilon(\mathcal{E}, \mathcal{Q}) > 0$ *and a non-decreasing function* $M(\mathcal{E}, \mathcal{Q})$ *such that if* $A_{t,x}$ *is a solution (in the sense of Theorem 1.13) to the Yang–Mills equation in caloric gauge on* I *with energy* $\leq \mathcal{E}$ *and with initial caloric size* $\mathcal{Q}$ *that obeys*

$$\|F\|_{ED[I]} = \sup_{k \in \mathbb{Z}} 2^{-2k} \|P_k F\|_{L^\infty(I \times \mathbb{R}^4)} \leq \epsilon(\mathcal{E}, \mathcal{Q}),$$

then it satisfies the a-priori bound

$$\|A_x\|_{S^1[I]} \leq M(\mathcal{E}, \mathcal{Q}),$$

as well as

$$\sup_{t \in I} \mathcal{Q}(A(0)) \ll 1.$$

By finiteness of the S^1-norm, $A_{t,x}$ may be continued as a solution to the Yang–Mills equation in the caloric gauge past finite endpoints of I, and scatters in some sense towards the infinite endpoints; see Remarks 5.2 and 5.3.

Remark 1.17. In contrast to Theorem 1.13, in Theorem 1.16 the dependence on $\mathcal{Q}$ is very mild. This feature is due to the fact that small energy dispersion, combined with the energy bound, implies that $\mathcal{Q}$ must be either very large or very small; see Lemma 5.10 below. In particular if $\mathcal{E}$ is subthreshold then the dependence on $\mathcal{Q}$ above can be omitted altogether.

While powerful conclusions about the solution (represented by the S^1-norm bound) can be made in the caloric gauge, it has the disadvantage that the causality (or the finite speed of propagation) property is lost. To remedy this, we also establish small data well-posedness result in the temporal gauge $A_0 = 0$:

Theorem 1.18. *If the energy of the initial data set is smaller than* ϵ_*^2 *(as in Corollary 1.14), then the corresponding solution* $(A_{t,x}, \partial_t A_{t,x})$ *in the temporal gauge* $A_0 = 0$ *exists globally in* $C_t(\mathbb{R}; \dot{H}^1 \times L^2)$*. The solution is unique among the local-in-time limits of smooth solutions, and it depends continuously on data* $(a, e) \in \dot{H}^1 \times L^2$*.*

In fact, Theorem 1.18 is a consequence of Corollary 1.14, after the observation that the gauge transformation from the caloric gauge to the temporal gauge obeys optimal regularity bounds; see Theorem 5.1 (10) below. We note that the strong dispersive S^1-norm bound for A is generally lost in the temporal gauge, as some part of the solution is merely transported (instead of solving a wave equation).

Theorems 1.18 is used in the third paper [19] of the sequence to establish the large data local theory for the $(4+1)$-dimensional Yang–Mills equation in arbitrary topological classes. Then in the fourth paper [20], this theory is put together with Theorems 1.13 and 1.16 to establish global well-posedness and scattering in the caloric gauge for data with subthreshold energy (often called the *threshold theorem* in the literature), as well as a bubbling vs. scattering dichotomy for arbitrary finite-energy solutions, formulated in a gauge covariant sense.

Remark 1.19. Within the setup of this paper, one could in effect easily relax the hypothesis of the above theorem, and show that temporal gauge solutions exist for as long as caloric solutions exist. We do not pursue this, as our primary interest in terms of the temporal gauge is to use it for solutions which are not necessarily caloric. These matters are further discussed in our third and fourth papers [19, 20].

The overall strategy for the proofs originated from the work of Sterbenz and the second author on the energy critical wave maps [23, 24], and was adapted to the case of the energy critical Maxwell–Klein–Gordon (MKG) equation, which is a simpler model for Yang–Mills, in the authors' previous works [16, 17, 15]. We also note an alternative independent approach for the energy critical wave maps [8] and MKG [7] based on the Kenig–Merle method [4, 3]. A more extensive historical perspective is provided in the fourth paper [20].

In [16] and [17], the analogues of Theorems 1.13 and 1.16 (respectively) were proved using distinct strategies. However, here we derive both main results (see Section 7 for details) from the following single a-priori estimate concerning regular solutions, whose proof is the central goal of this paper:

Theorem 1.20. *There exist non-increasing functions $\epsilon(\mathcal{E}, \mathcal{Q}), T(\mathcal{E}, \mathcal{Q}) > 0$ as well as a non-decreasing function $M(\mathcal{E}, \mathcal{Q})$ such that if $A_{t,x}$ is a regular solution to the Yang–Mills equation in caloric gauge on I with energy $\leq \mathcal{E}$ such that $A_x \in \mathcal{C}_{\mathcal{Q}}$ for all $t \in I$, and moreover*

$$\sup_{k \geq m} 2^{-2k} \|P_k F\|_{L^\infty(I \times \mathbb{R}^4)} \leq \epsilon(\mathcal{E}, \mathcal{Q}) \quad \text{and} \quad |I| \leq 2^{-m} T(\mathcal{E}, \mathcal{Q})$$

for some $m \in \mathbb{Z}$, then it satisfies the a-priori bound

$$\|A_x\|_{S^1[I]} \leq M(\mathcal{E}, \mathcal{Q}).$$

In words, for a regular solution with small energy dispersion only at certain frequency 2^m and above, an a-priori S^1-norm bound holds on time intervals of the corresponding scale $O(2^{-m})$.

1.4. Overview of the paper.

- **Section 2.** In this section, we collect some notation and conventions used throughout this paper for the reader's convenience. Some basic concepts, such as disposability, dyadic function spaces, frequency envelopes, etc, are also described.

After Section 2, the paper is organized into two tiers. The first tier consists of Sections 3 to 7, and its goal is to describe the large-scale proof of the main results, assuming the validity of certain linear and multilinear estimates collected in Section 4.

- **Section 3.** Here, we recall from [18] further results concerning the Yang–Mills heat flow and the caloric gauge. First, we state some quantitative bounds for the Yang–Mills heat flow and its linearization in the caloric gauge, using the language of frequency envelopes (Section 3.1). Next, we derive the wave equation satisfied by A_x and $A_x(s)$ $(s > 0)$ in the caloric gauge (Section 3.2). In this process we use the *dynamic Yang–Mills heat flow* (3.5), which is the Yang–Mills heat flow augmented with a heat evolution (in s) for the temporal component.
- **Section 4.** We first describe the fine function space framework for analyzing the hyperbolic Yang–Mills equation in the caloric gauge (Section 4.1). The main function spaces are identical to those in [10, 17, 11], which in turn have their roots in the works on wave maps [32, 25]. We also explain the three main sources of smallness in our analysis: divisibility, small energy dispersion and short time interval. Then we state the linear and multilinear estimates needed for the proof of the main theorems (Sections 4.2 and 4.3); it is the goal of the second tier of the paper (described below) to prove them. The primary estimates here are the bilinear null form estimates, which in the context of our function spaces have their origin in [10, 17, 11]. The bilinear null structure of the

Yang–Mills nonlinearities was first described in [5]; a secondary trilinear null structure, which also play a role here, was discovered in [12] in the (MKG) context.

- **Section 5.** We prove a strong *structure theorem* for a solution to the hyperbolic Yang–Mills equation in the caloric gauge with finite S^1-norm (Section 5.1). In particular, it reduces the tedious task of controlling various parts of a solution $A_{t,x}$ to proving a single S^1-norm bound for the spatial components A_x. We also consider the effect of small inhomogeneous energy dispersion on a correspondingly short time interval (Section 5.2). The analysis is repeated for the dynamic Yang–Mills heat flow of a solution (Section 5.3).
- **Section 6.** We prove the central result, Theorem 1.20, by an induction on energy argument. The argument is similar to [17], which in turn was based on the work [23], with modifications to handle the low frequencies with possibly large energy dispersion with the short length of the time interval (see, in particular, Scenario (1) in Section 6.2).
- **Section 7.** Here, we derive the main theorems stated in Section 1.3 from Theorem 1.20. The key point in the derivation of Theorem 1.13 is the simple fact that energy dispersion is small for frequencies above the inverse of the energy-concentration scale (Section 7.2). Theorem 1.16 follows essentially by scaling (Section 7.3).

The second tier consists of Sections 8 to 11. Here, we provide proofs of the estimates stated in Section 4.

- **Section 8.** The goal of this section is to prove all multilinear estimates stated in Section 4. The proofs proceed in two stages: In the first stage, we assume global-in-time dyadic (in spatial frequency) estimates (Section 8.2), and derive the interval-localized frequency envelope bounds stated in Section 4 (Section 8.3). A key technical issue in interval localization is to deal with modulation projections, which are non-local in time. In the second stage, we establish the global-in-time dyadic estimates (Section 8.4). Much is borrowed from the previous works [10, 17, 11].
- **Section 9.** We begin this section by reducing the proof of the key linear estimates in Section 4 to construction of a parametrix for the paradifferential d'Alembertian $\Box + 2\sum_k ad(P_{<k-\kappa}\mathbf{P}_\alpha A)\partial^\alpha P_k$ (Section 9.1). As in [11], the parametrix is constructed via conjugation of the free wave propagator by a pseudodifferential renormalization operator. We define and state the key properties of the renormalization operator (Section 9.3), and establish the desired estimates for the parametrix assuming these properties (Section 9.4).
- **Section 10.** Here, we prove the mapping properties of the renormalization operator claimed in Section 9. The key difference from [11] lies in the source of smallness: Whereas smallness of the S^1-norm of A was used in [11], in this paper we rely instead on largeness of the frequency gap κ in the paradifferential d'Alembertian. The idea of exploiting a large frequency gap was used in [23, 17].
- **Section 11.** Finally, we estimate the error for conjugation of the paradifferential d'Alembertian by the renormalization operator claimed in Section 9, thereby completing our parametrix construction. One aspect of our proof that differs from the previous works [23, 17] is that, in addition to the large frequency gap κ, we need to use smallness of a divisible norm (weaker than S^1) of A, which requires a careful interval localization procedure (Sections 11.3 and 11.4).

Acknowledgments. Part of the work was carried out during the semester program "New Challenges in PDE" held at MSRI in Fall 2015. S.-J. Oh was supported by the Miller Research Fellowship from the Miller Institute, UC Berkeley and the TJ Park Science Fellowship from the POSCO TJ Park Foundation. D. Tataru was partially supported by the NSF grant DMS-1266182 as well as by the Simons Investigator grant from the Simons Foundation.

2. Notation, conventions and other preliminaries

2.1. **Notation and conventions.** Here we collect some notation and conventions used in this paper.

- The symbols $\lesssim$, $\gtrsim$, $\ll$ and $\gg$ are defined with their usual meanings, where the implicit constants in these notations are allowed to vary from line to line.
- By $A \lesssim_E B$ and $A \ll_E B$, we mean that $A \leq C_E B$ and $A \leq c_E B$, respectively, where $C_E = C_0(1+E)^{C_1}$ and $c_E = C_0^{-1}(1+E)^{-C_1}$ for some constants $C_0, C_1 > 0$ that are again allowed to vary from line to line.
- For $u \in \mathfrak{g}$ and $O \in \mathbf{G}$, define $ad(u) = [u, \cdot]$ and $Ad(O) = O(\cdot)O^{-1}$, both of which are in $\mathrm{End}(\mathfrak{g})$. Recall the minus Killing form, which is invariant under $Ad(O)$ and $ad(X)$. On $\mathfrak{g}$, define $|\cdot|_{\mathfrak{g}}$ on $\mathfrak{g}$ by the minus Killing form. On $\mathrm{End}(\mathfrak{g})$, use the induced metric $|a|_{\mathrm{End}(\mathfrak{g})} = \sup_{|u|_{\mathfrak{g}} \leq 1} |au|_{\mathfrak{g}}$. By Ad-invariance, $|Ad(O)a|_{\mathrm{End}(\mathfrak{g})} = |aAd(O^{-1})|_{\mathrm{End}(\mathfrak{g})} = |a|_{\mathrm{End}(\mathfrak{g})}$.
- We use the notation $B_r(x)$ for the ball of radius r centered at x. We write $|\angle(\xi,\eta)|$ for the angular distance $|\frac{\xi}{|\xi|} - \frac{\eta}{|\eta|}|$, and $|\angle(\mathcal{C},\mathcal{C}')|$ for $\inf_{\xi\in\mathcal{C},\eta\in\mathcal{C}'} |\angle(\xi,\eta)|$.
- We use the notation $\nabla = \partial_{t,x}$, $D_\mu = i^{-1}\partial_\mu$. Also, for D and A we often suppress the subscript x and write $D = D_x$ and $A = A_x$.
- We say that a multilinear operator $\mathcal{O}(u_1,\dots,u_m)$ is *disposable* if its kernel is translation invariant and has mass $\lesssim 1$. In particular, we have
$$\|\mathcal{O}(u_1,\dots,u_m)\|_Y \lesssim \|u_1\|_{X_1}\cdots\|u_m\|_{X_m}$$
for any translation invariant spaces $X_1,\dots,X_m,Y$ provided that a product estimate
$$\|u_1\cdots u_m\|_Y \lesssim \|u_1\|_{X_1}\cdots\|u_m\|_{X_m}$$
holds for any functions $u_1 \in X_1,\dots,u_m \in X_m$.
- We often use the 'duality' pairing
$$\iint u_0\mathcal{O}(u_1,\dots,u_m)\,dxdt$$
so as to have symmetry among u_0 and the inputs. Indeed, we have
$$\iint u_0\mathcal{O}(u_1,\dots,u_m)\,dxdt = \iint_{\Xi^0+\Xi^1+\cdots+\Xi^m=0} \mathcal{O}(\Xi^1,\dots,\Xi^m)\widetilde{u_0}(\Xi^0)\widetilde{u_1}(\Xi^1)\cdots\widetilde{u_m}(\Xi^m)\,d\Xi dt$$
- We define $\mathcal{O}^{*_i}$ as
$$\iint u_0\mathcal{O}^{*_i}(u_1,\dots,u_i,\dots,u_m)\,dtdx = \iint u_i\mathcal{O}(u_1,\dots,\overbrace{u_0}^{i\text{-th entry}},\dots,u_m)\,dtdx$$
- By a *bilinear operator (of $\mathfrak{g}$-valued functions) with symbol $m(\xi,\eta) = m^{\mathbf{ab}}(\xi,\eta)$* (which is a complex-valued 4×4-matrix), we mean an expression of the form
$$\mathfrak{L}(a,b) = \iint \left(m^{\mathbf{ab}}(\xi,\eta)[\hat{a}_{\mathbf{a}}(\xi),\hat{b}_{\mathbf{b}}(\eta)]\right) e^{i(\xi+\eta)\cdot x}\,\frac{d\xi\,d\eta}{(2\pi)^8}.$$

For a scalar-valued symbol $m(\xi,\eta)$, we implicitly associate the corresponding multiple of the identity $m^{\mathbf{ab}}(\xi,\eta) = m(\xi,\eta)\delta^{\mathbf{ab}}$.

If $\mathfrak{L}$ were symmetric, then the symbol $m(\xi,\eta)$ is *anti-symmetric in* ξ,η, in the sense that $m^{\mathbf{ab}}(\xi,\eta) = -m^{\mathbf{ba}}(\eta,\xi)$; this is due to the antisymmetry of the Lie bracket.

2.2. **Basic multipliers and function spaces.** Here we provide the definitions of basic multipliers and function spaces. For the more elaborate frequency projections and function spaces for the hyperbolic Yang–Mills equation, see Section 4.1.

- Given a function space X (on either $\mathbb{R}^d$ or $\mathbb{R}^{1+d}$), we define the space $\ell^p X$ by
$$\|u\|_{\ell^p X}^p = \sum_k \|P_k u\|_X^p$$
(with the usual modification for $p=\infty$), where P_k ($k \in \mathbb{Z}$) are the usual Littlewood–Paley projections to dyadic frequency annuli.
- For a spatial 1-form A, we define $\mathbf{P}A$ to be its *Leray projection*, i.e., the L^2-projection to divergence-free vector fields:
$$\mathbf{P}_j A = A_j + (-\Delta)^{-1}\partial_j \partial^\ell A_\ell.$$
We write $\mathbf{P}_j^\perp A = A_j - \mathbf{P}_j A$.
- For a space-time 1-form A_α, we introduce the notation $\mathbf{P}_\alpha A = (\mathbf{P}A)_\alpha$ by defining
$$\mathbf{P}_\alpha A = \begin{cases} \mathbf{P}_j A_x & \alpha = j \in \{1,\dots,4\}, \\ A_0 & \alpha = 0. \end{cases}$$
We also define $\mathbf{P}_\alpha^\perp A = (\mathbf{P}^\perp A)_\alpha = A_\alpha - \mathbf{P}_\alpha A$.
- We denote by $\dot{W}^{\sigma,p}$ the homogeneous L^p-Sobolev space with regularity σ. In the case $p=2$, we simply write $\dot{H}^\sigma = \dot{W}^{\sigma,2}$.
- The mixed space-time norm $L^q_t \dot{W}^{\sigma,r}_x$ of functions on $\mathbb{R}^{1+d}$ is often abbreviated as $L^q \dot{W}^{\sigma,r}$.

2.3. **Frequency envelopes.** To provide more accurate versions of many of our estimates and results we use the language of frequency envelopes.

Given a sequence c_k ($k \in \mathbb{Z}$) of positive numbers and a translation invariant norm $\|\cdot\|_X$, we introduce the shorthand
$$\|u\|_{X_c} := \sup_k \frac{\|P_k u\|_X}{c_k}.$$

Definition 2.1. Given a translation invariant space of functions X, we say that a sequence c_k of positive numbers is a frequency envelope for a function $u \in X$ if

(i) The dyadic pieces of u satisfy
$$\|u\|_{X_c} \leq 1, \text{ or equivalently, } \|P_k u\|_X \leq c_k$$
(ii) The sequence c_k is slowly varying,
$$2^{-\delta(j-k)} \lesssim \frac{c_k}{c_j} \lesssim 2^{\delta(j-k)}, \qquad j > k.$$

Here δ is a small positive universal constant. For some of the results we need to relax the slowly varying property in a quantitative way. Fixing a universal small constant $0 < \epsilon \ll 1$, we set

Definition 2.2. Let $\sigma_1, \sigma_2 > 0$. A frequency envelope c_k is called $(-\sigma_1, \sigma_2)$-admissible if

$$2^{-\sigma_1(1-\epsilon)(j-k)} \lesssim \frac{c_k}{c_j} \lesssim 2^{\sigma_2(1-\epsilon)(j-k)}, \qquad j > k.$$

When $\sigma_1 = \sigma_2$, we simply say that c_k is σ-admissible.

Another situation that will occur frequently is that where we have a reference frequency envelope c_k, and then a secondary envelope d_k describing properties which apply on a background controlled by c_k. In this context the envelope d_k often cannot be chosen arbitrarily but instead must be in a constrained range depending on c_k. To address such matters we set:

Definition 2.3. We say that the envelope d_k is σ*-compatible* with c_k if we have

$$c_k \sum_{j<k} 2^{\sigma(1-\epsilon)(j-k)} d_j \lesssim d_k.$$

We will often replace envelopes d_k which do not satisfy the above compatibility condition by slightly larger envelopes that do:

Lemma 2.4 ([18, Lemma 3.5])**.** *Assume that c_k and d_k are $(-\sigma_1, S)$ envelopes, and also that c_k is bounded. Then for $\tilde{\sigma} < \sigma(1-\epsilon)$ the envelope*

$$e_k = d_k + c_k \sum_{j<k} 2^{\tilde{\sigma}(j-k)} d_j$$

is σ-compatible with c_k. The implicit constant in Definition 2.3 is bounded above by $1 + C_{\sigma(1-\epsilon)-\tilde{\sigma}} \|c\|_{\ell^\infty}$.

Finally we need the following additional frequency envelope notation:

$$(c \cdot d)_k = \ c_k d_k, \qquad\qquad a_{\leq k} = \sum_{j \leq k} a_j,$$

$$c_k^{[\sigma]} = \ \sup_{j<k} 2^{(1-\epsilon)\sigma(j-k)} c_j \qquad (\sigma > 0).$$

2.4. **Global small constants.** In this paper, we use a string of global small constants $\delta_1, \ldots, \delta_6, \delta_7$ with the following hierarchy:

$$0 < \delta_* = \delta_7 \ll \delta_6 \ll \delta_5 \ll \delta_4 \ll \delta_3 \ll \delta_2 \ll \delta_1 \ll \delta_0 \ll 1. \tag{2.1}$$

These are fixed from right to left, so that

$$\delta_{i+1} \ll \delta_i^{100}.$$

The role of each constant is roughly as follows:

- δ_0: For definition of functions spaces, such as Str^1 and b_0, b_1, p_0 in Section 4.
- δ_1: For all bounds from other papers, such as [18, 11, 17]; also for all dyadic gains in explicit nonlinearities (Section 8) and for energy dispersion gains in the Str^1 norm (4.21).
- δ_2: For energy dispersion, frequency gap and off-diagonal gains in Sections 4.
- δ_3: For frequency envelope admissibility range in Sections 4.
- δ_4: For energy dispersion and frequency gap gains in Sections 5.
- δ_5: For frequency envelope admissibility range in Sections 5.

- δ_6: For energy dispersion and frequency gap gains in Sections 6.
- δ_*: For frequency envelope admissibility range in Sections 6.

We use an additional set of small constants in our parametrix construction (Sections 9–11), which are fixed after δ_1 but before δ_2.

3. Yang–Mills heat flow and the caloric gauge

In this section, which is a continuation of Section 1.2, we recall the results from the first paper [18] that are needed in the present paper.

In Section 3.1, we state quantitative bounds for the Yang–Mills heat flow (and its linearization) in the caloric gauge, using the language of frequency envelopes. Section 3.2 is concerned with the task of interpreting the hyperbolic Yang–Mills equation in the caloric gauge as a system of nonlinear wave equations for A_x.

3.1. **Frequency envelope bounds in the caloric gauge.** We begin with frequency envelope bounds for the caloric gauge Yang–Mills heat flow and its linearization.

Proposition 3.1 ([18, Proposition 7.27])**.** *Let $(a,b) \in T^{L^2}\mathcal{C}_{\mathcal{Q}}$ with $\mathcal{E} = \mathcal{E}_e(a)$, and let (A,B) be the solution to* (1.5) *and* (1.13) *with (a,b) as data. Let c_k be a $(-\delta_1, S)$-frequency envelope in $\dot{H}^1 \times L^2$ for (a,b), and let $c_k^{\sigma,p}$ be a $(-\delta_1, S)$-frequency envelope in $\dot{W}^{\sigma,p} \times \dot{W}^{\sigma-1,p}$ for (a,b) which is δ_1-compatible with c_k. Define*

$$\mathbf{A}(s) = A(s) - e^{s\Delta}a, \qquad \mathbf{B}(s) = B(s) - e^{s\Delta}b. \tag{3.1}$$

Then the following properties hold.

(1) We have

$$\|P_k \mathbf{A}(s)\|_{\dot{H}^1} + \|P_k \mathbf{B}(s)\|_{L^2} \lesssim_{\mathcal{E},\mathcal{Q},N} \langle 2^{-2k}s^{-1}\rangle^{-\delta_1} \langle 2^{2k}s\rangle^{-N} c_k^2 \tag{3.2}$$

(2) For (σ,p) and (σ_1,p_1) satisfying

$$c_{\delta_1} \leq \sigma \leq \frac{4}{p} - c_{\delta_1}, \quad 2 + c_{\delta_1} \leq p \leq c_{\delta_1}^{-1}, \quad 0 \leq \sigma_1 \leq \sigma - c_{\delta_1}, \quad \frac{4}{p_1} - \sigma_1 = 2\left(\frac{4}{p} - \sigma\right), \tag{3.3}$$

we have

$$\|P_k \mathbf{A}(s)\|_{\dot{W}^{\sigma_1+1,p_1}} + \|P_k \mathbf{B}(s)\|_{\dot{W}^{\sigma_1,p_1}} \lesssim_{\mathcal{E},\mathcal{Q},N} \langle 2^{-2k}s^{-1}\rangle^{-\delta_1} \langle 2^{2k}s\rangle^{-N} (c_k^{\sigma,p})^2. \tag{3.4}$$

A central object of the remainder of this section is the *dynamic Yang–Mills heat flow* for space-time connections, which is an augmentation of (1.5) with an equation for the temporal component. More precisely, we say that a pair (A_0, A) of a $\mathfrak{g}$-valued function A_0 and a connection A on $\mathbb{R}^4 \times J$ (where J is a subinterval of $[0,\infty)$) is the dynamic Yang–Mills heat flow development of (a_0, a) if

$$F_{s\alpha} = \mathbf{D}^\ell F_{\ell\alpha}, \qquad (A_0, A)(s=0) = (a_0, a). \tag{3.5}$$

This flow is well-defined as long as the spatial and s-components A are well-defined as a solution to (1.5). In particular, if $a \in \mathcal{C}$, then (A_0, A) exists on $[0,\infty)$, $\lim_{s\to\infty} A_0 = 0$ in $\dot{H}^1$ and $\lim_{s\to\infty} F_{0j} = 0$ in L^2. Moreover, the following proposition holds.

Proposition 3.2 ([18, Propositions 7.7 and 8.9])**.** *Let $a \in \mathcal{C}_{\mathcal{Q}}$ and $e \in L^2$ satisfy $\|(f,e)\|_{L^2}^2 \leq \mathcal{E}$. Consider also $a_0 \in \dot{H}^1$ and $b \in T_a^{L^2}\mathcal{C}$ which obeys $e = b - \mathbf{D}a_0$ (cf. Theorem 1.9), and let (A_0, A) be a caloric gauge solution to* (3.5) *with data (a_0, a). Then the following properties hold.*

(1) The spatial 1-form $B_j(s) = F_{0j}(s) - \mathbf{D}_j A_0(s)$ obeys the linearized Yang–Mills heat flow in the caloric gauge with $B_j(0) = b_j$. Moreover,

$$\|A(s)\|_{\dot{H}^1} + \|B(s)\|_{L^2} \lesssim_{\mathcal{E},\mathcal{Q}} \|(f,e)\|_{L^2}. \tag{3.6}$$

(2) Let d_k be a δ_1-frequency envelope for (f,e) in $\dot{W}^{-2,\infty}$. Then

$$2^{-k}\|P_k A(s)\|_{L^\infty} + 2^{-2k}\|P_k B(s)\|_{L^\infty} \lesssim_{\mathcal{E},\mathcal{Q},N} \langle 2^{2k}s\rangle^{-N}(d_k)^{\frac{1}{2}}. \tag{3.7}$$

(3) Let c_k be a $(-\delta_1, S)$-frequency envelope for (a,b) in $\dot{H}^1 \times L^2$. Then

$$\|P_k\mathbf{A}(s)\|_{\dot{H}^1} + \|P_k\mathbf{B}(s)\|_{L^2} \lesssim_{\mathcal{E},\mathcal{Q},N} \langle 2^{-2k}s^{-1}\rangle^{-\delta_1}\langle 2^{2k}s\rangle^{-N}(d_k)^{\frac{1}{2}}c_k, \tag{3.8}$$

$$\|P_k\partial^j A_j(s)\|_{L^2} + \|P_k\partial^j B_j(s)\|_{\dot{H}^{-1}} \lesssim_{\mathcal{E},\mathcal{Q},N} \langle 2^{-2k}s^{-1}\rangle^{-\delta_1}\langle 2^{2k}s\rangle^{-N}(d_k)^{\frac{1}{2}}c_k, \tag{3.9}$$

where $\mathbf{A}$, $\mathbf{B}$ are as in (3.1).

3.2. **Wave equation for A in caloric gauge.** Here, and in the rest of this paper, we shift the notation and denote by $A_{t,x} = A_{t,x}(t,x)$, instead of (a_0, a), the space-time connection on $I \times \mathbb{R}^4$ (viewed as $\{s = 0\}$). For the spatial components, we omit the subscript x and write $A_x(t,x) = A(t,x)$. We write $A_{t,x,s}(s) = A_{t,x,s}(t,x,s)$ for the dynamic Yang–Mills heat flow of $A_{t,x}(t,x)$.

In this subsection, we recall from [18] the interpretation of the hyperbolic Yang–Mills equations for a space-time connection $A_{t,x}$ in the caloric gauge as a hyperbolic evolution for the spatial components A augmented with nonlinear expressions of $\partial^\ell A_\ell$, A_0 and $\partial_0 A_0$ in terms of $(A, \partial_t A)$; see Theorem 3.5. An analogous hyperbolic equation holds for the dynamic Yang–Mills heat flow development $A_{t,x}(s)$ of $A_{t,x}$ in the caloric gauge, which may be thought of as a gauge-covariant regularization of A; see Theorem 3.6.

We present explicit expressions for the quadratic nonlinearities, for which we need to reveal the null structure in order to handle them, and state stronger bounds for the remaining higher order nonlinearities. For economy of notation in the latter task, we introduce the following definition:

Definition 3.3. Let X, Y be dyadic norms.

- A map $\mathbf{F}: X \to Y$ is said to be *envelope-preserving of order* $\geq n$ ($n \in \mathbb{N}$ with $n \geq 2$) if the following property holds: Let c be a $(-\delta_1, S)$ frequency envelope for a in X. Then

$$\|\mathbf{F}(a)\|_{Y_{(c^{[\delta_1]})^{n-1}c}} \lesssim_{\|a\|_X} 1.$$

- A map $\mathbf{F}: X \to Y$ is said to be *Lipschitz envelope-preserving of order* $\geq n$ if, in addition to being envelope preserving of order $\geq n$, the following additional property holds: Let c be a common δ_1-frequency envelopes for a_1 and a_2 in X, and let d be a δ_1-frequency envelope for $a_1 - a_2$ in X that is δ_1-compatible with c. Then

$$\|P_k(\mathbf{F}(a_1) - \mathbf{F}(a_2))\|_{Y_k} \lesssim_{\|a_1\|_X, \|a_2\|_X} c_k^{n-2} e_k,$$

where $e_k = d_k + c_k(c \cdot d)_{\leq k}$.

Remark 3.4. The modified envelope e appears since the maps $\mathbf{F}$ that arise below are defined on a nonlinear manifold, namely, spatial connections a on a time interval I such that $(a, \partial_t a)(t) \in T^{L^2}\mathcal{C}$ for each fixed time. We remark moreover that if the frequency envelopes c and d are ℓ^2-summable, which is usually the case in practice, then $\mathbf{F}(a)$ and $\mathbf{F}(a_1) - \mathbf{F}(a_2)$ belong to $\ell^1 Y$.

We also need to introduce the non-sharp Strichartz spaces Str and Str^1, which scale like $L^\infty L^2$ and $L^\infty \dot{H}^1$, respectively. We define

$$\|u\|_{\mathrm{Str}} = \sup\{\|u\|_{L^p \dot{W}^{\sigma,q}} : \tfrac{1}{q} + \tfrac{4}{p} = 2,\ \delta_0 \leq \tfrac{1}{p} \leq \tfrac{1}{2} - \delta_0,\ \tfrac{2}{p} + \tfrac{3}{q} \leq \tfrac{3}{2} - \delta_0\}, \tag{3.10}$$

as well as

$$\|u\|_{\mathrm{Str}^1} = \|\nabla u\|_{\mathrm{Str}}. \tag{3.11}$$

Conditions in (3.10) insure that the (p, q, σ)'s are Strichartz exponents, but away from the sharp endpoints. These norms have two key properties:

- They are divisible in time, i.e. can be made small by subdividing the time interval.
- Saturating the associated Strichartz inequalities requires strong pointwise concentration (i.e., small energy dispersion).

In [18], we have shown that the spatial components of the Yang–Mills equation $\mathbf{D}^\alpha F_{j\alpha} = 0$ ($j \in \{1, 2, 3, 4\}$) may be interpreted as a system of wave equation for the spatial components $A = A_x$, where the temporal component A_0 is determined in terms of $(A, \partial_t A)$, as follows:

Theorem 3.5 ([18, Theorem 9.1]). *Let $A_{t,x} = (A_0, A) \in C_t(I; \dot{H}^1 \times \mathcal{C}_\mathcal{Q})$ with $(\partial_t A_0, \partial_t A) \in C_t(I; L^2 \times T^{L^2}_{A(t)} \mathcal{C}_\mathcal{Q})$ be a solution to (1.1) with energy $\mathcal{E}$. Then its spatial components $A = A_x$ satisfy an equation of the form*

$$\Box_A A_j = \mathbf{P}_j[A, \partial_x A] + 2\Delta^{-1}\partial_j \mathbf{Q}(\partial^\alpha A, \partial_\alpha A) + R_j(A), \tag{3.12}$$

together with a compatibility condition

$$\partial^\ell A_\ell = \mathbf{DA}(A) := \mathbf{Q}(A, A) + \mathbf{DA}^3(A). \tag{3.13}$$

Moreover, the temporal component A_0 and its time derivative $\partial_t A_0$ admit the expressions

$$A_0 = \mathbf{A}_0(A) := \Delta^{-1}[A, \partial_t A] + 2\Delta^{-1}\mathbf{Q}(A, \partial_t A) + \mathbf{A}_0^3(A), \tag{3.14}$$

$$\partial_t A_0 = \mathbf{DA}_0(A) := -2\Delta^{-1}\mathbf{Q}(\partial_t A, \partial_t A) + \mathbf{DA}_0^3(A). \tag{3.15}$$

Here $\mathbf{P}$ is the Leray projector, and $\mathbf{Q}$ is a symmetric[3] *bilinear form with symbol*

$$\mathbf{Q}(\xi, \eta) = \frac{|\xi|^2 - |\eta|^2}{2(|\xi|^2 + |\eta|^2)}. \tag{3.16}$$

Moreover, $R_j(t)$, $\mathbf{DA}^3(t)$, $\mathbf{A}_0^3(t)$ and $\mathbf{DA}_0^3(t)$ are uniquely determined by $(A, \partial_t A)(t) \in T^{L^2}\mathcal{C}$, and are Lipschitz envelope preserving maps of order ≥ 3 on the following spaces:

$$R_j(t) : \dot{H}^1 \to \dot{H}^{-1}, \tag{3.17}$$

$$\mathbf{DA}^3(t) : \dot{H}^1 \to L^2, \tag{3.18}$$

$$\mathbf{A}_0^3(t) : \dot{H}^1 \to \dot{H}^1, \tag{3.19}$$

$$\mathbf{DA}_0^3(t) : \dot{H}^1 \to L^2. \tag{3.20}$$

[3] Observe here that the symbol of $\mathbf{Q}$ is odd, but this is combined with the antisymmetry of the Lie brackets appearing in the bilinear form.

Finally, on any interval $I \subseteq \mathbb{R}$, R_j, $\mathbf{DA}^3$, $\mathbf{A}_0^3$ *and* $\mathbf{DA}_0^3$ *are Lipschitz envelope preserving maps of order* ≥ 3 *(with bounds independent of* I*) on the following spaces:*

$$R_j : \mathrm{Str}^1[I] \to L^1L^2 \cap L^2\dot{H}^{-\frac{1}{2}}[I], \tag{3.21}$$

$$\mathbf{DA}^3 : \mathrm{Str}^1[I] \to L^1\dot{H}^1 \cap L^2\dot{H}^{\frac{1}{2}}[I], \tag{3.22}$$

$$\mathbf{A}_0^3 : \mathrm{Str}^1[I] \to L^1\dot{H}^2 \cap L^2\dot{H}^{\frac{3}{2}}[I], \tag{3.23}$$

$$\mathbf{DA}_0^3 : \mathrm{Str}^1[I] \to L^1\dot{H}^1 \cap L^2\dot{H}^{\frac{1}{2}}[I]. \tag{3.24}$$

All implicit constants depend on $\mathcal{Q}$ *and* $\mathcal{E}$.

Next, we consider the dynamic Yang–Mills heat flow $A_{t,x}(s)$ of $A_{t,x}$ in the caloric gauge. For $s > 0$, we have $\mathbf{D}^\beta F_{\alpha\beta}(s) = w_\alpha \neq 0$ in general. We expect the "heat-wave commutator" w_α (called the Yang–Mills tension field) to be concentrated primarily at frequency comparable to $s^{-\frac{1}{2}}$. Indeed, the following theorem holds.

Theorem 3.6 ([18, Theorem 9.3])**.** *Let* $A_{t,x} = (A_0, A) \in C_t(I; \dot{H}^1 \times \mathcal{C}_\mathcal{Q})$ *with* $(\partial_t A_0, \partial_t A) \in C_t(I; L^2 \times T^{L^2}_{A(t)}\mathcal{C}_\mathcal{Q})$ *be a solution to (1.1) with energy* $\mathcal{E}$. *Let* $A_{t,x}(s) = A_{t,x}(t,x,s)$ *be the dynamic Yang–Mills heat flow development of* $A_{t,x}$ *in the caloric gauge. Then the spatial components* $A(s) = A_x(s)$ *of* $A_{t,x}(s)$ *satisfy an equation of the form*

$$\begin{aligned}\Box_{A(s)} A_j(s) =& \mathbf{P}_j[A(s), \partial_x A(s)] + 2\Delta^{-1}\partial_j \mathbf{Q}(\partial^\alpha A(s), \partial_\alpha A(s)) + R_j(A(s)) \\ &+ \mathbf{P}_j \mathbf{w}_x^2(\partial_t A, \partial_t A, s) + R_{j;s}(A)\end{aligned} \tag{3.25}$$

together with the compatibility condition

$$\partial^\ell A_\ell(s) = \mathbf{DA}(A(s)). \tag{3.26}$$

Moreover, the temporal component $A_0(s)$ *and its time derivative* $\partial_t A_0(s)$ *admit the expansions*

$$\begin{aligned}A_0(s) =& \mathbf{A}_0(A(s)) + \mathbf{A}_{0;s}(A) \\ :=& \mathbf{A}_0(A(s)) + \Delta^{-1}\mathbf{w}_0^2(A, A, s) + \mathbf{A}_{0;s}^3(A),\end{aligned} \tag{3.27}$$

$$\partial_t A_0(s) = \mathbf{DA}_0(A(s)) + \mathbf{DA}_{0;s}(A) \tag{3.28}$$

Here $\mathbf{P}$, $\mathbf{Q}$, R_j, $\mathbf{DA}$, $\mathbf{A}_0$ *and* $\mathbf{DA}_0$ *are as before, and* $\mathbf{w}_\alpha^2$ *are defined as*

$$\mathbf{w}_0^2(A, B, s) = -2\mathbf{W}(\partial_t A, \Delta B, s), \tag{3.29}$$

$$\mathbf{w}_j^2(A, B, s) = -2\mathbf{W}(\partial_t A, \partial_j \partial_t B - 2\partial_x \partial_t B_j, s), \tag{3.30}$$

where $\mathbf{W}(\cdot, \cdot, s)$ *is a bilinear form with symbol*

$$\mathbf{W}(\xi, \eta, s) = -\frac{1}{2\xi \cdot \eta} e^{-s|\xi+\eta|^2}\left(1 - e^{2s(\xi\cdot\eta)}\right). \tag{3.31}$$

Moreover, $R_{j;s}(t)$, $\mathbf{A}_{0;s}^3(t)$ *and* $\mathbf{DA}_{0;s}(t)$ *are uniquely determined by* $(A, \partial_t A)(t) \in T^{L^2}\mathcal{C}$ *for each* $s > 0$*, and satisfy the following properties*

- $R_{j;s}(t) : \dot{H}^1 \to \dot{H}^{-1}$ *is a Lipschitz map with output concentrated at frequency* $s^{-\frac{1}{2}}$. *More precisely,*

$$(1 - s\Delta)^N R_{j;s}(t) : \dot{H}^1 \to 2^{-\delta_1 k(s)} \dot{H}^{-1-\delta_1}. \tag{3.32}$$

- $\mathbf{A}_{0;s}^3(t) : \dot{H}^1 \to \dot{H}^1$ *is a Lipschitz map with output concentrated at frequency* $s^{-\frac{1}{2}}$, *i.e.,*

$$(1 - s\Delta)^N \mathbf{A}_{0;s}^3(t) : \dot{H}^1 \to 2^{-\delta_1 k(s)} \dot{H}^{1-\delta_1} \tag{3.33}$$

- $\mathbf{DA}_{0;s}(t) : \dot H^1 \to L^2$ *is a Lipschitz map with output concentrated at frequency* $s^{-\frac{1}{2}}$*, i.e.,*
$$(1 - s\Delta)^N \mathbf{DA}_{0;s}(t) : \dot H^1 \to 2^{-\delta_1 k(s)} \dot H^{-\delta_1}. \tag{3.34}$$

Finally, on any time interval $I \subseteq \mathbb{R}$ *(with bounds independent of* I*),* $R_{j;s}$*,* $\mathbf{A}^3_{0;s}$ *and* $\mathbf{DA}_{0;s}$ *satisfy the following properties:*

- $R_{j;s} : \mathrm{Str}^1[I] \to L^1 L^2 \cap L^2 \dot H^{-\frac{1}{2}}[I]$ *is a Lipschitz map with output concentrated at frequency* $s^{-\frac{1}{2}}$*, i.e.,*
$$(1 - s\Delta)^N R_{j;s} : \mathrm{Str}^1[I] \to 2^{-\delta_1 k(s)} (L^1 \dot H^{-\delta_1} \cap L^2 \dot H^{-\frac{1}{2}-\delta_1})[I] \tag{3.35}$$
- $\mathbf{A}^3_{0;s} : \mathrm{Str}^1[I] \to L^1 \dot H^2 \cap L^2 \dot H^{\frac{3}{2}}[I]$ *is a Lipschitz map with output concentrated at frequency* $s^{-\frac{1}{2}}$*, i.e.,*
$$(1 - s\Delta)^N \mathbf{A}^3_{0;s} : \mathrm{Str}^1[I] \to 2^{-\delta_1 k(s)} (L^1 \dot H^{2-\delta_1} \cap L^2 \dot H^{\frac{3}{2}-\delta_1})[I] \tag{3.36}$$
- $\mathbf{DA}_{0;s} : \mathrm{Str}^1[I] \to L^2 \dot H^{\frac{1}{2}}[I]$ *is a Lipschitz map with output concentrated at frequency* $s^{-\frac{1}{2}}$*, i.e.,*
$$(1 - s\Delta)^N \mathbf{DA}_{0;s} : \mathrm{Str}^1[I] \to 2^{-\delta_1 k(s)} L^2 \dot H^{\frac{1}{2}-\delta_1}[I] \tag{3.37}$$

All implicit constants depend on $\mathcal{Q}$ *and* $\mathcal{E}$*.*

Remark 3.7. Some notable features of Theorem 3.6 are as follows.

- Compared with the prior result, here we have additional contributions $R_{k;s}$, $\mathbf{A}_{0;s}$ and $\mathbf{DA}_{0;s}$ as well as the $\mathbf{w}$ terms. These have the downside that they depend on A and $\partial_t A$ at $s = 0$ rather than $A(s)$ and $\partial_t A(s)$. The redeeming feature is that these terms will not only be small due to the energy dispersion, but also, critically, concentrated at frequency $s^{-\frac{1}{2}}$.
- The other change here is due to the inhomogeneous terms $\mathbf{w}^2_\alpha$; these are matched in the $A_k(s)$ and the $A_0(s)$ equations, and will interact in the trilinear analysis (see Proposition 4.29 below).
- For the new error terms here we do not need to worry about difference bounds; see Section 6 below.

4. Summary of function spaces and estimates

In this section, we summarize the properties of the function spaces and the estimates needed to analyze the hyperbolic Yang–Mills equation in the caloric gauge, as given by Theorems 3.5 and 3.6.

4.1. **Function spaces.**

The aim of this subsection is to give precise definitions of the fine functions spaces used to analyze caloric Yang–Mills waves.

4.1.1. *Frequency projections.* We start with a brief discussion of various frequency projections. Let $m_0 : \mathbb{R} \to \mathbb{R}$ be a smooth non-negative even bump function supported on $\{x \in \mathbb{R} : |x| \in (2^{-1}, 2^2)\}$ such that $\{m_k = m_0(\cdot/2^k)\}_{k \in \mathbb{Z}}$ is a partition of unity on $\mathbb{R}$. For $k \in \mathbb{Z}$, recall that P_k was defined as the multiplier on $\mathbb{R}^4$ with symbol $P_k(\xi) = m_k(|\xi|)$. Given $j \in \mathbb{Z}$ and a sign $\pm$, we introduce the modulation projections $Q^\pm_j$ and Q_j, which are multipliers on $\mathbb{R}^{1+4}$ with symbols
$$Q^\pm_j(\tau, \xi) = m_j(\tau \mp |\xi|), \quad Q_j(\tau, \xi) = m_j(|\tau| - |\xi|).$$

We also define $Q^{\pm}_{<j}$, $Q^{\pm}_{\geq j}$, $Q_{<j}$, $Q_{\geq j}$ etc. in the obvious manner. To connect $Q^{\pm}_j$ with Q_j, we introduce the sharp time-frequency cutoffs $Q^{\pm}$, which are multipliers on $\mathbb{R}^{1+4}$ with symbols

$$Q^{\pm}(\tau, \xi) = \chi_{(0,\infty)}(\pm\tau).$$

Note that $P_k Q^{\pm} Q_j = P_k Q^{\pm}_j$ for $j < k$.

For $\ell \in -\mathbb{N}$, consider a collection of directions $\omega \in \mathbb{S}^3 \subseteq \mathbb{R}^4$, which are maximally separated with distance $\simeq 2^{\ell}$. To each such an ω, we associate a smooth cutoff function m^{ω}_{ℓ} supported on a cap of radius $\simeq 2^{\ell}$ centered at ω, with the property that $\sum_{\omega} m_{\omega} = 1$. Let P^{ω}_{ℓ} be the multiplier on $\mathbb{R}^4$ with symbol

$$P^{\omega}_{\ell}(\xi) = m^{\omega}_{\ell}\left(\frac{\xi}{|\xi|}\right).$$

Given $k' \in \mathbb{Z}$ and $\ell' \in -\mathbb{N}$, consider rectangular boxes $\mathcal{C}_{k'}(\ell')$ of dimensions $2^{k'} \times (2^{k'+\ell'})^3$ (where the $2^{k'}$-side lies along the radial direction), which cover $\mathbb{R}^4 \setminus \{|x| \lesssim 2^{k'}\}$ and have finite overlap with each other. Let $m_{\mathcal{C}_{k'}(\ell')}$ b a partition of unity adapted to $\{\mathcal{C}_{k'}(\ell')\}$, and we define the multiplier $P_{\mathcal{C}_{k'}(\ell')}$ on $\mathbb{R}^4$ with symbol

$$P_{\mathcal{C}_{k'}(\ell')}(\xi) = m_{\mathcal{C}_{k'}(\ell')}(\xi).$$

For convenience, when $k' = k$, we choose the covering and the partition of unity so that $P_k P^{\omega}_{\ell} = P_k P_{\mathcal{C}_k(\ell)}$.

We now discuss the boundedness properties of the frequency projections. For any $k \in \mathbb{Z}$, let $P_{k/<k}$ denote one of the dyadic frequency projections $\{P_k, P_{<k}\}$. Let $Q^{\square}_{j/<j}$ denote one of the modulation projections $Q^{\pm}_j$, $Q^{\pm}_{<j}$, Q_j or $Q_{<j}$. Let ω be an angular sector of size $\simeq 2^{\ell}$ ($\ell \in -\mathbb{N}$), and $\mathcal{C}$ a rectangular box of the form $\mathcal{C}_{k'}(\ell')$ ($k' \in \mathbb{Z}$, $\ell' \in -\mathbb{N}$). Then the following statements hold:

- The multipliers $P_{k/<k}$, $P_{k/<k} P^{\omega}_{\ell}$ and $P_{\mathcal{C}}$ are disposable.
- The multiplier $P_{k/<k} Q^{\square}_{j/<j}$ is disposable if $j \geq k + O(1)$; see [25, Lemma 3]. For general $j, k \in \mathbb{Z}$, it is straightforward to check that $P_{k/<k} Q^{\square}_{j/<j}$ has a kernel with mass $O(2^{4(k-j)_+})$.
- The multiplier $P_{k/<k} Q^{\square}_{j/<j}$ is bounded on $L^p L^2$ for any $1 \leq p \leq \infty$; see [25, Lemma 4].
- The multiplier $P_{k/<k} P^{\omega}_{\ell} Q^{\square}_{j/<j}$ is disposable if $j \geq k + 2\ell + O(1)$; see [25, Lemma 6].

4.1.2. *Function spaces on the whole space-time.* Here, we define the global-in-time function spaces used in this work. Unless otherwise stated, all spaces below are defined for functions on $\mathbb{R}^{1+4}$. We remark that all of them are translation-invariant.

We first define the space $X^{\sigma,b}_r$, equipped with the norm

$$\|u\|^2_{X^{\sigma,b}_r} = \sum_k 2^{2\sigma k} \Big(\sum_j (2^{bj} \|P_k Q_j u\|_{L^2 L^2})^r \Big)^{\frac{2}{r}}$$

when $1 \leq r < \infty$. As usual, we replace the ℓ^r-sum by the supremum in j when $r = \infty$. The spaces $X^{\sigma,b}_{\pm,r}$ are defined similarly, with Q_j replaced by $Q^{\pm}_j$.

We are now ready to introduce the function spaces in earnest, which are all defined in terms of (semi-)norms.

Core nonlinearity norm N. We define

$$N = L^1 L^2 + X^{0,-\frac{1}{2}}_1.$$

This norm scales like L^1L^2. We also define $N_\pm = L^1L^2 + X^{0,-\frac{1}{2}}_{\pm,1}$. Note that $N = N_+ \cap N_-$. Moreover, we have the embeddings

$$X^{0,-\frac{1}{2}}_1 \subseteq N \subseteq X^{0,-\frac{1}{2}}_\infty, \quad X^{0,-\frac{1}{2}}_{\pm,1} \subseteq N \subseteq X^{0,-\frac{1}{2}}_{\pm,\infty}.$$

The inclusions on the left are obvious, whereas the inclusions on the right follow from Bernstein in time. We omit the proofs.

Core solution norm S. We define

$$\|u\|_S^2 = \sum_k \|P_k u\|_{S_k}^2, \quad S_k = S_k^{str} \cap X^{0,\frac{1}{2}}_\infty \cap S_k^{ang} \cap S_k^{sq},$$

where S_k^{sq} is related to square function bounds,

$$\|u\|_{S_k^{sq}} = 2^{-\frac{3}{10}k} \|u\|_{L_x^{\frac{10}{3}} L_t^2}$$

and S_k^{str} and S_k^{ang} are essentially as in [10, Eqs. (6)–(8)]:

$$\begin{aligned}
\|u\|_{S_k^{str}} =& \sup_{(p,q): \frac{1}{p} + \frac{3}{2q} \leq \frac{3}{4}} 2^{-(2-\frac{1}{p}-\frac{4}{q})k} \|u\|_{L^pL^q}, \\
\|u\|_{S_k^{ang}}^2 =& \sup_{\ell<0} \sum_\omega \|P_\ell^\omega Q_{<k+2\ell} u\|_{S_k^\omega(\ell)}^2, \\
\|u\|_{S_k^\omega(\ell)}^2 =& \|u\|_{S_k^{str}}^2 + 2^{-2k}\|u\|_{NE}^2 + 2^{-3k} \sum_\pm \|Q^\pm u\|_{PW_\omega^\mp(\ell)}^2 \\
&+ \sup_{\substack{k' \leq k,\ \ell' \leq 0 \\ k+2\ell \leq k'+\ell' \leq k+\ell}} \sum_{\mathcal{C}_{k'}(\ell')} \Big(\|P_{\mathcal{C}_{k'}(\ell')} u\|_{S_k^{str}}^2 + 2^{-2k} \|P_{\mathcal{C}_{k'}(\ell')} u\|_{NE}^2 \\
&+ 2^{-2k'-k} 2^{-\ell'} \|P_{\mathcal{C}_{k'}(\ell')} u\|_{L^2L^\infty}^2 + 2^{-3(k'+\ell')} \sum_\pm \|Q^\pm P_{\mathcal{C}_{k'}(\ell')} u\|_{PW_\omega^\mp(\ell)}^2 \Big).
\end{aligned}$$

Here, the NE and $PW_\omega^\mp(\ell)$ are the *null frame spaces* [32, 25], defined by

$$\begin{aligned}
\|u\|_{PW_\omega^\mp(\ell)} =& \inf_{u = \int u^{\omega'}} \int_{|\omega-\omega'| \leq 2^\ell} \|u^{\omega'}\|_{L^2_{\pm\omega'} L^\infty_{(\pm\omega')^\perp}} \, d\omega', \\
\|u\|_{NE} =& \sup_\omega \|\not\nabla_\omega u\|_{L_\omega^\infty L_{\omega^\perp}^2},
\end{aligned}$$

where the L^q_ω norm is with respect to the variable $t_\omega^\pm = t \pm \omega \cdot x$, the $L^r_{\omega^\perp}$ norm is defined on each $\{t_\omega^\pm = const\}$, and $\not\nabla_\omega$ denotes the tangential derivatives to $\{t_\omega^\pm = const\}$.

In the last two lines of the definition of $S_k^\omega(\ell)$, the restrictions $k' \leq k$, $\ell' \leq 0$ and $k' + \ell' \leq k + \ell$ ensure that rectangular boxes of the form $\mathcal{C}_{k'}(\ell')$ fit in the frequency support of P_ℓ^ω. The restriction $k + 2\ell \leq k' + \ell'$ is imposed by the main parametrix estimate (see Section 10.8 or [10, Section 11]), to ensure square-summability in $\mathcal{C}_{k'}(\ell')$.

The null frame spaces in $S_k^\omega(\ell)$ allow one to exploit transversality in frequency space, and play an important role in the proof of the trilinear null form estimate; see [10, Eqs. (136)–(138)] and Proposition 8.18 below. On the other hand, the L^2L^∞-norm for $P_{\mathcal{C}_{k'}(\ell')}u$ allows us to gain the dimensions of $\mathcal{C}_{k'}(\ell')$.

Remark 4.1. For the reader who is familiar with the function space framework in [10], we point out that our $S_k^\omega(\ell)$ is slightly stronger compared to that in [10]. More precisely, instead of $2^{-k'-\frac{1}{2}k}2^{-\frac{1}{2}\ell'}\|P_{\mathcal{C}_{k'}(\ell')}u\|_{L^2L^\infty}$ as in our definition, it is $2^{-k'-\frac{1}{2}k}\|P_{\mathcal{C}_{k'}(\ell')}u\|_{L^2L^\infty}$ in [10]. However, we note that the extra factor $2^{-\frac{1}{2}\ell'}$ is actually present in the main parametrix estimate in [10, Subsection 11.3].

Remark 4.2. The square function norm S_k^{sq} is new here in the structure of the S norm. It plays no role in the study of the solutions for the hyperbolic Yang–Mills equation in the caloric gauge, i.e. in Theorems 1.13 and 1.16. Instead, it is only needed in order to justify the transition to the temporal gauge in Theorem 1.18.

This norm scales like $L^\infty L^2$. Moreover, it obeys the embeddings

$$P_k X_1^{0,\frac{1}{2}} \subseteq S_k, \quad S_k \subseteq X_\infty^{0,\frac{1}{2}}.$$

Indeed, the latter embedding is trivial. The former embedding has essentially been proved in [32, 25]; we sketch its proof as follows. It suffices to show that any $u = P_k Q_j u$ satisfies $\|u\|_{S_k} \lesssim 2^{\frac{j}{2}}\|u\|_{L^2L^2}$. We claim that

$$\|P_\ell^\omega u\|_{S_k} \lesssim 2^{\frac{j}{2}}\|P_\ell^\omega u\|_{L^2L^2} \quad \text{for } \ell = \tfrac{j-k}{2} + O(1) \text{ and } 2^\ell\text{-separated } \omega\text{'s on } \mathbb{S}^3.$$

Recalling that $S_k = S_k^{str} \cap X_\infty^{0,\frac{1}{2}} \cap S_k^{ang} \cap S_k^{sq}$, the desired conclusion would follow from the claim after square-summing in ω.

Note that $P_\ell^\omega P_k Q_j$ with the above value of ℓ is multiplication on the Fourier side by a bump function adapted to a parallelepiped of dimensions $2^j \times 2^k \times (2^{k+\ell})^3$, where the 2^j- and 2^k-sides lie along the τ- and the radial (in ξ) directions, respectively. The claim is straightforward for $S_k^{str} \cap X_\infty^{0,\frac{1}{2}} \cap S_k^{sq}$ by appropriate versions of Bernstein's inequality. For S_k^{ang}, by orthogonality, we need to show that

$$\begin{aligned}
&\|P_\ell^\omega P_{\mathcal{C}_{k'}(\ell')}u\|_{S_k^{str}} + 2^{-k}\|P_\ell^\omega P_{\mathcal{C}_{k'}(\ell')}u\|_{NE} \\
&+ 2^{-k'-\frac{1}{2}k-\frac{1}{2}\ell'}\|P_\ell^\omega P_{\mathcal{C}_{k'}(\ell')}u\|_{L^2L^\infty} + 2^{-\frac{3}{2}(k'+\ell')}\|P_\ell^\omega P_{\mathcal{C}_{k'}(\ell')}u\|_{PW_\omega^\mp(\ell)} \lesssim 2^{\frac{j}{2}}\|P_\ell^\omega P_{\mathcal{C}_{k'}(\ell')}u\|_{L^2L^2}.
\end{aligned}$$

for each $\mathcal{C}_{k'}(\ell')$ arising in $S_k^\omega(\ell)$. We remark that the parts of S_k^{ang} that do not involve $\mathcal{C}_{k'}(\ell')$ are handled in a similar but simpler manner. The S_k^{str}, NE and L^2L^∞ norms are handled via Bernstein's inequality as before, where we note that $P_{\mathcal{C}_{k'}(\ell')}P_k Q_j$ is multiplication on the Fourier side by a bump function adapted to a parallelepiped of dimensions $2^j \times 2^{k'} \times (2^{k'+\ell'})^3$ with the same orientation as before. For the $PW_\omega^\mp(\ell)$ norm, we decompose $Q^\pm P_\ell^\omega P_{\mathcal{C}_{k'}(\ell')}u = \int u^{\pm,\omega'}\,d\omega'$ with

$$u^{\pm,\omega'} = (2\pi)^{-5}\int_{|a|=O(2^j)} e^{iat}\int_0^\infty \mathcal{F}(P_\ell^\omega P_{\mathcal{C}_{k'}(\ell')}u)(\lambda + a, \lambda\omega')e^{i\lambda(\pm t - \omega'\cdot x)}\lambda^3\,d\lambda da.$$

Indeed, this decomposition is nothing but the Fourier inversion formula written in polar coordinates. Note that, thanks to the projections P_ℓ^ω and $P_{\mathcal{C}_{k'}(\ell')}$, $u^{\omega'}$ is zero for ω' outside either $\{\omega' : |\omega' - \omega| \leq 2^{\ell}\}$ or an angular sector of radius $O(2^{k'+\ell'-k})$. Therefore, by Cauchy–Schwarz and the Fourier inversion formula (in λ and in τ, ξ), we have

$$\begin{aligned}
&\| P^{\omega}_{\ell} P_{\mathcal{C}_{k'}(\ell')} u\|_{PW^{\mp}_{\omega}(\ell)} \\
&\lesssim \int \| \int_{|a|=O(2^j)} e^{iat} \int_0^\infty \mathcal{F}(P^{\omega}_{\ell} P_{\mathcal{C}_{k'}(\ell')} u)(\lambda + a, \lambda\omega') e^{i\lambda(\pm t - \omega'\cdot x)} \lambda^3 \, d\lambda da \|_{L^2_{\mp\omega'} L^\infty_{(\mp\omega')^\perp}} \, d\omega' \\
&\lesssim 2^{\frac{3}{2}(k'+\ell'-k)} 2^{\frac{j}{2}} \| \int_0^\infty \mathcal{F}(P^{\omega}_{\ell} P_{\mathcal{C}_{k'}(\ell')} u)(\lambda + a, \lambda\omega') e^{\pm i\lambda(t \mp \omega'\cdot x)} \lambda^3 \, d\lambda \|_{L^2(d(t\mp\omega'\cdot x) da d\omega')} \\
&\lesssim 2^{\frac{3}{2}(k'+\ell')} 2^{\frac{j}{2}} \| \mathcal{F}(P^{\omega}_{\ell} P_{\mathcal{C}_{k'}(\ell')} u)(\lambda + a, \lambda\omega') \|_{L^2(\lambda^3 d\lambda da d\omega')} = 2^{\frac{3}{2}(k'+\ell')} 2^{\frac{j}{2}} \| P^{\omega}_{\ell} P_{\mathcal{C}_{k'}(\ell')} u\|_{L^2 L^2},
\end{aligned}$$

as desired.

For $k, k' \in \mathbb{Z}$ satisfying $k' \leq k$ and $\ell' < -5$, we define

$$\begin{aligned}
\|u\|^2_{S_k[\mathcal{C}_{k'}(\ell')]} =& 2^{-\frac{5}{3}k} \|u\|^2_{L^2 L^6} + 2^{-2k'-k} 2^{-\ell'} \|u\|^2_{L^2 L^\infty} \\
&+ \sup_{j:\, |j-(k'+2\ell')| \leq 5} \Big(\|Q_{<j} u\|^2_{L^\infty L^2} + 2^{-2k} \|Q_{<j} u\|^2_{NE} \\
&\qquad + 2^{-3(k'+\ell')} \sum_{\omega} \sum_{\pm} \|P^{\omega}_{\frac{j-k}{2}} Q^{\pm}_{<j} u\|^2_{PW^{\mp}_{\omega}(\frac{j-k}{2})} \Big),
\end{aligned}$$

where the ω-summation runs over the $O(2^{\frac{j-k}{2}})$-separated subset of $\mathbb{S}^3$ associated with the projections $P^{\omega}_{\frac{j-k}{2}}$. We note that $S_k[\mathcal{C}_{k'}(\ell')]$ depends only on the parameters k, k', ℓ' (in particular, no particular choice of a rectangular box is involved), and the notation $\mathcal{C}_{k'}(\ell')$ is meant to suggest that it will be measured for $P_{\mathcal{C}} u$ with $\mathcal{C}$ a rectangular box of the form $\mathcal{C}_{k'}(\ell')$. The virtue of this norm is that it is square-summable in boxes of the form $\mathcal{C}_{k'}(\ell')$:

Lemma 4.3. *For any k, k', ℓ' such that $k' \leq k$ and $\ell' \leq 0$, we have*

$$\sum_{\mathcal{C} \in \{\mathcal{C}_{k'}(\ell')\}} \|P_{\mathcal{C}} u\|^2_{S_k[\mathcal{C}_{k'}(\ell')]} \lesssim \|u\|^2_{S_k}. \tag{4.1}$$

Proof. The desired square-summability estimate for the $L^\infty L^2$, NE and $PW^{\mp}_{\omega}$ components follow immediately from the definition of $S^{ang}_k \supseteq S_k$. For the $L^2 L^6$ and $L^2 L^\infty$ components, we split

$$u = Q_{<k'+2\ell'} u + Q_{\geq k'+2\ell'} u.$$

For the former we use S^{ang}_k, and for the latter we simply note that, by Bernstein,

$$2^{-\frac{5}{6}k} \|Q_{\geq k'+2\ell'} P_{\mathcal{C}_{k'}(\ell')} u\|_{L^2 L^6} + 2^{-k'-\frac{1}{2}k} 2^{-\frac{1}{2}\ell'} \|Q_{\geq k'+2\ell'} P_{\mathcal{C}_{k'}(\ell')} u\|_{L^2 L^\infty} \lesssim \|P_{\mathcal{C}_{k'}(\ell')} u\|_{X^{0,\frac{1}{2}}_\infty},$$

which is clearly square-summable. □

Sharp solution norm $S^\sharp$. We define

$$\begin{aligned}
\|u\|_{S^\sharp_k} =& 2^{-k} (\|\nabla u\|_{L^\infty L^2} + \|\Box u\|_N), \\
\|u\|_{(S^\sharp_\pm)_k} =& \|u\|_{L^\infty L^2} + \|(D_t \mp |D|) u\|_{N_\pm}.
\end{aligned}$$

both of which scale like $L^\infty L^2$. These norms are used in the parametrix construction in Section 9.

Remark 4.4. Again for the reader familiar with [10], we note that our definition of $S_k^\sharp$ differs from that in [10] by a factor of 2^k (in [10], $S_k^\sharp$ scales like $L^\infty \dot{H}^1$).

Scattering (or controlling) norm S^1. Given any $\sigma \in \mathbb{R}$, we define $S^\sigma = \ell^2 S^\sigma$, i.e.,

$$\|u\|_{S^\sigma}^2 = \sum_k \|P_k u\|_{S_k^\sigma}^2, \qquad \|u\|_{S_k^\sigma} = 2^{(\sigma-1)k} \left(\|\nabla u\|_S + \|\Box u\|_{L^2 \dot{H}^{-\frac{1}{2}}} \right). \tag{4.2}$$

This norm scales like $L^\infty \dot{H}^\sigma$. The norm S^1 will be the main scattering (or controlling) norm, in the sense that finiteness of this norm for a caloric Yang–Mills wave would imply finer properties of the solution itself and those nearby (see Theorem 5.1 below).

$X_r^{\sigma,b,p}$-type norms. To close the estimates for caloric Yang–Mills waves, we need norms which give additional control[4] off the characteristic cone (i.e., "high" modulation regime). We use an $L^p L^{p'}$ generalization of the usual $L^2 L^2$-based $X^{\sigma,b}$-norm, defined as follows: For $\sigma, b \in \mathbb{R}$, $1 \leq p, r < \infty$, let

$$\|u\|_{(X_r^{\sigma,b,p})_k} = 2^{\sigma k} \left(\sum_j \left(2^{bj} \Big(\sum_\omega \|P_k Q_j P_{\frac{j-k}{2}}^\omega u\|_{L^p L^{p'}}^2 \Big)^{\frac{1}{2}} \right)^r \right)^{\frac{1}{r}}, \tag{4.3}$$

where $p' = \frac{p}{p-1}$ is the dual Lebesgue exponent of p. The cases $p = \infty$ or $r = \infty$ are defined in the obvious manner. We also define the dyadic norm $(X_{\pm,r}^{\sigma,b,p})_k$ by replacing Q_j by $Q_j^\pm$ in the above definition.

When $p = 2$, by orthogonality we have

$$\|u\|_{(X_r^{\sigma,b,2})_k} = 2^{\sigma k} \Big(\sum_j \big(2^{bj} \|P_k Q_j u\|_{L^2 L^2} \big)^r \Big)^{\frac{1}{r}}.$$

Analogous identities hold for $X_{\pm,r}^{\sigma,b,2}$. To be consistent with the usual notation, we will often omit the exponents p and r when they are equal to 2, i.e., $X_r^{\sigma,b} = X_r^{\sigma,b,2}$, $X^{\sigma,b} = X_2^{\sigma,b,2}$, $X_{\pm,r}^{\sigma,b} = X_{\pm,r}^{\sigma,b,2}$ and $X_\pm^{\sigma,b} = X_{\pm,2}^{\sigma,b,2}$.

Before we introduce the specific norms we use, for logical clarity, we first fix the parameters that will be used. We introduce b_0, b_1 and p_0, which are smaller than but close to $\frac{1}{4}$, $\frac{1}{2}$ and ∞, respectively. More precisely, we fix

$$b_0 = \frac{1}{4} - \delta_0, \qquad b_1 = \frac{1}{2} - 10\delta_0, \qquad 1 - \frac{1}{p_0} = 5\delta_0,$$

so that

$$0 < \frac{1}{4} - b_0 < \frac{1}{48}, \quad 2\left(\frac{1}{4} - b_0\right) < 1 - \frac{1}{p_0} < \frac{1}{24}, \tag{4.4}$$

$$\frac{1}{4} < b_1 < \frac{1}{2} - \left(1 - \frac{1}{p_0}\right). \tag{4.5}$$

[4] In particular, with ℓ^1-summability in dyadic frequencies.

We define

$$\|f\|_{\Box Z^1_k} = \|Q_{<k+C} f\|_{X_1^{-\frac{5}{4}-b_0, -\frac{3}{4}+b_0, 1}},$$
$$\|u\|_{Z^1_k} = \|\Box u\|_{\Box Z^1_k} = \|Q_{<k+C} u\|_{X_1^{-\frac{1}{4}-b_0, \frac{1}{4}+b_0, 1}}.$$

Note that the Z^1_k-norm scales like $L^\infty \dot{H}^1$. As in [10, 11], this norm is used as an auxiliary device to control the bulk of nonlinearities (i.e., the part where the secondary null structure is *not* necessary) when re-iterating the Yang–Mills equations; see the proofs of Propositions 4.23–4.29 in Section 8.

Remark 4.5. The Z^1-norm used in [10] corresponds to the case $b_0 = 0$. Therefore, our Z^1-norm is *weaker* than the Z^1-norm in [10]. This modification is made to handle the contribution of $\Box^{-1}\mathbf{P}[A^\alpha, \partial_\alpha A]$ in the re-iteration procedure; see Proposition 4.22.

Next, we also define

$$\|f\|_{(\Box Z^1_{p_0})_k} = \|Q_{<k+C} f\|_{X_\infty^{\frac{3}{2}-\frac{3}{p_0}+(\frac{1}{4}-b_0)\theta_0, -\frac{1}{2}-(\frac{1}{4}-b_0)\theta_0, p_0}}$$

where $\theta_0 = 2(\frac{1}{p_0} - \frac{1}{2})$, as well as the intermediate norm

$$\|f\|_{(\Box \tilde{Z}^1_{p_0})_k} = \|Q_{<k+C} f\|_{X_1^{\frac{5}{4}-\frac{3}{p_0}+(\frac{1}{4}-b_0)\theta_0, -\frac{1}{4}-(\frac{1}{4}-b_0)\theta_0, p_0}}.$$

These norms scale like $L^1 L^2$. Clearly, $(\Box Z^1_{p_0})_k \subseteq (\Box \tilde{Z}^1_{p_0})_k$. Given any caloric Yang–Mills wave A with a finite S^1-norm, we will put $\Box \mathbf{P} A$ in $\ell^1 \Box \tilde{Z}^1_{p_0}$ and $\Box \mathbf{P} A \in \ell^1 \Box Z^1_{p_0}$; see Proposition 5.4.

Note that the following embeddings hold:

$$P_k Q_j L^1 L^2 \subseteq 2^{\frac{1}{4}(j-k)} \Box Z^1_k, \tag{4.6}$$
$$X_\infty^{0,-\frac{1}{2}} \cap \Box Z^1_k \subseteq (\Box Z^1_{p_0})_k \subseteq (\Box \tilde{Z}^1_{p_0})_k. \tag{4.7}$$

Estimate (4.6) follows from Bernstein, whereas the first embedding in (4.7) follows by a simple interpolation argument. We omit the straightforward proofs.

Finally, as in [11], we also need to use the function space

$$\ell^1 X^{-\frac{1}{2}+b_1, -b_1},$$

which also scales like $L^1 L^2$. Given any caloric Yang–Mills wave A with a finite S^1-norm, we will be able to place $\Box \mathbf{P} A$ in $\ell^1 X^{-\frac{1}{2}+b_1, -b_1}$. This bound, in turn, is used crucially in the parametrix construction.

High modulation norms $\underline{X}^1$ and $\tilde{X}^1$ for 1-forms. In our analysis below, we need to use different high modulation norms for the Leray projection $\mathbf{P} A$ than for the general components of a caloric Yang–Mills wave. Hence it is convenient to define norms for 1-forms with this distinction built in.

Let A and G be spatial 1-forms on $\mathbb{R}^{1+4}$. We define

$$\|G\|_{\Box \underline{X}^1_k} = \|G\|_{L^2 \dot{H}^{-\frac{1}{2}}} + \|G\|_{L^{\frac{9}{5}} \dot{H}^{-\frac{4}{9}}} + \|\mathbf{P} G\|_{(\Box Z^1_{p_0})_k}.$$

For any $\sigma \in \mathbb{R}$, we define

$$\|G\|_{\Box \underline{X}^\sigma_k} = 2^{(\sigma-1)k} \|G\|_{\Box \underline{X}^1_k}, \qquad \|A\|_{\underline{X}^\sigma_k} = \|\Box A\|_{\Box \underline{X}^\sigma_k}.$$

Similarly, we define

$$\|G\|_{\Box \tilde{X}^1_k} = \|G\|_{L^2 \dot{H}^{-\frac{1}{2}}} + \|G\|_{L^{\frac{9}{5}} \dot{H}^{-\frac{4}{9}}} + \|\mathbf{P}G\|_{(\Box \tilde{Z}^1_{p_0})_k},$$

as well as $\Box \tilde{X}^\sigma_k$ and $\tilde{X}^\sigma_k$. Given any caloric Yang–Mills wave A with a finite S^1-norm, we will place $\Box A$ successively in $\ell^1 \Box \tilde{X}^1$ and $\Box A \in \ell^1 \Box \underline{X}^1$; see Proposition 5.4.

We have the embeddings

$$P_k (L^1 L^2 \cap L^2 \dot{H}^{-\frac{1}{2}}) \subseteq (\Box \underline{X}^1)_k \subseteq (\Box \tilde{X}^1)_k.$$

Since $L^1 L^2 \subseteq N$, it follows that

$$\|G\|_{N \cap \Box \underline{X}^1} \lesssim \|G\|_{L^1 L^2 \cap L^2 \dot{H}^{-\frac{1}{2}}}. \tag{4.8}$$

Strengthened solution norm $\underline{S}^1$. Putting together S^1 and $\underline{X}^1$, for a 1-form A on $\mathbb{R}^{1+4}$, we define

$$\|A\|_{\underline{S}^\sigma_k} = \|A\|_{S^\sigma_k} + \|\Box A\|_{\Box \underline{X}^\sigma_k}.$$

Core elliptic norm Y. We return to functions u on $\mathbb{R}^{1+4}$. We define

$$\|u\|_{Y_k} = \|u\|_{L^2 \dot{H}^{\frac{1}{2}}} + \|u\|_{L^{p_0} \dot{W}^{2-\frac{3}{p_0}, p_0'}},$$

where p_0 was fixed in (4.4) above. This norm scales like $L^\infty L^2$.

Main elliptic norm Y^1. For $\sigma \in \mathbb{R}$, we define

$$\|u\|^2_{Y^\sigma} = \sum_k \|P_k u\|^2_{Y^\sigma_k}, \quad \|u\|_{Y^\sigma_k} = 2^{\sigma k} \left(\|u\|_{Y_k} + 2^{-k} \|\partial_t u\|_{L^2 \dot{H}^{\frac{1}{2}}} \right).$$

This norm scales like $L^\infty \dot{H}^\sigma$. We will put the elliptic components A_0 and $\mathbf{P}^\perp A = \Delta^{-1} \partial_x \partial^\ell A_\ell$ of a caloric Yang–Mills wave in Y^1.

4.1.3. *Interval localization and extension.* So far, the function spaces have been defined over the whole space-time $\mathbb{R}^{1+4}$. In our analysis, we also need to consider localization of these spaces on finite time intervals. We use the same set-up as [17, 11].

For most of our function spaces (with the important exceptions of $Z^1_{p_0}$, $\tilde{Z}^1_{p_0}$, $\underline{X}^1$ and $\tilde{X}^1$; see below), we take a simple route and define the interval-localized counterparts by restriction. In particular, given a time interval $I \subseteq \mathbb{R}$, we define

$$\|u\|_{S^\sigma[I]} = \inf_{\tilde{u} \in S^\sigma : u = \tilde{u}\restriction_I} \|\tilde{u}\|_{S^\sigma}, \quad \|u\|_{S[I]} = \inf_{\tilde{u} \in S : u = \tilde{u}\restriction_I} \|\tilde{u}\|_S, \quad \|f\|_{N[I]} = \inf_{\tilde{f} \in N : f = \tilde{f}\restriction_I} \|\tilde{f}\|_N, \tag{4.9}$$

An important technical question then is that of finding a common extension procedure outside I which preserve these norms. The following proposition provides an answer.

Proposition 4.6. *Let I be a time interval.*

(1) Let χ_I be the characteristic function of I. Then we have the bounds

$$\|\chi_I u\|_S \lesssim \|u\|_S, \quad \|\chi_I f\|_N \lesssim \|f\|_N. \tag{4.10}$$

For a fixed function f on $\mathbb{R}^{1+4}$, the norms $\|\chi_I f\|_N$ and $\|f\|_{N[I]}$ are also continuous as a function of the endpoints of I. We also have the linear estimates

$$\|\nabla u\|_{S[I]} \lesssim \|\nabla u(0)\|_{L^2} + \|\Box u\|_{N[I]}, \tag{4.11}$$

$$\|u\|_{S^1[I]} \lesssim \|\nabla u(0)\|_{L^2} + \|\Box u\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]}. \tag{4.12}$$

(2) *Consider any partition $I = \cup_k I_k$. Then the N and $L^2\dot{H}^{-\frac{1}{2}}$ are interval square divisible, i.e.,*

$$\sum_k \|f\|_{N[I_k]}^2 \lesssim \|f\|_{N[I]}^2, \quad \sum_k \|f\|^2_{L^2\dot{H}^{-\frac{1}{2}}[I_k]} \lesssim \|f\|^2_{L^2\dot{H}^{-\frac{1}{2}}[I]}, \tag{4.13}$$

and the S and S^1 are interval square summable, i.e.,

$$\|u\|_{S[I]}^2 \lesssim \sum_k \|u\|_{S[I_k]}^2, \quad \|u\|_{S^1[I]}^2 \lesssim \sum_k \|u\|_{S^1[I_k]}^2. \tag{4.14}$$

For a proof, we refer to [17, Proposition 3.3].

Remark 4.7. As a consequence of part (1), up to equivalent norms, we can replace the arbitrary extension in (4.9) by the zero extension in the case of S and N, and by the homogeneous waves with $(\phi, \partial_t \phi)$ at each endpoint as data outside I in the case of S^1.

The elliptic norms Y and Y^1 only involve spatial multipliers and norms of the form $L^p L^q$, so their interval-localization $Y[I]$ and $Y^1[I]$ are obviously defined (either by restriction, or using the $L^p L^q[I]$-norm; both are equivalent). In particular, in the case of Y, observe that

$$\|u\|_{Y[I]} = \|\chi_I u\|_Y \leq \|u\|_Y,$$

so the zero extension can be used.

On the other hand, given a function u on I, we *directly* define the $\|u\|_{(Z^1_{p_0})_k[I]}$ [resp. $\|u\|_{(\tilde{Z}_{p_0})^1_k[I]}$] to be $\|u^{ext}\|_{(Z^1_{p_0})_k[I]}$ [resp. $\|u^{ext}\|_{(\tilde{Z}_{p_0})^1_k[I]}$, where u^{ext} is the extension of u outside I by homogeneous waves. Equivalently, for $(\Box Z^1_{p_0})_k$ and $(\Box \tilde{Z}^1_{p_0})_k$, we define

$$\|f\|_{(\Box Z^1_{p_0})_k[I]} = \|\chi_I f\|_{(\Box Z^1_{p_0})_k}, \quad \|f\|_{(\Box \tilde{Z}^1_{p_0})_k[I]} = \|\chi_I f\|_{(\Box Z^1_{p_0})_k}.$$

Accordingly, we define

$$\|G\|_{\Box \underline{X}^1_k[I]} = \|G\|_{L^2\dot{H}^{-\frac{1}{2}}[I]} + \|G\|_{L^{\frac{9}{5}}\dot{H}^{-\frac{4}{9}}[I]} + \|\chi_I \mathbf{P} G\|_{(\Box Z^1_{p_0})_k}, \quad \|A\|_{\underline{X}^1_k[I]} = \|\Box A\|_{\Box \underline{X}^1_k[I]},$$

and similarly for $\Box \tilde{X}^1[I]$ and $\tilde{X}^1[I]$.

The advantage of this definition is clear: We may thus use a common extension procedure (namely, by homogeneous waves) for S^1 and $\underline{X}^1$. The price we pay is that in estimating the $\Box Z^1_{p_0}$- and the $\Box \tilde{Z}^1_{p_0}$-norms, we need to carefully absorb the sharp time cutoff χ_I.

4.1.4. *Sources of smallness: Divisibility, energy dispersion and short time interval.* In this work, we rely on several sources of smallness for analysis of caloric Yang–Mills waves.

One important source of smallness is divisibility, which refers to the property of a norm on an interval that it can be made arbitrarily small by splitting the interval into a controlled number of pieces. Unfortunately, our main function space $S^1[I]$ is far from satisfying such a property (see, however, Theorem 5.1.(6) below), which causes considerable difficulty. Our workaround, as in [17], is to utilize a weaker yet divisible norm

$$\|u\|_{DS^1[I]} = \||D|^{-\frac{5}{6}} \nabla u\|_{L^2 L^6[I]} + \|\nabla u\|_{\mathrm{Str}^0[I]} + \|\Box u\|_{L^2\dot{H}^{-\frac{1}{2}}[I]}. \tag{4.15}$$

Another important source of smallness is *energy dispersion*:

Definition 4.8. Given any $m \in \mathbb{Z}$, we define the *energy dispersion below scale* 2^{-m} (or above frequency 2^m) of u of order 0 and 1 to be, respectively,

$$\|u\|_{ED_{\geq m}[I]} := \sup_{k \in \mathbb{Z}} 2^{-\delta_2(m-k)_+} 2^{-2k} \|P_k u\|_{L^\infty L^\infty[I]}, \tag{4.16}$$

and

$$\|u\|_{ED^1_{\geq m}[I]} := \sup_{k \in \mathbb{Z}} 2^{-\delta_2(m-k)_+} 2^{-2k} \|\nabla P_k u\|_{L^\infty L^\infty[I]}. \tag{4.17}$$

The quantity $\| \cdot \|_{ED_{\geq m}[I]}$ (resp. $\| \cdot \|_{ED^1_{\geq m}[I]}$) is used at the level of the curvature F (resp. the connection A). As we work mostly at the level of the connection, unless stated otherwise, by energy dispersion we usually refer to the order 1 case.

Clearly, $ED^1_{\geq m}[I]$ fails to be useful at frequencies below $O(2^m)$. In this regime, we exploit instead the *length* $|I|$ of the time interval as a source of smallness. Due to the scaling property of $\Box$, we must require $2^m |I|$ to be sufficiently small. To conveniently pack together the previous two concepts, we introduce the notion of an (ε, M)*-energy dispersed function on an interval.*

Definition 4.9 ((ε, M)-energy dispersed function on an interval)**.** Let I be a time interval, and let $u \in S^1[I]$. For $\varepsilon > 0$ and $M > 0$, we will say that the pair (u, I) is (ε, M)*-energy dispersed* if there exists some $m \in \mathbb{Z}$ such that the following properties hold:

- (S^1-norm bound)
$$\|u\|_{S^1[I]} \leq M; \tag{4.18}$$
- (small energy dispersion)
$$\|u\|_{ED^1_{\geq m}[I]} \leq \varepsilon M; \tag{4.19}$$
- (high modulation bound)
$$\|\Box u\|_{L^2 \dot{H}^{-\frac{1}{2}}[I]} \leq \varepsilon M; \tag{4.20}$$
- (short time interval) $|I| \leq \varepsilon 2^{-m}$.

Observe (by interpolation) that if (u, I) is (ε, M)-energy dispersed, then

$$\sup_k \|P_k u\|_{\mathrm{Str}^1[I]} \leq C \varepsilon^{\delta_1} M. \tag{4.21}$$

Finally, we state a proposition showing how the norms $DS^1[I]$ and $ED^1_{\geq m}[I]$ behave under the extension procedure described above. Given an interval I, we denote by χ_I^k a generalized cutoff function adapted to the scale 2^{-k}:

$$\chi_I^k(t) = (1 + 2^k \mathrm{dist}(t, I))^{-N}, \tag{4.22}$$

where N is a sufficiently large number. Let us recall [17, Proposition 3.4][5]:

Proposition 4.10. *Let $k \in \mathbb{Z}$, $\kappa \geq 0$ and I be a time interval such that $|I| \geq 2^{-k-\kappa}$. Consider a function u_I on I localized at frequency 2^k, and denote by u_I^{ext} its extension outside I as homogeneous waves. Then we have*

$$2^{-k} \|\chi_I^k \nabla u_I^{ext}\|_{L^q L^r} \lesssim_N 2^{C\kappa} \left(\|u_I\|_{L^q L^r[I]} + 2^{(\frac{1}{2} - \frac{1}{q} - \frac{4}{r})} \|\Box u_I\|_{L^2 L^2[I]} \right), \tag{4.23}$$

$$2^{-2k} \|\chi_I^k \nabla u_I^{ext}\|_{L^\infty L^\infty} \lesssim_N 2^{-2k} \|\nabla u_I\|_{L^\infty L^\infty[I]}, \tag{4.24}$$

[5]To be pedantic, [17, Proposition 3.4] only corresponds to the case $\kappa = 0$. However, the required modification of the proof is straightforward.

where (q, r) is any pair of admissible Strichartz exponents on $\mathbb{R}^{1+4}$.

Remark 4.11. Since $2^{-k}[\chi_I^k, \nabla] = 2^{-k}(\nabla \chi_I^k)$ is simply multiplication by another generalized cutoff function adapted to the frequency scale 2^k, the conclusions of Proposition 4.10 also hold with $\chi_I^k 2^{-k} \nabla u_I^{ext}$ replaced by $2^{-k}\nabla(\chi_I^k u_I^{ext})$ on the LHSs.

4.2. **Estimates for quadratic nonlinearities.** Here we state estimates for the quadratic nonlinearities in Theorems 3.5 and 3.6. All estimates stated here are proved in Section 8.3.

Throughout this and the next subsections, we will denote by A a $\mathfrak{g}$-valued *spatial* 1-form $A = A_j\, dx^j$ on $I \times \mathbb{R}^4$ for some time interval I. To denote a $\mathfrak{g}$-valued *space-time* 1-form, we use the notation $A_{t,x} = A_\alpha\, dx^\alpha$. We will use B [resp. $B_{t,x}$] to denote[6] another $\mathfrak{g}$-valued spatial [resp. space-time] 1-form on $I \times \mathbb{R}^4$. Unless otherwise stated, all frequency envelopes will be assumed to be δ_3-admissible.

We begin with the quadratic nonlinearities in the equations for A_0, $\partial_t A_0$ and $\partial^\ell A_\ell$. We introduce the notation

$$\mathcal{M}_0^2(A, B) = [A_\ell, \partial_t B^\ell], \tag{4.25}$$

$$\mathcal{D}\mathcal{M}_0^2(A, B) = -\, 2\mathbf{Q}(\partial_t A, \partial_t B). \tag{4.26}$$

These are the main quadratic nonlinearities in the ΔA_0 and $\Delta \partial_t A_0$ equations, respectively. The estimates that we need for these nonlinearities are as follows.

Proposition 4.12. *We have the fixed-time bounds*

$$\| |D|^{-1} \mathcal{M}_0^2(A, B)(t) \|_{L^2_{cd}} \lesssim \|A(t)\|_{\dot{H}^1_c} \|\partial_t B(t)\|_{L^2_d}, \tag{4.27}$$

$$\| |D|^{-2} \mathcal{D}\mathcal{M}_0^2(A, B)(t) \|_{L^2_{cd}} \lesssim \|\partial_t A(t)\|_{L^2_c} \|\partial_t B(t)\|_{L^2_d}, \tag{4.28}$$

and the space-time bounds

$$\| |D|^{-1} \mathcal{M}_0^2(A, B) \|_{Y_{cd}[I]} \lesssim \|A\|_{S^1_c[I]} \|B\|_{S^1_d[I]}, \tag{4.29}$$

$$\| |D|^{-1} \mathcal{M}_0^2(A, B) \|_{L^2 \dot{H}^{\frac{1}{2}}_{cd}[I]} + \| |D|^{-2} \mathcal{D}\mathcal{M}_0^2(A, B) \|_{L^2 \dot{H}^{\frac{1}{2}}_{cd}[I]} \lesssim \|A\|_{\mathrm{Str}^1_c[I]} \|B\|_{\mathrm{Str}^1_d[I]}. \tag{4.30}$$

Moreover, for any $\kappa > 0$, the nonlinearity $\mathcal{M}_0^2(A, B)$ admits the splitting

$$\mathcal{M}_0^2(A, B) = \mathcal{M}_{0,small}^{\kappa,2}(A, B) + \mathcal{M}_{0,large}^{\kappa,2}(A, B)$$

where the small part obeys the improved bound

$$\| |D|^{-1} \mathcal{M}_{0,small}^{\kappa,2}(A, B) \|_{Y_{cd}[I]} \lesssim 2^{-\delta_2 \kappa} \|A\|_{S^1_c[I]} \|B\|_{S^1_d[I]}, \tag{4.31}$$

and the large part is bounded by divisible norms of A and B:

$$\| |D|^{-1} \mathcal{M}_{0,large}^{\kappa,2}(A, B) \|_{Y_{cd}[I]} \lesssim 2^{C\kappa} \|A\|_{DS^1_c[I]} \|B\|_{DS^1_d[I]}. \tag{4.32}$$

Finally, if either

$$\|A\|_{S^1_c[I]} \le 1 \quad \textit{and} \quad (B, I) \textit{ is } (\varepsilon, M)\textit{-energy dispersed, or}$$
$$\|B\|_{S^1_c[I]} \le 1 \quad \textit{and} \quad (A, I) \textit{ is } (\varepsilon, M)\textit{-energy dispersed,}$$

[6]Note that this convention is different from [18] and Section 3, where B was reserved for caloric gauge linearized Yang–Mills heat flows.

then we have

$$\| |D|^{-1} \mathcal{M}_0^2(A,B)\|_{Y_c[I]} \lesssim \varepsilon^{\delta_2} M, \tag{4.33}$$

$$\| |D|^{-2} \mathcal{D}\mathcal{M}_0^2(A,B)\|_{L^2 \dot{H}_c^{\frac{1}{2}}[I]} \lesssim \varepsilon^{\delta_2} M. \tag{4.34}$$

The remaining quadratic nonlinearities in the equations for A_0 and $\partial^\ell A_\ell$ involve $\mathbf{Q}$, and they obey simpler estimates.

Proposition 4.13. *For $\sigma = 0$ or 1, we have the fixed-time bound*

$$\| |D|^{-\sigma} \mathbf{Q}(A, \partial_t^\sigma B)(t)\|_{L^2_{cd}} \lesssim \|A(t)\|_{\dot{H}^1_c} \|\partial_t^\sigma B(t)\|_{\dot{H}^{1-\sigma}_d}, \tag{4.35}$$

and the space-time bounds

$$\| |D|^{-\sigma} \mathbf{Q}(A, \partial_t^\sigma B)\|_{L^2 \dot{H}^{\frac{1}{2}}_{cd}[I]} \lesssim \|A\|_{\mathrm{Str}^1_c[I]} \|B\|_{\mathrm{Str}^1_d[I]}, \tag{4.36}$$

$$\| |D|^{-\sigma} \mathbf{Q}(A, \partial_t^\sigma B)\|_{Y_{cd}[I]} + \| |D|^{-\sigma-1} \mathbf{Q}(A, \partial_t^\sigma B)\|_{L^1 L^\infty_{cd}[I]} \lesssim \|A\|_{DS^1_c[I]} \|B\|_{DS^1_d[I]}. \tag{4.37}$$

Finally, if either

$$\|A\|_{S^1_c[I]} \leq 1 \quad \text{and} \quad (B, I) \text{ is } (\varepsilon, M)\text{-energy dispersed, or}$$
$$\|B\|_{S^1_c[I]} \leq 1 \quad \text{and} \quad (A, I) \text{ is } (\varepsilon, M)\text{-energy dispersed,}$$

then

$$\| |D|^{-\sigma} \mathbf{Q}(A, \partial_t^\sigma B)\|_{Y_c[I]} \lesssim \varepsilon^{\delta_2} M. \tag{4.38}$$

Also for the quadratic part $\mathbf{A}_0^2$ of A_0, given by

$$\mathbf{A}_0^2(A, A) = \Delta^{-1}([A, \partial_t A] + 2Q(A, \partial_t A)$$

we have the following additional property, which will be used in the proof of Theorem 1.18:

Proposition 4.14. *For the quadratic form $\mathbf{A}_0^2$ we have*

$$\| |D|^2 \mathbf{A}_0^2(A, B)\|_{(L^2_x L^1_t)_{cd}[I]} \lesssim \|\nabla A\|_{S^{sq}_c} \|\nabla B\|_{S^{sq}_d}. \tag{4.39}$$

For the quadratic nonlinearity in the $\Box_A A_j$ equation, we introduce the notation

$$\mathbf{P}_j \mathcal{M}^2(A, B) = \mathbf{P}_j [A_\ell, \partial_x B^\ell],$$
$$\mathbf{P}_j^\perp \mathcal{M}^2(A, B) = 2\Delta^{-1} \partial_j \mathbf{Q}(\partial^\alpha A, \partial_\alpha A),$$

so that (3.12) becomes

$$\Box_A A_j = \mathbf{P}_j \mathcal{M}(A, A) + \mathbf{P}_j^\perp \mathcal{M}(A, A) + R_j(A, \partial_t A).$$

Proposition 4.15. *We have the fixed time bounds*

$$\|\mathbf{P}\mathcal{M}^2(A, B)(t)\|_{\dot{H}^{-1}_{cd}} \lesssim \|A(t)\|_{\dot{H}^1_c} \|B(t)\|_{\dot{H}^1_d}, \tag{4.40}$$

$$\|\mathbf{P}^\perp \mathcal{M}^2(A, B)(t)\|_{\dot{H}^{-1}_{cd}} \lesssim \|\nabla A(t)\|_{L^2_c} \|\nabla B(t)\|_{L^2_d}. \tag{4.41}$$

and space-time bounds

$$\|\mathbf{P}\mathcal{M}^2(A, B)\|_{(N \cap \Box \underline{X}^1)_{cd}[I]} \lesssim \|A\|_{S^1_c[I]} \|B\|_{S^1_d[I]}, \tag{4.42}$$

$$\|\mathbf{P}^\perp \mathcal{M}^2(A, B)\|_{(N \cap \Box \underline{X}^1)_{cd}[I]} \lesssim \|A\|_{S^1_c[I]} \|B\|_{S^1_d[I]}. \tag{4.43}$$

In particular, the $L^2\dot{H}^{-\frac{1}{2}}$-norms are bounded by the Str^1*-norms of A and B:*

$$\|\mathbf{P}\mathcal{M}^2(A,B)\|_{L^2\dot{H}^{-\frac{1}{2}}_{cd}[I]} \lesssim \|A\|_{\mathrm{Str}^1_c[I]}\|B\|_{\mathrm{Str}^1_d[I]}, \tag{4.44}$$

$$\|\mathbf{P}^\perp\mathcal{M}^2(A,B)\|_{L^2\dot{H}^{-\frac{1}{2}}_{cd}[I]} \lesssim \|A\|_{\mathrm{Str}^1_c[I]}\|B\|_{\mathrm{Str}^1_d[I]}. \tag{4.45}$$

Moreover, for any $\kappa > 0$, the terms $\mathbf{P}_j\mathcal{M}^2(A,B)$ and $\mathbf{P}_j^\perp\mathcal{M}^2(A,B)$ admit the splittings

$$\mathbf{P}_j\mathcal{M}^2(A,B) = \mathbf{P}_j\mathcal{M}^{\kappa,2}_{small}(A,B) + \mathbf{P}_j\mathcal{M}^{\kappa,2}_{large}(A,B),$$
$$\mathbf{P}_j^\perp\mathcal{M}^2(A,B) = \mathbf{P}_j^\perp\mathcal{M}^{\kappa,2}_{small}(A,B) + \mathbf{P}_j^\perp\mathcal{M}^{\kappa,2}_{large}(A,B),$$

so that the N-norm of the small parts obey the improved bounds

$$\|\mathbf{P}\mathcal{M}^{\kappa,2}_{small}(A,B)\|_{N_{cd}[I]} \lesssim 2^{-\delta_2\kappa}\|A\|_{S^1_c[I]}\|B\|_{S^1_d[I]}, \tag{4.46}$$

$$\|\mathbf{P}^\perp\mathcal{M}^{\kappa,2}_{small}(A,B)\|_{N_{cd}[I]} \lesssim 2^{-\delta_2\kappa}\|A\|_{S^1_c[I]}\|B\|_{S^1_d[I]}, \tag{4.47}$$

and that of the large parts are bounded by divisible norms of A and B:

$$\|\mathbf{P}\mathcal{M}^{\kappa,2}_{large}(A,B)\|_{N_{cd}[I]} \lesssim 2^{C\kappa}\|A\|_{DS^1_c[I]}\|B\|_{DS^1_d[I]}, \tag{4.48}$$

$$\|\mathbf{P}^\perp\mathcal{M}^{\kappa,2}_{large}(A,B)\|_{N_{cd}[I]} \lesssim 2^{C\kappa}\|A\|_{DS^1_c[I]}\|B\|_{DS^1_d[I]}. \tag{4.49}$$

Finally, if either

$$\|A\|_{\underline{S}^1_c[I]} \leq 1 \quad \text{and} \quad (B,I) \text{ is } (\varepsilon,M)\text{-energy dispersed, or}$$
$$\|B\|_{\underline{S}^1_c[I]} \leq 1 \quad \text{and} \quad (A,I) \text{ is } (\varepsilon,M)\text{-energy dispersed,}$$

then

$$\|\mathbf{P}\mathcal{M}^2(A,B)\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})_c[I]} \lesssim \varepsilon^{\delta_2}M, \tag{4.50}$$

$$\|\mathbf{P}^\perp\mathcal{M}^2(A,B)\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})_c[I]} \lesssim \varepsilon^{\delta_2}M. \tag{4.51}$$

We end this subsection with bilinear estimates for $\mathbf{w}_0^2$ and $\mathbf{w}_x^2$, which arise in the equation for a dynamic Yang–Mills heat flow of a caloric Yang–Mills wave.

Proposition 4.16. *For any $s > 0$, we have the fixed-time bound*

$$\||D|^{-1}P_k\mathbf{w}_0^2(A,B,s)(t)\|_{L^2} \lesssim \langle 2^{2k}s\rangle^{-10}\langle 2^{-2k}s^{-1}\rangle^{-\delta_2}c_kd_k\|\partial_tA(t)\|_{L^2_c}\|B(t)\|_{\dot{H}^1_d}, \tag{4.52}$$

and the space-time bounds

$$\||D|^{-1}P_k\mathbf{w}_0^2(A,B,s)\|_{L^2\dot{H}^{\frac{1}{2}}[I]} \lesssim \langle 2^{2k}s\rangle^{-10}\langle 2^{-2k}s^{-1}\rangle^{-\delta_2}c_kd_k\|A\|_{\mathrm{Str}^1_c[I]}\|B\|_{\mathrm{Str}^1_d[I]}, \tag{4.53}$$

$$\||D|^{-1}P_k\mathbf{w}_0^2(A,B,s)\|_{Y[I]} \lesssim \langle 2^{2k}s\rangle^{-10}\langle 2^{-2k}s^{-1}\rangle^{-\delta_2}c_kd_k\|A\|_{S^1_c[I]}\|B\|_{S^1_d[I]}. \tag{4.54}$$

Moreover, if (B,I) is (ε,M)-energy dispersed, then

$$\||D|^{-1}P_k\mathbf{w}_0^2(A,B,s)\|_{Y[I]} \lesssim \varepsilon^{\delta_2}\langle 2^{2k}s\rangle^{-10}\langle 2^{-2k}s^{-1}\rangle^{-\delta_2}c_k\|A\|_{S^1_c[I]}M. \tag{4.55}$$

Proposition 4.17. *For any $s > 0$, we have the fixed-time bound*

$$\|P_k\mathbf{P}\mathbf{w}_x^2(A,B,s)(t)\|_{\dot{H}^{-1}} \lesssim \langle 2^{2k}s\rangle^{-10}\langle 2^{-2k}s^{-1}\rangle^{-\delta_2}c_kd_k\|\nabla A(t)\|_{L^2_c}\|\nabla B(t)\|_{L^2_d}. \tag{4.56}$$

and the space-time bounds

$$\begin{aligned} \|P_k \mathbf{Pw}_x^2(A,B,s)\|_{L^2 \dot{H}^{-\frac{1}{2}}[I]} \\ \lesssim \langle 2^{2k}s\rangle^{-10}\langle 2^{-2k}s^{-1}\rangle^{-\delta_2} c_k d_k \|(\nabla A, \nabla \mathbf{P}^\perp A)\|_{(\mathrm{Str}^0\times L^2\dot{H}^{\frac{1}{2}})_c[I]} \|B\|_{\mathrm{Str}^1_d[I]}, \end{aligned} \tag{4.57}$$

$$\begin{aligned} \|P_k \mathbf{Pw}_x^2(A,B,s)\|_{N\cap \Box \underline{X}^1[I]} \\ \lesssim \langle 2^{2k}s\rangle^{-10}\langle 2^{-2k}s^{-1}\rangle^{-\delta_2} c_k d_k \|(A, \mathbf{P}^\perp A)\|_{(S^1\times Y^1)_c[I]} \|B\|_{S^1_d[I]}. \end{aligned} \tag{4.58}$$

Moreover, if (B,I) is (ε, M)-energy dispersed, then

$$\begin{aligned} \|P_k \mathbf{Pw}_x^2(A,B,s)\|_{N\cap L^2\dot{H}^{-\frac{1}{2}}[I]} \\ \lesssim \langle 2^{2k}s\rangle^{-10}\langle 2^{-2k}s^{-1}\rangle^{-\delta_2} c_k (\varepsilon^{\delta_2}\|A\|_{\underline{S}^1_c[I]} + \|\nabla \mathbf{P}^\perp A\|_{L^2\dot{H}^{\frac{1}{2}}_c[I]})M. \end{aligned} \tag{4.59}$$

4.3. **Estimates for the covariant wave operator.** We now state estimates concerning the covariant wave operator $\Box_A$. All estimates stated here without proofs are proved in Section 8.3, with the exceptions of Theorem 4.24 and Proposition 4.25, which are proved in Section 9.

We begin by expanding $\Box_A B$ to

$$\Box_A B = \Box B + 2[A_\alpha, \partial^\alpha B] + [\partial^\alpha A_\alpha, B] + [A^\alpha, [A_\alpha, B]].$$

We have the following simple fixed-time estimates for $\Box_A - \Box$.

Proposition 4.18. *For any $\alpha, \beta, \gamma \in \{0, 1, \dots, 4\}$, we have the fixed-time bounds*

$$\|[A_\alpha, \partial^\alpha B](t)\|_{\dot{H}^{-1}_{cd}} \lesssim \|(A_0, A)(t)\|_{\dot{H}^1_c} \|\nabla B(t)\|_{L^2_d}, \tag{4.60}$$

$$\|[\partial^\alpha A_\alpha, B](t)\|_{\dot{H}^{-1}_{cd}} \lesssim (\|A(t)\|_{\dot{H}^1_c} + \|\partial_t A_0(t)\|_{L^2_c})\|B(t)\|_{\dot{H}^1_d}, \tag{4.61}$$

$$\|[A^{(1)}_\alpha, [A^{(2)\alpha}, B]](t)\|_{\dot{H}^{-1}_{cde}} \lesssim \|(A^{(1)}_0, A^{(1)})(t)\|_{\dot{H}^1_c} \|(A^{(2)}_0, A^{(2)})(t)\|_{\dot{H}^1_d} \|B(t)\|_{\dot{H}^1_e}, \tag{4.62}$$

and the space-time bounds

$$\|[A_\ell, \partial^\ell B]\|_{L^2\dot{H}^{-\frac{1}{2}}_{cd}[I]} \lesssim \|A\|_{\mathrm{Str}^1_c[I]} \|B\|_{\mathrm{Str}^1_d[I]}, \tag{4.63}$$

$$\|[A_0, \partial_0 B]\|_{L^2\dot{H}^{-\frac{1}{2}}_{cd}[I]} \lesssim \|\nabla A_0\|_{L^2\dot{H}^{\frac{1}{2}}_c[I]} \|B\|_{\mathrm{Str}^1_d[I]}, \tag{4.64}$$

$$\|[\partial^\alpha A_\alpha, B]\|_{L^2\dot{H}^{-\frac{1}{2}}_{cd}[I]} \lesssim \|(\nabla A_0, \nabla \mathbf{P}^\perp A)\|_{L^2\dot{H}^{\frac{1}{2}}_c[I]} \|B\|_{\mathrm{Str}^1_d[I]}, \tag{4.65}$$

$$\begin{aligned} \|[A^{(1)}_\alpha, [A^{(2)\alpha}, B]](t)\|_{L^2\dot{H}^{-\frac{1}{2}}_{cde}[I]} \lesssim& \|(\nabla A^{(1)}_0, \nabla A^{(1)})(t)\|_{L^2\dot{H}^{\frac{1}{2}}\times \mathrm{Str}^0_c[I]} \\ &\times \|(\nabla A^{(2)}_0, \nabla A^{(2)})(t)\|_{L^2\dot{H}^{\frac{1}{2}}\times \mathrm{Str}^0_c[I]} \|B\|_{\mathrm{Str}^1_e[I]}. \end{aligned} \tag{4.66}$$

In order to proceed, we recall the notation $\mathbf{P}_\alpha A = (\mathbf{P}A)_\alpha$ for a space-time 1-form $A_{t,x}$:

$$\mathbf{P}_\alpha A = \begin{cases} \mathbf{P}_j A_x & \alpha = j \in \{1, \dots, 4\}, \\ A_0 & \alpha = 0. \end{cases}$$

We also write $\mathbf{P}^\perp_\alpha A = (\mathbf{P}^\perp A)_\alpha = A_\alpha - \mathbf{P}_\alpha A$.

Given a parameter $\kappa \in \mathbb{N}$, we furthermore decompose $2[A_\alpha, \partial^\alpha B]$ so that

$$\begin{aligned} \Box_A B &= \Box B + 2[A_\alpha, \partial^\alpha B] + \mathrm{Rem}^3_A B \\ &= \Box B + \mathrm{Diff}^\kappa_{\mathbf{P}A} B + \mathrm{Diff}^\kappa_{\mathbf{P}^\perp A} B + \mathrm{Rem}^{\kappa,2}_A B + \mathrm{Rem}^3_A B, \end{aligned} \tag{4.67}$$

where[7]

$$\mathrm{Diff}^{\kappa}_{\mathbf{P}A} = \sum_k 2[P_{<k-\kappa}\mathbf{P}_\alpha A, \partial^\alpha P_k B], \tag{4.68}$$

$$\mathrm{Diff}^{\kappa}_{\mathbf{P}^\perp A} = \sum_k 2[P_{<k-\kappa}\mathbf{P}^\perp_\alpha A, \partial^\alpha P_k B], \tag{4.69}$$

$$\mathrm{Rem}^{\kappa,2}_{A} = \sum_k 2[P_{\geq k-\kappa} A_\alpha, \partial^\alpha P_k B], \tag{4.70}$$

$$\mathrm{Rem}^3_A B = [\partial^\alpha A_\alpha, B] + [A^\alpha, [A_\alpha, B]]. \tag{4.71}$$

We now turn to the bounds for each part of the decomposition (4.67). For a fixed $B \in S^1[I]$, we introduce the nonlinear maps

$$\begin{aligned}\mathrm{Rem}^3(A)B = & - [\mathbf{DA}_0(A), B] + [\mathbf{DA}(A), B] \\ & - [\mathbf{A}_0(A), [\mathbf{A}_0(A), B]] + [A^\ell, [A_\ell, B]],\end{aligned} \tag{4.72}$$

$$\mathrm{Rem}^3_s(A)B = -[\mathbf{DA}_{0;s}(A), B] - [\mathbf{A}_{0;s}(A), [\mathbf{A}_{0;s}(A), B]], \tag{4.73}$$

defined for spatial connections A on I such that $(A, \partial_t A)(t) \in T^{L^2}\mathcal{C}$ for each fixed time $t \in I$. In view of Theorems 3.5 and 3.6, for a caloric Yang–Mills wave A we have

$$\begin{aligned}\mathrm{Rem}^3_A B = & \mathrm{Rem}^3(A)B, \\ \mathrm{Rem}^3_{A(s)} B = & \mathrm{Rem}^3(A(s))B + \mathrm{Rem}^3_s(A)B.\end{aligned}$$

The nonlinear maps $\mathrm{Rem}^3(A)B$ and $\mathrm{Rem}^3_s(A)B$ are well-behaved:

Proposition 4.19. *Suppose that $A(t) \in \mathcal{C}_\mathcal{Q}$ for every $t \in I$. Then the following properties hold with bounds depending on $\mathcal{Q}$, but otherwise independent of I:*

- *Let c and d be $(-\delta_2, S)$-frequency envelopes for A and B in $\mathrm{Str}^1[I]$, respectively. Then*

$$\| P_k(\mathrm{Rem}^3(A)B)\|_{L^1L^2 \cap L^2 \dot{H}^{-\frac{1}{2}}[I]} \lesssim_{\mathcal{Q}, \|A\|_{\mathrm{Str}^1[I]}} (c^{[\delta_2]}_k)^2 d_k + c_k c^{[\delta_2]}_k d^{[\delta_2]}_k. \tag{4.74}$$

- *For a fixed $A \in \mathrm{Str}^1[I]$, $\mathrm{Rem}^3(A)B$ is linear in B. On the other hand, for a fixed B with $\|B\|_{\mathrm{Str}^1[I]} \leq 1$, $\mathrm{Rem}^3(\cdot)B : \mathrm{Str}^1[I] \to L^1L^2 \cap L^2\dot{H}^{-\frac{1}{2}}[I]$ is Lipschitz envelope-preserving.*
- *For a fixed $A \in \mathrm{Str}^1[I]$, $\mathrm{Rem}^3_s(A)B$ is linear in B. On the other hand, for a fixed $B \in S^1[I]$ with $\|B\|_{\mathrm{Str}^1[I]} \leq 1$, $\mathrm{Rem}^3_s(A)B$ is a Lipschitz map*

$$\mathrm{Rem}^3_s(A)B : \mathrm{Str}^1[I] \to L^1L^2 \cap L^2\dot{H}^{-\frac{1}{2}}[I] \tag{4.75}$$

with output concentrated at frequency $s^{-\frac{1}{2}}$,

$$(1 - s\Delta)^N \mathrm{Rem}^3_s(A)B : \mathrm{Str}^1[I] \to 2^{-\delta_2 k(s)} L^1 \dot{H}^{-\delta_2} \cap L^2 \dot{H}^{-\frac{1}{2}-\delta_2}[I]. \tag{4.76}$$

Next, we consider the term $2[A_\alpha, \partial^\alpha B] = \mathrm{Diff}^{\kappa}_{\mathbf{P}A}B + \mathrm{Diff}^{\kappa}_{\mathbf{P}^\perp A}B + \mathrm{Rem}^{\kappa,2}_A B$. We begin with $\mathrm{Rem}^{\kappa,2}_A B$, which obeys analogous bounds as $\mathbf{P}\mathcal{M}^2(A, B)$ and $\mathbf{P}^\perp\mathcal{M}^2(A, B)$ (cf. Proposition 4.15).

[7] Although the definition depends on the whole space-time connection $A_{t,x}$, we deviate from our convention and simply write $\mathrm{Diff}^{\kappa}_{\mathbf{P}A}$, $\mathrm{Diff}^{\kappa}_{\mathbf{P}^\perp A}$, $\mathrm{Rem}^{\kappa,2}_A$ etc. to avoid cluttered notation.

Proposition 4.20. *For any $\kappa > 0$, the term $\mathrm{Rem}^{\kappa,2}_A B$ obeys the bound*

$$\|\mathrm{Rem}^{\kappa,2}_A B\|_{(N\cap\Box\underline{X}^1)_{cd}[I]} \lesssim 2^{C\kappa}\left(\|A\|_{S^1_c[I]} + \|(\mathbf{P}^\perp A, A_0)\|_{Y^1_c[I]}\right)\|B\|_{S^1_d[I]}. \tag{4.77}$$

In particular, its $L^2\dot{H}^{-\frac{1}{2}}$-norm is bounded by:

$$\|\mathrm{Rem}^{\kappa,2}_A B\|_{L^2\dot{H}^{-\frac{1}{2}}_{cd}[I]} \lesssim \left(\|A\|_{\mathrm{Str}^1_c[I]} + \|(\nabla\mathbf{P}^\perp A, \nabla A_0)\|_{(L^2\dot{H}^{\frac{1}{2}})_c[I]}\right)\|B\|_{\mathrm{Str}^1_d[I]}. \tag{4.78}$$

Furthermore, $\mathrm{Rem}^{\kappa,2}_A B$ admits the splitting

$$\mathrm{Rem}^{\kappa,2}_A B = \mathrm{Rem}^{\kappa,2}_{A,small} B + \mathrm{Rem}^{\kappa,2}_{A,large} B$$

so that the N-norm of the small part obeys the improved bound

$$\|\mathrm{Rem}^{\kappa,2}_{A,small} B\|_{N_{cd}[I]} \lesssim 2^{-\delta_2\kappa}\|A\|_{S^1_c[I]}\|B\|_{S^1_d[I]}, \tag{4.79}$$

and that of the large part is bounded by a divisible norm of (A_0, A):

$$\|\mathrm{Rem}^{\kappa,2}_{A,large} B\|_{N_{cd}[I]} \lesssim 2^{C\kappa}\left(\|A\|_{DS^1_c[I]} + \|(\nabla\mathbf{P}^\perp A, \nabla A_0)\|_{(L^2\dot{H}^{\frac{1}{2}})_c[I]}\right)\|B\|_{S^1_d[I]}. \tag{4.80}$$

Finally, if (B, I) is (ε, M)-energy dispersed, then

$$\begin{aligned}\|\mathrm{Rem}^{\kappa,2}_A B\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})_c[I]} \lesssim\, & (2^{-\delta_2\kappa} + 2^{C\kappa}\varepsilon^{\delta_2})\|A\|_{\underline{S}^1_c[I]} M \\ & + 2^{C\kappa}\|(\nabla\mathbf{P}^\perp A, \nabla A_0)\|_{(L^2\dot{H}^{\frac{1}{2}})_c[I]} M.\end{aligned} \tag{4.81}$$

It remains to consider the paradifferential terms. The term $\mathrm{Diff}^\kappa_{\mathbf{P}^\perp A} B$ can be handled using the following estimate, in combination with (3.22) and Proposition 4.12:

Proposition 4.21. *For any $\kappa > 0$, we have*

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}^\perp A} B\|_{(X^{-\frac{1}{2}+b_1,-b_1}\cap\Box\underline{X}^1)_{cd}[I]} \lesssim \|\mathbf{P}^\perp A\|_{Y^1_c[I]}\|B\|_{S^1_d[I]}. \tag{4.82}$$

Moreover, we have

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}^\perp A} B\|_{L^1 L^2_f[I]} \lesssim \|\mathbf{P}^\perp A\|_{L^1 L^\infty_a[I]}\|B\|_{S^1_e[I]} \tag{4.83}$$

where $f_k = \left(\sum_{k'<k-\kappa} a_{k'}\right) e_k$.

The only remaining term is the paradifferential term $\mathrm{Diff}^\kappa_{\mathbf{P}A} B$. We first state the high modulation bounds.

Proposition 4.22. *For any $\kappa > 0$, consider the splitting $\mathrm{Diff}^\kappa_{\mathbf{P}A} = \mathrm{Diff}^\kappa_{A_0} + \mathrm{Diff}^\kappa_{\mathbf{P}_x A}$, where*

$$\mathrm{Diff}^\kappa_{A_0} B = -\sum_k 2[P_{<k-\kappa} A_0, \partial_t P_k B], \qquad \mathrm{Diff}^\kappa_{\mathbf{P}_x A} B = \sum_k 2[P_{<k-\kappa}\mathbf{P}_\ell A, \partial^\ell P_k B].$$

For $\mathrm{Diff}_{A_0} B$, we have the bound

$$\|\mathrm{Diff}^\kappa_{A_0} B\|_{(X^{-\frac{1}{2}+b_1,-b_1}\cap\Box\underline{X}^1)_{cd}[I]} \lesssim \|A_0\|_{Y^1_c[I]}\|B\|_{S^1_d[I]}. \tag{4.84}$$

On the other hand, for $\mathrm{Diff}_{\mathbf{P}_x A} B$, we have the bounds

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}_x A} B\|_{(\Box\tilde{X}^1)_{cd}[I]} \lesssim \|A_x\|_{S^1_c[I]}\|B\|_{S^1_d[I]}, \tag{4.85}$$

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}_x A} B\|_{(\Box\underline{X}^1)_{cd}[I]} \lesssim \|A_x\|_{(S^1\cap\tilde{X}^1)_c[I]}\|B\|_{S^1_d[I]}, \tag{4.86}$$

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}_x A} B\|_{(X^{-\frac{1}{2}+b_1,-b_1})_{cd}[I]} \lesssim \|A_x\|_{(S^1\cap\underline{X}^1)_c[I]}\|B\|_{S^1_d[I]}. \tag{4.87}$$

Next, we consider the $N \cap L^2 \dot{H}^{\frac{1}{2}}$ norm of $\mathrm{Diff}_{\mathbf{P}A} B$. The contribution of each Littlewood-Paley projection $P_{k_0} \mathbf{P} A$ is perturbative, as the following proposition states:

Proposition 4.23. *Let $A_{t,x}$ be a caloric Yang–Mills wave on an interval I obeying*

$$\|A\|_{S^1[I]} \leq M. \tag{4.88}$$

Then for any $\kappa > 0$ and $k_0 \in \mathbb{Z}$, we have

$$\|\mathrm{Diff}^{\kappa}_{P_{k_0} \mathbf{P} A} B\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_d[I]} \lesssim_M \|B\|_{S^1_d[I]}. \tag{4.89}$$

However, we *cannot* sum up in k_0. The proper way to handle $\mathrm{Diff}^{\kappa}_{\mathbf{P}A}$ is not to regard it as a perturbative nonlinearity, but rather as a part of the underlying linear operator. Indeed, for the operator $\Box + \mathrm{Diff}^{\kappa}_{\mathbf{P}A}$, we have the following well-posedness result:

Theorem 4.24. *Let $A_{t,x}$ be a caloric Yang–Mills wave on an interval I obeying* (4.88). *Consider the following initial value problem on $I \times \mathbb{R}^4$:*

$$\begin{cases} \Box B + \mathrm{Diff}^{\kappa}_{\mathbf{P}A} B = G, \\ (B, \partial_t B)(t_0) = (B_0, B_1), \end{cases} \tag{4.90}$$

for some $\mathfrak{g}$-valued spatial 1-form $G \in N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]$, $(B_0, B_1) \in \dot{H}^1 \times L^2$ and $t_0 \in I$.

Then for $\kappa \geq \kappa_1(M)$, where $\kappa_1(M) \gg 1$ is some function independent of $A_{t,x}$, there exists a unique solution $B \in S^1[I]$ to (4.90). *Moreover, for any admissible frequency envelope c, the solution obeys the bound*

$$\|B\|_{S^1_c[I]} \lesssim_M \|(B_0, B_1)\|_{(\dot{H}^1 \times L^2)_c} + \|G\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_c[I]}. \tag{4.91}$$

As a quick corollary of Propositions 4.19–4.20 and Theorem 4.24, we obtain well-posedness of the initial value problem associated to $\Box_A$; see Theorem 5.1.(1) below.

Theorem 4.24 is proved in Sections 9, 10 and 11. The main ingredient for the proof is construction of a parametrix for $\Box + \mathrm{Diff}^{\kappa}_{\mathbf{P}A}$ by renormalization with a pseudodifferential gauge transformation; for a more detailed discussion, see Section 9.

The paradifferential wave equation (4.90) leads to the following *weak divisibility* property of the S^1 norm, which will later play an important role in the energy induction argument.

Proposition 4.25. *Let $A_{t,x}$ be a caloric Yang–Mills wave on an interval I which obeys* (4.88) *for some $M > 0$. Let $B \in S^1[I]$ be a solution to the paradifferential wave equation* (4.90) *with the source $G \in N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]$, which obeys the bound*

$$\sup_{t \in I} \|(B, \partial_t B)(t)\|_{L^2} \leq E \tag{4.92}$$

for some $E > 0$. Then there exists a partition $I = \cup_{i \in \mathcal{I}} I_i$ such that

$$\|B\|_{S^1[I_i]} \lesssim_E 1 \qquad \text{for } i \in \mathcal{I} \tag{4.93}$$

where

$$\# \mathcal{I} \lesssim_{E, M, \|B\|_{S^1[I]}, \|G\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]}} 1.$$

The proof of this proposition also involves the parametrix construction (cf. Sections 9, 10 and 11), as well as Proposition 4.23.

We now state additional estimates satisfied by $\mathrm{Diff}^{\kappa}_{\mathbf{P}A}$, which are needed to analyze the difference of two solutions (or even approximate solutions). For this purpose, it is necessary to exploit the so-called secondary null structure of the Yang–Mills equation, which becomes available after reiterating the equations for $\mathbf{P}A$.

We begin with simple bilinear estimates, which allows us to peel off the non-essential parts (in particular, the contribution of the cubic and higher order nonlinearities) of A_0 and $\mathbf{P}A$.

Proposition 4.26. *We have*

$$\|\mathrm{Diff}^{\kappa}_{A_0} B\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_f[I]} \lesssim \|A_0\|_{(L^1 L^\infty \cap L^2 \dot{H}^{\frac{3}{2}})_a[I]} \|B\|_{S^1_e[I]}, \tag{4.94}$$

$$\|\mathrm{Diff}^{\kappa}_{\mathbf{P}_x A} B\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_f[I]} \lesssim (\|\mathbf{P}A[t_0]\|_{(\dot{H}^1 \times L^2)_a} + \|\Box \mathbf{P}A\|_{L^1 L^2_a[I]}) \|B\|_{S^1_e[I]}, \tag{4.95}$$

where

$$f_k = \Big(\sum_{k' < k - \kappa} a_{k'} \Big) e_k.$$

The contribution of the quadratic nonlinearities $\mathcal{M}^2_0$ and $\mathcal{M}^2$ in the equations for A_0 and A_x, respectively, cannot be treated separately. This is precisely where we exploit the secondary null structure, which only manifests itself after combining the contribution of these nonlinearities in $\mathrm{Diff}^{\kappa}_{\mathbf{P}A}$.

Proposition 4.27. *Let*

$$\Delta A_0 = [B^{(1)\ell}, \partial_t B^{(2)}_{\ell}], \tag{4.96}$$

$$\Box \mathbf{P}A = \mathbf{P}[B^{(1)\ell}, \partial_x B^{(2)}_{\ell}], \qquad \mathbf{P}A[t_0] = 0. \tag{4.97}$$

where $B^{(1)}, B^{(2)} \in S^1[I]$. Then we have

$$\|\mathrm{Diff}^{\kappa}_{\mathbf{P}A} B\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_f[I]} \lesssim_{\tilde{M}} \|B^{(1)}\|_{S^1_c[I]} \|B^{(2)}\|_{S^1_d[I]} \|B\|_{S^1_e[I]} \tag{4.98}$$

where

$$f_k = \Big(\sum_{k' < k - \kappa} c_{k'} d_{k'} \Big) e_k.$$

Next, we turn to the contribution of terms of the form $[A_\alpha, \partial^\alpha A]$ in the equation for $\mathbf{P}_x A$. The frequency envelope bound for this term is slightly involved, because it does not obey a good N-norm estimate.

Proposition 4.28. *Let $A_0 = 0$ and*

$$\Box \mathbf{P}A_j = \sum_{n=1}^{N} \mathbf{P}[B^{n(1)}_\alpha, \partial^\alpha B^{n(2)}_j], \quad \mathbf{P}A[t_0] = 0, \tag{4.99}$$

where

$$\|B^{n(1)}\|_{\underline{S}^1_{c^n}[I]} + \|(B^{n(1)}_0, \mathbf{P}^\perp B^{n(1)})\|_{Y^1_{c^n}[I]} \leq 1, \quad \|B^{n(2)}\|_{S^1_{d^n}[I]} \leq 1. \tag{4.100}$$

Assume furthermore that

$$\|\mathbf{P}A\|_{S^1_a[I]} \leq 1, \quad \|B\|_{S^1_e[I]} \leq 1. \tag{4.101}$$

Then we have

$$\|\mathrm{Diff}^{\kappa}_{\mathbf{P}_x A} B\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_f[I]} \lesssim 1, \tag{4.102}$$

where

$$f_k = \Big(\sum_{k' < k - \kappa} (a_{k'} + \sum_{n=1}^{N} c^n_{k'} d^n_{k'}) \Big) e_k.$$

Next, we state a trilinear estimate for $\mathrm{Diff}^{\kappa}_{\mathbf{P}A}$ in the presence of $\mathbf{w}^2_{\mu}$, which is analogous to Proposition 4.27. This is needed for analyzing the dynamic Yang–Mills heat flow of a caloric Yang–Mills wave.

Proposition 4.29. *Let*

$$\Delta A_0 = \mathbf{w}^2_0(B^{(1)}, B^{(2)}, s), \tag{4.103}$$

$$\Box \mathbf{P} A = \mathbf{P} \mathbf{w}^2_x(B^{(1)}, B^{(2)}, s), \quad \mathbf{P} A[t_0] = 0, \tag{4.104}$$

where $B^{(1)} \in S^1[I]$, $\mathbf{P}^{\perp} B^{(1)} \in Y^1[I]$ *and* $B^{(2)} \in S^1[I]$. *Then we have*

$$\|\mathrm{Diff}^{\kappa}_{\mathbf{P}A} B\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_f[I]} \lesssim_{\tilde{M}} (\|B^{(1)}\|_{S^1_c[I]} + \|\mathbf{P}^{\perp} B^{(1)}\|_{Y^1_c[I]}) \|B^{(2)}\|_{S^1_d[I]} \|B\|_{S^1_e[I]}, \tag{4.105}$$

where

$$f_k = \Big(\sum_{k' < k - \kappa} \langle s 2^{2k'} \rangle^{-10} \langle s^{-1} 2^{-2k'} \rangle^{-\delta_2} c_{k'} d_{k'} \Big) e_k.$$

Finally, we end this subsection with auxiliary estimates for $\mathrm{Diff}^{\kappa}_{\mathbf{P}A}$, which are needed to justify approximate linear energy conservation for the paradifferential wave equation.

Proposition 4.30. *Let* $\kappa \geq 10$. *We have*

$$\| |D|^{-1} [\nabla, \mathrm{Diff}^{\kappa}_{\mathbf{P}A}] B \|_{N_{cd}} \lesssim 2^{-\delta_2 \kappa} (\|\mathbf{P} A_x\|_{S^1_c[I]} + \|D A_0\|_{L^2 \dot{H}^{\frac{1}{2}}_c[I]}) \|B\|_{S^1_d[I]}. \tag{4.106}$$

Moreover, consider the L^2*-adjoint of* $\mathrm{Diff}^{\kappa}_{\mathbf{P}A}$, *which is given by*

$$(\mathrm{Diff}^{\kappa}_{\mathbf{P}A})^* B = \sum_k P_k \partial^{\alpha} [\mathbf{P}_{\alpha} A_{<k-\kappa}, B].$$

Then we have

$$\|(\mathrm{Diff}^{\kappa}_{\mathbf{P}A})^* B - \mathrm{Diff}^{\kappa}_{\mathbf{P}A} B\|_{N_{cd}[I]} \lesssim 2^{-\delta_2 \kappa} (\|\mathbf{P} A_x\|_{S^1_c[I]} + \|D A_0\|_{L^2 \dot{H}^{\frac{1}{2}}_c[I]}) \|B\|_{S^1_d[I]} \tag{4.107}$$

5. Structure of caloric Yang–Mills waves

In this section, we use the results stated in Section 4 to study properties of subthreshold caloric Yang–Mills waves satisfying an a-priori S^1-norm bound on an interval.

5.1. Structure of a caloric Yang–Mills wave with finite S^1-norm.

The following theorem provides detailed properties of a caloric Yang–Mills wave with finite S^1-norm. It will be useful for the proof of the key regularity result (Theorem 6.1), as well as the main results stated in Section 1.3.

For a regular solution to the Yang–Mills equation in the caloric gauge, we have seen in Theorem 3.5 that (3.12), (3.13), (3.14) and (3.15) are satisfied. More generally, we say that a one-parameter family $A(t)$ ($t \in I$) of connections in $\mathcal{C}$ (which is quite rough in general) solves the Yang–Mills equation in the caloric gauge, or in short that A is a *caloric Yang–Mills wave*, if $(A, \partial_t A) \in L^{\infty}(I; T^{L^2} \mathcal{C})$ and satisfies (3.12), (3.13), (3.14) and (3.15).

Theorem 5.1. *Let* A *be a caloric Yang–Mills wave on a time interval* I *with energy* $\mathcal{E}$ *obeying*

$$A(t) \in \mathcal{C}_{\mathcal{Q}} \quad \text{for all } t \in I, \tag{5.1}$$

$$\|A\|_{S^1[I]} \leq M \tag{5.2}$$

for some $0 < \mathcal{Q}, M < \infty$. *Let* c *be a* δ_5*-frequency envelope for the initial data* $(A, \partial_t A)(t_0)$ $(t_0 \in I)$ *in* $\dot{H}^1 \times L^2$. *Then the following properties hold:*

(1) *(Linear well-posedness for* $\Box_A$*) The initial value problem for the linear equation*

$$\Box_A u = f \tag{5.3}$$

is well-posed. Moreover,

$$\|u\|_{S^1_d[I]} \lesssim_{M,\mathcal{Q}} \|(u, \partial_t u)(t_0)\|_{(\dot{H}^1 \times L^2)_d} + \|f\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_d[I]} \tag{5.4}$$

for any δ_5*-frequency envelope* d.

(2) *(Frequency envelope bound)*

$$\|A\|_{S^1_c[I]} + \|\Box_A A\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_{c^2}[I]} \lesssim_{M,\mathcal{Q}} 1. \tag{5.5}$$

(3) *(Elliptic component bounds)*

$$\|A_0\|_{Y^1_{c^2}[I]} + \|\mathbf{P}^\perp A\|_{Y^1_{c^2}[I]} \lesssim_{M,\mathcal{Q}} 1, \tag{5.6}$$

(4) *(High modulation bounds)*

$$\|\Box A\|_{\Box \underline{X}^1_{c^2}[I]} + \|\Box A\|_{X^{-\frac{1}{2}+b_1, -b_1}_{c^2}[I]} \lesssim_{M,\mathcal{Q}} 1. \tag{5.7}$$

(5) *(Paradifferential formulation) For any* $\kappa \geq 10$,

$$\|\Box A + \mathrm{Diff}^\kappa_{\mathbf{P}A} A\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_{c^2}[I]} \lesssim_{M,\mathcal{Q}} 2^{C\kappa}. \tag{5.8}$$

(6) *(Weak divisibility) There exists a partition* $I = \cup_{i \in \mathcal{I}} I_i$ *so that* $\#\mathcal{I} \lesssim_{M,\mathcal{Q}} 1$ *and*

$$\|A\|_{S^1[I_i]} \lesssim_{\mathcal{E}} 1. \tag{5.9}$$

(7) *(Persistence of regularity) If* $(A, \partial_t A)(t_0) \in \dot{H}^N \times \dot{H}^{N-1}$ $(N \geq 1)$, *then* $A \in S^N \cap S^1[I]$ *and* $A_0 \in Y^N \cap Y^1[I]$. *Moreover,*

$$\|A\|_{S^N \cap S^1[I]} + \|A_0\|_{Y^N \cap Y^1[I]} \lesssim_{M,\mathcal{Q},N} \|(A, \partial_t A)(t_0)\|_{(\dot{H}^N \times \dot{H}^{N-1}) \cap (\dot{H}^1 \times L^2)}. \tag{5.10}$$

For the subsequent properties, let $\tilde{A}$ *be another caloric Yang–Mills wave on* I *obeying the same conditions* (5.1) *and* (5.2).

(8) *(Weak Lipschitz dependence on data) For* $\sigma < 1$ *sufficiently close to* 1, *we have*

$$\|A - \tilde{A}\|_{S^\sigma[I]} \lesssim_{M,\mathcal{Q}} \|(A - \tilde{A}, \partial_t(A - \partial_t \tilde{A}))(t_0)\|_{\dot{H}^\sigma \times \dot{H}^{\sigma-1}}. \tag{5.11}$$

(9) *(Elliptic component bound for the transport equation)*

$$\|A_0\|_{(|D|^{-2} L^2_x L^1_t)_{c^2}[I]} \lesssim_{M,\mathcal{Q}} 1. \tag{5.12}$$

Moreover, if d_k *is a* δ_5*-frequency envelope for* $A - \tilde{A}$ *in* $S^1[I]$, *then*

$$\|A_0 - \tilde{A}_0\|_{(|D|^{-2} L^2_x L^1_t)_{ce}[I]} \lesssim_{M,\mathcal{Q}} 1, \tag{5.13}$$

where $e_k = c_k + c_k (c \cdot d)_{\leq k}$.

Remark 5.2. The frequency envelope bound (5.5) implies a uniform-in-time positive lower bound on the energy concentration scale r_c; see Lemma 7.8 below. As a consequence, once Theorem 1.13 is proved, finiteness of the S^1-norm would imply that solution can be continued past finite endpoints of I (We note, however, that Theorem 5.1 will be used in the proof of Theorem 1.13).

Remark 5.3. Combination of (1), (2) and divisibility of the norm $N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]$ (cf. Proposition 4.6) show that a finite S^1-norm Yang–Mills wave on I exhibits some modified scattering behavior, i.e., that each A_j tends to a homogeneous solution to the equation $\Box_A u = 0$ towards infinite endpoints of I.

We start by establishing some weaker derived bounds.

Proposition 5.4. *Let A be a caloric Yang–Mills wave on a time interval I, which obeys $A(t) \in \mathcal{C}_\mathcal{Q}$ for all $t \in I$ and $\|A\|_{S^1[I]} \leq M$. Let c be a $C\delta_5$-frequency envelope for A in $S^1[I]$, i.e., $\|A\|_{S^1_c[I]} \leq 1$.*

(1) The following derived bounds for $A_{t,x}$ hold:

$$\|A_0\|_{Y^1_{c^2}[I]} + \|\mathbf{P}^\perp A\|_{Y^1_{c^2}[I]} \lesssim_{M,\mathcal{Q}} 1, \tag{5.14}$$

$$\|\Box A\|_{\Box \underline{X}^1_{c^2}[I]} + \|\Box A\|_{X^{-\frac{1}{2}+b_1,-b_1}_{c^2}[I]} \lesssim_{M,\mathcal{Q}} 1. \tag{5.15}$$

(2) Let $\tilde{A}$ be another caloric Yang–Mills wave on I that also obeys $\|\tilde{A}\|_{S^1[I]} \leq M$. Let d be a δ_5-frequency envelope for the difference $A - \tilde{A}$ in $S^1[I]$, i.e., $\|A - \tilde{A}\|_{S^1_d[I]} \leq 1$. Then we have

$$\|A_0 - \tilde{A}_0\|_{Y^1_e[I]} + \|\mathbf{P}^\perp A - \mathbf{P}^\perp \tilde{A}\|_{Y^1_e[I]} \lesssim_{M,\mathcal{Q}} 1, \tag{5.16}$$

$$\|\Box(A - \tilde{A})\|_{\Box \underline{X}^1_e[I]} + \|\Box(A - \tilde{A})\|_{X^{-\frac{1}{2}+b_1,-b_1}_e[I]} \lesssim_{M,\mathcal{Q}} 1, \tag{5.17}$$

where $e_k = d_k + c_k (c \cdot d)_{\leq k}$.

As a quick consequence of Proposition 5.4, we see that any caloric Yang–Mills wave A with $A(t) \in \mathcal{C}_\mathcal{Q}$ for all $t \in I$ and $\|A\|_{S^1[I]} \leq M$ obeys

$$\|A\|_{\underline{S}^1[I]} \lesssim_{M,\mathcal{Q}} 1.$$

Remark 5.5. The reason why we state these weaker bounds as a separate proposition is for logical clarity. As it will be evident, the proof of Proposition 5.4 depends *only* on Propositions 4.12–4.22. In fact, after these propositions are established in Section 8, Proposition 5.4 will be used in the proofs of Proposition 4.23, Theorem 4.24 and Proposition 4.25 in Sections 8 and 9.

Proof of Proposition 5.4. Since A is a caloric Yang–Mills wave, Theorem 3.5 determines A_0, $\partial_0 A_0$ and $\mathbf{P}^\perp_j A = \Delta^{-1} \partial_j \partial^\ell A_\ell$ in terms of A. To derive the equation for $\partial_t \mathbf{P}^\perp A$, we first compute

$$\begin{aligned} \partial_t \mathbf{P}^\perp A =& \partial_t \frac{\partial_x \partial^\ell}{\Delta} A_\ell = \Delta^{-1} \partial_x \partial^\ell (F_{0\ell} + \partial_\ell A_0 + [A_\ell, A_0]) \\ =& \Delta^{-1} \partial_x (\mathbf{D}^\ell F_{0\ell} + \Delta A_0 + \partial^\ell [A_\ell, A_0] - [A^\ell, F_{0\ell}]). \end{aligned}$$

By the constraint equation, we have $\mathbf{D}^\ell F_{0\ell} = 0$. Expanding $F_{0\ell}$ in terms of $A_{t,x}$, we arrive at

$$\partial_t \mathbf{P}^\perp_j A = \partial_j A_0 + \Delta^{-1} \partial_j (\partial^\ell [A_\ell, A_0] - [A^\ell, \partial_t A_\ell] + [A^\ell, \partial_\ell A_0] - [A^\ell, [A_0, A_\ell]]). \tag{5.18}$$

The rest of the proof consists of combining Theorem 3.5 with Propositions 4.12, 4.13 and 4.22 in the right order. We first sketch the proof of the non-difference bounds (5.14)–(5.15).

We begin by verifying that

$$\||D|A_0\|_{Y_{c^2}[I]} + \||D|\mathbf{P}^\perp A\|_{Y_{c^2}[I]} \lesssim_{M,\mathcal{Q}} 1.$$

Indeed, by the mapping properties in Theorem 3.5 and the embeddings

$$L^1\dot{H}^1 \cap L^2\dot{H}^{\frac{1}{2}} \subseteq Y,$$

the contribution of $\mathbf{A}_0^3$ in A_0 and $\mathbf{DA}^3$ in $\mathbf{P}^\perp A$ are handled easily. For the quadratic nonlinearities, we apply (4.29) for A_0, (4.37) with $\sigma = 0$ for $\mathbf{P}^\perp A$ and $\sigma = 1$ for A_0.

Next, we show that

$$\|\partial_t A_0\|_{L^2\dot{H}^{\frac{1}{2}}_{c^2}[I]} + \|\partial_t \mathbf{P}^\perp A\|_{L^2\dot{H}^{\frac{1}{2}}_{c^2}[I]} \lesssim_{M,\mathcal{Q}} 1.$$

For $\partial_t A_0$, we use Theorem 3.5 for $\mathbf{DA}_0^3$ and (4.30) for the quadratic nonlinearity. For $\partial_t \mathbf{P}^\perp A$, we estimate the RHS of (5.18), where we use the $Y[I]$-norm bound for A_0 that was just established.

We now consider $\Box A$. We first prove the weaker bound

$$\|\Box A\|_{\Box\tilde{X}^1_{c^2}[I]} \lesssim_{M,\mathcal{Q}} 1. \tag{5.19}$$

By the mapping properties in Theorem 3.5 and the embeddings

$$L^1L^2 \cap L^2\dot{H}^{-\frac{1}{2}} \subseteq \Box\underline{X}^1 \cap X^{-\frac{1}{2}+b_1,-b_1} \subseteq \Box\tilde{X}^1$$

the contribution of R_j is acceptable in both cases. For the quadratic nonlinearities $\mathbf{P}\mathcal{M}^2 + \mathbf{P}^\perp\mathcal{M}^2$, and the contribution of $\Box A - \Box_A A$, we apply (4.42), (4.43), (4.74), (4.77), (4.84) and (4.85); note that we need to use (5.14) in both (4.77) and (4.84).

We are ready to prove (5.17). The desired estimate for the $\Box\underline{X}^1[I]$-norm follows by repeating the preceding argument with (4.85) replaced by (4.86), and using (5.19). On the other hand, for the $\Box X^{-\frac{1}{2}+b_1,-b_1}[I]$-norm, we replace (4.85) by (4.87) instead, and use the $\Box\underline{X}^1[I]$-norm bound that we have just proved.

Finally, the proof of the difference bounds (5.16)–(5.17) proceeds similarly, taking the difference of each of the equations (3.12)–(3.15). We leave the details to the reader. □

We now prove Theorem 5.1, using the estimates stated in Section 4.

Proof of Theorem 5.1. Throughout this proof, we omit the dependence of constants on $\mathcal{Q}$.

Proof of (1). We begin with a $\Box_A$ decomposition which will be repeatedly used in the sequel. Given $\kappa > 10$, we write

$$\Box_A = \Box + \mathrm{Diff}^\kappa_{\mathbf{P}A} - R^\kappa_A$$

where, using the decomposition in (4.67), the remainder R^κ_A is given by

$$R^\kappa_A = \mathrm{Diff}^\kappa_{\mathbf{P}^\perp A} - \mathrm{Rem}^{\kappa,2}_A - \mathrm{Rem}^{\kappa,3}_A$$

Lemma 5.6. *Let $J \subset I$. Let d be a δ_5-frequency envelope for u in $S^1[J]$. Then we have*

$$\|R^\kappa_A u\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})_d[J]} \lesssim_M \left(2^{-\delta_2\kappa}\|A\|_{S^1[J]} + 2^{C\kappa}C(A,J)\right)\|u\|_{S^1_d[J]} \tag{5.20}$$

with

$$C(A,J) = \|\mathbf{P}^\perp A\|_{Y^1[J]} + \|\mathbf{P}^\perp A\|_{\ell^1L^1L^\infty[J]} + \|A\|_{\mathrm{Str}^1[J]} + \|(\nabla\mathbf{P}^\perp A, \nabla A_0)\|_{L^2\dot{H}^{\frac{1}{2}}[J]} \tag{5.21}$$

Proof. We successively bound the three terms in R^κ_A as follows. For the first of them we have

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}^\perp A} u\|_{(N\cap L^2\dot H^{-\frac12})_d[J]} \lesssim_M (\|\mathbf{P}^\perp A\|_{Y^1[J]} + \|\mathbf{P}^\perp A\|_{\ell^1 L^1 L^\infty[J]})\|u\|_{S^1_d[J]}$$

using the bounds (4.82) and (4.83), and noting that the second norm of A is estimated using (4.37) for the quadratic part and (3.22) by

$$\|\mathbf{P}^\perp A\|_{\ell^1 L^1 L^\infty[J]} \lesssim_M 1$$

For the second term in R^κ_A in (5.22) we have

$$\|\mathrm{Rem}^{\kappa,2}_A u\|_{(N\cap L^2\dot H^{-\frac12})_d[J]} \lesssim_M (2^{-\delta_2\kappa}\|A\|_{S^1[J]} + 2^{C\kappa}C(A,J))\|u\|_{S^1_d[J]},$$

as a consequence of (4.78), (4.79) and (4.80).

Finally, for the third term in R^κ_A we have

$$\|\mathrm{Rem}^{\kappa,3}_A u\|_{(N\cap L^2\dot H^{-\frac12})_d[J]} \lesssim_M \|A\|_{\mathrm{Str}^1[J]}\|u\|_{S^1_d[J]}$$

due to (4.74). □

To prove (1) we rewrite the equation (5.3) in the form

$$(\Box + \mathrm{Diff}^\kappa_{\mathbf{P}A})u = f - R^\kappa_A u \tag{5.22}$$

The important fact is that all the A norms in $C(A,J)$ except for S^1 are divisible norms, and also controlled by M. On the other hand the S^1 norm of A has the redeeming $2^{-\delta_2\kappa}$ factor. To proceed we choose κ large enough,

$$\kappa \ll_{M,\mathcal{Q}} 1$$

Then we can subdivide the interval $I = \cup_{j\in\mathcal{J}} J_k$ so that $\#\mathcal{J} \lesssim_M 1$, and so that in each interval J_j we have smallness,

$$\|R^\kappa_A u\|_{(N\cap L^2\dot H^{-\frac12})_d[J_j]} \ll_M \|u\|_{S^1_d[J_j]} \tag{5.23}$$

A second consequence of our choice for κ is that Theorem 4.24 applies. Then we can successively apply Theorem 4.24 in each interval J_k, treating R^κ_A perturbatively.

Proof of (2). The argument here is similar to the previous one. For any interval $J \subset I$ and any $(-\delta_5, N)$ frequency envelope d for A in $S^1[J]$ we can use the bounds (4.44)-(4.49) and (3.21) to estimate

$$\|\Box_A A\|_{(N\cap L^2\dot H^{-\frac12})_d[J]} \lesssim_M \left(2^{-\delta_2\kappa}\|A\|_{S^1[J]} + 2^{C\kappa}\|A\|_{DS^1[J]}\right)\|A\|_{S^1_d[J]} \tag{5.24}$$

As before we use the divisibility of the DS^1 norm to partition the interval I into finitely many subintervals J_k, whose number depends only on M, and so that in each subinterval we have

$$2^{-\delta_2\kappa}\|A\|_{S^1[J]} + 2^{C\kappa}\|A\|_{DS^1[J]} \le \epsilon \ll_{M,\mathcal{Q}} 1.$$

We now specialize the choice of d, choosing it to be a minimal δ_5-frequency envelope for A in the first interval J_1. Applying the result in part (1) in J_1 we conclude that

$$d \lesssim_{M,\mathcal{Q}} c + \epsilon d$$

which by the smallness of ϵ implies that $d \lesssim_{M,\mathcal{Q}} c$. Then we reiterate.

Proofs of (3) and (4). These follow from (5.5) and Proposition 5.4.

Proof of (5). This is obtained by combining the bound (5.20) for $J = I$ and $u = A$ with the bound (5.24).

Proof of (6). In view of (5), this is a direct consequence of Proposition 4.25.

Proof of (7). We use frequency envelopes. It suffices to show that if c_k is a $(-\delta_5, S)$-frequency envelope for the initial data in the energy space then $C(M)c_k$ is a frequency envelope for A in S^1 and A_0 in Y^1. We begin with a version of Lemma 5.6:

Lemma 5.7. *Let $J \subset I$. Let $d = d(J)$ be a $(-\delta_5, S)$-frequency envelope for A in $S^1[J]$. Then we have*

$$\|R^\kappa_A A\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})_d[J]} \lesssim_M \left(2^{-\delta_2\kappa}\|A\|_{S^1[J]} + 2^{C\kappa}C(A,J)\right)\|A\|_{S^1_d[J]}. \tag{5.25}$$

Proof. The same argument as in the proof of (5.8) applies for the first term in R^κ_A, as there the output frequency and the u input frequency are the same. On the other hand for the two remaining terms, the frequency envelope d is inherited from the highest frequency input, see Propositions 4.19, 4.20. □

Combining the bound in the lemma with (5.24) we obtain the estimate

$$\|\Box A + \mathrm{Diff}^\kappa_{\mathbf{P}A} A\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})_d[J]} \lesssim_M \left(2^{-\delta_2\kappa}\|A\|_{S^1[J]} + 2^{C\kappa}C(A,J)\right)\|A\|_{S^1_d[J]}. \tag{5.26}$$

Now we can conclude as in the proof of (2). We first choose κ large enough so that Theorem 4.24 applies, and also so that

$$2^{-\delta_2\kappa}\|A\|_{S^1[I]} \ll_M 1.$$

Then we divide the interval I into finitely many subintervals (again, depending only on M and $\mathcal{Q}$) so that for each subinterval J we have

$$2^{C\kappa}\|A\|_{DS^1[J]} \ll_M 1.$$

Thus, for each subinterval J we have insured that

$$\|\Box A + \mathrm{Diff}^\kappa_{\mathbf{P}A} A\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})_d[J]} \ll_M \|A\|_{S^1_d[J]}.$$

Let c_k be a $(-\delta_5, S)$-frequency envelope for the initial data in the energy space, Then applying Theorem 4.24 in the first interval J_1 we conclude that

$$\|P_k A\|_{S^1[J_1]} \lesssim_{M,\mathcal{Q}} c_k + \epsilon d_k, \qquad \epsilon \ll_M 1. \tag{5.27}$$

for any $(-\delta_5, S)$ frequency envelope d_k for A in $S^1[J_1]$. In particular if d_k is a minimal $(-\delta_5, S)$ frequency envelope for A in $S^1[J_1]$ then we obtain

$$d_k \lesssim_M c_k + \epsilon d_k,$$

which leads to

$$d_k \lesssim_{M,\mathcal{Q}} c_k,$$

i.e., the desired bound in J_1. We now reiterate this bound in successive intervals J_j. Finally, the Y bound follows as in (3).

Proof of (8). Assume $0 < 1 - \sigma \ll \delta_5$. We write the equation for $\delta A = A - \tilde{A}$ in the form

$$(\Box + \mathrm{Diff}^\kappa_{\mathbf{P}\tilde{A}})\delta A = F^\kappa,$$

where

$$F^\kappa = \mathrm{Diff}^\kappa_{\mathbf{P}A-\mathbf{P}\tilde{A}} A + (R^\kappa_A A - R^\kappa_{\tilde{A}} \tilde{A}) + (\Box_A A - \Box_A \tilde{A}). \tag{5.28}$$

We claim that we can estimate the terms in F^κ as follows:

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}A-\mathbf{P}\tilde{A}} A\|_{N^{\sigma-1}\cap L^2\dot{H}^{\sigma-1-\frac{1}{2}}[J]} \lesssim_M 2^{-c_\sigma \kappa}(\|A\|_{S^1} + \|\tilde{A}\|_{S^1})\|\delta A\|_{S^\sigma[J]}, \tag{5.29}$$

$$\|R^\kappa_A A - R^\kappa_{\tilde{A}} \tilde{A}\|_{N^{\sigma-1}\cap L^2\dot{H}^{\sigma-1-\frac{1}{2}}[J]} \lesssim_M 2^{C\kappa}(C(A,J) + C(\tilde{A},J))\|\delta A\|_{S^\sigma[J]}, \tag{5.30}$$

$$\|\Box_A A - \Box_A \tilde{A}\|_{N^{\sigma-1}\cap L^2\dot{H}^{\sigma-1-\frac{1}{2}}[J]} \lesssim_M (C(A,J) + C(\tilde{A},J))\|\delta A\|_{S^\sigma[J]}. \tag{5.31}$$

We first show how to conclude the proof of (8) using (5.29), (5.30) and (5.31). As in the proofs of (1),(2) and (7), we first choose κ large enough, $\kappa \gg_M 1$. Then we use divisibility for the expressions $C(A,J)$ and $C(\tilde{A},J)$ in order to divide the interval I into subintervals J_j so that on each subinterval F^κ is perturbative, i.e.

$$\|F^\kappa\|_{N^{\sigma-1}\cap L^2\dot{H}^{\sigma-1-\frac{1}{2}}[J_j]} \ll_{M,\kappa} \|\delta A\|_{S^\sigma[J_j]}$$

Finally, we apply Theorem 4.24 successively on the intervals J_j; then (8) follows.

It remains to prove the bounds (5.29), (5.30) and (5.31). The bounds (5.30) and (5.31) are the difference counterparts of (5.25), respectively (5.24), and are proved in a very similar fashion. Details are omitted. We only remark that the requirement $\sigma < 1$ is not needed here, and that these bounds hold for any δ_5-admissible frequency envelope c_k for δA in S^1.

We now turn our attention to the novel part of the argument, which is the bound for $\mathrm{Diff}^\kappa_{\mathbf{P}A-\mathbf{P}\tilde{A}} A$. It is here that the condition $\sigma < 1$ pays a critical role. This is done in the next lemma. For later use we state the result in a more general fashion. This will be needed again in the proof of Proposition 6.4. A variation of the same argument will also be needed in Proposition 6.3.

Lemma 5.8. *Let $J \subset I$. Let c_k, d_k, b_k be frequency envelopes for $A, \tilde{A}$, respectively δA and B in $S^1[J]$. Then the expression $\mathrm{Diff}^\kappa_{\mathbf{P}A-\mathbf{P}\tilde{A}} B$ can be estimated as follows:*

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}A-\mathbf{P}\tilde{A}} B\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})_f[J]} \lesssim_{M,\mathcal{Q}} 2^{-c_\sigma \kappa}\|\delta A\|_{S^\sigma_d[J]}\|B\|_{S^1_b[J]}, \tag{5.32}$$

where f_k is given by

$$f_k = \left(\sum_{k' \leq k-\kappa} d_{k'} + c_{k'}(c\cdot d)_{\leq k'}\right) b_k. \tag{5.33}$$

Before proving the lemma we show that it implies (5.29). To measure δA in S^σ we can choose the frequency envelope d_k with the property that $2^{(\sigma-1)k}d_k$ is a $(-\delta, 1-\sigma+\delta)$ admissible envelope with $\delta < \frac{1}{2}(1-\sigma)$, $\delta \ll \delta_5$, and so that

$$\|\delta A\|^2_{S^\sigma[J]} \approx \sum_k (2^{(\sigma-1)k}d_k)^2.$$

Then we have

$$f_k \lesssim_M d_{k-\kappa}c_k \lesssim_M 2^{-\frac{1}{2}(1-\sigma)\kappa} d_k,$$

and (5.29) follows. We return to the proof of the lemma:

Proof of Lemma 5.8. We first recall the equations for $\mathbf{P}A_x$ and A_0. Following Theorem 3.5, these have the form:

$$\begin{aligned} \Box \mathbf{P} A_x &= \mathbf{P}[A^\ell, \partial_x A_\ell] - 2\mathbf{P}[A_\ell, \partial^{ell} A_x] + \mathbf{P}(R(A) + [A_\ell, [A^\ell, A_x]]), \\ \Delta A_0 &= [A^\ell, \partial_x A_\ell] + \mathbf{Q}(A, \partial_0 A) + \Delta \mathbf{A}_0^3. \end{aligned} \tag{5.34}$$

Based on this equations we consider the following decomposition of $\mathbf{P}A = (\mathbf{P}A_x, A_0)$:

$$\mathbf{P}A = (A_x^{main}, A_0^{main}) + (A_x^2, 0) + (A_x^3, A_0^3),$$

where the three components are determined by the following three sets of equations:

$$\begin{aligned} \Box A_x^{main} &= \mathbf{P}[A^\ell, \partial_x A_\ell], \qquad A_x^{main}[0] = 0, \\ \Delta A_0^{main} &= [A^\ell, \partial_x A_\ell], \end{aligned}$$

respectively $A_0^2 = 0$ and

$$\Box A_x^2 = -2\mathbf{P}[A_\ell, \partial^\ell A_x] \qquad A_x^2[0] = 0,$$

and finally

$$\begin{aligned} \Box A_x^3 &= \mathbf{P}(R(A) + \mathbf{P}[A_\ell, [A^\ell, A_x]]), \qquad A_x^3[0] = \mathbf{P}A[0], \\ \Delta A_0^3 &= \mathbf{Q}(A, \partial_0 A) + \Delta \mathbf{A}_0^3. \end{aligned} \tag{5.35}$$

We also use the same set of equations and the same decomposition for $\mathbf{P}\tilde{A}$, and take the differences δA^{main}, δA^2 respectively δA^3. We are now ready to estimate the three contributions.

The contribution of δA^{main}. For this we use the estimates in Proposition 4.27, which yield

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}A^{main} - \mathbf{P}\tilde{A}^{main}} B\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}}[J])_f} \lesssim_M 2^{-\sigma\kappa} \|\delta A\|_{S^1_d[J]} \|B\|_{S^1_b[J]}, \tag{5.36}$$

where

$$f_k = \left(\sum_{k' \leq k - \kappa} c_{k'} d_{k'} \right) b_k.$$

which suffices. For later use, we also record the following consequence of Proposition 4.15, which provides a bound for $\|\Box \delta A_x^{main}\|_{N \cap L^2 \dot{H}^{\frac{1}{2}}}$:

$$\|\delta A_x^{main}\|_{S^1_{cd}[J]} \lesssim \|\delta A\|_{S^1_d[J]} (\|A\|_{S^1_c[J]} + \|\tilde{A}\|_{S^1_c[J]}). \tag{5.37}$$

The contribution of δA^3. This is more easily dealt with using instead Proposition 4.26. We start with $A_0^3 - \tilde{A}_0^3$, which is estimated using the bounds (4.36) and (4.37) in Proposition (4.13) for the first term, respectively (3.23) for the second, by

$$\|A_0^3 - \tilde{A}_0^3\|_{(L^1 L^\infty \cap L^2 \dot{H}^{\frac{3}{2}})_{cd}[J]} \lesssim_M \|\delta A\|_{S^1_d[J]} (\|A\|_{S^1_c[J]} + \|\tilde{A}\|_{S^1_c[J]}). \tag{5.38}$$

Similarly, for $A_x^3 - \tilde{A}_x^3$ we can apply the difference bound associated to (3.21) for R_x and Strichartz estimates for the remaining cubic term to obtain

$$\|\Box(A_x^3 - \tilde{A}_x^3)\|_{(L^1 L^2 \cap L^2 \dot{H}^{-\frac{1}{2}})_{cd}[J]} \lesssim_M \|\delta A\|_{S^1_d[J]} (\|A\|_{S^1_c[J]} + \|\tilde{A}\|_{S^1_c[J]}). \tag{5.39}$$

As a consequence this also gives

$$\|A_x^3 - \tilde{A}_x^3\|_{S^1_{cd}[J]} \lesssim_M \|\delta A\|_{S^\sigma_d[J]} (\|A\|_{S^1_c[J]} + \|\tilde{A}\|_{S^1_c[J]}). \tag{5.40}$$

Using (5.38) and (5.40) in Proposition 4.26 yields the desired bound

$$\|\mathrm{Diff}^{\kappa}_{\delta A^3} B\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})_f[J]} \lesssim_{M,\mathcal{Q}} \|\delta A\|_{S^1_d[J]}\|B\|_{S^1_b[J]}(\|A\|_{S^1_c[J]} + \|\tilde{A}\|_{S^1_c[J]}) \tag{5.41}$$

with the same f_k as in the previous case.

The contribution of A^2. Here we will use Proposition 4.28. For this we need to verify its hypotheses. We begin with (4.101), for which we combine (5.37) and (5.40) to conclude that

$$\|\delta A^2_x\|_{S^1_d[J]} \lesssim_M \|\delta A\|_{S^1_d[J]}, \tag{5.42}$$

Next we consider (4.100). Using the second part of Proposition 5.4 we obtain

$$\|\delta A\|_{\underline{S}^1_e[J]} + \|(\delta A_0, \mathbf{P}^{\perp}\delta A)\|_{Y^1_e[J]} \lesssim_M \|\delta A\|_{S^1_d[J]}, \tag{5.43}$$

with

$$e_k = d_k + c_k(c\cdot d)_{<k}.$$

The last two bounds allow us to use Proposition 4.28. This yields

$$\|\mathrm{Diff}^{\kappa}_{\delta A^2} B\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})f[J])} \lesssim_{M,\mathcal{Q}} \|\delta A\|_{S^1_d[J]}\|B\|_{S^1_c[J]}(\|A\|_{S^1_c[J]} + \|\tilde{A}\|_{S^1_c[J]}) \tag{5.44}$$

where

$$f_k = \left(\sum_{k'\leq k-\kappa} d_{k'} + e_{k'}d_{k'}\right) b_k.$$

The proof of the lemma is now concluded. ☐

Proof of (9). This is a direct consequence of the bounds (4.39) and (3.23) for the quadratic part $\mathbf{A}^2_0$ of A_0, respectively its cubic and higher part $\mathbf{A}^3_0$. ☐

5.2. **Caloric Yang–Mills waves with small energy dispersion on a short interval.** Next, we consider the effect of small inhomogeneous energy dispersion on a time interval with compatible scale.

Theorem 5.9. *Let A be a caloric Yang–Mills wave on a time interval I with energy $\mathcal{E}$, obeying* (5.1), (5.2) *as well as the smallness relations*

$$\|F\|_{ED_{\geq 0}[I]} \leq \epsilon, \qquad |I| \leq \epsilon. \tag{5.45}$$

Let c be a δ_5-frequency envelope for A in $S^1[I]$. Then for sufficiently small $\epsilon > 0$ depending on M and $\mathcal{Q}$, the following properties hold:

(1) (Small energy dispersion below scale 1 *for A)*

$$\|A\|_{ED^1_{\geq 0}[I]} \lesssim_{\mathcal{E},\mathcal{Q}} \epsilon^{\delta_2} \tag{5.46}$$

(2) (Elliptic component bounds)

$$\|A_0\|_{Y^1_c[I]} + \|\mathbf{P}^{\perp}A\|_{Y^1_c[I]} \lesssim_{M,\mathcal{Q}} \epsilon^{\delta_2}. \tag{5.47}$$

(3) (High modulation bounds)

$$\|\Box A\|_{L^2\dot{H}^{-\frac{1}{2}}_c[I]} \lesssim_{M,\mathcal{Q}} \epsilon^{\delta_2} \tag{5.48}$$

(4) (Paradifferential formulation)

$$\|\Box A + \mathrm{Diff}^{\kappa}_{\mathbf{P}A} A\|_{(N\cap L^2\dot{H}^{-\frac{1}{2}})_c[I]} \lesssim_{M,\mathcal{Q}} \epsilon^{\delta_4} 2^{C\kappa} \tag{5.49}$$

(5) (Approximate linear energy conservation) For any $t_1, t_2 \in I$,

$$\big|\|\nabla A(t_1)\|_{L^2}^2 - \|\nabla A(t_2)\|_{L^2}^2\big| \lesssim_{M,\mathcal{Q}} \epsilon^{\delta_4} \tag{5.50}$$

(6) (Approximate conservation of $\mathcal{Q}$*) For any* $t_1, t_2 \in I$,

$$|\mathcal{Q}(A(t_1) - \mathcal{Q}(A(t_2))| \lesssim_{\mathcal{E},\mathcal{Q}} \epsilon^{\delta_4} \tag{5.51}$$

Proof. Again, we omit the dependence of constants on $\mathcal{Q}$. The property that will be used here repeatedly is (4.21), which asserts that all non-sharp Strichartz norms are small. We recall it here for convenience:

$$\sup_k \|P_k F\|_{\mathrm{Str}} \lesssim_M \epsilon^{\delta_1} \lesssim \epsilon^{\delta_2}. \tag{5.52}$$

Proof of (1). This is a consequence of the caloric bound (3.7) applied with $d_k = \epsilon$.

Proof of (2). We repeat the arguments in the proof of Proposition 5.4.(1). The bounds for the cubic and higher terms in Theorem 3.5 use only the Strichartz Str^1 norms, so the contributions of $\mathbf{A}_0^3$ in A_0, $\mathbf{DA}^3$ in $\mathbf{P}^\perp A$ and $\mathbf{DA}_0^3$ in $\partial_t A_0$ are easily estimated. For the quadratic terms we replace (4.29) with (4.33) in the case of A_0, and then (4.37) with (4.38) in the case of $\mathbf{P}^\perp A$ and $\partial_t A_0$; again the smallness comes from Str^1.

Proof of (3). We consider the terms in the A_x equation in Theorem 3.5. The cubic terms R_x and $[A_\ell, [A^\ell, A]]$ are estimated only in terms of $\|A\|_{\mathrm{Str}^1}$. For the quadratic terms we use instead the bounds (4.30), (4.36), (4.63) and (4.65); all smallness come from Str^1.

Proof of (4). We first establish the similar bound for $\Box_A A$, which is given by the equation (3.12). For the quadratic terms we use (4.50) and (4.51). For the cubic term we use (3.21). Hence it remains to estimate the difference

$$R_A^\kappa A = \mathrm{Diff}^\kappa_{\mathbf{P}^\perp A} A - \mathrm{Rem}_A^{\kappa,2} A - \mathrm{Rem}_A^{\kappa,3} A.$$

For the first term we use (4.83), where the ϵ smallness comes from the $L^1 L^\infty$ norm of $\mathbf{P}^\perp A$ due to the bounds (4.38), respectively (3.22) for the quadratic, respectively the cubic part of $A^\perp$.

For the second term we use the bound (4.81). The second term on the right is small due to (5.47), so we obtain

$$\|\mathrm{Rem}_A^{\kappa,2} A\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_c} \lesssim_M (2^{-\delta_2 \kappa} + 2^{C\kappa} \epsilon^{\delta_2}) \|A\|_{\underline{S}^1_c}.$$

Now we observe that on the right we can replace κ with any $\kappa' > \kappa$ without any change in the proof. Then it suffices to optimize with respect to κ'.

For the third term we use directly (4.74).

Proof of (5). This statement is a corollary of (5.49). For the proof, we introduce the *linear energy*

$$E_{lin}(A)(t) = \frac{1}{2} \int_{\mathbb{R}^4} \sum_{\mu=0}^{4} |\partial_\mu A(t)|^2 \, dx.$$

Given any interval $I' = (t_1, t_2) \subseteq I$, we consider

$$\mathcal{I} = \int_{\mathbb{R} \times \mathbb{R}^4} \chi_{I'} \langle (\Box + \mathrm{Diff}^\kappa_{\mathbf{P}A}) A, \partial_t A \rangle \, dt dx.$$

Integrating by parts, we may rewrite

$$\begin{aligned}\mathcal{I} =& E_{lin}(A)(t_1) - E_{lin}(A)(t_2)\\ &+ \frac{1}{2}\int \langle \mathrm{Diff}^{\kappa}_{\mathbf{P}A} A, A\rangle(t_2)\, dx - \frac{1}{2}\int \langle \mathrm{Diff}^{\kappa}_{\mathbf{P}A} A, A\rangle(t_1)\, dx\\ &- \frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}^4} \chi_{I'} \langle [\partial_t, \mathrm{Diff}^{\kappa}_{\mathbf{P}A}] A, A\rangle \, dtdx\\ &+ \frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}^4} \chi_{I'} \langle (\mathrm{Diff}^{\kappa}_{\mathbf{P}A} - (\mathrm{Diff}^{\kappa}_{\mathbf{P}A})^*) A, \partial_t A\rangle \, dtdx.\end{aligned}$$

By Proposition 4.30 and the straightforward bound

$$\int \langle \mathrm{Diff}^{\kappa}_{\mathbf{P}A} A, A\rangle(t) \lesssim 2^{-\kappa} \|(A, A_0)(t)\|_{\dot{H}^1} \|\nabla A(t)\|_{L^2}^2 \lesssim_M 2^{-\kappa},$$

we see that

$$|\mathcal{I} - (E_{lin}(A)(t_1) - E_{lin}(A)(t_2))| \lesssim_M 2^{-c\kappa}. \tag{5.53}$$

On the other hand, by duality, we may put $\chi_{I'}(\Box + \mathrm{Diff}^{\kappa}_{\mathbf{P}A})A$ and $\chi_{I'}\partial_t A$ in N and N^*, respectively. Then by Proposition 4.6, (5.2) and (5.49), we have

$$|\mathcal{I}| \lesssim_M \epsilon^{\delta_4} 2^{C\kappa}. \tag{5.54}$$

Optimizing the choice of κ, (5.50) follows.

Proof of (6). We will use the caloric flow in order to compare $\mathcal{Q}(A(t_1))$ and $\mathcal{Q}(A(t_2))$. Denote by $A(t,s)$ the caloric flow of A. We will split the difference in three as

$$\mathcal{Q}(A(t_1)) - \mathcal{Q}(A(t_2)) = \mathcal{Q}(A(t_1,1) - \mathcal{Q}(A(t_2,1) + \mathcal{Q}(A(t_1)) - \mathcal{Q}(A(t_1,1)) - \mathcal{Q}(A(t_2)) + \mathcal{Q}(A(t_2,1))$$

For the first difference we estimate at parabolic time $s = 1$ as follows:

$$\begin{aligned}|\mathcal{Q}(A(t_1,1)) - \mathcal{Q}(A(t_2,1))| &\lesssim \int_{t_1}^{t_2} \int_{\mathbb{R}^4} \frac{d}{dt} |F(s,t,x)|^3 dxdt\\ &\lesssim \int_{t_1}^{t_2} \int_{\mathbb{R}^4} |F(1,t,x)|^2 |\partial_t F(1,x,t)| dxdt\\ &\lesssim \int_{t_1}^{t_2} \int_{\mathbb{R}^4} |F(s,t,x)|^2 |\partial_t F| dxdt\\ &\lesssim_{\mathcal{E},\mathcal{Q}} |t_1 - t_2| c_1^3,\end{aligned}$$

where at the last step we have simply used the fixed time L^2 bounds given by Proposition 3.1(1) and Bernstein's inequality. Now we gain smallness from the time interval.

For the remaining two differences we only need fixed time estimates, which for reference we state in the following

Lemma 5.10. *Let $a \in \mathcal{C}$ be a caloric connection with energy $\mathcal{E}$ and $\mathcal{Q}(A) = \mathcal{Q}$, and A its caloric Yang–Mills flow.*

a) Assume that a is energy dispersed at high frequencies,

$$\|f\|_{ED_{\geq m}} \leq \epsilon. \tag{5.55}$$

Then for its caloric Yang–Mills heat flow $A(s)$ we have

$$\mathcal{Q}(a) - \mathcal{Q}(A(2^{-2m})) \lesssim_{\mathcal{E},\mathcal{Q}} \epsilon^c. \tag{5.56}$$

b) If a is fully energy dispersed,

$$\|f\|_{ED} \leq \epsilon, \tag{5.57}$$

then we have

$$\mathcal{Q}(a) \lesssim_{\mathcal{E},\mathcal{Q}} \epsilon^c. \tag{5.58}$$

Proof. a) By scaling we can set $m = 0$. Denote by c_k a frequency envelope for f in L^2, and by d_k a frequency envelope for f in $\dot{W}^{-2,\infty}$. By the energy dispersion bound we have $d_k \leq \epsilon$ for $k \geq 0$. By Proposition 3.2 we have the L^2 bound

$$\|P_k F\|_{L^2} \lesssim_{\mathcal{E},\mathcal{Q}} c_k \langle 2^{2k} s\rangle^{-N},$$

respectively the L^∞ bound

$$\|P_k F\|_{L^\infty} \lesssim_{\mathcal{E},\mathcal{Q}} 2^{2k} d_k^{\frac{1}{2}} \langle 2^{2k} s\rangle^{-N}.$$

We use these bounds to estimate the difference

$$\begin{aligned}
\mathcal{Q}(a) - \mathcal{Q}(A(1)) &= \int_0^1 \int_{\mathbb{R}^4} |F(s,t,x)|^3 dx ds \\
&\lesssim \sum_{k_1 \leq k_2 \leq k_3} \int_0^1 \int_{\mathbb{R}^4} |P_{k_1} F(s,t,x)| |P_{k_2} F(s,t,x)| |P_{k_3} F(s,t,x)| dx ds \\
&\lesssim_{\mathcal{E},\mathcal{Q}} \sum_{k_1 \leq k_2 \leq k_3} \frac{1}{1 + 2^{2k_3}} 2^{2k_1} d_{k_1}^{\frac{1}{2}} c_{k_2} c_{k_3} \\
&\lesssim \sum_{1 \leq k_3} d_{k_3}^{\frac{1}{2}} c_{k_3}^2 \\
&\lesssim \epsilon^{\frac{1}{2}}
\end{aligned}$$

where at the next to last step we have used both the low frequency decay and the off-diagonal decay for the summation in k_1 and k_2.

b) This follows by letting $m \to -\infty$ in part (a). The proof of the Lemma is concluded. □

The proof of (5.51) is also concluded. □

5.3. **The dynamic Yang–Mills heat flow of a caloric Yang–Mills wave.** Here we investigate the structure of the dynamic Yang–Mills heat flow of a caloric Yang–Mills wave A with finite S^1-norm. As before, we consider two cases: (1) when A only obeys a finite S^1-norm bound; and (2) when A has small inhomogeneous energy dispersion on a short time interval of compatible scale.

In the general case, we have the following structure theorem.

Theorem 5.11. *Let A be a caloric Yang–Mills wave with energy $\mathcal{E}$ on a time interval I, obeying* (5.1) *and* (5.2). *Let $A_{t,x}(s)$ be the dynamic Yang–Mills heat flow of $A_{t,x}$ at heat-time $s > 0$ in the caloric gauge. Then the following properties hold:*

(1) *(Fixed-time bounds) For any $t \in I$, let $c^{(0)}(t)$ be a δ_5-frequency envelope for $\nabla A(t)$ in L^2. Then*

$$\|P_k(\nabla A(s) - \nabla e^{s\Delta} A)(t)\|_{L^2} \lesssim_{\mathcal{E},\mathcal{Q}} \langle 2^{-2k} s^{-1}\rangle^{-\delta_4} \langle 2^{2k} s\rangle^{-10} c_k^{(0)}(t)^2, \tag{5.59}$$

$$\|P_k \partial^\ell A_\ell(t,s)\|_{L^2} \lesssim_{\mathcal{E},\mathcal{Q}} \langle 2^{2k} s\rangle^{-10} c_k^{(0)}(t)^2, \tag{5.60}$$

$$\|P_k \nabla A_0(t,s)\|_{L^2} \lesssim_{\mathcal{E},\mathcal{Q}} \langle 2^{2k} s\rangle^{-10} c_k^{(0)}(t)^2, \tag{5.61}$$

$$\|P_k \Box A(t,s)\|_{\dot{H}^{-1}} \lesssim_{\mathcal{E},\mathcal{Q}} \langle 2^{2k} s\rangle^{-10} c_k^{(0)}(t)^2. \tag{5.62}$$

(2) *(Frequency envelope bounds) Let c be a δ_5-frequency envelope for A in $S^1[I]$. Then*

$$\|P_k(A(s) - e^{s\Delta} A)\|_{\underline{S}^1[I]} \lesssim_{M,\mathcal{Q}} \langle 2^{-2k} s^{-1}\rangle^{-\delta_4} \langle 2^{2k} s\rangle^{-10} c_k^2, \tag{5.63}$$

$$\|P_k A_0(s)\|_{Y^1[I]} \lesssim_{M,\mathcal{Q}} \langle 2^{2k} s\rangle^{-10} c_k^2, \tag{5.64}$$

$$\|P_k \mathbf{P}^\perp A(s)\|_{Y^1[I]} \lesssim_{M,\mathcal{Q}} \langle 2^{2k} s\rangle^{-10} c_k^2. \tag{5.65}$$

(3) *(Derived difference bounds) Let $\tilde{A}$ be a caloric Yang–Mills wave on I obeying $\|\tilde{A}\|_{S^1[I]} \leq \tilde{M}$, and let d be a δ_5 frequency envelope for the difference $A(s) - \tilde{A}$ in $S^1[I]$. Then*

$$\begin{aligned} &\|P_k(A_0(s) - \tilde{A}_0)\|_{Y^1[I]} + \|P_k(\mathbf{P}^\perp A(s) - \mathbf{P}^\perp \tilde{A})\|_{Y^1_d[I]} \\ &\lesssim_{M,\tilde{M},\mathcal{Q}} e_k + \min\{1, (s^{-\frac{1}{2}}|I|)^{\delta_4}\} \langle 2^{-2k} s^{-1}\rangle^{-\delta_4} \langle 2^{2k} s\rangle^{-10} c_k^2, \end{aligned} \tag{5.66}$$

$$\begin{aligned} &\|P_k \Box(A(s) - \tilde{A})\|_{\Box \underline{X}^1[I]} + \|P_k \Box(A(s) - \tilde{A})\|_{X^{-\frac{1}{2}+b_1,-b_1}[I]} \\ &\lesssim_{M,\tilde{M},\mathcal{Q}} e_k + \min\{1, (s^{-\frac{1}{2}}|I|)^{\delta_4}\} \langle 2^{-2k} s^{-1}\rangle^{-\delta_4} \langle 2^{2k} s\rangle^{-10} c_k^2, \end{aligned} \tag{5.67}$$

where $e_k = d_k + c_k (c \cdot d)_{\leq k}$.

Remark 5.12. Combining (5.63) with the obvious bound for $e^{s\Delta} A$, we get the simple bound

$$\|P_k A(s)\|_{\underline{S}^1[I]} \lesssim_{M,\mathcal{Q}} \langle 2^{2k} s\rangle^{-10} c_k. \tag{5.68}$$

Next, we consider the effect of small inhomogeneous energy dispersion on a time interval of compatible scale.

Theorem 5.13. *Let A be a caloric Yang–Mills wave with energy $\mathcal{E}$ on a time interval I, obeying* (5.1), (5.2) *and* (5.45), *and $A_{t,x}(s)$ be the dynamic Yang–Mills heat flow of $A_{t,x}$ at heat-time $s > 0$ in the caloric gauge. Let c be a δ_5-frequency envelope for A in $S^1[I]$. Then the following properties hold:*

(1) *(Fixed-time smallness bound)*

$$\|\nabla P_k(A(s) - e^{s\Delta} A)(t)\|_{L^2} \lesssim_{\mathcal{E},\mathcal{Q}} 2^{\delta_4 (m-k)_+} \epsilon^{\delta_4} \langle 2^{-2k} s^{-1}\rangle^{-\delta_4} \langle 2^{2k} s\rangle^{-10} c_k^{(0)}(t), \tag{5.69}$$

$$\|P_k \partial^\ell A_\ell(t,s)\|_{L^2} \lesssim_{\mathcal{E},\mathcal{Q}} 2^{\delta_4 (m-k)_+} \epsilon^{\delta_4} \langle 2^{2k} s\rangle^{-10} c_k^{(0)}(t). \tag{5.70}$$

(2) *(Small energy dispersion below scale 1 for $A(s)$)*

$$\|A(s)\|_{ED^{-1}_{\geq 0}[I]} \lesssim_{\mathcal{E},\mathcal{Q}} \epsilon^{\delta_4}. \tag{5.71}$$

(3) *(Frequency envelope bounds)*

$$\|P_k(A(s) - e^{s\Delta}A)\|_{\underline{S}^1[I]} \lesssim_{M,\mathcal{Q}} \epsilon^{\delta_4} \langle 2^{-2k}s^{-1}\rangle^{-\delta_4} \langle 2^{2k}s\rangle^{-10} c_k, \tag{5.72}$$

$$\|P_k A_0(s)\|_{Y^1[I]} \lesssim_{M,\mathcal{Q}} \epsilon^{\delta_4} \langle 2^{2k}s\rangle^{-10} c_k, \tag{5.73}$$

$$\|P_k \mathbf{P}^{\perp} A(s)\|_{Y^1[I]} \lesssim_{M,\mathcal{Q}} \epsilon^{\delta_4} \langle 2^{2k}s\rangle^{-10} c_k. \tag{5.74}$$

(4) *(Derived difference bounds) Let $\tilde{A}$ be a caloric Yang–Mills wave on I with $\|\tilde{A}\|_{S^1[I]} \leq \tilde{M}$, and let d be a δ_5-frequency envelope for the difference $A(s) - \tilde{A}$ in $S^1[I]$. Then*

$$\begin{aligned} &\|P_k(A_0(s) - \tilde{A}_0)\|_{Y^1[I]} + \|P_k(\mathbf{P}^{\perp}A(s) - \mathbf{P}^{\perp}\tilde{A})\|_{Y^1_d[I]} \\ &\lesssim_{M,\tilde{M},\mathcal{Q}} e_k + \epsilon^{\delta_4} \langle 2^{-2k}s^{-1}\rangle^{-\delta_4} \langle 2^{2k}s\rangle^{-10} c_k, \end{aligned} \tag{5.75}$$

$$\begin{aligned} &\|P_k \Box(A(s) - \tilde{A})\|_{\Box \underline{X}^1[I]} + \|P_k \Box(A(s) - \tilde{A})\|_{X^{-\frac{1}{2}+b_1, -b_1}[I]} \\ &\lesssim_{M,\tilde{M},\mathcal{Q}} e_k + \epsilon^{\delta_4} \langle 2^{-2k}s^{-1}\rangle^{-\delta_4} \langle 2^{2k}s\rangle^{-10} c_k, \end{aligned} \tag{5.76}$$

where $e_k = d_k + c_k(c \cdot d)_{\leq k}$.

We now turn to the proof of each theorem.

Proof of Theorem 5.11. In the proof, we omit the dependence of constants on M and $\mathcal{Q}$. We introduce the notation

$$\mathbf{A}(t,s) = A(t,s) - e^{s\Delta}A(t).$$

Proof of (1). By (3.2) in Proposition 3.1 (note that $\partial_t A$ here corresponds to B in the the proposition) we get

$$\|\nabla P_k \mathbf{A}(t,s)\|_{L^2[I]} \lesssim \langle 2^{-2k}s^{-1}\rangle^{-\delta_1} \langle 2^{2k}s\rangle^{-10} (c_k^{(0)})^2. \tag{5.77}$$

Now the second bound follows from (3.18) for $\mathbf{DA}^3$ and Proposition 4.13 for $Q(A,A)$.

Proof of (2). We proceed in several substeps.

Step (2).1. Our first (and main) goal is to prove

$$\|P_k \mathbf{A}(s)\|_{S^1[I]} \lesssim \langle 2^{-2k}s^{-1}\rangle^{-c\delta_3} \langle 2^{2k}s\rangle^{-10} c_k^2. \tag{5.78}$$

We begin by invoking (3.4) with $(\sigma, p) = (\frac{1}{4}, 4)$ and $(\sigma_1, p_1) = (\frac{1}{2}, 2)$. Since $S^1[I] \subseteq \mathrm{Str}^1[I] \subseteq L^4 \dot{W}^{\frac{1}{4},4}[I]$, we also obtain (after taking $L^2_t[I]$)

$$\|\nabla P_k \mathbf{A}(s)\|_{L^2 \dot{H}^{\frac{1}{2}}[I]} \lesssim \langle 2^{-2k}s^{-1}\rangle^{-\delta_1} \langle 2^{2k}s\rangle^{-10} c_k^2. \tag{5.79}$$

In view of the embedding $P_k L^2 \dot{H}^{\frac{1}{2}}[I] \subseteq P_k X_1^{0,\frac{1}{2}}[I] \subseteq 2^{-k} S_k[I]$, we have

$$\|\nabla P_k \mathbf{A}(s)\|_{S_k[I]} \lesssim \langle 2^{-2k}s\rangle^{-\delta_1} \langle 2^{2k}\rangle^{-10} c_k^2. \tag{5.80}$$

To complete the proof of (5.78), it only remains to establish (recall (4.2))

$$\|\Box P_k \mathbf{A}(s)\|_{L^2 \dot{H}^{-\frac{1}{2}}[I]} \lesssim \langle 2^{-2k}s\rangle^{-\delta_1} \langle 2^{2k}\rangle^{-10} c_k^2. \tag{5.81}$$

We argue differently depending on $s2^{2k} \gtrsim 1$ or $s2^{2k} \ll 1$. In the former case, we consider $e^{s\Delta}A$ and $A(s)$ separately. In view of (5.7), note that

$$\|\Box P_k e^{s\Delta} A\|_{L^2 \dot{H}^{-\frac{1}{2}}[I]} \lesssim \langle 2^{2k}\rangle^{-10} c_k^2,$$

so it suffices to prove

$$\|\Box P_k A(s)\|_{L^2 \dot{H}^{-\frac{1}{2}}[I]} \lesssim \langle 2^{2k} \rangle^{-10} c_k^2.$$

For this, we need to use the wave equation for $A(s)$ (cf. Theorem 3.6):

$$\begin{aligned} \Box A(s) =& (\Box - \Box_{A(s)}) A(s) + \mathcal{M}^2(A(s), A(s)) + R_j(A(s)) \\ &+ \mathbf{P}\mathbf{w}_x^2(A, A, s) + R_{j;s}(A) \end{aligned} \tag{5.82}$$

As in the proof of Proposition 5.4, we note that $\Box - \Box_{A(s)}$ contains the terms $A_0(s)$, $\partial^\ell A(s)$ and $\partial_0 A_0(s)$ that are in turn determined by $A, A(s)$ (cf. Theorem 3.6). By (5.80) and an obvious bound for $e^{s\Delta} A$, we see that $\langle 2^{2k} s \rangle^{-10} c_k$ is a frequency envelope for $A(s)$ in $\mathrm{Str}^1[I]$. The desired estimate is proved by applying the $L^2 L^2$-type estimates in Section 4 (observe that they only involve the Str^1-norm of A!) and Theorem 3.6.

In the case $s 2^{2k} \ll 1$, we begin by writing $\mathbf{A}(s) = (A(s) - A) + (1 - e^{s\Delta})A$. For the second term, again by (5.7), we have

$$\|\Box P_k (1 - e^{s\Delta}) A\|_{L^2 \dot{H}^{-\frac{1}{2}}[I]} \lesssim \langle 2^{-2k} s^{-1} \rangle^{-\delta_1} c_k^2.$$

Thus, for $s 2^{2k} \ll 1$, it suffices to establish

$$\|\Box P_k (A(s) - A)\|_{L^2 \dot{H}^{-\frac{1}{2}}[I]} \lesssim \langle 2^{-2k} s^{-1} \rangle^{-c\delta_3} c_k^2. \tag{5.83}$$

Here, we use the equation $\Box(A(s) - A)$ obtained by taking the difference of the equations in Theorems 3.5 and 3.6:

$$\begin{aligned} \Box(A(s) - A) =& (\Box - \Box_{A(s)}) A(s) - (\Box - \Box_A) A \\ &+ \mathcal{M}^2(A(s), A(s)) - \mathcal{M}^2(A, A) \\ &+ R_j(A(s)) - R_j(A) \\ &+ \mathbf{P}_j \mathbf{w}_x^2(A, A, s) + R_{j;s}(A). \end{aligned} \tag{5.84}$$

We note that $(\Box - \Box_{A(s)}) A(s) - (\Box - \Box_A) A$ contains the differences $A_0(s) - A_0$, $\partial_\ell^\ell A(s) - \partial^\ell A_\ell$ and $\partial_0 A_0(s) - \partial_0 A_0$, for which similar difference equations may be derived from Theorems 3.5 and 3.6.

As before, c_k is a δ_5-frequency envelope for A and $A(s)$ in $\mathrm{Str}^1[I]$, whereas $d_k = \langle 2^{-2k} s^{-1} \rangle^{-c\delta_3} c_k$ is a δ_3-frequency envelope for $A(s) - A$ in $\mathrm{Str}^1[I]$ by (5.80) and an obvious bound for $(1 - e^{s\Delta})A$. Hence the difference envelope e_k in Theorem 3.5 obeys the bound

$$e_k = d_k + c_k (c \cdot d)_{\leq k} \lesssim \langle 2^{-2k} s^{-1} \rangle^{-c\delta_3} c_k.$$

The desired estimate (5.83) is proved by applying the $L^2 L^2$-type estimates in Section 4 (again, they only involve the Str^1-norm of ∇A, $\nabla A(s)$ and $\nabla(A(s) - A)$) and Theorem 3.6.

Step (2).2. To complete the proof, it remains to show that (5.78) implies (5.63)–(5.65). This is proved in a completely analogous way as Proposition 5.4.(1), replacing Theorem 3.5 by Theorem 3.6 (where we use Propositions 4.16, 4.17 for $\mathbf{w}_0$ and $\mathbf{w}_x$, respectively).

Proof of (3). This is analogous to the proof of Proposition 5.4.(1). The only difference in the analysis arises from the extra terms

(i) $\mathbf{P}_j w_x^2(\partial_t A, \partial_t A, s) + R_{j;s}(A)$ in $\Box_{A(s)} A(s)$,
(ii) $\mathbf{A}_{0;s} = \Delta^{-1} \mathbf{w}_0^2(A, A, s) + \mathbf{A}_{0;s}^3(A)$ in $A_0(s)$,
(iii) $\mathbf{D}\mathbf{A}_{0;s}(A)$ in $\partial_t A_0(s)$.

For the first term in (5.75) we need to estimate

$$\||D|^{-1}\mathbf{w}_0^2(A,A,s)\|_Y + \||D|\mathbf{A}^3_{0;s}(A)\|_Y + \|\mathbf{D}\mathbf{A}_{0;s}(A)\|_Y$$

The last two terms are estimated directly using (3.36) and (3.37) and Bernstein's inequality. The first term is estimated via (4.54).

For the extra gain when $s^{\frac{1}{2}} > |I|$ we rebalance by using Holder in time t and Bernstein in x. Because of this, in that range it suffices to use $L^\infty L^2$ bounds instead of Y, and thus rely instead on (3.33) and (3.34), respectively (4.52).

For the second term in (5.75) we follow the computation for $\partial_t \mathbf{P}^\perp A(s)$ in the proof of Proposition 5.4. The extra contributions there are

$$\Delta^{-1}\partial_j(\partial^\ell[A_\ell(s),\mathbf{A}_{0;s}] + [A^\ell(s),\partial_\ell \mathbf{A}_{0;s}] + [A^\ell,[A_\ell,\mathbf{A}_{0;s}]]).$$

For these it suffices to use (4.53) and (3.36) for long intervals I, respectively (4.52) and (4.52) and (3.33) for short intervals.

Finally, for the two terms in (5.76) we need to bound

$$\|\mathbf{P}_j w_x^2(\partial_t A,\partial_t A,s)\|_{\Box \underline{X}^1 \cap X^{-\frac{1}{2}+b+1,-b_1}} + \|R_{j;s}(A)\|_{\Box \underline{X}^1 \cap X^{-\frac{1}{2}+b+1,-b_1}}$$

For this it suffices to use the bounds (4.58) and (3.35) in the range $|I| > s^{\frac{1}{2}}$, respectively (4.56) and (3.32) in the range $|I| \le s^{\frac{1}{2}}$. □

Proof of Theorem 5.13. As before, we omit the dependence of constants on M and $\mathcal{Q}$.

Proof of (1) and (2). The three bounds follow directly from Proposition 3.2, precisely in order from the estimates (3.8), (3.9) and (3.7).

Proof of (3). We repeat the arguments in the proof of Theorem 5.11.(2). The bound (5.79) for $P_k\mathbf{A}(s)$ goes through the Str^1 norm so by the same proof we also obtain for $k \ge 0$

$$\|\nabla P_k \mathbf{A}(s)\|_{L^2\dot{H}^{\frac{1}{2}}[I]} \lesssim \langle 2^{-2k}s^{-1}\rangle^{-c\delta_3}\langle 2^{2k}s\rangle^{-10}\epsilon^{\delta_2}c_k. \tag{5.85}$$

On the other hand for $k \le 0$ we can use (5.69) and Holder's inequality in time to gain smallness.

Similarly, the bound (5.81) also uses only Str^1 norms so it can be replaced by

$$\|\Box P_k \mathbf{A}(s)\|_{L^2\dot{H}^{-\frac{1}{2}}[I]} \lesssim \langle 2^{-2k}s^{-1}\rangle^{-c\delta_3}\langle 2^{2k}\rangle^{-10}\epsilon^{\delta_2}c_k. \tag{5.86}$$

for $k \ge 0$. Again for $k \le 0$ we can use a simpler $L^\infty \dot{H}^{-1}$ bound and then Holder's inequality in time. Together, the bounds (5.85) and (5.86) imply (5.72).

Finally, it remains to establish (5.73) and (5.74). Here the same considerations as in the proof of (5.47) apply, but using Theorem 3.6 instead of Theorem 3.5, as well as Proposition 4.16.

Proof of (4). This repeats the proof of Theorem 5.11.(3), but taking advantage of the Str^1 norm in estimating $\mathbf{A}^3_{0;s}$ and $\mathbf{D}\mathbf{A}_{0;s}$ and using (4.55) instead of (4.54). As before, the ϵ gain is due to energy dispersion if $k \ge 0$ and to the interval size otherwise. □

6. Energy dispersed caloric Yang–Mills waves

The goal of this section is to prove the following key theorem for energy dispersed subthreshold caloric Yang–Mills waves, which is essentially a restatement of Theorem 1.20 in terms of the linear energy:

Theorem 6.1. *There exist a non-decreasing positive functions $M(E, \mathcal{Q})$ and non-increasing positive functions $\epsilon(E, \mathcal{Q})$ and $T(E, \mathcal{Q})$ so that the following holds. Let A be a regular caloric Yang–Mills wave on a time interval I satisfying*

$$\inf_{t \in I} \|\nabla A(t)\|_{L^2}^2 \leq E, \quad A(t) \in \mathcal{C}_{\mathcal{Q}} \quad \textit{for all } t \in I. \tag{6.1}$$

If A moreover obeys the smallness bounds

$$\|F\|_{ED_{\geq m}[I]} \leq \epsilon(E, \mathcal{Q}), \quad |I| \leq 2^{-m} T(E, \mathcal{Q}), \tag{6.2}$$

then we have

$$\|A\|_{S^1[I]} \leq M(E, \mathcal{Q}). \tag{6.3}$$

We next show that Theorem 1.16 immediately follows. Indeed, for caloric waves we have (see Theorem 1.6)

$$\|\nabla A\|_{L^2} \lesssim_{\mathcal{E}, \mathcal{Q}} 1$$

as well as

$$\mathcal{E} \lesssim_{\|\nabla A\|_{L^2}} 1.$$

Thus the linear and nonlinear energy are interchangeable in the statement of the theorem. The (minor) difference is that the nonlinear energy is exactly conserved, whereas the linear energy is only approximately conserved for energy dispersed Yang–Mills waves, see Theorem 5.9.(5).

For the remainder of this section, we fix $\mathcal{Q}$. We omit any dependence of constants on $\mathcal{Q}$ and write $\epsilon(E) = \epsilon(E, \mathcal{Q})$, $T(E) = T(E, \mathcal{Q})$, $M = M(E, \mathcal{Q})$ etc.

Theorem 6.1 is proved by an induction on energy argument of similar structure to [23] and [17]. For the initial step, we show that it holds for small E (Proposition 6.2). For the induction step, we assume that the result holds for all solutions with $\inf_I E_{lin}(A) \leq E$, and we seek to show that it holds up to $\inf_I E_{lin}(A) \leq E + c(E)$ for some small $c(E) > 0$. Notably, in order to continue the induction argument, we do not want $c(E)$ to depend on $F(E)$ or $\epsilon(E)$.

6.1. **Induction on energy argument.**

As remarked earlier, the initial step of the proof of Theorem 6.1 is essentially small energy global regularity for the Yang–Mills equation in the caloric gauge, which is a quick consequence of Theorem 5.1.

Proposition 6.2. *There exists a small universal constant $E_* > 0$ (in particular, independent of I) such that if a classical caloric Yang–Mills connection satisfies*

$$\inf_{t \in I} \|\nabla A(t)\|_{L^2}^2 \leq E_*, \tag{6.4}$$

then we have

$$\|A\|_{S^1[I]} \lesssim \sqrt{E_*}. \tag{6.5}$$

Proof. We will follow a standard continuity argument, similar to the one used in the Coulomb gauge in [11]. Start from a near minimum t_0 for $\|\nabla A(t)\|^2_{L^2}$. Denote by c a frequency envelope for $A[t_0]$ in $\dot{H}^1 \times L^2$. For a short time, there exists a classical solution, which satisfies

$$\|A\|_{S^1[I]} \lesssim E_*$$

We now consider the maximal interval I containing t_0 and where the solution A exists as a classical solution and satisfies

$$\|A\|_{S^1[I]} \leq 1 \tag{6.6}$$

This in particular implies

$$Q(A) \lesssim 1$$

Hence by Theorem 5.1.(2) it follows that

$$\|A\|_{S^1_c[I]} \lesssim 1$$

and in particular

$$\|A\|_{S^1[I]} \lesssim E_* \tag{6.7}$$

Assume now by contradiction that I has a finite end T. The S^1 (6.6) bound implies that A is uniformly bounded near $t = T$ and has a limit as a classical solution. Hence it can be extended further as a classical solution (for a precise statement see in particular Theorem 7.6). However, in view of (6.7), if E_* is sufficiently small then by continuity we can find a larger interval $I \subsetneq J$ where (6.6) holds. This is a contradiction. It follows that the solution A is global and satisfies (6.7). □

For the induction step, consider a regular caloric Yang–Mills wave A on I such that

$$E < \inf_{t \in I} \|\nabla A(t)\|^2_{L^2} \leq E + c(E), \qquad \|F\|_{ED_{\geq 0}(I)} \leq \epsilon, \qquad |I| \leq T. \tag{6.8}$$

Our goal is to establish a uniform bound

$$\|A\|_{S^1[I]} \leq M \tag{6.9}$$

for appropriately chosen $c(E) > 0$ (depending *only* on E), ϵ, T and M (which may depend on E, $\epsilon(E)$, $T(E)$, $M(E)$ and $c(E)$).

Once this goal is achieved, we may extend $M(E)$, $\epsilon(E)$ and $T(E)$ to $[0, E + c(E)]$ so that $M(E + c(E)) = M$, $\epsilon(E + c(E)) = \epsilon$ and $T(E + c(E)) = T$, while keeping validity of Theorem 6.1 in this range of energy. Since $c(E)$ is a positive number depending only on E, this procedure can be continued until Theorem 6.1 holds for all regular subthreshold caloric Yang–Mills waves.

We now turn to the proof of (6.9). By translating and reversing t, we may assume without any loss of generality that $I = [0, T_+)$ for some $T_+ > 0$ and

$$E < \|\nabla A(0)\|^2_{L^2} \leq E + 2c(E).$$

Since A is regular, it can be easily seen that $\|A\|_{S^1[0,T)}$ is a continuous function of T satisfying

$$\limsup_{T \to 0+} \|A\|_{S^1[0,T)} \lesssim \|\nabla A(t)\|_{L^2} \lesssim E^{\frac{1}{2}}.$$

Therefore, on a subinterval $J = [0, T) \subseteq I$, we may make the bootstrap assumption

$$\|A\|_{S^1[J]} \leq 2M. \tag{6.10}$$

In order to improve (6.10) to (6.9), we compare A with a caloric Yang–Mills wave $\tilde{A}$ with $S^1[I]$-norm $\leq M(E)$ (eventually), which we construct as follows.

To begin with, we view the space-time connection $A_{t,x}$ on $I \times \mathbb{R}^4$ as a caloric initial data and solve the dynamic Yang–Mills heat flow in the local caloric gauge, i.e.,

$$\begin{aligned}\partial_s A_\mu(t,x,s) =& \mathbf{D}^k F_{k\mu}(t,x,s),\\ A_\mu(t,x,0) =& A_\mu(t,x).\end{aligned}$$

From the results in Section 3, we obtain a global-in-heat-time solution $A_{t,x}(t,x,s)$ on $I \times \mathbb{R}^4 \times [0,\infty)$. Note that $\partial_t A$ solves the linearized Yang–Mills heat flow in local caloric gauge, and we have $(A, \partial_t A)(t,s) \in T^{L^2}\mathcal{C}$ for every $(t,s) \in I \times [0,\infty)$.

By the caloric gauge condition, the linear energy $\|(A,\partial_t A)(t,s)\|^2_{\dot{H}^1 \times L^2} = \|\nabla A(t,s)\|^2_{L^2}$ eventually tends to zero as $s \to \infty$. Thus there exists a heat-time $s'_* > 0$ such that

$$\|(A,\partial_t A)(0,s)\|^2_{\dot{H}^1 \times L^2} = E.$$

To eliminate ambiguity, we take s'_* to be the minimum such heat-time. In order to choose the cut-off heat-time s_*, we distinguish two scenarios:

(1) If $s'_* \geq 1$, then we define $s_* = 1$.
(2) If $s'_* < 1$, then we define $s_* = s'_*$.

With s_* chosen as above, we define $\tilde{A}$ to be the caloric Yang–Mills wave with initial data

$$(\tilde{A}, \partial_t \tilde{A})(0) = (A, \partial_t A)(0, s_*).$$

In both scenarios, we aim to prove that $\tilde{A}$ exists on J and is well-approximated by $A(s_*)$. Moreover, by the induction hypothesis, $\tilde{A}$ should obey a nice S^1-norm bound.

Proposition 6.3. *Let $\tilde{A}$ be defined as above. For sufficiently small $\epsilon, T > 0$ depending on M, $M(E)$, $T(E)$, $\epsilon(E)$ and $c(E)$, the regular caloric Yang–Mills wave $\tilde{A}$ exists on the interval J and obeys*

$$\|\tilde{A}\|_{S^1[J]} \leq M(E) + C_0\sqrt{E}, \tag{6.11}$$

$$\|A(s_*) - \tilde{A}\|_{\underline{S}^1_{c^*}[J]} \lesssim_M \epsilon^{\delta_6}, \tag{6.12}$$

$$\|A_0(s_*) - \tilde{A}_0\|_{Y^1_{c^*}[J]} \lesssim_M \epsilon^{\delta_6}, \tag{6.13}$$

$$\|\mathbf{P}^\perp A(s_*) - \mathbf{P}^\perp \tilde{A}\|_{Y^1_{c^*}[J]} \lesssim_M \epsilon^{\delta_6}, \tag{6.14}$$

where C_0 is a universal constant and c^* is a frequency envelope defined as

$$c^*_k = 2^{-\delta_*|k-k(s_*)|}. \tag{6.15}$$

On the other hand, viewing A as a "high frequency perturbation" of $\tilde{A}$, we show below that A stays close to $\tilde{A}$ in the space S^1.

Proposition 6.4. *Let $\tilde{A}$ be defined as above on the interval J. Provided that $c = c(E) > 0$ is chosen small enough compared to E (but independent of $M(E)$, $T(E)$ or $\epsilon(E)$) and $T, \epsilon > 0$ are also sufficiently small depending on M, $M(E)$, $T(E)$, $\epsilon(E)$ and $c(E)$, we have*

$$\|A - \tilde{A}\|_{S^1[J]} \lesssim_{M(E),E} 1. \tag{6.16}$$

Assuming the preceding two propositions, we may choose M sufficiently large compared to $M(E)$ and E, then choose ϵ and T accordingly, so that the desired estimate (6.9) follows from (6.11) and (6.16).

It remains to prove Propositions 6.3 and 6.4, which are the subjects of Sections 6.2 and 6.3, respectively.

6.2. **Control of $\tilde{A} - A(s_*)$: Proof of Proposition 6.3.** We introduce the notation

$$\delta A^{low} = \tilde{A} - A(s_*). \tag{6.17}$$

We proceed differently depending on how s_* was chosen.

Scenario (1): $s_* = 1 (\leq s_*')$. This scenario is simpler to handle, and we do not need to invoke the induction hypothesis.

Step (1).1: S^1-norm bound for $\tilde{A}$. We first prove the S^1-norm bound (6.11). The idea is to exploit smoothing property of the Yang–Mills heat flow, which implies control of higher Sobolev norms of $(\tilde{A}, \partial_t \tilde{A})(0) = (A, \partial_t A)(0, 1)$ in terms of $\sqrt{E}$, and use subcritical local regularity of Yang–Mills in the caloric gauge, which works in a time interval of length $O_E(1)$.

Fix a large integer N (say $N = 10$). We claim that $\tilde{A}$ exists on J and

$$\|\tilde{A}\|_{S^N \cap S^1[J]} \lesssim \sqrt{E}, \tag{6.18}$$

provided that T is sufficiently small depending only on E (so that $|J| \ll_E 1$).

By the smoothing property for the Yang–Mills heat flow and its linearization in the caloric gauge (see Section 3), we have

$$\|(\tilde{A}, \partial_t \tilde{A})(0)\|_{(\dot{H}^N \times \dot{H}^{N-1}) \cap (\dot{H}^1 \times L^2)} \lesssim \sqrt{E}.$$

For T sufficiently small (depending only on E), the following local-in-time a-priori estimates at subcritical regularity hold:

$$\sup_{t \in J} \|(\tilde{A}, \partial_t \tilde{A})(t)\|_{(\dot{H}^N \times \dot{H}^{N-1}) \cap (\dot{H}^1 \times L^2)} + |J| \|\Box \tilde{A}\|_{L^\infty (\dot{H}^{N-1} \cap L^2)[J]} \lesssim \sqrt{E},$$
$$\sup_{t \in J} \|(\tilde{A}_0, \partial_t \tilde{A}_0)(t)\|_{(\dot{H}^N \times \dot{H}^{N-1}) \cap (\dot{H}^1 \times L^2)} \lesssim \sqrt{E}.$$

The proof is via Theorem 3.5 and, as usual, the Sobolev embedding into L^∞; we omit the details.

As a consequence of the preceding a-priori bounds, we obtain (6.18) as desired. Moreover, by Theorem 3.5 and the fixed-time bounds in Section 4, we have

$$\|\Box \tilde{A}\|_{L^\infty \dot{H}^{-1}[J]} \lesssim_E 1. \tag{6.19}$$

Step (1).2: S^1-norm bound for $A(s_*) - \tilde{A}$. As a preparation for the proof of (6.12), we claim that

$$\|A(s_*) - \tilde{A}\|_{S^1_{c^*}[J]} \lesssim_M \epsilon^c. \tag{6.20}$$

In the present case, $2^{k(s_*)} = 1$. For frequencies higher than 1, we simply use (6.18) with smoothing estimates for $A(s_*)$ in S^1. For frequencies lower than 1, we control $\Box(\tilde{A} - A(s_*))$ in $L^\infty \dot{H}^{-1}$ and integrate in time.

By Theorem 5.11, we have

$$\|P_k A(s_*)\|_{S^1[J]} \lesssim_M 2^{-20k_+}, \tag{6.21}$$

$$\|P_k \Box A(t,s)\|_{\dot{H}^{-1}} \lesssim_M 2^{-20k_+}. \tag{6.22}$$

Let $\kappa_0 \geq k(s_*)$ be a parameter to be fixed below. By (6.20) and (6.21), we have

$$\|P_k \delta A^{low}\|_{S^1[J]} \leq \|P_k \tilde{A}\|_{S^1[J]} + \|P_k A(s_*)\|_{S^1[J]} \lesssim_M 2^{-c\kappa_0} c_k^* \qquad \text{for } k \geq \kappa_0, \tag{6.23}$$

where $0 < c \ll 1$ is a universal constant. Since

$$P_k(L^\infty \dot{H}^{-1}[J]) \hookrightarrow |J| 2^k N \cap (|J|2^k)^{\frac{1}{2}} L^2 \dot{H}^{-\frac{1}{2}},$$

for $k \leq \kappa_0$ it follows from (6.19) and (6.22) that

$$\begin{aligned}\|P_k \Box \delta A^{low}\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})[J]} &\leq \|P_k \Box \tilde{A}\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})[J]} + \|P_k \Box A(s_*)\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})[J]} \\ &\lesssim_M ((|J|2^{\kappa_0})^{\frac{1}{2}} + (|J|2^{\kappa_0}) + \epsilon^c) c_k^*.\end{aligned}$$

Since $\delta A^{low}[0] = 0$, we arrive at

$$\|P_k \delta A^{low}\|_{S^1[J]} \lesssim_M ((|J|2^{\kappa_0})^{\frac{1}{2}} + (|J|2^{\kappa_0}) + \epsilon^c) c_k^* \qquad \text{for } k \leq \kappa_0. \tag{6.24}$$

Step (1).3: Completion of proof. Finally, the bounds (6.12)–(6.14) follow from (6.20) and Theorem 5.11.(3) with $d_k = c_k^*$ provided that $|J| \leq T$ is sufficiently small. Here, note that

$$e_k = c_k^* + c_k (c \cdot c^*)_{\leq k} \lesssim_M c_k^*.$$

Scenario (2): $s_* = s_*' > 1$. In the second scenario, we analyze the equation satisfied by the difference $\delta A^{low} = A(s_*) - \tilde{A}$ to prove (6.12), then make use of the induction hypothesis to derive (6.11). By another continuous induction in time, we may make the following extra bootstrap assumptions:

$$\|\tilde{A}\|_{S^1[J]} \leq 2(M(E) + C_0 \sqrt{E}), \tag{6.25}$$

as well as

$$\|\delta A^{low}\|_{S^1_{c^*}[J]} \leq \epsilon^{c\delta_6}. \tag{6.26}$$

Here we use a smaller power of ϵ, so this last bound will only serve to insure some a-priori smallness of δA^{low} in $S^1_{c^*}$.

By Theorem 5.13, we have

$$\|P_k A(s_*)\|_{S^1[J]} \lesssim_M c_k \langle 2^{2k} s_* \rangle^{-10}, \tag{6.27}$$

$$\|A(s_*)\|_{ED^1_{\geq 0}[J]} \lesssim_E \epsilon^{\delta_4}, \tag{6.28}$$

$$\|\Box A(s_*)\|_{L^2 \dot{H}^{-\frac{1}{2}}[J]} \lesssim_M \epsilon^{\delta_4}. \tag{6.29}$$

Therefore, $(A(s_*), J)$ is (ε, M_*)-energy dispersed for $M_* \lesssim_M 1$ and $\varepsilon \leq \epsilon^{\delta_4}$.

Step (2).1: Bounds for δA^{low}. Here we establish (6.12). We write an equation for δA^{low} of the form

$$\Box_{\tilde{A}} \delta A^{low} = F, \qquad \delta A^{low}[0] = 0$$

We claim that in each subinterval J_1 of J and for each $\kappa > 10$ we have the bound

$$\|F\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}})_{c^*}[J_1]} \lesssim_M (2^{-c\delta_* \kappa} \|\tilde{A}\|_{S^1[J_1]} + 2^{C\kappa} C(\tilde{A}, J_1)) \|\delta A^{low}\|_{S^1_{c^*}[J_1]} + \epsilon^{\delta_6}, \tag{6.30}$$

where $C(\tilde{A}, J_1)$ contains only divisible norms of $\tilde{A}$, see (5.21).

We first verify that the bound (6.30) implies (6.12). Using the well-posedness for the $\Box_{\tilde{A}}$ equation, given by Theorem 5.1, in the time interval $J_1 = [t_1, t_2]$, we obtain the bound

$$\|\delta A^{low}\|_{S^1_{c^*}[J_1]} \leq C(M)(\|\delta A^{low}[t_1]\|_{\mathcal{H}_{c^*}} + (2^{-c\delta_*\kappa}\|\tilde{A}\|_{S^1[J_1]} + 2^{C\kappa}C(\tilde{A}, J_1))\|\delta A^{low}\|_{S^1_{c^*}[J_1]} + \epsilon^{\delta_6}).$$

For this to be useful we need to insure that the coefficient of $\|\delta A^{low}\|_{S^1_{c^*}[J_1]}$ on the right is small. To achieve that we first choose κ large enough, $k \gg_M 1$, depending only on M, so that

$$C(M)2^{-c\delta_*\kappa}\|\tilde{A}\|_{S^1[J]} \ll 1$$

Then we divide the interval J into subintervals J_j so that

$$C(M)2^{C\kappa}C(\tilde{A}, J_j) \ll 1$$

The number of such intervals depends only on M. On each subinterval $J_j = [t_{j-1}, t_j]$ we have the bound

$$\|\delta A^{low}\|_{S^1_{c^*}[J_1]} + \|\delta A^{low}[t_j]\|_{\mathcal{H}_{c^*}} \leq C(M)(\|\delta A^{low}[t_{j-1}]\|_{\mathcal{H}_{c^*}} + \epsilon^{\delta_6}).$$

Reiterating this we obtain (6.12).

If remains to prove the bound (6.30). We relabel J_1 by J for simplicity. As a preliminary step, we observe that, by Theorem 5.13 and the bootstrap assumption (6.26), we have

$$\|\delta A^{low}\|_{\underline{S}^1_{c^*}[J]} + \|\delta A^{low}_0\|_{Y^1_{c^*}[J]} + \|\mathbf{P}^{\perp}\delta A^{low}\|_{Y^1_{c^*}[J]} \lesssim_M \|\delta A^{low}\|_{S^1_{c^*}[J]}. \tag{6.31}$$

In particular, this proves the bounds (6.13) and (6.14) once (6.12) is known.

The expression for F is obtained from Theorems 3.5 and 3.6,

$$F := \Box_{\tilde{A}}\delta A^{low} = \Box_{\tilde{A}}\tilde{A} - \Box_{A(s_*)}A(s_*) + (\Box_{A(s_*)} - \Box_{\tilde{A}})A(s_*),$$

where we further expand the two terms as

$$\begin{aligned}\Box_{\tilde{A}}\tilde{A} - \Box_{A(s_*)}A(s_*) =& \mathcal{M}^2(\tilde{A}, \tilde{A}) - \mathcal{M}^2(A(s_*), A(s_*)) + R(\tilde{A}) - R(A(s_*)) \\ &+ \mathbf{P}w^2_x(\partial_t A, \partial_t A, s) + R_{j;s}(A),\end{aligned}$$

respectively

$$\begin{aligned}(\Box_{A(s_*)} - \Box_{\tilde{A}})A(s_*) =& -\mathrm{Diff}^{\kappa}_{\mathbf{P}\delta A^{low}}A(s_*) - \mathrm{Diff}^{\kappa}_{\mathbf{P}^{\perp}\delta A^{low}}A(s_*) - \mathrm{Rem}^{\kappa,2}_{\delta A^{low}}A(s_*) \\ &+ (\mathrm{Rem}^3(A(s_*)) - \mathrm{Rem}^3(\tilde{A}))A(s_*) + \mathrm{Rem}^3_{s_*}(A)A(s_*).\end{aligned}$$

We successively estimate the terms above as in (6.30).

(a) For $\mathcal{M}^2(\tilde{A}, \tilde{A}) - \mathcal{M}^2(A(s_*), A(s_*))$ we use the estimate (4.50). We inherit the envelope c_* from δA^{low} but we also gain an additional power of ϵ from the energy dispersion of $A(s_*)$.

(b) For $R(\tilde{A}) - R(A(s_*))$ we use the difference version of the bound (3.21), with a similar gain.

(c) For $\mathbf{P}w^2_x(\partial_t A, \partial_t A, s)$ we use (4.59), taking advantage of the energy dispersion for A.

(d) For $R_{j;s}(A)$ we use (3.35), gaining a power of ϵ from the Str^1 norm.

(e) For $\mathrm{Diff}^{\kappa}_{\mathbf{P}^{\perp}\delta A^{low}}A(s_*)$ we use (4.82) combined with (6.31) for the high modulations, respectively (4.83) combined with (4.37) and (3.22) for low modulations.

(f) For $\mathrm{Rem}^{\kappa,2}_{\delta A^{low}}A(s_*)$ we use (4.81).

(g) For $(\mathrm{Rem}^3(A(s_*)) - \mathrm{Rem}^3(\tilde{A}))A(s_*)$ we use (4.74).

(h) For $\mathrm{Rem}^3_{s_*}(A)A(s_*)$ we use (4.76).

This leaves us with the most difficult term $\mathrm{Diff}^\kappa_{\mathbf{P}\delta A^{low}} A(s_*)$, for which we claim that

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}\delta A^{low}} A(s_*)\|_{(N\cap L^2\dot H^{-\frac12})_{c^*}[J]} \lesssim_M 2^{-c\delta_*\kappa}\|\delta A^{low}\|_{S^1[J]}. \tag{6.32}$$

For $\mathbf{P}\delta A^{low}$ we consider the same type of decomposition as in the proof of Lemma 5.8,

$$\mathbf{P}\delta A^{low} = \mathbf{P}\delta A^{low,main} + \mathbf{P}\delta A^{low,main,2} + \mathbf{P}\delta A^{low,rem,2} + \mathbf{P}\delta A^{low,rem,3}$$

where

$$\delta A_0^{low,main} = \Delta^{-1}\left([\tilde A, \partial_t \tilde A] - [A(s_*), \partial_t A(s_*)]\right).$$

$$\delta A_0^{low,main,2} = \Delta^{-1}\mathbf{w}_0(A, A, s),$$

$$\delta A_0^{low,rem,2} = 2\Delta^{-1}\left(\mathbf{Q}(\tilde A, \partial_t \tilde A) - \mathbf{Q}(A(s_*), \partial_t A(s_*))\right),$$

$$\delta A_0^{low,rem,3} = A_0^3(\tilde A, \partial_t \tilde A) - A_0^3(A(s_*), \partial_t A(s_*)) + A^3_{0;s}(A, \partial_t A)$$

respectively

$$\delta A_x^{low,main} = \Box^{-1}\left(\mathbf{P}\mathcal{M}^2(\tilde A, \tilde A) - \mathbf{P}\mathcal{M}^2(A(s_*), A(s_*))\right)$$

$$\delta A_x^{low,main,2} = \Box^{-1}\mathbf{P}\mathbf{w}_x(A, A, s),$$

$$\delta A_x^{low,rem,2} = \Box^{-1}\mathbf{P}\left([\tilde A_\alpha, \partial^\alpha \tilde A] - [A_\alpha(s_*), \partial^\alpha A(s_*)]\right),$$

$$\begin{aligned}\delta A_x^{low,rem,3} = {} & \Box^{-1}\mathbf{P}\left(R(\tilde A) - R(A(s_*)) - \mathrm{Rem}^3(\tilde A)\tilde A + \mathrm{Rem}^3(A(s_*))A(s_*)\right) \\ & + \Box^{-1}\mathbf{P}\left(R_{j;s}(A) - \mathrm{Rem}^3_s(A)A(s_*)\right).\end{aligned}$$

where $\Box^{-1}$ is the wave parametrix with zero Cauchy data at $t = 0$.

As a preliminary observation we note that

$$\|\delta A_x^{low,main}\|_{S^1_{c^*}} + \|\delta A_x^{low,main,2}\|_{S^1_{c^*}} + \|\delta A_x^{low,rem,2}\|_{S^1_{c^*}} + \|\delta A_x^{low,rem,3}\|_{S^1_{c^*}} \lesssim_M \|\delta A^{low}\|_{S^1_{c^*}} + \epsilon^{\delta_2} \tag{6.33}$$

This is a consequence of (4.42) for the first term, (4.59) and (5.47) for the second, respectively (3.21), (3.35), (4.74) and (4.76) for the last term. The bound for the third term follows indirectly since they all add up to δA^{low}.

Now we consider the contributions of each of these terms to $\mathrm{Diff}^\kappa_{\mathbf{P}\delta A^{low}} A(s_*)$.

a) The contributions of $\delta A_x^{low,main}$ and $\delta A_0^{low,main}$. These are considered together, and estimated using Proposition 4.27. This yields the frequency envelope

$$f_k = \left(\sum_{k' < k-\kappa} c^*_{k'} c_{k'} \langle 2^{2k'} s_*\rangle^{-N}\right) c_k \langle 2^{2k'} s_*\rangle^{-N} \|\delta A^{low}\|_{S^1_{c^*}[J]} \lesssim_M 2^{-c\delta_*\kappa} c^*_k \|\delta A^{low}\|_{S^1_{c^*}[J]},$$

as needed.

b) The contributions of $\delta A_x^{low,main,2}$ and $\delta A_0^{low,main,2}$. These are also considered together, but now we want to use Proposition 4.29. As they involve no δA^{low} differences, we need to estimate these contributions by ϵ^{δ_6}. Unfortunately Proposition 4.29 provides no source for an energy dispersion gain, so we use a subterfuge, decomposing

$$\mathrm{Diff}^{\kappa}_{\delta A^{low,main,2}} A(s_*) = \mathrm{Diff}^{\kappa'}_{\delta A^{low,main,2}} A(s_*) + \mathrm{Diff}^{[\kappa',\kappa]}_{\delta A^{low,main,2}} A(s_*)$$

where $\kappa' > \kappa$ is a secondary parameter to be chosen shortly. For the first term we apply Proposition 4.29, which yields

$$\|\mathrm{Diff}^{\kappa'}_{\delta A^{low,main,2}} A(s_*)\|_{(N\cap L^2 \dot H^{-\frac12})_{c^*}[J]} \lesssim_M 2^{-c\delta_* \kappa'}$$

For the second term, on the other hand, we use instead the bounds (4.55) and (4.59), which capture both the c^* decay and the energy dispersion. The price to pay is that this way we only have access to the S^1 norm of $\delta A^{low,main,2}$, so we are only allowed to use (4.77). This yields

$$\|\mathrm{Diff}^{[\kappa',\kappa]}_{\delta A^{low,main,2}} A(s_*)\|_{(N\cap L^2 \dot H^{-\frac12})_{c^*}[J]} \lesssim_M \epsilon^{c\delta_c \delta_g} 2^{C\kappa'}$$

We now add the last two bounds and then optimize in κ' to obtain the desired estimate

$$\|\mathrm{Diff}^{\kappa}_{\delta A^{low,main,2}} A(s_*)\|_{(N\cap L^2 \dot H^{-\frac12})_{c^*}[J]} \lesssim_M \epsilon^{\delta_h}.$$

c) The contribution of $\delta A^{low,rem,2}$. The $\delta A_x^{low,rem,2}$ part is estimated using Proposition 4.28, with (6.33) serving to verify the hypothesis. For the output this yields the frequency envelope

$$f_k = \left(\sum_{k'<k-\kappa} c^*_{k'}\right) c_k \langle 2^{2k'} s_*\rangle^{-N} \lesssim_M 2^{-c\delta_* \kappa} c^*_k.$$

A simpler analysis applies for the contribution of $\delta A_0^{low,rem,2}$ where we can use Proposition 4.13.

d) The contribution of $\delta A^{low,rem,3}$. For the contribution of $\delta A_0^{low,rem,3}$ we use (3.23) respectively (3.36), while for the contribution of $\delta A_x^{low,rem,3}$ where we use (3.21), (3.35), (4.74) and (4.76), all combined with Proposition 4.26.

Step (2).2: S^1-norm bound for $\tilde A$ via induction hypothesis.

Taking ϵ sufficiently small and using the bootstrap assumption (6.26), we may ensure that

$$\|\tilde F\|_{ED_{\geq 0}[J]} \leq \epsilon(E). \tag{6.34}$$

By the induction hypothesis, we may thus assume that

$$\|\tilde A\|_{S^1[J]} \leq M(E). \tag{6.35}$$

6.3. **Control of $A - \tilde A$: Proof of Proposition 6.4.** Here, we seek to bound

$$\delta A^{high} = A - \tilde A.$$

We begin by observing that

$$\|\tilde A\|_{ED^{-1}_{\geq 0}[J]} + \|\Box \tilde A\|_{L^2 \dot H^{-\frac12}[J]} \lesssim_M \epsilon^{\delta_6}.$$

Therefore, both (A, J) and $(\tilde A, J)$ are (ε, M)-dispersed, where $\varepsilon \lesssim_M \epsilon^{\delta_6}$.

Step 1: Consequence of approximate linear energy conservation. We claim that

$$\sup_{t\in J} \|(\delta A^{high}, \partial_t \delta A^{high})(t)\|^2_{\dot H^1\times L^2} \lesssim c(E) + C_M \varepsilon^{\delta_6}. \tag{6.36}$$

Note that

$$\delta A^{high} = (1 - e^{s_*\Delta})A + e^{s_*\Delta}A - A(s_*) + A(s_*) - \tilde{A}.$$

We begin with the inequality

$$\|\nabla A(t)\|_{L^2}^2 \geq \|\nabla(1 - e^{s_*\Delta})A(t)\|_{L^2}^2 + \|e^{s_*\Delta}A(t)\|_{L^2}^2,$$

which follows from Plancherel and non-negativity of the symbol of $(1 - e^{s_*\Delta})e^{s_*\Delta}$. On the one hand, by Theorem 5.13.(1) and (6.12), we have

$$\|\nabla e^{s_*\Delta}A(t)\|_{L^2}^2 = \|\nabla \tilde{A}(t)\|_{L^2}^2 + C_M\epsilon^{\delta_6}, \tag{6.37}$$

$$\|\nabla(1 - e^{s_*\Delta})A(t)\|_{L^2}^2 = \|\nabla(A - \tilde{A})(t)\|_{L^2}^2 + C_M\epsilon^{\delta_6}. \tag{6.38}$$

Hence, by Theorem 5.9.(5), we have

$$\begin{aligned}\|\nabla(A - \tilde{A})(t)\|_{L^2}^2 \leq& \|\nabla A(t)\|_{L^2}^2 - \|\nabla\tilde{A}(t)\|_{L^2}^2 + C_M\epsilon^{\delta_6}\\ \leq& \|\nabla A(0)\|_{L^2}^2 - \|\nabla\tilde{A}(0)\|_{L^2}^2 + C_M\epsilon^{\delta_6}\\ \leq& c(E) + C_M\epsilon^{\delta_6}.\end{aligned}$$

Step 2: Weak divisibility and reinitialization. By Theorem 5.1.(7) there exists a partition $J = \cup_{k=1}^K J_k$ such that $K \lesssim_{M(E)} 1$ and

$$\|\tilde{A}\|_{S^1[J_k]} \lesssim_E 1, \tag{6.39}$$

so that the number of such intervals is also controlled $K \lesssim_{M(E)} 1$. Using the uniform control of the energy of δA^{high} in Step 1, it suffices to estimate δA^{high} in S^1 separately in each of these intervals.

We will make a bootstrap assumption

$$\|\delta A^{high}\|_{S^1[J_k]} \leq 2. \tag{6.40}$$

Then our goal is to improve (6.40) to

$$\|\delta A^{high}\|_{S^1[J_k]} \leq 1. \tag{6.41}$$

by taking $c \ll_E 1$, $\epsilon \ll_M 1$ and $T \ll_{M,\epsilon} 1$.

In view of (6.39) and (6.40), in all the estimates below within a single interval J_k, all implicit constants will depend on E rather than $M(E)$. To simplify the notations we drop the subscript and replace J_k by J in what follows.

Step 3: Frequency envelope bounds. Let c_k be a frequency envelope for A in $S^1[J]$. Then by Proposition 3.1, the initial data in J_k for $A(s)$ has the frequency envelope $2^{-(k-k^*)_+}c_k$. By Theorem 5.1, we have a similar envelope in S^1,

$$\|P_k\tilde{A}(s)\|_{S^1[J[} \lesssim_E 2^{-(k-k^*)_+}c_k. \tag{6.42}$$

On the other hand, by the estimate (6.12) we have, under the assumption $\epsilon \ll_E 1$, the bound

$$\|P_k(\tilde{A} - A(s))\|_{S^1[J]} \lesssim_E 2^{-\delta_*|k-k^*|}c_k. \tag{6.43}$$

Hence for the high frequency difference A^h we have the bound

$$\|P_k\delta A^{high}\|_{S^1[J]} \lesssim_E 2^{-\delta_*(k-k^*)_-}c_k. \tag{6.44}$$

Step 4: Control of nonlinearity. By Theorem 5.9.(4) applied separately to A and $\tilde{A}$ we have

$$\|(\Box + \mathrm{Diff}^{\kappa}_{\mathbf{P}A})\delta A^{high} + \mathrm{Diff}^{\kappa}_{\mathbf{P}\delta A^{high}} \tilde{A}\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[J]} \lesssim_E 2^{C\kappa} \epsilon^{\delta_4 \delta_6}. \tag{6.45}$$

where the parameter $\kappa \geq 10$ is arbitrary for now, to be chosen later. We claim that the second term can be estimated separately as

$$\|\mathrm{Diff}^{\kappa}_{\mathbf{P}\delta A^{high}} \tilde{A}\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[J]} \lesssim_E 2^{-c\delta_* \kappa}. \tag{6.46}$$

This is a consequence of Lemma 5.8. To see that we use the bounds (6.42) and (6.44) to compute the frequency envelope f_k in Lemma 5.8. We have

$$f_k \lesssim_E \left(\sum_{k' < k - \kappa} 2^{-c\delta_* (k' - k^*)_-} c_{k'} + 2^{-(k' - k^*)_+} c_{k'} (c^2 c^*)_{<k'} \right) 2^{-(k-k^*)_+} c_k \lesssim_E 2^{-c\delta_* |k - k^*|} c_k,$$

and thus (6.46) follows. Combining (6.45) with (6.46) yields

$$\|(\Box + \mathrm{Diff}^{\kappa}_{\mathbf{P}A})\delta A^{high}\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[J]} \lesssim_E 2^{-c\delta_* \kappa} + 2^{C\kappa} \epsilon^{\delta_4 \delta_6}. \tag{6.47}$$

Hence by Theorem 5.1.(1) we conclude that

$$\|\delta A^{high}\|_{S^1[J_k]} \lesssim_E c + 2^{-c\delta_* \kappa} + 2^{C\kappa} \epsilon^{\delta_4 \delta_6}.$$

Hence by taking $\kappa \gg_E 1$, $c \ll_E 1$, $\epsilon \ll_{E,\kappa} 1$ and $T \ll_{E,\epsilon,\kappa} 1$ the desired conclusion (6.41) follows.

7. Proof of the main results

The purpose of this short section is to deduce Theorems 1.13, 1.20 and 1.18 from Theorem 6.1.

7.1. **Higher regularity local well-posedness.**

In this subsection, we sketch the proof of higher regularity local well-posedness of the hyperbolic Yang–Mills equation. We first use the temporal gauge, which works for general connections, and then turn to the caloric gauge, which works for data satisfying (1.12).

7.1.1. *Temporal gauge.* Here we write the Yang–Mills equations in the temporal gauge,

$$A_0 = 0 \tag{7.1}$$

They take the form

$$\Box_A A_j = \mathbf{D}^k \partial_j A_k \tag{7.2}$$

with the additional constraint equation

$$\mathbf{D}^j \partial_0 A_j = 0 \tag{7.3}$$

This can be viewed as a semilinear system of wave equations for the curl of A, coupled with a second order transport equation for the divergence of A.

We consider the Cauchy problem with initial data

$$A[0] = (A_j(0), \partial_t A_j(0)).$$

The initial data is uniquely determined by the Yang–Mills initial data and the gauge condition (7.1).

The system (7.2) together with the constraint equation (7.3) is well-posed in regular Sobolev spaces. Precisely, we have

Theorem 7.1. *The system* (7.2) *is locally well-posed in* $H^N \times H^{N-1}$ *for* $N \geq 2$*, with Lipschitz dependence on the initial data.*

We further remark that the temporal gauge fully describes all classical solutions to the Yang–Mills system:

Theorem 7.2. *Let* A *be a solution to the Yang–Mills system which has local in time regularity* $(A, \partial_t A) \in C([0,T]; H^N \times H^{N-1})$ *for* $N \geq 3$*. Then* A *has a temporal gauge equivalent* $\tilde{A}$ *with the same regularity* $(\tilde{A}, \partial_t \tilde{A}) \in C([0,T]; H^N \times H^{N-1})$*.*

To see this, it suffices to solve an equation for the gauge transformation O, namely

$$O^{-1}\partial_0 O = A_0, \qquad O(0,x) = I,$$

which is an ODE on the Lie group $\mathbf{G}$. If $A \in C(H^N)$ then this yields a unique solution $O \in C(H^N)$. This in turn yields a temporal gauge equivalent solution

$$(\tilde{A}, \partial_t \tilde{A}) \in C([0,T]; H^{N-1} \times H^{N-2}).$$

This argument loses one derivative. However, the initial data is in $H^N \times H^{N-1}$, which by the well-posedness result yields a $C([0,T]; H^N \times H^{N-1})$ solution. But by the $H^{N-1} \times H^{N-2}$ well-posedness the two must agree, so we obtain a unique representation in the temporal gauge with the same data and without loss of derivatives.

Remark 7.3. Analogues of Theorems 7.1 and 7.2 hold for the space $H^N_{loc} \times H^{N-1}_{loc}$ instead of $H^N \times H^{N-1}$, where H^N_{loc} is equipped with the norm $\sup_{x \in \mathbb{R}^4} \| \cdot \|_{H^N(B_1(x))}$.

7.1.2. *Caloric gauge.* In view of Theorem 1.11 we can fully describe caloric Yang–Mills waves as continuous functions

$$I \ni t \to (A_x(t), \partial_0 A_x(t)) \in T^{L^2} C$$

For higher regularity Yang–Mills waves we have the following:

Theorem 7.4. *Let* A *be a solution to the Yang–Mills system which has local in time regularity* $(A, \partial_t A) \in C([0,T]; H^N \times H^{N-1})$ *for* $N \geq 2$*. Assume in addition that the bound* (1.12) *is uniformly satisfied by its caloric extension, globally in parabolic time. Then* A *has a caloric gauge equivalent* $\tilde{A}$ *with the same regularity* $(\tilde{A}, \partial_t \tilde{A}) \in C([0,T]; H^N \times H^{N-1})$*.*

This result is a direct consequence of Theorem 1.11, with one minor exception. Precisely, Theorem 1.11 does not directly yield the $C_t L^2_x$ regularity for $\partial_0 A_0$. For that we instead need to refer to the expression (3.15) and the bounds (3.18) respectively (4.28) for the two terms in (3.15).

Remark 7.5. The same result will easily hold for $(A, \partial_t A) \in C([0,T]; \mathbf{H} \times L^2)$. However, if we only assume that $A, \partial_t A) \in C([0,T]; \dot{H}^1 \times L^2)$ then one would also need to resolve the remaining gauge freedom. For that it suffices to observe tat if two A's have a small difference in L^2, then the two O's can be chosen in tandem so that they agree at infinity.

In particular this says that a caloric gauge solution exists for as long as a regular solution exists and the L^3 bound in (1.12) remains finite. This will allow us to bootstrap the existence time for as long as we have good bounds in the caloric gauge. Precisely, for[8] $N \geq 3$ suppose that an H^N solution exists in the caloric gauge up to time T. If this solution has uniform H^N bounds up to time T, then its temporal gauge representation has uniform H^N bounds up to time T. Thus it can be extended further in the temporal gauge, hence also in the caloric gauge. This shows that a maximal caloric gauge solution must either explode in H^N at the (finite) end if its lifespan, or the L^3 norm in (1.12) must explode. The latter cannot happen for subthreshold solutions. Thus we have

Theorem 7.6. *The Yang–Mills system in the caloric gauge is locally well-posed in $H^N \times H^{N-1}$ for $N \geq 2$. Further, the solution extends for as long as the $H^N \times H^{N-1}$ norm remains bounded and the L^3 norm in* (1.12) *remains bounded.*

For regular data, this result reduces the problem of global well-posedness to that of obtaining uniform bounds for caloric solutions.

7.2. **Local well-posedness in the caloric manifold $\mathcal{C}$: Proof of Theorem 1.13.** For $\epsilon_* > 0$, recall that the energy concentration scale $r_c^{\epsilon_*}$ was defined as

$$\begin{aligned} r_c^{\epsilon_*}[a,e] =& \sup\{r > 0 : \mathcal{E}_{B_r(x)}(a,e) \leq \epsilon_*^2 \text{ for all } x \in \mathbb{R}^4\} \\ =& \sup\{r > 0 : \sup_{x \in \mathbb{R}^4} \frac{1}{2} \sum_{\alpha < \beta} \|f_{\alpha\beta}\|^2_{L^2(B_r(x))} \leq \epsilon_*^2\}, \end{aligned}$$

where f_{jk} is the curvature form corresponding to a_j, $f_{0j} = -f_{j0} = e_j$ and $f_{00} = 0$. Since the definition only involves $f_{\alpha\beta}$, we will slightly abuse the notation and simply write $r_c^{\epsilon_*}[f]$ for $r_c^{\epsilon_*}[a,e]$.

Lemma 7.7. *Let A be a regular caloric Yang–Mills wave on $I = (-T_0, T_0)$. For any $\epsilon > 0$, if ϵ_* is sufficiently small compared to ϵ and*

$$T_0 \leq r_c^{\epsilon_*}[a,e],$$

then we have

$$\|F\|_{ED_{\geq m}[I]} \leq \epsilon \qquad \textit{with } 2^m = \epsilon (r_c^{\epsilon_*}[a,e])^{-1}$$

Proof. By our notation, $f_{\alpha\beta} = F_{\alpha\beta}(0)$. After rescaling, we may set $r_c^{\epsilon_*}(F(0)) = 1$. We begin with the observation that

$$\|P_k F(t)\|_{L^\infty} \lesssim 2^{ck_-} 2^{-2k} \sup_{x \in \mathbb{R}^4} \|F(t)\|_{L^2(B_1(x))}, \tag{7.4}$$

which follows from the properties of the convolution kernel of P_k; in particular, it is rapidly decaying on the scale 2^{-k} and its L^2-norm is bounded by 2^{-2k}. Then, by the localized energy estimate for the hyperbolic Yang–Mills equation, i.e.,

$$\mathcal{E}_{\{t\} \times B_{R-|t|}}(F) \leq \mathcal{E}_{\{0\} \times B_R}(F) \qquad (0 < |t| < R), \tag{7.5}$$

the lemma follows. ∎

[8]The requirement $N \geq 3$ is so that there is no loss of regularity in the transition to the temporal gauge. Precisely, we want to insure that $A_0 \in C(\dot{H}^1 \cap \dot{H}^{N+1})$.

Proof of Theorem 1.13. We prove the theorem in several steps:

1. Regular solutions. Let A be a regular caloric Yang–Mills wave with energy $\mathcal{E}$ and initial caloric size $\mathcal{Q}$. For ϵ_* small enough, to be chosen later, let $r_c := r_c^{\epsilon_*}$ be the corresponding energy concentration scale for the initial data.

Our goal is to prove that if ϵ_* is small enough, depending only on $\mathcal{E}$ and $\mathcal{Q}$, then the solution A persists as a regular caloric solution up to time r_c. Precisely, we will to apply Theorem 6.1 to the solution A in order to show that the solution A exists in $[-r_c, r_c]$ and satisfies the bound

$$\|A\|_{S^1[-r_c,r_c]} \leq M(\mathcal{E}, 3\mathcal{Q}). \tag{7.6}$$

We use a continuity argument. Let $T_0 \leq r_c$ be a maximal time with the property that the solution A given by Theorem 7.4 exists as a classical caloric solution in $(-T_0, T_0)$, and further satisfies the bound

$$\sup_{t\in[-T_0,T_0]} \mathcal{Q}(A(t)) \leq 3\mathcal{Q}. \tag{7.7}$$

For $0 < T < T_0$ we seek to apply Theorem 6.1 to A in $I = [-T, T]$. To verify the hypothesis of Theorem 6.1 we need to insure that for a suitable choice of m we have

$$\|F\|_{ED_{\geq m}} \leq \epsilon(\mathcal{E}, 3\mathcal{Q}), \qquad |I| \leq 2^{-m} T(\mathcal{E}, 3\mathcal{Q}).$$

For this it suffices to apply Lemma 7.7 with

$$\epsilon = \min\{\epsilon(\mathcal{E}, 3\mathcal{Q}), T(\mathcal{E}, 3\mathcal{Q})\}.$$

which yields the appropriate choice of ϵ_*.

Now by Theorem 6.1 we obtain the uniform bound

$$\|A\|_{S^1[-T,T]} \leq M(\mathcal{E}, 3\mathcal{Q}), \qquad 0 < T < T_0.$$

By the Structure Theorem 5.1 it follows that higher regularity bounds are also uniformly propagated,

$$\sup_{t\in(-T_0,T_0)} \|(A, \partial_t A)(t)\|_{H^N} < \infty.$$

Thus by the local result for regular solutions in Theorem 7.6 we can continue the regular caloric Yang–Mills connection A beyond the time interval $[-T_0, T_0]$.

Finally, we consider the bounds for $\mathcal{Q}(A)$. These we can propagate using Theorem 5.9, which implies that

$$\sup_{t\in[-T_0,T_0]} \mathcal{Q}(A(t)) - \mathcal{Q} \lesssim_{\mathcal{Q},\mathcal{E}} \epsilon^{\delta_4}.$$

Readjusting ϵ if needed, it follows that

$$\sup_{t\in[-T_0,T_0]} \mathcal{Q}(A(t)) \leq 2\mathcal{Q} \tag{7.8}$$

This implies that the bound (7.7) also can be propagated beyond $\pm T_0$. This contradicts the maximality of T_0 unless $T_0 = r_c$. Hence the classical caloric Yang–Mills wave exists in $[-r_c, r_c]$ and (7.6) holds.

2. Rough solutions. Given any caloric initial data (a, b) with finite energy $\mathcal{E}$ and caloric size $\mathcal{Q}$, we consider the corresponding regularized data $(a(s), b(s))$ obtained using the Yang–Mills heat flow. We have the uniform bounds

$$\mathcal{E}(a(s), b(s)) \leq \mathcal{E}(a, b), \qquad \mathcal{Q}(a(s), b(s)) \leq \mathcal{Q}(a, b).$$

In particular, we have $(f(s), e(s)) \to (f, e)$ in $\dot{H}^1 \times L^2$. This implies that the energy concentration scales for $(a(s), e(s))$ converge to those for (a, e). Thus, by the analysis in the smooth case above, for small enough s the corresponding solutions $A(s)$ exist as smooth caloric Yang–Mills waves in $I = [-r_c, r_c]$ and satisfy the uniform S^1 bound (7.6).

Now we use the Structure Theorem 5.1 to consider the limit as $s \to 0$. If c_k is a frequency envelope for (a, e), then by Proposition 3.1 it follows that

(i) For $(a(s), b(s))$ we have the frequency envelope in $\dot{H}^1 \times L^2$

$$c_k(s) = c_k \langle 2^{2k} s\rangle^{-c\delta_5}.$$

(ii) For the difference $(a, b) - (a(s), b(s))$ we have the envelope in $\dot{H}^1 \times L^2$

$$\delta c_k(s) = c_k \langle 2^{-2k} s^{-1}\rangle^{-c\delta_5}.$$

(iii) For the difference $(a(s), b(s)) - (a(2s), b(2s))$ we have the envelope in $\dot{H}^1 \times L^2$

$$c_k^*(s) = c_{k(s)} 2^{-c\delta_5 |k-k(s)|}.$$

By Theorem 5.1.(2), it follows that $c_k(s)$ is a frequency envelope for $A(s)$ in S_1. Combining this with Theorem 5.1.(8), it follows that $c_k^*(s)$ is a frequency envelope for $A(s) - A(2s)$. Summing up such differences, we obtain the general difference bound

$$\|A(s_1) - A(s_2)\|_{S^1} \lesssim_{\mathcal{E},\mathcal{Q}} c_{[k(s_1), k(s_2)]}. \tag{7.9}$$

This implies that the limit

$$A = \lim_{s\to 0} A(s)$$

exists in s. We define A to be the caloric Yang–Mills wave associated to the (a, b) data. We remark that by (7.9) we have the difference bound

$$\|A - A(s)\|_{S^1} \lesssim_{\mathcal{E},\mathcal{Q}} c_{\geq k(s)}. \tag{7.10}$$

3. Difference bound. The difference bound in part (4) of the theorem is a direct consequence of the difference bound in Theorem 5.1.(8).

4. Continuous dependence. We consider a convergent sequence of caloric initial data

$$(a^{(n)}, b^{(n)}) \to (a, b) \qquad \text{in } \dot{H}^1 \times L^2. \tag{7.11}$$

Let $A^{(n)}(s)$, respectively $A(s)$ be the corresponding solutions with regularized data.

Denote by c_k^n a corresponding sequence of frequency envelopes for the initial data $(a^{(n)}, b^{(n)})$ in $\dot{H}^1 \times L^2$. By Theorem 5.1.(2), these are also frequency envelopes for the solutions $A^{(n)}(s)$.

By Theorem 7.4 we know that for each s we have

$$A^{(n)}(s) \to A(s) \qquad \text{in } S^1$$

and in effect in stronger topologies. Then we estimate

$$\begin{aligned} \limsup_{n\to\infty} \|A^{(n)} - A\|_{S^1} &\lesssim \lim_{s\to\infty} \limsup_{n\to\infty} \|A^{(n)}(s) - A(s)\|_{S^1} + c^n_{\geq k(s)} + c_{\geq k(s)} \\ &\lesssim \lim_{s\to\infty} \limsup_{n\to\infty} c^n_{\geq k(s)} \end{aligned}$$

But the last limit is zero in view of the convergence in (7.11). The continuous dependence follows. □

We end this subsection with a lemma that bounds the energy concentration scale from below by an L^2-frequency envelope for F, which proves Remark 5.2.

Lemma 7.8. *Let c be a frequency envelope for $F_{\alpha\beta}$ in L^2 for all $\alpha, \beta \in \{0, 1, \dots, 4\}$. Suppose that $\|c\|_{\ell^2_{\geq m}} < C^{-1}\epsilon_*$ for some $m \in \mathbb{Z}$ and a sufficiently large universal constant $C > 0$. Then $r_c^{\epsilon_*}(F) \geq 2^{-m}$.*

Proof. It suffices to establish the bound

$$\|F\|_{L^2(B(x,2^{-k})} \lesssim c_{\geq k}.$$

To see this we use Bernstein's inequality to estimate

$$\begin{aligned}\|F\|_{L^2(B(x,2^{-k})} &\lesssim \|F_{\geq k}\|_{L^2} + \sum_{j<k} 2^{-2k}\|F_j\|_{L^\infty} \\ &\lesssim c_{\geq k} + \sum_{j<k} 2^{2j-2k} c_j \approx c_{\geq k}.\end{aligned}$$

□

7.3. **Regularity of energy-dispersed solutions: Proof of Theorem 1.20.** Consider a time t_0 where $\mathcal{Q}(A(t))$ is nearly minimal. From Lemma 5.10 we have the estimate

$$\mathcal{Q}(A(t_0)) \lesssim_{\mathcal{E}} \epsilon^c.$$

If ϵ is small enough this allows us to conclude first that $\mathcal{Q} \leq 1$, and then that

$$\mathcal{Q} \lesssim_E \epsilon^c.$$

Now a straightforward continuity argument shows that

$$Q(A(t)) \leq 1, \qquad t \in I,$$

which again by Lemma 5.10 yields

$$\mathcal{Q}(A(t)) \lesssim_{\mathcal{E}} \epsilon^c, \qquad t \in I.$$

Then we can apply directly the result in Theorem 6.1 for any $m \in \mathbb{Z}$. This eliminates any restriction on the size of the interval I.

7.4. **Gauge transformation into temporal gauge: Proof of Theorem 1.18.** To produce a temporal gauge solution to (1.1) from the caloric gauge solution we use a gauge transformation O defined as the solution to the following ODE:

$$O^{-1}\partial_t O = A_0, \qquad O(0) = I. \tag{7.12}$$

Here for A_0 we have the regularity given by Theorem 5.1.(9), namely

$$A_0 \in \ell^1 |D|^{-2} L^2_x L^1_t. \tag{7.13}$$

We use this to compute the regularity of O:

Lemma 7.9. *a) Assume that A_0 is as in* (7.13)*. Then the solution O to the ODE has the following properties:*

(i) $O_{;x} \in C_t(\dot{H}^1)$.

(ii) O is continuous in both x and t.

b) Consider two solutions O and $\tilde{O}$ arising from A_0 and $\tilde{A}_0$. Then we have *(i) $\dot H^1$ bound:*

$$\|O^{-1}\partial_x O - \tilde O^{-1}\partial_x \tilde O\|_{\dot H^1} \lesssim \|A_0 - \tilde A_0\|_{\ell^1|D|^{-2}L^2_x L^1_t}.$$

(ii) Uniform bound:

$$\|d(O,\tilde O)\|_{L^\infty} \lesssim \|A_0 - \tilde A_0\|_{\ell^1|D|^{-2}L^2_x L^1_t}.$$

Proof. a) We first consider the ODE

$$O^{-1}\partial_t O = F, \qquad O(0) = I, \tag{7.14}$$

and observe that for smooth F this is easily solvable.

Next we consider a smooth one parameter family of solutions $O(h)$. For this we compute

$$\frac{d}{dt}(O^{-1}\partial_h O) = \partial_h F - [F, O^{-1}\partial_h O],$$

which immediately leads to

$$|O^{-1}\partial_h O(t)| \leq \int_0^t |\partial_h F(s)| ds.$$

Comparing two solutions O and $\tilde O$ generated by F and $\tilde F$ using the straight line between them, it follows that

$$d(O(t),\tilde O(t)) \leq \int_0^t |F(s) - \tilde F(s)| ds. \tag{7.15}$$

This yields a Lipschitz property for the map

$$L^1_t \ni F \to O \in C_t$$

which is thus by density extended to all $F \in L^1_t$.

Next we turn our attention to A_0, which by Bernstein's inequality satisfies

$$A_0 \in C_x L^1_t.$$

This implies the desired continuity of O.

Finally we consider the evolution of $O^{-1}\partial_x O$,

$$\frac{d}{dt}(O^{-1}\partial_x O) = \partial_x A_0 - [A_0, O^{-1}\partial_x O].$$

Since $\partial_x A_0 \in L^4_x L^1_t$, this immediately gives

$$O^{-1}\partial_x O \in L^4_x C_t \subset C L^4.$$

A second differentiation yields as well

$$\partial_x(O^{-1}\partial_x O) \in L^2_x C_t \subset C L^2.$$

b) The uniform bound for the difference follows directly from (7.15). For the difference of the derivatives we compute

$$\partial_t(O^{-1}\partial_j O - \tilde O^{-1}\partial_j \tilde O) + [A_0, O^{-1}\partial_j O - \tilde O^{-1}\partial_j \tilde O] = \partial_j A_0 - \partial_j \tilde A_0 - [A_0 - \tilde A_0, \tilde O \partial_j \tilde O].$$

As above, we can estimate this first in L^4 and then in $\dot H^1$. □

To conclude the proof of Theorem 1.18 it remains to verify (i) that gauge transformations O having the properties in the above lemma yield temporal connections $A^{[t]} \in C(\dot{H}^1)$, and (ii) these connections depend continuously on the initial data.

For the continuity in time we write

$$A^{[t]} = O(A - O^{-1}\partial_x O)O^{-1}.$$

The second term above is in $C_t\dot{H}^1$ due to the previous lemma. For the first term we differentiate, then use again the lemma combined with the continuity of O and dominated convergence.

For the continuous dependence of the temporal solutions with caloric data the same argument as above applies. However, we also need to consider general finite energy initial data sets. Here the construction of the temporal gauge solutions starting from a general initial data (a, e) goes as follows:

(1) Given the initial position $a \in \dot{H}^1$, we consider the gauge transformation $O = O(a)$ which turns a into $(\tilde{a}, \tilde{e})$, its caloric gauge counterpart.
(2) Given the caloric data $(\tilde{a}, \tilde{e})$ we have as above an unique temporal solution $\tilde{A}$.
(3) To return to the data (a, e) we apply to $\mathbf{A}$ the inverse gauge transformation O^{-1} to obtain the temporal solution A.

The regularity of the gauge transformation O is $O^{-1}\partial_x O \in \dot{H}^1$, which suffices in order for it to map $C(\dot{H}^1)$ connections into $C(\dot{H}^1)$ connections. It remains to prove the continuous dependence. Consider a convergent sequence of data $(a^{(n)}, e^{(n)}) \to (a, e)$ in $\dot{H}^1 \times L^2$. Without any restriction in generality we can assume that (a, e) is caloric. Denote by $O^{(n)}$ the corresponding gauge transformations, which, we recall, are only unique up to constant gauge transformations. Then we need to show that for a well chosen (sub)sequence of representatives $O^{(n)}$ we have the following properties:

(1) $(O^{(n)})^{-1}\partial_x O^{(n)} \to 0$ in $\dot{H}^1$.
(2) $O^{(n)}(x) \to I$ a.e. in x.

But this is a consequence of Theorem 1.2, see also Remark 1.3 (recall also that $O_{;x} = Ad(O)(O^{-1}\partial_x O)$).

8. Multilinear estimates

The purpose of this section is to prove most of the results stated without proof in Section 4. The exceptions are Theorem 4.24 and Proposition 4.25, which involve construction of a parametrix for $\Box + \mathrm{Diff}^{\kappa}_{\mathbf{P}A}$; their proofs are given in the next section.

8.1. **Disposable operators and null forms.**

In this subsection we collect preliminary materials that are needed for analysis of the multilinear operators in the nonlinearity of the Yang–Mills equation in the caloric gauge.

8.1.1. *Disposable operators.* Boundedness properties of the multilinear operators arising in caloric gauge (see Section 3) can be conveniently phrased in terms of *disposability* (after multiplication with appropriate weights) of these operators.

We begin by considering the multilinear operator $\mathbf{Q}$ with the symbol

$$\mathbf{Q}(\xi, \eta) = \frac{|\xi|^2 - |\eta|^2}{2(|\xi|^2 + |\eta|^2)} = \frac{(\xi + \eta) \cdot (\xi - \eta)}{2(|\xi|^2 + |\eta|^2)},$$

which arose in the wave equation for A_x (most notably through the expression for $\partial^\ell A_\ell$) in the caloric gauge.

Lemma 8.1. *For any $k, k_1, k_2 \in \mathbb{Z}$, the bilinear operator*

$$2^{k_{\max}-k} P_k \mathbf{Q}(P_{k_1}(\cdot), P_{k_2}(\cdot))$$

is disposable.

Proof. To begin with, note the symbol bound

$$|\mathbf{Q}(\xi, \eta)| \lesssim \frac{|\xi + \eta|}{(|\xi|^2 + |\eta|^2)^{\frac{1}{2}}},$$

which implies that the symbol of $2^{k_{\max}-k} P_k \mathbf{Q}(P_{k_1}(\cdot), P_{k_2}(\cdot))$ is uniformly bounded. In the case $k_2 < k_1 - 5$ so that $|k_{\max} - k| \leq 3$, it can also be checked that

$$2^{n_1 k_1} 2^{n_2 k_2} |\partial_\xi^{(n_1)} \partial_\eta^{(n_2)} (P_k(\xi + \eta) \mathbf{Q}(\xi, \eta) P_{k_1}(\xi) P_{k_2}(\eta))| \lesssim_{n_1, n_2} 1,$$

which proves the desired disposability property. By symmetry, the case $k_1 < k_2 - 5$ follows as well. In the case $|k_1 - k_2| < 5$ (so that $|k_{\max} - k_1| < 10$), making the change of variables $(\xi, \zeta) = (\xi, \xi + \eta)$, it can be seen that

$$2^{k_1 - k} 2^{n_1 k_1} 2^{n_2 k} |\partial_\xi^{(n_1)} \partial_\zeta^{(n_2)} (P_k(\zeta) \mathbf{Q}(\xi, \zeta - \xi) P_{k_1}(\xi) P_{k_2}(\zeta - \xi))| \lesssim_{n_1, n_2} 1,$$

which implies disposability of $2^{k_{\max}-k} P_k \mathbf{Q}(P_{k_1}(\cdot), P_{k_2}(\cdot))$. □

Next, we consider the multilinear operator $\mathbf{W}(s)$ with the symbol

$$\mathbf{W}(\xi, \eta, s) = -\frac{1}{2\xi \cdot \eta} e^{-s|\xi+\eta|^2} (1 - e^{2s\xi \cdot \eta}),$$

which arose in the wave equation for the Yang–Mills heat flow development $A_x(s)$ of a caloric Yang–Mills wave.

Lemma 8.2. *For any $k, k_1, k_2 \in \mathbb{Z}$ and $s > 0$, the bilinear operator*

$$\langle s 2^{2k} \rangle^{10} \langle s^{-1} 2^{-2k_{\max}} \rangle 2^{2k_{\max}} P_k \mathbf{W}(P_{k_1}(\cdot), P_{k_2}(\cdot), s) \tag{8.1}$$

is disposable.

Proof. Without loss of generality, we may assume that $s = 1$ by scaling. We distinguish two scenarios:

Case 1: (High–Low or Low–High: $k = \max\{k_1, k_2\} + O(1)$). To prove disposability of (8.1), it suffices to show that

$$\langle 2^{2k_{\max}} \rangle^{11} 2^{n_1 k_1} 2^{n_2 k_2} \left| \partial_\xi^{(n_1)} \partial_\eta^{(n_2)} \left(P_k(\xi + \eta) e^{-|\xi+\eta|^2} \frac{1 - e^{2\xi \cdot \eta}}{\xi \cdot \eta} P_{k_1}(\xi) P_{k_2}(\eta) \right) \right| \lesssim_{n_1, n_2} 1$$

for any $n_1, n_2 \in \mathbb{N}$. Since the derivatives of $P_k(\xi + \eta) P_{k_1}(\xi) P_{k_2}(\eta)$ already obey desirable bounds, it only remains to prove

$$\langle 2^{2k_{\max}} \rangle^{11} 2^{n_1 k_1} 2^{n_2 k_2} \left| \partial_\xi^{(n_1)} \partial_\eta^{(n_2)} \left(e^{-|\xi+\eta|^2} \frac{1 - e^{2\xi \cdot \eta}}{\xi \cdot \eta} \right) \right| \lesssim_{n_1, n_2} 1 \tag{8.2}$$

for ξ, η in the support of the symbol (8.1).

Since $k = \max\{k_1, k_2\} + O(1)$, we have $2^{2k_{\max}} \simeq |\xi|^2 + |\eta|^2 \simeq |\xi+\eta|^2$. On the one hand, it is straightforward to verify

$$\begin{aligned}
2^{n_1 k_1} 2^{n_2 k_2} |\partial_\xi^{(n_1)} \partial_\eta^{(n_2)} e^{-|\xi+\eta|^2}| &\lesssim_{n_1,n_2} 2^{n_1 k_1} 2^{n_2 k_2} (1 + |\xi+\eta|^2)^{\frac{n_1+n_2}{2}} e^{-|\xi+\eta|^2} \\
&\lesssim_{n_1,n_2} 2^{(n_1+n_2)k_{\max}} \langle 2^{2k_{\max}} \rangle^{\frac{n_1+n_2}{2}} e^{-|\xi+\eta|^2}.
\end{aligned} \tag{8.3}$$

On the other hand, we also have

$$\begin{aligned}
2^{n_1 k_1} 2^{n_2 k_2} \left| \partial_\xi^{(n_1)} \partial_\eta^{(n_2)} \left(\frac{1 - e^{2\xi\cdot\eta}}{\xi\cdot\eta} \right) \right| &\lesssim_{n_1,n_2} 2^{n_1 k_1} 2^{n_2 k_2} (1 + |\xi|^2 + |\eta|^2)^{\frac{n_1+n_2}{2}} (1 + e^{2\xi\cdot\eta}) \\
&\lesssim_{n_1,n_2} 2^{(n_1+n_2)k_{\max}} \langle 2^{2k_{\max}} \rangle^{\frac{n_1+n_2}{2}} (1 + e^{2\xi\cdot\eta}).
\end{aligned} \tag{8.4}$$

The key point here is that when $|\xi\cdot\eta| \ll 1$, the denominator $\xi\cdot\eta$ cancels with the first term in the Taylor expansion of the numerator $1 - \xi\cdot\eta$; we omit the details. Combining (8.3) and (8.4), it follows that

$$2^{n_1 k_1} 2^{n_2 k_2} \left| \partial_\xi^{(n_1)} \partial_\eta^{(n_2)} \left(e^{-|\xi+\eta|^2} \frac{1 - e^{2\xi\cdot\eta}}{\xi\cdot\eta} \right) \right| \lesssim_{n_1,n_2} \langle 2^{2k_{\max}} \rangle^{n_1+n_2} e^{-|\xi+\eta|^2} (1 + e^{2\xi\cdot\eta}).$$

Since $e^{-|\xi+\eta|^2}(1 + e^{2\xi\cdot\eta}) = e^{-|\xi+\eta|^2} + e^{-(|\xi|^2+|\eta|^2)} \lesssim e^{-C^{-1} 2^{2k_{\max}}}$, (8.2) follows.

Case 2: (High–High: $k < \max\{k_1, k_2\} - C$). As usual, we make the change of variables $(\xi, \zeta) = (\xi, \xi + \eta)$. It suffices to prove

$$\langle 2^{2k} \rangle^{10} \langle 2^{2k_{\max}} \rangle 2^{n_1 k_1} 2^{n_2 k} \left| \partial_\xi^{(n_1)} \partial_\zeta^{(n_2)} \left(P_k(\zeta) e^{-|\zeta|^2} \frac{1 - e^{2\xi\cdot(\zeta-\xi)}}{\xi\cdot(\zeta-\xi)} P_{k_1}(\xi) P_{k_2}(\xi - \zeta) \right) \right| \lesssim_{n_1,n_2} 1.$$

Note that the derivatives of $\langle 2^{2k} \rangle^{10} P_k(\zeta) e^{-|\zeta|^2} P_{k_1}(\xi) P_{k_2}(\xi-\zeta)$ already obey desirable bounds. Hence we are only left to show

$$\langle 2^{2k_{\max}} \rangle 2^{n_1 k_1} 2^{n_2 k} \left| \partial_\xi^{(n_1)} \partial_\zeta^{(n_2)} \left(\frac{1 - e^{2\xi\cdot(\zeta-\xi)}}{\xi\cdot(\zeta-\xi)} \right) \right| \lesssim_{n_1,n_2} 1, \tag{8.5}$$

for ξ, ζ in the support of (8.1).

Note that $k_1 = k_{\max} + O(1)$. In the case $2^{2k_{\max}} \lesssim 1$, (8.5) follows from

$$|\partial_\xi^{(n_1)} \partial_\zeta^{(n_2)} \left((2\xi\cdot(\zeta-\xi))^{-1} (1 - e^{2\xi\cdot(\zeta-\xi)}) \right)| \lesssim_{n_1,n_2} 1,$$

which follows by Taylor expansion at $\xi\cdot(\zeta-\xi) = 0$. In the case $2^{2k_{\max}} \gtrsim 1$, we use

$$\begin{aligned}
2^{n_1 k_1} 2^{n_2 k} |\partial_\xi^{(n_1)} \partial_\zeta^{(n_2)} (\xi\cdot(\zeta-\xi))^{-1}| &\lesssim 2^{-2k_{\max}}, \\
2^{n_1 k_1} 2^{n_2 k} |\partial_\xi^{(n_1)} \partial_\zeta^{(n_2)} (1 - e^{2\xi\cdot(\zeta-\xi)})| &\lesssim 1,
\end{aligned}$$

both of which follow from simple computation, whose details we omit. □

8.1.2. *Null forms.* We now discuss the null forms that arise in caloric gauge, which occur in conjunction with various (disposable) translation-invariant operators. To treat these in a systematic fashion, it is useful to define null forms in terms of an appropriate decomposition property of the symbol.

Definition 8.3 (Null forms)**.** Let $\mathcal{T}$ be a translation-invariant bilinear operator on $\mathbb{R}^{1+4}$ and let $\pm \in \{+, -\}$ be a sign. Given $k_1, k_2 \in \mathbb{Z}$, $\ell, \ell' \in -\mathbb{N}$, $\omega, \omega' \in \mathbb{S}^3$, define

$$\theta_\pm = \max\{|\angle(\omega, \pm\omega')|, 2^\ell, 2^{\ell'}\}.$$

(1) We say that $\mathcal{T}$ is a *null form of type* $\mathcal{N}_\pm$ and write

$$\mathcal{T}(\cdot,\cdot) = \mathcal{N}_\pm(\cdot,\cdot),$$

if for every $k_1, k_2 \in \mathbb{Z}$, $\ell, \ell' \in -\mathbb{N}$ and $\omega, \omega' \in \mathbb{S}^3$, $\mathcal{T}$ admits a decomposition of the form

$$\mathcal{T}((\tau,\xi),(\sigma,\eta))(P_{k_1}P_\ell^\omega)(\xi)(P_{k_2}P_{\ell'}^{\omega'})(\eta) = \theta_\pm 2^{k_1+k_2}\mathcal{O}((\tau,\xi),(\sigma,\eta)) \sum_{i_1,i_2\in\mathbb{N}} a_{i_1}(\xi)b_{i_2}(\eta),$$

where the Fourier multipliers

$$(1+|i_1|)^{100}a_{i_1}, \quad (1+|i_2|)^{100}b_{i_2} \tag{8.6}$$

are disposable, and the translation invariant bilinear operator with symbol

$$\mathcal{O}((\tau,\xi),(\sigma,\eta))$$

is disposable as well.

(2) We say that $\mathcal{T}$ is a *null form of type* $\mathcal{N}$ if $\mathcal{T}(\cdot,\cdot) = \mathcal{N}_+(\cdot,\cdot)$ and $\mathcal{T}(\cdot,\cdot) = \mathcal{N}_-(\cdot,\cdot)$.

(3) We say that $\mathcal{T}$ is a *null form of type* $\mathcal{N}_{0,\pm}$ and write

$$\mathcal{T}(\cdot,\cdot) = \mathcal{N}_{0,\pm}(\cdot,\cdot),$$

if for every $k_1, k_2 \in \mathbb{Z}$, $\ell, \ell' \in -\mathbb{N}$ and $\omega, \omega' \in \mathbb{S}^3$, $\mathcal{T}$ admits a decomposition of the form

$$\mathcal{T}(\xi,\eta)(P_{k_1}P_\ell^\omega)(\xi)(P_{k_2}P_{\ell'}^{\omega'})(\eta) = \theta_\pm^2 2^{k_1+k_2}\mathcal{O}((\tau,\xi),(\eta,\sigma)) \sum_{i_1,i_2\in\mathbb{N}} a_{i_1}(\xi)b_{i_2}(\eta),$$

where the Fourier multipliers

$$(1+|i_1|)^{100}a_{i_1}, \quad (1+|i_2|)^{100}b_{i_2} \tag{8.7}$$

are disposable, and also the translation-invariant bilinear operator which has symbol $\mathcal{O}((\tau,\xi),(\sigma,\eta))$ is disposable as well.

In particular, $\mathcal{O}$, a_{i_1} and b_{i_2} may depend on $k_1, k_2, \ell, \ell', \omega, \omega'$, but the disposability bounds stated above do not.

Remark 8.4 (Null form gain). To exploit the null form, it is convenient to make the following observation: As a immediate consequence of the definition, we may write

$$\mathcal{N}_\pm(P_{k_1}P_\ell^\omega u, P_{k_2}P_{\ell'}^{\omega'}v) = C\theta_\pm 2^{k_1+k_2}\tilde{\mathcal{O}}(P_{k_1}P_\ell^\omega u, P_{k_2}P_{\ell'}^{\omega'}v)$$

for a universal constant $C > 0$ and some disposable $\tilde{\mathcal{O}}$. Analogous statements hold for $\mathcal{N}$ and $\mathcal{N}_{0,\pm}$.

Remark 8.5 (Behavior under symbol multiplication). The properties in Definition 8.3 seem complicated at first, but its usefulness comes from the fact that it is well-behaved under symbol-multiplication with a disposable multilinear operator. More precisely, if $\mathcal{O}(\cdot,\cdot)$ is a disposable translation-invariant bilinear operator and $\mathcal{T}(\cdot,\cdot)$ is a null form in the sense of Definition 8.3, then the translation-invariant bilinear operator with symbol $\mathcal{O}(\xi,\eta)\mathcal{T}(\xi,\eta)$ is clearly also a null form of the same type.

We now verify that the standard null forms are indeed null forms according to Definition 8.3. We have the following separation-of-variables result for the symbols of the standard null forms.

Lemma 8.6 (Standard null forms)**.** *Consider the symbols*

$$\mathbf{N}_{ij}(\xi,\eta) = \xi_i\eta_j - \xi_j\eta_i, \quad \mathbf{N}_{0,\pm}(\xi,\eta) = \pm|\xi||\eta| - \xi\cdot\eta.$$

These symbols admit the decompositions

$$|\xi|^{-1}|\eta|^{-1}\mathbf{N}_{ij}(\xi,\eta)(P_{k_1}P_\ell^\omega)(\xi)(P_{k_2}P_{\ell'}^{\omega'})(\eta) = \min\{\theta_+,\theta_-\}\sum_{i_1,i_2\in\mathbb{N}} a_{i_1}(\xi)b_{i_2}(\eta), \tag{8.8}$$

$$|\xi|^{-1}|\eta|^{-1}\mathbf{N}_{0,\pm}(\xi,\eta)(P_{k_1}P_\ell^\omega)(\xi)(P_{k_2}P_{\ell'}^{\omega'})(\eta) = \theta_\pm^2\sum_{i_1,i_2\in\mathbb{N}} a'_{i_1}(\xi)b'_{i_2}(\eta), \tag{8.9}$$

where

$$(1+|i_1|)^{100}a_{i_1}, \quad (1+|i_1|)^{100}a'_{i_1}, \quad (1+|i_2|)^{100}b_{i_2}, \quad (1+|i_2|)^{100}b'_{i_2} \tag{8.10}$$

are disposable.

As a corollary, it follows that $\mathbf{N}_{ij}$ is a null form of type $\mathcal{N}$, whereas $\mathbf{N}_{0,\pm}$ are null forms of type $\mathcal{N}_\pm$.

As before, a_{i_1}, a'_{i_1}, b_{i_2} and b'_{i_2} depend on $k_1, k_2, \ell, \ell', \omega, \omega'$, but the disposability bounds stated in (8.10) do not.

This lemma can be proved by performing separation of variables using Fourier series on an appropriate rectangular box containing the support of $P_{k_1}P_\ell^\omega(\xi)P_{k_2}P_\omega^{\ell'}(\xi')$. For the details in the case of $|\xi|^{-1}|\eta|^{-1}\mathbf{N}_{ij}(\xi,\eta)$, we refer to [2, Proof of Proposition 7.8]. For $\mathbf{N}_{0,\pm}$, observe that $\tilde{\mathbf{N}}_{0,\pm}(\xi,\eta) := |\xi|^{-1}|\eta|^{-1}\mathbf{N}_{0,\pm}(\xi,\eta)$ obeys

$$\begin{aligned}
&|\tilde{\mathbf{N}}_{0,\pm}(\xi,\eta)| \lesssim \theta_\pm^2, \quad |\partial_\xi\tilde{\mathbf{N}}_{0,\pm}(\xi,\eta)| \lesssim 2^{-k_1}\theta_\pm, \quad |\partial_\eta\tilde{\mathbf{N}}_{0,\pm}(\xi,\eta)| \lesssim 2^{-k_2}\theta_\pm,\\
&|\partial_\xi^{(n_1)}\partial_\eta^{(n_2)}\tilde{\mathbf{N}}_{0,\pm}(\xi,\eta)| \lesssim 2^{-n_1k_1}2^{-n_2k_2} \quad (n_1+n_2\geq 2).
\end{aligned}$$

for ξ,η in the support of $P_{k_1}P_\ell^\omega(\xi)P_{k_2}P_\omega^{\ell'}(\eta)$. Using these symbol bounds, the case of $\mathbf{N}_{0,\pm}$ can be handled by essentially the same proof as in [2, Proof of Proposition 7.8]. See also [1, Section 8].

We now present algebraic lemmas, which are used to identify null forms in the Yang–Mills equation in the caloric gauge. The following lemma identifies all bilinear null forms.

Lemma 8.7. *Let $\mathcal{O}$ be a disposable bilinear operator on $\mathbb{R}^{1+4}$. Let A be a spatial 1-form and let u, v be functions in the Schwartz class on $\mathbb{R}^{1+4}$. Then we have*

$$\mathcal{O}(\mathbf{P}^\ell A, \partial_\ell u) = \sum_j \mathcal{N}(|D|^{-1}A_j, u), \tag{8.11}$$

$$\mathbf{P}_x\mathcal{O}(u,\partial_x v) = |D|^{-1}\mathcal{N}(u,v). \tag{8.12}$$

Moreover, we also have

$$\begin{aligned}
\mathcal{O}(\partial^\alpha u, \partial_\alpha v) =& \mathcal{N}_{0,+}(Q^+u, Q^+v) + \mathcal{N}_{0,+}(Q^-u, Q^-v)\\
&+ \mathcal{N}_{0,-}(Q^+u, Q^-v) + \mathcal{N}_{0,-}(Q^-u, Q^+v) + \mathcal{R}_0(u,v)
\end{aligned} \tag{8.13}$$

where

$$\begin{aligned}
\mathcal{R}_0(u',v') =& \mathcal{O}((D_t - |D|)Q^+u' + (D_t + |D|)Q^-u, D_t v')\\
&+ \mathcal{O}(|D|(Q^+u' - Q^-u'), (D_t - |D|)Q^+v' + (D_t + |D|)Q^-v').
\end{aligned} \tag{8.14}$$

Remark 8.8. As it is evident from the proof below, Lemma 8.7 readily generalizes to a disposable multilinear operator $\mathcal{O}$ that has one of the above structures with respect to two inputs. We omit the precise statement, as the notation gets unnecessarily involved. However, we point out that this is all we need in order to handle the trilinear secondary null structure.

Remark 8.9. An alternative way to make use of the null form $\mathcal{O}(\partial^\alpha u, \partial_\alpha v)$ is to rely on the simple algebraic identity

$$2\mathcal{O}(\partial^\alpha u, \partial_\alpha v) = \Box \mathcal{O}(u, v) - \mathcal{O}(\Box u, v) - \mathcal{O}(u, \Box v). \tag{8.13$'$}$$

We have elected to use the decomposition (8.13) to unify the treatment of null forms.

Proof. We begin with (8.11) and (8.12). By Remark 8.5, it suffices to consider the case when $\mathcal{O}(u, v)$ is the product uv. Then it is a well-known fact (going back to [5, 6]) that $\mathbf{P}^\ell A \partial_\ell u$ and $\mathbf{P}_j(u \partial_x v)$ are standard null forms, i.e.,

$$\mathbf{P}^\ell A \partial_\ell u = \mathbf{N}_{\ell j}((-\Delta)^{-1} \partial^\ell A^j, u), \tag{8.15}$$

$$\mathbf{P}_j(u \partial_x v) = (-\Delta)^{-1} \partial^\ell \mathbf{N}_{\ell j}(u, v). \tag{8.16}$$

We omit the simple symbol computation. Hence (8.11) and (8.12) follow.

Next, we prove (8.13), which is essentially the well-known fact that $\partial^\alpha u \partial_\alpha v = -D^\alpha u D_\alpha v$ is a null form. To verify (8.13), we first decompose $u = Q^+ u + Q^- u$ and $v = Q^+ v + Q^-$, then we substitute

$$D_t Q^\pm u = \pm |D| Q^\pm u + (D_t \mp |D|) Q^\pm u, \quad D_t Q^{\pm'} v = \pm' |D| Q^{\pm'} v + (D_t \mp' |D|) Q^{\pm'} v.$$

When $\mathcal{O}(u, v) = uv$, the contribution of the first terms give

$$\sum_{\pm, \pm'} \Big(\pm \pm' \, |D| Q^\pm u |D| Q^{\pm'} v - D^\ell Q^\pm u D_\ell Q^{\pm'} v \Big) = \sum_{\pm, \pm'} \mathbf{N}_{0, \pm\pm'}(Q^\pm u, Q^{\pm'} v).$$

By Remark 8.5, the same contribution constitutes the first four terms in (8.13) in general. Note moreover that the remainder makes up $\mathcal{R}_0(u, v)$, which proves (8.13). $\Box$

Next, we present an algebraic computation, which will be used to reveal the trilinear secondary null form of the caloric Yang–Mills wave equation.

Lemma 8.10. *Let $\mathcal{O}, \mathcal{O}'$ be disposable bilinear operators on $\mathbb{R}^{1+4}$. Then we have*

$$\begin{aligned}
&\mathcal{O}'(\Delta^{-1} \mathcal{O}(u^{(1)}, \partial_0 u^{(2)}), \partial^0 u^{(3)}) + \mathcal{O}'(\Box^{-1} \mathbf{P}_i \mathcal{O}(u^{(1)}, \partial_x u^{(2)}), \partial^i u^{(3)}) \\
&= \mathcal{O}'(\Box^{-1} \mathcal{O}(u^{(1)}, \partial_\alpha u^{(2)}), \partial^\alpha u^{(3)}) - \mathcal{O}'(\Box^{-1} \Delta^{-1} \partial_t \partial_\alpha \mathcal{O}(u^{(1)}, \partial^\alpha u^{(2)}), \partial_t u^{(3)}) \\
&\quad - \mathcal{O}'(\Box^{-1} \Delta^{-1} \partial_\ell \partial_\alpha \mathcal{O}(u^{(1)}, \partial^\ell u^{(2)}), \partial^\alpha u^{(3)}),
\end{aligned}$$

provided that $\Delta^{-1}\mathcal{O}$, $\Box^{-1}\mathcal{O}$ and $\Box^{-1}\Delta^{-1}\mathcal{O}$ are well-defined in the sense that their kernels have finite masses.

Of course, the requirement that the kernels of $\Delta^{-1}\mathcal{O}$, $\Box^{-1}\mathcal{O}$ and $\Box^{-1}\Delta^{-1}\mathcal{O}$ have finite masses is excessively strong for the validity of the lemma, but it will be verified in the applications below.

Proof. The proof of this lemma is the same as in [10, Appendix]. Using the identities

$$\Delta^{-1} - \Box^{-1} = \Box^{-1} \Delta^{-1} (-\partial_t^2), \quad \mathbf{P}_i B = B_i - \Delta^{-1} \partial_i \partial^\ell B_\ell, \quad \partial^0 = -\partial_0 = -\partial_t$$

and adding and subtracting $\mathcal{O}'(\Box^{-1}\Delta^{-1}\partial_t\partial^\ell\mathcal{O}(u^{(1)},\partial_\ell u^{(2)}),\partial_t u^{(3)})$, we may write

$$\begin{aligned}
&\mathcal{O}'(\Delta^{-1}\mathcal{O}(u^{(1)},\partial_0 u^{(2)}),\partial^0 u^{(3)}) + \mathcal{O}'(\Box^{-1}\mathbf{P}_i\mathcal{O}(u^{(1)},\partial_x u^{(2)}),\partial^i u^{(3)}) \\
&= \mathcal{O}'(\Box^{-1}\mathcal{O}(u^{(1)},\partial_0 u^{(2)}),\partial^0 u^{(3)}) + \mathcal{O}'(\Box^{-1}\mathcal{O}(u^{(1)},\partial_i u^{(2)}),\partial^i u^{(3)}) \\
&\quad - \mathcal{O}'(\Box^{-1}\Delta^{-1}\partial_t\partial^0\mathcal{O}(u^{(1)},\partial_0 u^{(2)}),\partial_t u^{(3)}) - \mathcal{O}'(\Box^{-1}\Delta^{-1}\partial_i\partial^\ell\mathcal{O}(u^{(1)},\partial_\ell u^{(2)}),\partial^i u^{(3)}) \\
&\quad - \mathcal{O}'(\Box^{-1}\Delta^{-1}\partial_t\partial^\ell\mathcal{O}(u^{(1)},\partial_\ell u^{(2)}),\partial_t u^{(3)}) - \mathcal{O}'(\Box^{-1}\Delta^{-1}\partial_0\partial^\ell\mathcal{O}(u^{(1)},\partial_\ell u^{(2)}),\partial^0 u^{(3)}) \\
&= \mathcal{O}'(\Box^{-1}\mathcal{O}(u^{(1)},\partial_\alpha u^{(2)}),\partial^\alpha u^{(3)}) - \mathcal{O}'(\Box^{-1}\Delta^{-1}\partial_t\partial_\alpha\mathcal{O}(u^{(1)},\partial^\alpha u^{(2)}),\partial_t u^{(3)}) \\
&\quad - \mathcal{O}'(\Box^{-1}\Delta^{-1}\partial_\ell\partial_\alpha\mathcal{O}(u^{(1)},\partial^\ell u^{(2)}),\partial^\alpha u^{(3)}).
\end{aligned}$$

In the last equality, we paired the first and the second, the third and the fifth, and the fourth and the sixth terms, respectively, from the preceding lines. □

8.2. **Summary of global-in-time dyadic estimates.** In what follows, we denote by $\mathcal{O}$ a disposable translation-invariant bilinear operator on $\mathbb{R}^{1+4}$, and by $\mathcal{N}$ a bilinear null form as in Definition 8.3(2). Let u and v be test functions on $\mathbb{R}^{1+4}$. For convenience, we also introduce test functions u' and v', which stands for inputs of the form ∇u and ∇v, respectively, in the applications.

Given $k, k_1, k_2 \in \mathbb{Z}$, we define $k_{\max} = \max\{k, k_1, k_2\}$ and $k_{\min} = \min\{k, k_1, k_2\}$. We use the shorthands $u_{k_1} = P_{k_1} u$, $v_{k_2} = P_{k_2} v$ and $v'_{k_2} = P_{k_2} v'$.

8.2.1. *Bilinear estimates for elliptic components.* We start with simple bilinear bounds which do not involve any null forms.

Proposition 8.11. *We have*

$$\|P_k\mathcal{O}(u_{k_1},v'_{k_2})\|_{L^2\dot{H}^{-\frac{1}{2}}} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}\|Du_{k_1}\|_{\mathrm{Str}^0}\|v'_{k_2}\|_{\mathrm{Str}^0}, \tag{8.17}$$

$$\|P_k\mathcal{O}(u_{k_1},v'_{k_2})\|_{L^{\frac{9}{5}}\dot{H}^{-\frac{4}{9}}} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}\|Du_{k_1}\|_{\mathrm{Str}^0}\|v'_{k_2}\|_{\mathrm{Str}^0}, \tag{8.18}$$

$$\|P_k\mathcal{O}(u_{k_1},v'_{k_2})\|_{L^1\dot{W}^{-2,\infty}} \lesssim 2^{-\delta_1|k_1-k_2|}\|Du_{k_1}\|_S\|v'_{k_2}\|_S. \tag{8.19}$$

Furthermore, we have the following simpler variants of (8.17)*,* (8.18) *and* (8.19)*:*

$$\|P_k\mathcal{O}(u_{k_1},v'_{k_2})\|_{L^2\dot{H}^{-\frac{1}{2}}} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}\|u_{k_1}\|_{L^2\dot{H}^{\frac{3}{2}}}\|v'_{k_2}\|_S, \tag{8.20}$$

$$\|P_k\mathcal{O}(u_{k_1},v'_{k_2})\|_{L^{\frac{9}{5}}\dot{H}^{-\frac{4}{9}}} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}\|u_{k_1}\|_{L^2\dot{H}^{\frac{3}{2}}}\|v'_{k_2}\|_S, \tag{8.21}$$

$$\|P_k\mathcal{O}(u_{k_1},v'_{k_2})\|_{L^1\dot{W}^{-2,\infty}} \lesssim 2^{\frac{2}{3}k_{\min}}2^{-\frac{4}{3}k}2^{-\frac{1}{6}k_1}2^{\frac{5}{6}k_2}(2^{\frac{1}{6}k_1}\|u_{k_1}\|_{L^2L^6})(2^{-\frac{5}{6}k_2}\|v'_{k_2}\|_{L^2L^6}). \tag{8.22}$$

8.2.2. *Bilinear estimates concerning the N-norm.* Next, we state the N-norm estimates which will be used for the bilinear expressions arising from $\mathbf{P}\mathcal{M}$, $\mathbf{P}^\perp\mathcal{M}$ and $\mathrm{Rem}^{\kappa,2}$.

Proposition 8.12. *We have*

$$\|P_k\mathcal{N}(u_{k_1},v_{k_2})\|_N \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}2^k\|Du_{k_1}\|_S\|Dv_{k_2}\|_S, \tag{8.23}$$

$$\|P_k\mathcal{O}(\partial^\alpha u_{k_1},\partial_\alpha v_{k_2})\|_N \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}2^{k_{\max}}\|Du_{k_1}\|_S\|Dv_{k_2}\|_S, \tag{8.24}$$

$$\|P_k\mathcal{O}(u'_{k_1},v_{k_2})\|_{L^1L^2} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}\|u'_{k_1}\|_{L^2\dot{H}^{\frac{1}{2}}}(2^{\frac{1}{6}k_2}\|v_{k_2}\|_{L^2L^6}). \tag{8.25}$$

Furthermore, for any $\kappa \in \mathbb{N}$, we have the low modulation gain

$$\|P_k Q_{<k_{\min}-\kappa} \mathcal{N}(Q_{<k_{\min}-\kappa} u_{k_1}, Q_{<k_{\min}-\kappa} v_{k_2})\|_N \lesssim 2^{-\delta_1 \kappa} 2^k \|Du_{k_1}\|_S \|Dv_{k_2}\|_S, \tag{8.26}$$

$$\|P_k Q_{<k_{\min}-\kappa} \mathcal{O}(\partial^\alpha Q_{<k_{\min}-\kappa} u_{k_1}, \partial_\alpha Q_{<k_{\min}-\kappa} v_{k_2})\|_N \lesssim 2^{-\delta_1 \kappa} 2^{k_{\max}} \|Du_{k_1}\|_S \|Dv_{k_2}\|_S. \tag{8.27}$$

For the term $\mathrm{Diff}^{\kappa}_{\mathbf{P}A} B$, we need to distinguish the case when the low frequency input A has a dominant modulation. For this purpose, we borrow the bilinear operator $\mathcal{H}^*_k$ (and its "dual" $\mathcal{H}_k$) from [10].

Given a bilinear translation-invariant operator $\mathcal{O}$, we introduce the expression $\mathcal{H}_k \mathcal{O}$ [resp. $\mathcal{H}^*_k \mathcal{O}$], which essentially separates out the case when the modulation of the output [resp. the first input] is dominant. More precisely, we define

$$\mathcal{H}_k \mathcal{O}(u, v) = \sum_{j : j < k+C} Q_j \mathcal{O}(Q_{<j-C} u, Q_{<j-C} v),$$

$$\mathcal{H}^*_k \mathcal{O}(u, v) = \sum_{j : j < k+C} Q_{<j-C} \mathcal{O}(Q_j u, Q_{<j-C} v),$$

for some universal constant C such that $C < C_0$, where C_0 is the constant in Lemma 8.21. We also define

$$\mathcal{H} \mathcal{O}(u, v) = \sum_{k, k_1, k_2 : k < k_2 - C} P_k \mathcal{H}_k \mathcal{O}(P_{k_1} u, P_{k_2} v),$$

$$\mathcal{H}^* \mathcal{O}(u, v) = \sum_{k, k_1, k_2 : k_1 < k_2 - C} \mathcal{H}^*_{k_1} P_k \mathcal{O}(P_{k_1} u, P_{k_2} v).$$

We are now ready to state our estimates for the N-norm of the term $\mathrm{Diff}_{\mathbf{P}A} B$.

Proposition 8.13. *For $k_1 < k - 10$, we have*

$$\|P_k (1 - \mathcal{H}^*_{k_1}) \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})\|_N \lesssim \|Du_{k_1}\|_S \|Dv_{k_2}\|_S, \tag{8.28}$$

$$\|P_k (1 - \mathcal{H}^*_{k_1}) \mathcal{O}(u_{k_1}, v'_{k_2})\|_N \lesssim \|u_{k_1}\|_{L^2 \dot{H}^{\frac{3}{2}}} \|v'_{k_2}\|_S, \tag{8.29}$$

$$\|P_k \mathcal{H}^*_{k_1} \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})\|_N \lesssim \|u_{k_1}\|_{Z^1} \|Dv_{k_2}\|_S, \tag{8.30}$$

$$\|P_k \mathcal{H}^*_{k_1} \mathcal{O}(u_{k_1}, v'_{k_2})\|_N \lesssim \|u_{k_1}\|_{\Delta^{-\frac{1}{2}} \Box^{\frac{1}{2}} Z^1} \|v'_{k_2}\|_S. \tag{8.31}$$

Furthermore, for $k_1 < k - 10$ and any $\kappa \in \mathbb{N}$, we have

$$\|P_k \mathcal{H}^*_{k_1} \mathcal{N}(|D|^{-1} Q_{<k_1-\kappa} u_{k_1}, v_{k_2})\|_N \lesssim 2^{-\delta_1 \kappa} \|u_{k_1}\|_{Z^1} \|Dv_{k_2}\|_S, \tag{8.32}$$

$$\|P_k \mathcal{H}^*_{k_1} \mathcal{O}(Q_{<k_1-\kappa} u_{k_1}, v'_{k_2})\|_N \lesssim 2^{-\delta_1 \kappa} \|u_{k_1}\|_{\Delta^{-\frac{1}{2}} \Box^{\frac{1}{2}} Z^1} \|v'_{k_2}\|_S. \tag{8.33}$$

8.2.3. *Bilinear estimates concerning $X^{s,b,p}_r$-type norms.* We now state the Z^1-, $Z^1_{p_0}$- and $\tilde{Z}^1_{p_0}$-norm bounds. We begin with the ones for the bilinear expressions arising from $\mathbf{P}\mathcal{M}^2$, $\mathrm{Rem}^{\kappa,2}_A$ and $\mathcal{M}^2_0$.

Proposition 8.14. *We have*

$$\|P_k \mathcal{N}(u_{k_1}, v_{k_2})\|_{\Box Z^1_{p_0}} \lesssim 2^{-\delta_1 (k_{\max} - k_{\min})} 2^k \|Du_{k_1}\|_S \|Dv_{k_2}\|_S, \tag{8.34}$$

$$\|P_k \mathcal{N}(u_{k_1}, v_{k_2})\|_{\Box Z^1} \lesssim 2^{-\delta_1 |k_1 - k_2|} 2^k \|Du_{k_1}\|_S \|Dv_{k_2}\|_S, \tag{8.35}$$

Furthermore, for $k \leq k_1 - C$, we have

$$\|P_k(1-\mathcal{H}_k)\mathcal{N}(u_{k_1}, v_{k_2})\|_{\Box Z^1} \lesssim 2^{-\delta_1(k_1-k)} 2^k \|Du_{k_1}\|_S \|Dv_{k_2}\|_S, \tag{8.36}$$

$$\|P_k(1-\mathcal{H}_k)\mathcal{O}(u_{k_1}, v'_{k_2})\|_{\Delta^{\frac{1}{2}}\Box^{\frac{1}{2}} Z^1} \lesssim 2^{-\delta_1(k_1-k)} \|Du_{k_1}\|_S \|v'_{k_2}\|_S. \tag{8.37}$$

The following bounds are for the null form arising from $\mathrm{Diff}^{\kappa}_{\mathbf{P}_x A} B$; we remark that this is the only place where we need to use the intermediate $\tilde{Z}^1_{p_0}$-norm.

Proposition 8.15. *We have*

$$\|P_k \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})\|_{\Box \tilde{Z}^1_{p_0}} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})} \|u_{k_1}\|_{S^1} \|Dv_{k_2}\|_S, \tag{8.38}$$

$$\|P_k \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})\|_{\Box Z^1_{p_0}} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})} \|u_{k_1}\|_{S^1 \cap \tilde{Z}^1_{p_0}} \|Dv_{k_2}\|_S, \tag{8.39}$$

$$\|P_k \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})\|_{\Box Z^1} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})} \|u_{k_1}\|_{S^1 \cap Z^1_{p_0}} \|Dv_{k_2}\|_S, \tag{8.40}$$

$$\|P_k \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})\|_{X^{-\frac{1}{2}+b_1, -b_1}} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})} \|u_{k_1}\|_{S^1 \cap Z^1_{p_0}} \|Dv_{k_2}\|_S. \tag{8.41}$$

Finally, the following bounds are used to handle $\mathrm{Diff}^{\kappa}_{A_0} B$ and $\mathrm{Diff}^{\kappa}_{\mathbf{P}^\perp A} B$.

Proposition 8.16. *We have*

$$\|P_k \mathcal{O}(u_{k_1}, v'_{k_2})\|_{\Box Z^1_{p_0}} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})} \|Du_{k_1}\|_Y \|v'_{k_2}\|_S, \tag{8.42}$$

$$\|P_k \mathcal{O}(u_{k_1}, v'_{k_2})\|_{\Box Z^1} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})} \|Du_{k_1}\|_Y \|v'_{k_2}\|_S, \tag{8.43}$$

$$\|P_k \mathcal{O}(u_{k_1}, v'_{k_2})\|_{X^{-\frac{1}{2}+b_1, -b_1}} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})} \|Du_{k_1}\|_Y \|v'_{k_2}\|_S. \tag{8.44}$$

8.2.4. *Trilinear null form estimate.* Let $u^{(1)}, u^{(2)}, u^{(3)}$ be test function on $\mathbb{R}^{1+4}$. Given $k_i \in \mathbb{Z}$, we introduce the shorthand $u^{(i)}_{k_i} = P_{k_i} u^{(i)}$ $(i = 1, 2, 3)$.

Proposition 8.17. *Let $\mathcal{O}$ and $\mathcal{O}'$ be disposable bilinear operators on $\mathbb{R}^{1+4}$. Let $j < k - C$ and $k < \min\{k_0, k_1, \ldots, k_3\} - C$. Consider the expression*

$$\begin{aligned}\mathcal{N}^{cubic}_{k,j}(u^{(1)}_{k_1}, u^{(2)}_{k_2}, u^{(3)}_{k_3}) =& Q_{<j-C}\mathcal{O}'(\Delta^{-1} P_k Q_j \mathcal{O}(Q_{<j-C} u^{(1)}_{k_1}, \partial_0 Q_{<j-C} u^{(2)}_{k_2}), \partial^0 Q_{<j-C} u^{(3)}_{k_3}) \\ &+ Q_{<j-C}\mathcal{O}'(\Box^{-1} P_k Q_j \mathbf{P}_\ell \mathcal{O}(Q_{<j-C} u^{(1)}_{k_1}, \partial_x Q_{<j-C} u^{(2)}_{k_2}), \partial^\ell Q_{<j-C} u^{(3)}_{k_3}).\end{aligned}$$

Then we have

$$\|\mathcal{N}^{cubic}_{k,j}(u^{(1)}_{k_1}, u^{(2)}_{k_2}, u^{(3)}_{k_3})\|_{L^1 L^2} \lesssim 2^{-\delta_1(k_1-k)} 2^{-\delta_1(k-j)} \|Du^{(1)}_{k_1}\|_S \|Du^{(2)}_{k_2}\|_S \|Du^{(3)}_{k_3}\|_S. \tag{8.45}$$

In fact, for later use (in Section 11), it is convenient to also state a more atomic form of (8.45). Given $k_i \in \mathbb{Z}$ and a rectangular box $\mathcal{C}^{(i)}$, we use the shorthand $u^{(i)}_{k_i, \mathcal{C}^{(i)}} = P_{k_i} P_{\mathcal{C}^{(i)}} u^{(i)}$ $(i = 1, 2)$.

Proposition 8.18. *Let $\mathcal{O}$ and $\mathcal{O}'$ be translation-invariant bilinear operators on $\mathbb{R}^{1+4}$ such that $\mathcal{O}(P^{\omega}_{\ell} \cdot, P^{\omega'}_{\ell'} \cdot)$ and $\mathcal{O}'(P^{\omega}_{\ell} \cdot, P^{\omega'}_{\ell'} \cdot)$ are disposable for every $\ell, \ell' \in -\mathbb{N}$ and $\omega, \omega' \in \mathbb{S}^3$. Let*

$j < k - C$, $k < \min\{k_0, k_1, \dots, k_3\} - C$ *and* $\mathcal{C}^{(1)}, \mathcal{C}^{(2)} \in \{\mathcal{C}_k(\ell)\}$, *where* $\ell = \frac{j-k}{2}$. *We have*

$$\|P_{k_0} Q_{<j-C} \mathcal{O}'(\Box^{-1} P_k Q_j \mathcal{O}(Q_{<j-C} u^{(1)}_{k_1, \mathcal{C}^{(1)}}, \partial_\alpha Q_{<j-C} u^{(2)}_{k_2, \mathcal{C}^{(2)}}), \partial^\alpha Q_{<j-C} u^{(3)}_{k_3})\|_{L^1 L^2}$$
$$\lesssim 2^{-\delta_1(k_1 - k)} 2^{-\delta_1(k-j)} \|D u^{(1)}_{k_1, \mathcal{C}^{(1)}}\|_{S_{k_1}[\mathcal{C}_k(\ell)]} \|D u^{(2)}_{k_2, \mathcal{C}^{(2)}}\|_{S_{k_2}[\mathcal{C}_k(\ell)]} \|D u^{(3)}_{k_3}\|_S, \tag{8.46}$$

$$\|P_{k_0} Q_{<j-C} \mathcal{O}'(\Box^{-1} \Delta^{-1} P_k Q_j \partial_t \partial_\alpha \mathcal{O}(Q_{<j-C} u^{(1)}_{k_1, \mathcal{C}^{(1)}}, \partial^\alpha Q_{<j-C} u^{(2)}_{k_2, \mathcal{C}^{(2)}}), \partial_t Q_{<j-C} u^{(3)}_{k_3})\|_{L^1 L^2}$$
$$\lesssim 2^{-\delta_1(k_1 - k)} 2^{-\delta_1(k-j)} \|D u^{(1)}_{k_1, \mathcal{C}^{(1)}}\|_{S_{k_1}[\mathcal{C}_k(\ell)]} \|D u^{(2)}_{k_2, \mathcal{C}^{(2)}}\|_{S_{k_2}[\mathcal{C}_k(\ell)]} \|D u^{(3)}_{k_3}\|_S, \tag{8.47}$$

$$\|P_{k_0} Q_{<j-C} \mathcal{O}'(\Box^{-1} \Delta^{-1} P_k Q_j \partial_\ell \partial_\alpha \mathcal{O}(Q_{<j-C} u^{(1)}_{k_1, \mathcal{C}^{(1)}}, \partial^\ell Q_{<j-C} u^{(2)}_{k_2, \mathcal{C}^{(2)}}), \partial^\alpha Q_{<j-C} u^{(3)}_{k_3})\|_{L^1 L^2}$$
$$\lesssim 2^{-\delta_1(k_1 - k)} 2^{-\delta_1(k-j)} \|D u^{(1)}_{k_1, \mathcal{C}^{(1)}}\|_{S_{k_1}[\mathcal{C}_k(\ell)]} \|D u^{(2)}_{k_2, \mathcal{C}^{(2)}}\|_{S_{k_2}[\mathcal{C}_k(\ell)]} \|D u^{(3)}_{k_3}\|_S. \tag{8.48}$$

8.3. **Proof of the interval-localized estimates.** In this subsection, we prove all estimates claimed in Section 4 except Theorem 4.24 and Proposition 4.25, which are proved in the next section.

The key technical issue we address here is passage to interval-localized frequency envelope bounds (as stated in Section 4) from the global-in-time dyadic estimates stated in Section 8.2.

In what follows, we denote by $\mathcal{O}$ and $\mathbf{O}$ disposable multilinear operators on $\mathbb{R}^{1+4}$ and $\mathbb{R}^4$, respectively, which may vary from line to line. Similarly, χ^k_I indicates a generalized time cutoff adapted to the scale 2^{-k}, which may vary from line to line.

8.3.1. *Estimates that do not involve any null forms.* Here we establish Propositions 4.12, 4.13, 4.14 and 4.18, whose proofs do not involve any null forms.

Proofs of Propositions 4.12 and 4.13. We introduce the shorthand $A' = \partial_t A$ and $B' = \partial_t B$. Using (4.25) and Lemma 8.1 to write

$$|D|^{-1} P_k \mathcal{M}^2_0(P_{k_1} A, P_{k_2} B) = 2^{-k} P_k \mathbf{O}(P_{k_1} A, P_{k_2} B'), \tag{8.49}$$
$$P_k \mathbf{Q}(P_{k_1} A, P_{k_2} B) = 2^k 2^{-k_{\max}} P_k \mathbf{O}(P_{k_1} A, P_{k_2} B), \tag{8.50}$$
$$|D|^{-1} P_k \mathbf{Q}(P_{k_1} A, P_{k_2} \partial_t B) = 2^{-k_{\max}} P_k \mathbf{O}(P_{k_1} A, P_{k_2} B'), \tag{8.51}$$
$$|D|^{-2} P_k \mathcal{D}\mathcal{M}^2_0(P_{k_1} A, P_{k_2} B) = 2^{-k} 2^{-k_{\max}} P_k \mathbf{O}(P_{k_1} A', P_{k_2} B'). \tag{8.52}$$

Step 1: Fixed-time estimates. Applying Hölder and Bernstein (to one of the inputs or the output, whichever has the lowest frequency), we obtain

$$\|P_k \mathbf{O}(P_{k_1} u', P_{k_2} v')\|_{L^2} \lesssim 2^{2k_{\min}} \|u'\|_{L^2} \|v'\|_{L^2}. \tag{8.53}$$

Recalling (8.49)–(8.52), the fixed-time estimates (4.27), (4.28) and (4.35) follow.

Step 2: Space-time estimates. Here, we prove the remaining estimates in Propositions 4.12 and 4.13. In this step, we simply extend A, B, A', B' by zero outside I. Furthermore, we define

$$\mathcal{M}^{\kappa,2}_{0,small}(A, B) = \sum_{|k_{\max} - k_{\min}| \geq \kappa} P_k \mathcal{M}^2_0(P_{k_1} A, P_{k_2} B), \tag{8.54}$$
$$\mathcal{M}^{\kappa,2}_{0,large}(A, B) = \sum_{|k_{\max} - k_{\min}| < \kappa} P_k \mathcal{M}^2_0(P_{k_1} A, P_{k_2} B). \tag{8.55}$$

so that $\mathcal{M}^{\kappa,2}_0(A, B) = \mathcal{M}^{\kappa,2}_{0,small}(A, B) + \mathcal{M}^{\kappa,2}_{0,large}(A, B)$.

Step 2.1: $L^2\dot H^{\frac12}$-norm estimates. We first verify (4.29)–(4.34), (4.36) and (4.38) with the $L^2\dot H^{\frac12}$-norm (instead of the Y-norm) on the LHS. All of these estimates follow from (8.17) and (8.49)–(8.52). The small factor in (4.31) arises from the exponential gain in (8.17) and the frequency gap κ in (8.54), whereas the factor $\varepsilon^{\delta_2}M$ in (4.33), (4.34) and (4.38) arises from (4.21).

Step 2.2: L^1L^∞-norm estimates. By Hölder's inequality, we have

$$\|P_k u\|_{L^{p_0}\dot W^{2-\frac{3}{p_0},p_0'}} \lesssim \|P_k u\|_{L^2\dot H^{\frac12}}^{1-\theta_0}\|P_k u\|_{L^1\dot W^{-1,\infty}}^{\theta_0} \tag{8.56}$$

where $\theta_0 = 2(\frac{1}{p_0}-\frac12)\in(0,1)$. Therefore, (4.29), (4.31) and (4.33) follow by combining (8.19) with the $L^2\dot H^{\frac12}$-norm estimates from Step 2.1. On the other hand, for (4.32) we use (8.22) instead of (8.19), which allows us to use the DS^1-norm on the RHS at the expense of losing the exponential off-diagonal gain. Finally, for (4.37) and (4.38), observe that by (8.22), (8.50) and (8.51) we have

$$\||D|^{-\sigma-1}\mathbf{Q}(P_{k_1}A, P_{k_2}B')\|_{L^1L^\infty} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}\|P_{k_1}A\|_{DS^1}\||D|^{-\sigma}P_{k_2}B'\|_{DS^1},$$

for $\sigma=0,1$. Therefore, the L^1L^∞-norm bound in (4.37) follows directly, whereas the Y-norm bound in (4.37) and (4.38) follow after interpolating with the $L^2\dot H^{\frac12}$-norm estimates from Step 2.1. □

Proofs of Proposition 4.14. For this proof we use the square function $L_x^{\frac{10}{3}}L_t^2$ component of the S_k norm, for which we have

$$\|u\|_{S_k^{sq}} = 2^{-\frac{3}{10}k}\|u\|_{L_x^{\frac{10}{3}}L_t^2}.$$

We recall that the symbol of $\Delta\mathbf{A}_0^2$ is

$$\Delta\mathbf{A}_0^2(\xi,\eta) = \frac{|\xi|^2}{|\xi|^2+|\eta|^2}$$

Then we use Bernstein at the lowest frequency to estimate

$$\|P_k\Delta\mathbf{A}_0^2(A_{k_1},\partial_t A_{k_2})\|_{L^2L^1} \lesssim 2^{-2(k_2-k_1)_+}2^{-\frac{7}{10}k_1}2^{\frac{3}{10}k_2}2^{\frac{4}{10}k_{min}}c_{k_1}c_{k_2} \lesssim 2^{-\frac{3}{10}(k_{max}-k_{min})}c_{k_1}c_{k_2}.$$

Now the bound (4.39) immediately follows due to the off-diagonal decay.

□

Proof of Proposition 4.18. The bounds in this proposition are trivial consequences of Proposition 8.11, along with the observation that $\||D|u\|_{\mathrm{Str}^0} \lesssim \|\nabla u\|_{L^2\dot H^{\frac12}}$. We omit the details. □

8.3.2. *Estimates for $\mathbf{P}\mathcal{M}^2$, $\mathbf{P}^\perp\mathcal{M}^2$ and $\mathrm{Rem}^{2,\kappa}$.* We now present the proofs of Propositions 4.15 and 4.20, which require the bilinear null form estimates in Propositions 8.12, as well as the $X_r^{s,b,p}$-type norm estimates in Propositions 8.14, 8.15 and 8.16.

Proof of Proposition 4.15. Unless otherwise stated, we extend the inputs A, B by homogeneous waves outside I. For $k, k_1, k_2 \in \mathbb{Z}$, by Lemma 8.1, note that

$$P_k\mathbf{P}\mathcal{M}^2(P_{k_1}A, P_{k_2}B) = P_k\mathbf{P}\mathbf{O}(P_{k_1}A, \partial_x P_{k_2}B), \tag{8.57}$$

$$P_k\mathbf{P}^\perp\mathcal{M}^2(P_{k_1}A, P_{k_2}B) = 2^{-k_{\max}}P_k\mathbf{O}(\partial_\alpha P_{k_1}A, \partial^\alpha P_{k_2}B), \tag{8.58}$$

for some disposable operator $\mathbf{O}$ on $\mathbb{R}^4$. Note also that, by Lemma 8.7, the RHSs are null forms.

Step 0: Proofs of (4.40)**,** (4.40)**.** In view of (8.57) and (8.57), both follow easily using the standard Littlewood-Paley trichotomy and (8.53).

Step 1: Proofs of (4.42)**,** (4.43)**,** (4.44) **and** (4.45)**.** The N-norm bounds in (4.42) and (4.43) follow from the null form estimates (8.23)–(8.24). On the other hand, the $\Box \underline{X}^1$-norm bounds in (4.42) and (4.43) follow from (8.17), (8.18) and (8.34); we remark that the $\Box Z^1_{p_0}$-norm bound for $\mathbf{P}^\perp \mathcal{M}$ is unnecessary, since $\mathbf{P}\mathbf{P}^\perp \mathcal{M} = 0$. Estimates (4.44) and (4.45) immediately follow from (8.17), where we may simply extend $A, \partial_t A, B, \partial_t B$ by zero outside I as in the proofs of Propositions 4.12 and 4.13 above.

Step 2: Proofs of (4.46)**,** (4.47)**,** (4.48) **and** (4.49)**.** Since the case of $\mathbf{P}\mathcal{M}^2$ (i.e., estimates (4.46) and (4.48)) can be read off from [17, Proof of Proposition 4.1], we will only provide a detailed proof in the case of $\mathbf{P}^\perp \mathcal{M}^2$ (i.e., estimates (4.47), (4.49)).

Step 2.1: Off-diagonal dyadic frequencies. If $\max\{|k-k_1|, |k-k_2|\} \geq \kappa$, then by (8.24) we have

$$\begin{aligned} \|P_k \mathbf{P}^\perp \mathcal{M}^2(P_{k_1} A, P_{k_2} B)\|_N &\lesssim 2^{-\delta_1 (k_{\max} - k_{\min})} \|P_{k_1} A\|_{S^1} \|P_{k_2} B\|_{S^1} \\ &\lesssim 2^{-\frac{1}{2}\delta_1 \kappa} 2^{-\frac{1}{2}\delta_1 (k_{\max} - k_{\min})} \|P_{k_1} A\|_{S^1} \|P_{k_2} B\|_{S^1}. \end{aligned}$$

Hence the contribution in the case $\max\{|k-k_1|, |k-k_2|\} \geq \kappa$ can always be put in $\mathbf{P}^\perp \mathcal{M}^{\kappa,2}_{small}$.

Step 2.2: Balanced dyadic frequencies, short time interval. Next, we consider the case when $|k-k_1| < \kappa$, $|k-k_2| < \kappa$ and $|I| \leq 2^{-k+C\kappa}$. Then by Hölder and (8.58), we simply estimate

$$\begin{aligned} \|P_k \mathbf{P}^\perp \mathcal{M}^2(P_{k_1} A, P_{k_2} B)\|_{L^1 L^2[I]} &\lesssim |I|^{\frac{1}{2}} \|P_k \mathbf{P}^\perp \mathcal{M}^2(P_{k_1} A, P_{k_2} B)\|_{L^2 L^2[I]} \\ &\lesssim |I|^{\frac{1}{2}} 2^{-k_{\max}} \|\mathcal{O}(\partial^\alpha P_{k_1} A, \partial_\alpha P_{k_2} B)\|_{L^2 L^2} \\ &\lesssim 2^{C\kappa} \||D|^{-\frac{3}{4}} \nabla A_{k_1}\|_{L^4 L^4[I]} \||D|^{-\frac{3}{4}} \nabla B_{k_2}\|_{L^4 L^4[I]}. \end{aligned}$$

Therefore, when $|I| \leq 2^{-k+C\kappa}$, the contribution in the case $\max\{|k-k_1|, |k-k_2|\} < \kappa$ can be put in $\mathbf{P}^\perp \mathcal{M}^{\kappa,2}_{large}$.

Step 2.3: Balanced dyadic frequencies, long time interval. Finally, we consider the case when $|k-k_1| < \kappa$, $|k-k_2| < \kappa$ and $|I| \geq 2^{-k+C\kappa}$. We define $\mathbf{P}^\perp \mathcal{M}^{\kappa,2}_{large}$ by the relation

$$\begin{aligned} &\sum_{\max\{|k-k_1|, |k-k_2|\} < \kappa} P_k \mathbf{P}^\perp \mathcal{M}^2(P_{k_1} A, P_{k_2} B) \\ &= \sum_{\max\{|k-k_1|, |k-k_2|\} < \kappa} P_k Q_{<k_{\min}-\kappa} \mathbf{P}^\perp \mathcal{M}^2(P_{k_1} Q_{<k_{\min}-\kappa} A, P_{k_2} Q_{<k_{\min}-\kappa} B) + \mathbf{P}^\perp \mathcal{M}^{\kappa,2}_{large}(A, B). \end{aligned}$$

By (8.27), the first term on the RHS gains a factor of $2^{-c\delta_1 \kappa}$, and therefore can be put in $\mathbf{P}^\perp \mathcal{M}^{\kappa,2}_{small}$. Now it only remains to establish (4.49) for $\mathbf{P}^\perp \mathcal{M}^{\kappa,2}_{large}$ defined as above.

By definition, $\mathbf{P}^\perp \mathcal{M}^{\kappa,2}_{large}(A, B)$ is the sum over $\{(k, k_1, k_2) : \max\{|k-k_1|, |k-k_2|\} < \kappa\}$ of

$$P_k \mathbf{P}^\perp \mathcal{M}^2(P_{k_1} A, P_{k_2} B) - P_k Q_{<k_{\min}-\kappa} \mathbf{P}^\perp \mathcal{M}^2(P_{k_1} Q_{<k_{\min}-\kappa} A, P_{k_2} Q_{<k_{\min}-\kappa} B).$$

Since we are allowed to lose an exponential factor in κ in (4.49), it suffices to freeze k, k_1, k_2 and estimate the preceding expression. At this point, we divide into three subcases:

Step 2.3.a: Output has high modulation. When the output has modulation $\geq 2^{k_{\min}-\kappa}$, we use the $X_1^{0,-\frac{1}{2}}$-component of the N-norm. Since the kernel of $P_k Q_{\geq k_{\min}-\kappa}$ decays rapidly in t on the scale $\simeq 2^{-k} 2^{C\kappa}$, we have

$$\| P_k Q_{\geq k_{\min}-\kappa} \mathbf{P}^{\perp} \mathcal{M}^2 (P_{k_1} A, P_{k_2} B) \|_{X_1^{0,-\frac{1}{2}}[I]} \lesssim 2^{C\kappa} 2^{-\frac{1}{2}k} \| \chi_I^k \mathbf{P}^{\perp} \mathcal{M}^2 (P_{k_1} A, P_{k_2} A) \|_{L^2 L^2},$$

for some generalized cutoff function χ_I^k adapted to the scale 2^{-k}. Then, by Proposition 4.10,

$$\begin{aligned} 2^{C\kappa} 2^{-\frac{1}{2}k} \| \chi_I^k \mathbf{P}^{\perp} \mathcal{M}^2 (P_{k_1} A, P_{k_2} A) \|_{L^2 L^2} &\lesssim 2^{C\kappa} \| \chi_I^k |D|^{-\frac{3}{4}} \nabla P_{k_1} A \|_{L^4 L^4} \| \chi_I^k |D|^{-\frac{3}{4}} \nabla P_{k_2} B \|_{L^4 L^4} \\ &\lesssim 2^{C\kappa} \| |D|^{-\frac{3}{4}} \nabla P_{k_1} A \|_{L^4 L^4[I]} \| |D|^{-\frac{3}{4}} \nabla P_{k_2} B \|_{L^4 L^4[I]}, \end{aligned}$$

which is acceptable.

Step 2.3.b: A has high modulation. Next, we consider the case when the output has modulation $< 2^{k_{\min}-\kappa}$, yet A has modulation $\geq 2^{k_{\min}-\kappa}$. The kernel of $P_k Q_{< k_{\min}-\kappa}$ again decays rapidly in t on the scale $\simeq 2^{-k} 2^{C\kappa}$. For any $2 \leq q \leq \infty$, we have

$$\begin{aligned} &\| P_k Q_{< k_{\min}-\kappa} \mathbf{P}^{\perp} \mathcal{M}^2 (Q_{\geq k_{\min}-\kappa} P_{k_1} A, P_{k_2} B) \|_{L^1 L^2[I]} \\ &\quad \lesssim 2^{C\kappa} \| \chi_I^k \mathbf{P}^{\perp} \mathcal{M}^2 (Q_{\geq k_1-\kappa} P_{k_1} A, P_{k_2} B) \|_{L^1 L^2} \\ &\quad \lesssim 2^{C\kappa} \| |D|^{-\frac{1}{q}} \Box P_{k_1} A \|_{L^{q'} L^2} \| \chi_I^k |D|^{2-\frac{1}{q}} \nabla P_{k_2} B \|_{L^q L^\infty} \\ &\quad \lesssim 2^{C\kappa} \| |D|^{-\frac{1}{q}} \Box P_{k_1} A \|_{L^{q'} L^2[I]} \| |D|^{2-\frac{1}{q}} \nabla P_{k_2} B \|_{L^q L^\infty[I]}, \end{aligned}$$

where we used Proposition 4.10 on the last line. Taking $q = 2$, we see that the last line is bounded by $\lesssim 2^{C\kappa} \| \Box P_{k_1} A \|_{L^2 \dot{H}^{-\frac{1}{2}}[I]} \| P_{k_2} B \|_{DS^1[I]}$, which is acceptable.

Step 2.3.c: B has high modulation. Finally, the only remaining case is when the output and A have modulation $< 2^{k_{\min}-\kappa}$, but B has modulation $\geq 2^{k_{\min}-\kappa}$. Proceeding as in Step 2.3.b, and using the fact that the kernel of $P_{k_1} Q_{< k_{\min}-\kappa}$ decays rapidly in t on the scale $\simeq 2^{-k} 2^{C\kappa}$, we have

$$\begin{aligned} &\| P_k Q_{< k_{\min}-\kappa} \mathbf{P}^{\perp} \mathcal{M}^2 (Q_{< k_{\min}-\kappa} P_{k_1} A, Q_{\geq k_{\min}-\kappa} P_{k_2} B) \|_{L^1 L^2[I]} \\ &\quad \lesssim 2^{C\kappa} \| \chi_I^k \mathbf{P}^{\perp} \mathcal{M}^2 (Q_{< k_1-\kappa} P_{k_1} A, Q_{\geq k_2-\kappa} P_{k_2} B) \|_{L^1 L^2} \\ &\quad \lesssim 2^{C\kappa} \| \chi_I^k |D|^{-\frac{3}{2}} \nabla Q_{< k_{\min}-\kappa} P_{k_1} A \|_{L^2 L^\infty} \| |D|^{-\frac{1}{2}} \Box P_{k_2} B \|_{L^2 L^2} \\ &\quad \lesssim 2^{C\kappa} \| |D|^{-\frac{3}{2}} \nabla P_{k_1} A \|_{L^2 L^\infty[I]} \| \Box P_{k_2} B \|_{L^2 \dot{H}^{-\frac{1}{2}}[I]}, \end{aligned}$$

which is acceptable.

Step 3: Proofs of (4.50) and (4.51). Since the $L^2 \dot{H}^{-\frac{1}{2}}$-norm bounds follow from (4.21), (4.44) and (4.45), it remains to only consider the N-norm. The case of $\mathbf{P}\mathcal{M}^2$ can be read off from [17, Proof of Proposition 4.1]. Finally, for $\mathbf{P}^{\perp}\mathcal{M}^2$, we split into the small and large parts as in Step 2. For the small part, we already have

$$\| \mathbf{P}^{\perp} \mathcal{M}^{\kappa,2}_{small} (A, B) \|_{N_c[I]} \lesssim 2^{-c\delta_1 \kappa} \| A \|_{S^1_c[I]} M.$$

For the large part, we proceed as in Step 2, except we choose $q = \frac{9}{4}$ in Step 2.3.b. Then by (4.20), (4.21) and the embedding $\mathrm{Str}^1[I] \subseteq L^4 L^4[I] \cap L^{\frac{9}{4}} L^\infty[I]$, it follows that

$$\|\mathbf{P}^\perp \mathcal{M}^{\kappa,2}_{large}(A,B)\|_{N_c[I]} \lesssim 2^{C\kappa} \varepsilon^{\delta_1} \|A\|_{\underline{S}^1_c[I]} M.$$

Therefore, choosing $2^{-\kappa} = \varepsilon^c$ with $c > 0$ sufficiently small, (4.51) follows. □

Remark 8.19. As a corollary of the preceding proof in the case of $\mathbf{P}\mathcal{M}^2$, we obtain the following statement: Let $\mathbf{O}$ be a disposable operator on $\mathbb{R}^4$, and let A, B be $\mathfrak{g}$-valued functions (or 1-forms) on I. Then we have

$$\begin{aligned} &\|P_k(\mathbf{O}(\partial_i P_{k_1} A, \partial_j P_{k_2} B) - \mathbf{O}(\partial_j P_{k_1} A, \partial_i P_{k_2} B))\|_{N[I]} \\ &\quad \lesssim 2^{C(k_{\max} - k_{\min})} 2^k \|P_{k_1} A\|_{DS^1[I]} \|P_{k_2} B\|_{DS^1[I]}. \end{aligned} \tag{8.59}$$

Moreover, if (B, I) is (ε, M)-energy dispersed, then

$$\begin{aligned} &\|P_k(\mathbf{O}(\partial_i P_{k_1} A, \partial_j P_{k_2} B) - \mathbf{O}(\partial_j P_{k_1} A, \partial_i P_{k_2} B))\|_{N[I]} \\ &\quad \lesssim 2^{C(k_{\max} - k_{\min})} 2^k \varepsilon^{c\delta_1} \|P_{k_1} A\|_{\underline{S}^1[I]} M. \end{aligned} \tag{8.60}$$

Proof of Proposition 4.20. We decompose $\mathrm{Rem}^{\kappa,2}_A B$ into

$$\mathrm{Rem}^{\kappa,2}_A B = \mathrm{Rem}^{\kappa,2}_{\mathbf{P}_x A} B + \mathrm{Rem}^{\kappa,2}_{\mathbf{P}^\perp A} B + \mathrm{Rem}^{\kappa,2}_{A_0} B,$$

where

$$\mathrm{Rem}^{\kappa,2}_{\mathbf{P}_x A} B = \sum_{k, k_1, k_2 : k_1 \geq k_2 - \kappa} 2 P_k [\mathbf{P}_\ell P_{k_1} A, \partial^\ell P_{k_2} B] \tag{8.61}$$

$$\mathrm{Rem}^{\kappa,2}_{\mathbf{P}^\perp A} B = \sum_{k, k_1, k_2 : k_1 \geq k_2 - \kappa} 2 P_k [P_{k_1} \mathbf{P}^\perp_\ell A, \partial^\ell P_{k_2} B] \tag{8.62}$$

$$\mathrm{Rem}^{\kappa,2}_{A_0} B = - \sum_{k, k_1, k_2 : k_1 \geq k_2 - \kappa} 2 P_k [P_{k_1} A_0, P_{k_2} \partial_t B] \tag{8.63}$$

By Littlewood–Paley trichotomy, note that the summands on the RHSs of (8.61)–(8.63) vanish unless $k - k_1 \leq \kappa + C$.

Unless otherwise stated, we extend the in may not coincide with $\mathbf{P}^\perp$ of the extended A outside I in general.

Step 1: Proofs of (4.77) **and** (4.78)**.** The N-norm bound in (4.77) follows from Lemma 8.7 and (8.23) for $\mathrm{Rem}^{\kappa,2}_{\mathbf{P}_x A} B$, and (8.25) for $\mathrm{Rem}^{\kappa,2}_{\mathbf{P}^\perp A} B$, $\mathrm{Rem}^{\kappa,2}_{A_0} B$. On the other hand, for the $\Box \underline{X}^1$-norm bound in (4.77), we apply (8.17), (8.18), (8.34) to $\mathrm{Rem}^{\kappa,2}_{\mathbf{P}_x A} B$, and (8.20), (8.21) and (8.42) to $\mathrm{Rem}^{\kappa,2}_{\mathbf{P}^\perp A} B$, $\mathrm{Rem}^{\kappa,2}_{A_0} B$. Finally, (4.78) follows from (8.17) and (8.20).

Step 2: Proofs of (4.79)**,** (4.80) **and** (4.81)**.** The term $\mathrm{Rem}^{\kappa,2}_{A_0} B$ can be put in $\mathrm{Rem}^{\kappa,2}_{A,large} B$, since for each triple (k, k_1, k_2) within the range $k_1 \geq k_2 - \kappa$, by (8.25) we have

$$\begin{aligned} \|P_k [P_{k_1} A_0, P_{k_2} \partial_t B]\|_{L^1 L^2[I]} &= \|P_k \mathcal{O}(\chi_I P_{k_1} A_0, \chi_I P_{k_2} \partial_t B)\|_{L^1 L^2} \\ &\lesssim 2^{k_2 - k_1} \|P_k \mathcal{O}(\chi_I |D| P_{k_1} A_0, \chi_I |D|^{-1} P_{k_2} \partial_t B)\|_{L^1 L^2} \\ &\lesssim 2^\kappa 2^{-\delta_1 (k_{\max} - k_{\min})} \|P_{k_1} A_0\|_{L^2 \dot{H}^{\frac{3}{2}}[I]} \|P_{k_2} B\|_{DS^1[I]}. \end{aligned}$$

Similarly, the term $\mathrm{Rem}^{\kappa,2}_{\mathbf{P}^\perp A} B$ can be put in $\mathrm{Rem}^{\kappa,2}_{A,large} B$. Moreover, the contribution of these two terms to (4.81) are clearly acceptable, since they need not gain any small factor.

It remains to handle the term $\mathrm{Rem}^{\kappa,2}_{\mathbf{P}_x A} B$. We proceed differently according to the length of I. If $|I| \leq 2^{-k+C\kappa}$, we define

$$\mathrm{Rem}^{\kappa,2}_{A,small} B = \sum_{k,k_1,k_2 : k_1 \geq k_2 - \kappa,\, \max\{|k_1-k_2|,|k_1-k|\} \geq C_0\kappa} 2 P_k [\mathbf{P}_\ell P_{k_1} A, \partial^\ell P_{k_2} B],$$

and if $|I| \geq 2^{-k+C\kappa}$, we define

$$\begin{aligned} &\mathrm{Rem}^{\kappa,2}_{A,small} B \\ &= \sum_{k,k_1,k_2 : k_1 \geq k_2 - \kappa,\, \max\{|k_1-k_2|,|k_1-k|\} \geq C_0\kappa} 2 P_k [\mathbf{P}_\ell P_{k_1} A, \partial^\ell P_{k_2} B] \\ &\quad + \sum_{k,k_1,k_2 : \max\{|k_1-k_2|,|k_1-k|\} < C_0\kappa} 2 P_k Q_{<k_{\min} - C_0\kappa} [\mathbf{P}_\ell P_{k_1} Q_{<k_{\min} - C_0\kappa} A, \partial^\ell P_{k_2} Q_{<k_{\min} - C_0\kappa} B]. \end{aligned}$$

In both cases, we put the remainder $\mathrm{Rem}^{\kappa,2}_{\mathbf{P}_x A} B - \mathrm{Rem}^{\kappa,2}_{A,small} B$ in $\mathrm{Rem}^{\kappa,2}_{A,large} B$.

Choosing $C_0 > 0$ large enough (depending on δ_1), it follows from Lemma 8.7, (8.23) and (8.26) that $\mathrm{Rem}^{\kappa,2}_{A,small} B$ obeys the desired bound (4.79); this bound is also acceptable for (4.81). On the other hand, the contribution of $\mathrm{Rem}^{\kappa,2}_{\mathbf{P}_x A} B - \mathrm{Rem}^{\kappa,2}_{A,small} B$ in (4.80) and (4.81) can be handled by proceeding as in Steps 2.2–2.3 and 3 in Proof of Proposition 4.15; for the details, we refer to [17, Proof of Proposition 4.6]. □

8.3.3. *Estimates for* $\mathrm{Diff}^{\kappa}_{\mathbf{P}^\perp A} B$ *and high modulation estimates for* $\mathrm{Diff}^{\kappa}_{\mathbf{P} A} B$. Next, we prove Propositions 4.21 and 4.22, which mainly concern the $X^{-\frac{1}{2}+b_1,-b_1} \cap \Box \underline{X}^1$-norm of $\mathrm{Diff}^{\kappa}_{\mathbf{P}^\perp A} B$ and $\mathrm{Diff}^{\kappa}_{\mathbf{P} A} B$.

Proof of Proposition 4.21. We extend B by homogeneous waves outside I, and $\mathbf{P}^\perp A$ by zero outside I. Note that

$$\|D \mathbf{P}^\perp A\|_Y \lesssim \|\mathbf{P}^\perp A\|_{Y^1[I]}, \quad \|B\|_{S^1} \lesssim \|B\|_{S^1[I]}. \tag{8.64}$$

To prove (4.82), we need to estimate the $X^{-\frac{1}{2}+b_1,-b_1} \cap \Box \underline{X}^1$-norm of $\chi_I \mathrm{Diff}^{\kappa}_{\mathbf{P}^\perp A} B$. We may write

$$\chi_I \mathrm{Diff}^{\kappa}_{\mathbf{P}^\perp A} B = \sum_k 2[P_{<k-\kappa} \mathbf{P}^\perp_\ell A, \chi_I \partial^\ell P_k A] = \sum_k 2^k \mathcal{O}(P_{<k-\kappa} \mathbf{P}^\perp A, \chi_I P_k A).$$

Then by (8.20), (8.21), (8.42) and (8.44), as well as (8.64), we obtain (4.82). On the other hand, (4.83) simply follows from Hölder's inequality $L^1 L^\infty \times L^\infty L^2 \to L^1 L^2$. □

Proof of Proposition 4.22. We extend A, B by homogeneous waves outside I, and A_0 by zero outside I. In addition to $\|A\|_{S^1} \lesssim \|A\|_{S^1[I]}$, observe that we have

$$\|D A_0\|_Y \lesssim \|A_0\|_{Y^1[I]}, \quad \|\mathbf{P} A\|_{Z^1_{p_0}} \lesssim \|\mathbf{P} A\|_{Z^1_{p_0}[I]}, \quad \|\mathbf{P} A\|_{\tilde{Z}^1_{p_0}} \lesssim \|\mathbf{P} A\|_{\tilde{Z}^1_{p_0}[I]}. \tag{8.65}$$

Moreover, by (4.10), we have

$$\|\chi_I \nabla A\|_S \lesssim \|\nabla A\|_S \lesssim \|A\|_{S^1[I]}, \quad \|\chi_I \nabla B\|_S \lesssim \|\nabla B\|_S \lesssim \|B\|_{S^1[I]}. \tag{8.66}$$

We first prove (4.84), for which we need to estimate the $X^{-\frac{1}{2}+b_1,-b_1} \cap \Box \underline{X}^1$-norm of $\chi_I \mathrm{Diff}^{\kappa}_{A_0} B$. We may write

$$\chi_I \mathrm{Diff}^{\kappa}_{A_0} B = -\sum_k 2[P_{<k-\kappa} A_0, \chi_I \partial_t P_k B] = \sum_k \mathcal{O}(P_{<k-\kappa} A_0, \chi_I P_k \partial_t B).$$

Then by (8.20), (8.21), (8.42) and (8.44), as well as (8.65)–(8.66), we obtain (4.84).

For (4.85), (4.86) and (4.87), by Lemma 8.7, we may write

$$\chi_I \mathrm{Diff}^{\kappa}_{\mathbf{P}_x A} B = -\sum_k 2[P_{<k-\kappa}\mathbf{P}_\ell A, \chi_I \partial^\ell P_k B] = \sum_k \mathcal{N}(|D|^{-1}P_{<k-\kappa}\mathbf{P}A, \chi_I P_k B).$$

By (8.38), (8.39) and (8.41), combined with (8.17), (8.18) and the extension relations (8.65)–(8.66), we obtain the desired estimates. □

8.3.4. *Estimates for* $\mathrm{Diff}^{\kappa}_{\mathbf{P}A} B$. Here we prove Propositions 4.23, 4.26, 4.27, 4.28 and 4.30]. Note that, by the estimates proved so far in this subsection, we may now use Proposition 5.4 (see also Remark 5.5).

Before we embark on the proofs, we first establish some bilinear Z^1-norm bounds that will be used multiple times below.

Lemma 8.20. *We have*

$$\|P_k \mathbf{P}\mathcal{M}^2(\chi_I P_{k_1} A, P_{k_2} B)\|_{\Box Z^1} \lesssim 2^{-\delta_1|k_1-k_2|}\|P_{k_1}A\|_{S^1[I]}\|P_{k_2}B\|_{S^1[I]}, \tag{8.67}$$

$$\|P_k \mathcal{M}_0^2(\chi_I P_{k_1} A, P_{k_2} B)\|_{L^1L^\infty} \lesssim 2^{-\delta_1|k_1-k_2|}\|P_{k_1}A\|_{S^1[I]}\|P_{k_2}B\|_{S^1[I]}, \tag{8.68}$$

$$\|P_k [P_{k_1}\mathbf{P}_\ell A, \chi_I \partial^\ell P_{k_2} B]\|_{\Box Z^1} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}\|P_{k_1}A\|_{\underline{S}^1[I]}\|P_{k_2}B\|_{S^1[I]}, \tag{8.69}$$

$$\|P_k [P_{k_1} G, \chi_I \nabla P_{k_2} B]\|_{\Box Z^1} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}\|P_{k_1}G\|_{Y^1[I]}\|P_{k_2}B\|_{S^1[I]}. \tag{8.70}$$

Moreover, for $k < k_1 - 10$*, we have*

$$\|(1-\mathcal{H}_k)P_k \mathbf{P}\mathcal{M}^2(\chi_I P_{k_1} A, P_{k_2} B)\|_{\Box Z^1} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}\|P_{k_1}A\|_{S^1[I]}\|P_{k_2}B\|_{S^1[I]}, \tag{8.71}$$

$$\|(1-\mathcal{H}_k)P_k \mathcal{M}_0^2(\chi_I P_{k_1} A, P_{k_2} B)\|_{\Delta^{\frac12}\Box^{\frac12} Z^1} \lesssim 2^{-\delta_1(k_{\max}-k_{\min})}\|P_{k_1}A\|_{S^1[I]}\|P_{k_2}B\|_{S^1[I]}. \tag{8.72}$$

These bounds follow from Lemma 8.7, (8.19), (8.36), (8.37), (8.40) and (8.43), where we use (8.65) and (8.66) to absorb χ_I and return to interval-localized norms. We omit the straightforward details.

Proof of Proposition 4.23. As in the proof of Proposition 4.22, we extend A, B by homogeneous waves outside I, and A_0 by zero outside I. Furthermore, we extend $\mathbf{P}^\perp A$ by zero outside I, and denote the extension by G (we emphasize that, in general, G does not coincide with $\mathbf{P}^\perp A$ outside I). In addition to (8.65) and (8.66), by Proposition 5.4 (see also Remark 5.5) we have

$$\|A\|_{\underline{S}^1} \lesssim_M 1, \quad \|DA_0\|_{\ell^1 Y} \lesssim_M 1, \quad \|DG\|_{\ell^1 Y} \lesssim_M 1. \tag{8.73}$$

In the case of the $L^2\dot{H}^{-\frac12}$-norm on the LHS, (4.89) now follows easily from (8.17) and (8.20). It remains to estimate the N-norm of $\mathrm{Diff}^{\kappa}_{P_{k_0}\mathbf{P}A} B$.

By our extension procedure, note that $P_{k_0}A_0$ and $P_{k_0}\mathbf{P}_x A$ obey the equations

$$\begin{aligned}\Delta P_{k_0} A_0 =& P_{k_0}\left([\chi_I A^\ell, \partial_t A_\ell] + 2\mathbf{Q}(A, \chi_I \partial_t A) + \chi_I \Delta \mathbf{A}_0^3(A)\right)\\ \Box P_{k_0}\mathbf{P}_x A =& P_{k_0}\mathbf{P}\left(\mathbf{P}\mathcal{M}^2(\chi_I A, A) + 2[A_0, \chi_I\partial_t A] - 2[G_\ell, \chi_I \partial^\ell A] - 2[\mathbf{P}_\ell A, \chi_I \partial^\ell A]\right)\\ &+ P_{k_0}\mathbf{P}\left(\chi_I R(A) - \chi_I \mathrm{Rem}^3(A)A\right).\end{aligned}$$

For the cubic and higher order nonlinearities, by Theorem 3.5 and Proposition 4.19, we have

$$\|\chi_I P_{k_0} \Delta \mathbf{A}_0^3(A)\|_{L^1 L^2} \lesssim_M 1, \tag{8.74}$$

$$\|\chi_I P_{k_0} R(A)\|_{L^1 L^2} \lesssim_M 1, \tag{8.75}$$

$$\|\chi_I P_{k_0} \mathrm{Rem}^3(A) A\|_{L^1 L^2} \lesssim_M 1. \tag{8.76}$$

For the quadratic nonlinearities, we use (8.19) for $[\chi_I A^\ell, \partial_t A_\ell]$ and $\mathbf{Q}(A, \chi_I \partial_t A)$; Lemma 8.7 and (8.35) for $\mathbf{P}\mathcal{M}^2[\chi_I A, A)$; Lemma 8.7 and (8.40) for $-[\mathbf{P}_\ell A, \chi_I \partial^\ell A]$; and (8.43) for $[A_0, \chi_I \partial_t A]$ and $G_\ell \chi_I \partial^\ell A]$. Combining these with the cubic and higher order estimates and the embedding $L^1 L^2 \subseteq \Box Z^1 \cap \Delta^{-\frac{1}{2}} \Box^{\frac{1}{2}} Z^1$, we arrive at

$$\|P_{k_0} A_0\|_{L^1 L^\infty + L^2 \dot{H}^{\frac{3}{2}} \cap \Delta^{-\frac{1}{2}} \Box^{\frac{1}{2}} Z^1} \lesssim_M 1, \tag{8.77}$$

$$\|P_{k_0} \mathbf{P}_x A\|_{Z^1} \lesssim_M 1. \tag{8.78}$$

By Lemma 8.7, (8.28), (8.29), (8.30), (8.31) and Hölder's inequality $L^1 L^\infty \times L^\infty L^2 \to L^1 L^2$, it follows that

$$\begin{aligned}
\|P_k \mathrm{Diff}^\kappa_{P_{k_0} A_0} P_{k_2} B\|_N &\lesssim \|P_{k_0} A_0\|_{L^1 L^\infty + L^2 \dot{H}^{\frac{3}{2}} \cap \Delta^{-\frac{1}{2}} \Box^{\frac{1}{2}} Z^1} \|DB\|_S, \\
\|P_k \mathrm{Diff}^\kappa_{P_{k_0} \mathbf{P}_x A} P_{k_2} B\|_N &\lesssim \|P_{k_0} \mathbf{P}_x A\|_{S^1 \cap Z^1} \|DB\|_S.
\end{aligned}$$

Thanks to the frequency gap $\kappa \geq 5$, note furthermore that the LHSs vanish unless $k = k_2 + O(1)$. This completes the proof of Proposition 4.23. ☐

Proof of Proposition 4.26. Estimate (4.94) follows easily using Hölder and Bernstein. To prove (4.95), we extend $\mathbf{P}A, B$ by homogeneous waves outside I, so that $\|P_{k_1} \Box \mathbf{P}A\|_{L^1 L^2} \leq \|P_{k_1} \Box \mathbf{P}A\|_{L^1 L^2[I]}$ and $\|P_{k_2} B\|_{S^1} \lesssim \|P_{k_2} B\|_{S^1[I]}$. Moreover, by the embedding $L^1 L^2 \subseteq N \cap \Box Z^1$, we have $\|P_{k_1} \mathbf{P}A\|_{S^1 \cap Z^1} \lesssim \|P_{k_1} \nabla \mathbf{P}A(t_0)\|_{L^2} + \|P_{k_1} \Box \mathbf{P}A\|_{L^1 L^2[I]}$. Then (4.95) follows by Lemma 8.7, (8.28) and (8.30). ☐

Proof of Proposition 4.27. Here, in addition to the bilinear null forms (Lemma 8.7), we need to use the secondary null structure (Lemma 8.10).

Without loss of generality, we set $t_0 = 0$. We extend B, $B^{(1)}$ and $B^{(2)}$ by homogeneous waves outside I, then define A_0 and $\mathbf{P}A$ by solving the equations (4.96) and (4.97), respectively[9]. In A_0 and $\mathbf{P}A$, we separate out the (*high* × *high* → *low*) interaction terms by defining

$$\begin{aligned}
A_0^{hh} &= \sum_{k, k_1, k_2 : k < k_1 - 10} \Delta^{-1} P_k [P_{k_1} B^{(1)\ell}, P_{k_2} \partial_t B^{(2)}_\ell], \\
\mathbf{P}A^{hh} &= \sum_{k, k_1, k_2 : k < k_1 - 10} \Box^{-1} P_k \mathbf{P} [P_{k_1} B^{(1)\ell}, \partial_x P_{k_2} B^{(2)}_\ell],
\end{aligned}$$

where $\Box^{-1} f$ refers to the solution to the inhomogeneous wave equation $\Box u = f$ with $(u, \partial_t u)(0) = 0$. We also introduce

$$\begin{aligned}
\mathcal{H} A_0^{hh} &= \sum_{k, k_1, k_2 : k < k_1 - 10} \Delta^{-1} \mathcal{H}_k P_k [P_{k_1} B^{(1)\ell}, P_{k_2} \partial_t B^{(2)}_\ell], \\
\mathcal{H} \mathbf{P} A^{hh} &= \sum_{k, k_1, k_2 : k < k_1 - 10} \Box^{-1} \mathcal{H}_k P_k \mathbf{P} [P_{k_1} B^{(1)\ell}, \partial_x P_{k_2} B^{(2)}_\ell].
\end{aligned}$$

[9]We may put in χ_I on the RHSs of (4.96) and (4.97), but it is not necessary.

Accordingly, we split

$$\mathrm{Diff}^{\kappa}_{\mathbf{P}A}B = \sum_k \left(2[P_{<k-\kappa}(A_0 - \mathcal{H}A_0^{hh}), \partial^0 P_k B] + 2[P_{<k-\kappa}(\mathbf{P}_\ell A - \mathcal{H}\mathbf{P}_\ell A^{hh}), \partial^\ell P_k B]\right) \quad (8.79)$$

$$+ \sum_k \left(2[P_{<k-\kappa}\mathcal{H}A_0^{hh}, \partial^0 P_k B] + 2[P_{<k-\kappa}\mathcal{H}\mathbf{P}_\ell A^{hh}, \partial^\ell P_k B]\right). \quad (8.80)$$

By Propositions 4.12, 4.15 and Lemma 8.20, we have

$$\|A_0\|_{Y^1_{cd}} + \|A_0 - A_0^{hh}\|_{L^1 L^\infty_{cd}} + \|A_0^{hh}\|_{Y^1_{cd}} + \|A_0^{hh} - \mathcal{H}A_0^{hh}\|_{\Delta^{-\frac{1}{2}}\Box^{\frac{1}{2}} Z^1_{cd}} \lesssim \|B^{(1)}\|_{S^1_c}\|B^{(2)}\|_{S^1_d},$$

$$\|\mathbf{P}A\|_{S^1_{cd}} + \|\mathbf{P}A^{hh} - \mathcal{H}\mathbf{P}A^{hh}\|_{Z^1_{cd}} \lesssim \|B^{(1)}\|_{S^1_c}\|B^{(2)}\|_{S^1_d}.$$

Combining these bounds with Lemma 8.7, (8.28), (8.29), (8.30), (8.31) and Hölder's inequality $L^1 L^\infty \times L^\infty L^2 \to L^1 L^2$, it follows that

$$\|\sum_k [P_{<k-\kappa}(A_0 - \mathcal{H}A_0^{hh}), \partial^0 P_k B]\|_{N_f} \lesssim \|B^{(1)}\|_{S^1_c}\|B^{(2)}\|_{S^1_d}\|B\|_{S^1_e},$$

$$\|\sum_k [P_{<k-\kappa}(\mathbf{P}_\ell A - \mathcal{H}\mathbf{P}_\ell A^{hh}), \partial^\ell P_k B]\|_{N_f} \lesssim \|B^{(1)}\|_{S^1_c}\|B^{(2)}\|_{S^1_d}\|B\|_{S^1_e},$$

which handles the contribution of (8.79). On the other hand, unraveling the definitions, we may rewrite (8.80) as

$$(8.80) = \sum \Big(Q_{<j-C}\mathcal{O}'(\Delta^{-1}P_k Q_j \mathcal{O}(P_{k_1}Q_{<j-C}B^{(1)}, \partial_0 P_{k_2}Q_{<j-C}B^{(2)}), \partial^0 Q_{<j-C}P_{k_3}B)$$
$$+ Q_{<j-C}\mathcal{O}'(\Box^{-1}P_k Q_j \mathbf{P}_\ell \mathcal{O}(P_{k_1}Q_{<j-C}B^{(1)}, \partial_x P_{k_2}Q_{<j-C}B^{(2)}), \partial^\ell Q_{<j-C}P_{k_3}B)\Big)$$

for some disposable operators $\mathcal{O}$ and $\mathcal{O}'$, where the summation is taken over the range $\{(k, k_1, k_2, k_3) : k < k_1 - 10,\ k < k_3 - \kappa + 5\}$. By (8.45), it follows that

$$\|(8.80)\|_{L^1 L^2_f} \lesssim \|B^{(1)}\|_{S^1_c}\|B^{(2)}\|_{S^1_d}\|B\|_{S^1_e},$$

which is acceptable. Finally, for the $L^2\dot{H}^{-\frac{1}{2}}$-norm of $\mathrm{Diff}^{\kappa}_{\mathbf{P}A}B$, note that (8.17) and the preceding bounds imply

$$\|P_k(\mathrm{Diff}^{\kappa}_{\mathbf{P}A}B)\|_{L^2\dot{H}^{-\frac{1}{2}}} \lesssim c_{k-\kappa}d_{k-\kappa}e_k,$$

which is better than what we need. □

Proof of Proposition 4.28. As in the preceding proof, we extend B, $B^{(1)}$ and $B^{(2)}$ by homogeneous waves outside I. This time, however, we also extend $\mathbf{P}A$ by homogeneous waves outside I. We moreover extend B_0 and $\mathbf{P}^\perp B^{(1)}$ by zero outside I, where the latter is denoted by $G^{(1)}$. Note that $\mathbf{P}A$ solves the equation

$$\Box\mathbf{P}A = \mathbf{P}\left([\mathbf{P}_\ell B^{(1)}, \chi_I \partial^\ell B^{(2)}] + [B_0^{(1)}, \chi_I \partial^0 B^{(2)}] + [G_\ell^{(1)}, \chi_I \partial^\alpha B^{(2)}]\right).$$

By Lemma 8.20 and the frequency envelope bounds (4.100)–(4.101), it follows that

$$\|\mathbf{P}A\|_{Z^1_{cd}} \lesssim (\|B^{(1)}\|_{\underline{S}^1_c[I]} + \|(B_0^{(1)}, G^{(1)})\|_{Y^1_c[I]})\|B^{(2)}\|_{S^1_d[I]} \leq 1. \quad (8.81)$$

On the other hand, recall that $\|\mathbf{P}A\|_{S^1_a} \leq 1$ by (4.101). Therefore, by Lemma 8.7, (8.28) and (8.30), we have

$$\|\mathrm{Diff}^{\kappa}_{\mathbf{P}_x A}B\|_{N_f} \lesssim 1.$$

On the other hand, by (8.17), we also have

$$\|P_k(\mathrm{Diff}^{\kappa}_{\mathbf{P}_x A} B)\|_{L^2 \dot{H}^{-\frac{1}{2}}} \lesssim a_{k-\kappa} e_k,$$

which is better than what we need. The desired estimate (4.102) follows. □

Proof of Proposition 4.30. We move the problem to the entire real line using the free wave extension for $\mathbf{P} A_x$ and B, and the zero extension for A_0.

The expression $|D|^{-1}[\nabla, \mathrm{Diff}^{\kappa}_{\mathbf{P}A}]B$ is a translation invariant bilinear expression in $\mathbf{P}A$ and B, whose Littlewood-Paley pieces can be expressed in the form

$$|D|^{-1}[\nabla, \mathrm{Diff}^{\kappa}_{P_{k'}\mathbf{P}A}] P_k B = 2^{k'-k} \mathcal{O}(P_{k'} \mathbf{P} A_\alpha, \partial^\alpha P_k B), \qquad k' < k - \kappa \tag{8.82}$$

with $\mathcal{O}$ disposable. By (8.11) the spatial part is a null form, so we can rewrite the above expression as

$$2^{-k} \mathcal{N}(P_{k'} \mathbf{P} A_x, P_k B) + 2^{k'-k} \mathcal{O}(P_{k'} A_0, P_k \partial_t B)$$

We consider separately the spatial part and the temporal part. For the spatial part we use the bound (8.23) to estimate

$$\|2^{-k} \mathcal{N}(P_{k'} \mathbf{P} A_x, P_k B)\|_N \lesssim 2^{-\delta_1 |k-k'|} \|P'_k \mathbf{P} A\|_{S^1} \|B\|_{S^1}$$

which suffices after summation in $k' < k - \kappa$.

For the temporal part we use instead the bound (8.25), which yields

$$\|2^{k'} \mathcal{O}(P_{k'} A_0, P_k B)\|_{L^1 L^2} \lesssim 2^{-\delta_1 |k-k'|} \|P'_k D A_0\|_{L^2 \dot{H}^{\frac{1}{2}}} \|B\|_{S^1}$$

which again suffices.

The expression $\mathrm{Diff}^{\kappa}_{P_{k'}\mathbf{P}A} B - (\mathrm{Diff}^{\kappa}_{P_{k'}\mathbf{P}A})^* B$ is easily seen to have the same form as in (8.82), so the same estimate follows. □

8.3.5. *Estimates involving* $\mathbf{W}$. Here we prove Propositions 4.16, 4.17 and 4.29, which involve $\mathbf{w}_0^2$ and $\mathbf{w}_x^2$.

Proof of Proposition 4.16. By definition (3.29), we have

$$P_k \mathbf{w}_0^2(P_{k_1} A, P_{k_2} B, s) = -2 P_k \mathbf{W}(P_{k_1} \partial_t A, P_{k_2} \Delta B, s).$$

Applying Lemma 8.2 to the expression on the RHS, we have

$$P_k \mathbf{W}(P_{k_1} \partial_t A, P_{k_2} \Delta B, s) = -\langle s 2^{2k} \rangle^{-10} \langle s^{-1} 2^{-2k_{\max}} \rangle^{-1} 2^{-2k_{\max}} 2^{2k_2} P_k \mathbf{O}(P_{k_1} \partial_t A, P_{k_2} B), \tag{8.83}$$

for some disposable operator $\mathbf{O}$ on $\mathbb{R}^4$. The rest of the proof follows that of Proposition 4.12. First, by (8.53), it follows that

$$\begin{aligned} &\| |D|^{-1} P_k \mathbf{w}_0^2(P_{k_1} A, P_{k_2} B, s)\|_{L^2} \\ &\lesssim \langle s 2^{2k} \rangle^{-10} \langle s^{-1} 2^{-2k_{\max}} \rangle^{-1} 2^{2(k_{\min} - k_{\max})} 2^{k_2 - k} \|P_{k_1} \partial_t A\|_{L^2} \|P_{k_2} B\|_{\dot{H}^1}. \end{aligned}$$

From this dyadic bound, the frequency envelope bound (4.52) follows. Indeed, for any $0 < \delta' < 4\delta$ and any δ'-admissible frequency envelopes c, d, we compute

$$\begin{aligned} \langle s 2^{2k} \rangle^{-10} \langle s^{-1} 2^{-2k_{\max}} \rangle^{-1} 2^{-\delta(k_{\max} - k_{\min})} c_{k_1} d_{k_2} &\lesssim \langle s 2^{2k} \rangle^{-10} \langle s^{-1} 2^{-2k_{\max}} \rangle^{-1} 2^{-\frac{1}{2}\delta(k_{\max} - k_{\min})} c_k d_k \\ &\lesssim \langle s 2^{2k} \rangle^{-10} \langle s^{-1} 2^{-2k} \rangle^{-\frac{1}{4}\delta} c_k d_k. \end{aligned} \tag{8.84}$$

which proves (4.52). The estimate (4.53) follows in a similar manner from (8.53).

Next, extending $\partial_t A$ and B by zero outside I, then applying (8.17) and (8.19), it follows that

$$\begin{aligned}
&\| |D|^{-1} P_k \mathbf{w}_0^2(P_{k_1} A, P_{k_2} B, s)\|_{L^2 \dot{H}^{-\frac{1}{2}}[I]} \\
&\lesssim \langle s 2^{2k}\rangle^{-10} \langle s^{-1} 2^{-2k_{\max}}\rangle^{-1} 2^{-\delta_1(k_{\max}-k_{\min})} 2^{2(k_1-k_{\max})} \|P_{k_1} A\|_{\mathrm{Str}^1[I]} \|P_{k_2} B\|_{\mathrm{Str}^1[I]}, \\
&\| |D|^{-2} P_k \mathbf{w}_0^2(P_{k_1} A, P_{k_2} B, s)\|_{L^1 L^\infty[I]} \\
&\lesssim \langle s 2^{2k}\rangle^{-10} \langle s^{-1} 2^{-2k_{\max}}\rangle^{-1} 2^{2(k_1-k_{\max})} \|P_{k_1} A\|_{S^1[I]} \|P_{k_2} B\|_{S^1[I]}.
\end{aligned}$$

Using (4.21) and (8.56), these two bounds imply (4.54) and (4.55), as in Proof of Proposition 4.12, Step 2. ☐

Proof of Proposition 4.17. We begin with algebraic observations. By (3.30), we have

$$\begin{aligned}
P_k \mathbf{P}_j \mathbf{w}^2(P_{k_1} A, P_{k_2} B, s) =& -2 P_k \mathbf{P}_j \mathbf{W}(P_{k_1} \partial_t A^\ell, \partial_x P_{k_2} \partial_t B_\ell, s) \\
&+ 4 P_k \mathbf{P}_j \mathbf{W}(P_{k_1} \mathbf{P} \partial_t A^\ell, \partial_\ell P_{k_2} \partial_t B, s) \\
&+ 4 P_k \mathbf{P}_j \mathbf{W}(P_{k_1} \mathbf{P}^\perp \partial_t A^\ell, \partial_\ell P_{k_2} \partial_t B, s),
\end{aligned} \tag{8.85}$$

where, by Lemma 8.2, we may write

$$\begin{aligned}
&P_k \mathbf{P}_j \mathbf{W}(P_{k_1} \partial_t A^\ell, \partial_x P_{k_2} \partial_t B_\ell, s) \\
&= \langle s 2^{2k}\rangle^{-10} \langle s^{-1} 2^{-2k_{\max}}\rangle^{-1} 2^{-2k_{\max}} P_k \mathbf{P}_j \mathbf{O}(P_{k_1} \partial_t A^\ell, \partial_x P_{k_2} \partial_t B_\ell),
\end{aligned} \tag{8.86}$$

$$\begin{aligned}
&P_k \mathbf{P}_j \mathbf{W}(P_{k_1} \partial_t \mathbf{P} A^\ell, \partial_\ell P_{k_2} \partial_t B, s) \\
&= -2 \langle s 2^{2k}\rangle^{-10} \langle s^{-1} 2^{-2k_{\max}}\rangle^{-1} 2^{-2k_{\max}} P_k \mathbf{O}(\mathbf{P}_\ell P_{k_1} \partial_t A, \partial^\ell P_{k_2} \partial_t B),
\end{aligned} \tag{8.87}$$

$$\begin{aligned}
&P_k \mathbf{P}_j \mathbf{W}(P_{k_1} \partial_t \mathbf{P}^\perp A^\ell, \partial_\ell P_{k_2} \partial_t B, s) \\
&= -2 \langle s 2^{2k}\rangle^{-10} \langle s^{-1} 2^{-2k_{\max}}\rangle^{-1} 2^{-2k_{\max}} P_k \mathbf{O}(P_{k_1} \partial_t \mathbf{P}^\perp_\ell A, \partial^\ell P_{k_2} \partial_t B),
\end{aligned} \tag{8.88}$$

for some disposable operator $\mathbf{O}$ on $\mathbb{R}^4$. Note that (8.86) and (8.87) are null forms according to Lemma 8.7, and (8.88) is favorable since $\partial_t \mathbf{P}^\perp A$ is controlled in the $L^2 \dot{H}^{\frac{1}{2}}$-norm.

Given the above formulas for $\mathbf{w}_x$, the proof of the estimates (4.56) and (4.57) is almost identical to the proof of (4.52) (4.53), using the dyadic bounds (8.53),(8.53) and (8.84).

We now prove (4.58). We extend A, B by homogeneous waves outside I. By (8.17), (8.18), Lemma 8.7, (8.23) and (8.34), it follows that

$$\begin{aligned}
&\|P_k \mathbf{P}_j \mathbf{W}(P_{k_1} \partial_t A, \partial_x P_{k_2} \partial_t B, s)\|_{N \cap \Box \underline{X}^1} \\
&\lesssim \langle s 2^{2k}\rangle^{-10} \langle s^{-1} 2^{-2k_{\max}}\rangle^{-1} 2^{-\delta_1(k_{\max}-k_{\min})} 2^{k_1+k_2-2k_{\max}} \|P_{k_1} A\|_{S^1} \|P_{k_2} B\|_{S^1} \\
&\|P_k \mathbf{P}_j \mathbf{W}(P_{k_1} \partial_t \mathbf{P} A, \partial_x P_{k_2} \partial_t B, s)\|_{N \cap \Box \underline{X}^1} \\
&\lesssim \langle s 2^{2k}\rangle^{-10} \langle s^{-1} 2^{-2k_{\max}}\rangle^{-1} 2^{-\delta_1(k_{\max}-k_{\min})} 2^{k+k_2-2k_{\max}} \|P_{k_1} A\|_{S^1} \|P_{k_2} B\|_{S^1} \\
&\|P_k \mathbf{P}_j \mathbf{W}(P_{k_1} \partial_t \mathbf{P}^\perp A, \partial_x P_{k_2} \partial_t B, s)\|_{N \cap \Box \underline{X}^1} \\
&\lesssim \langle s 2^{2k}\rangle^{-10} \langle s^{-1} 2^{-2k_{\max}}\rangle^{-1} 2^{-\delta_1(k_{\max}-k_{\min})} 2^{2k_2-2k_{\max}} \|P_{k_1} \partial_t \mathbf{P}^\perp A\|_{L^2 \dot{H}^{\frac{1}{2}}} \|P_{k_2} B\|_{S^1}.
\end{aligned}$$

Clearly, $2^{k_1+k_2-2k_{\max}}$, $2^{k+k_2-2k_{\max}}$ and $2^{2k_2-2k_{\max}}$ are bounded, so they may be safely discarded. By the same frequency envelope computation (8.84) as before, we obtain (4.58).

In the energy dispersed case (4.59), we proceed as in the proofs of Propositions 4.15 and 4.20. The contribution of (8.88) is already acceptable, since we need not gain any smallness factor. Moreover, for the contribution of (8.86) and (8.87), the case of $L^2\dot H^{-\frac12}$ on the LHS can be easily handled using (8.17) and (4.21); we omit the details.

It remains to consider only the N-norm of (8.86) and (8.87). For a parameter $\kappa > 0$ to be chosen below, the preceding proof of (4.58) imply that in the case $k_{\max} - k_{\min} \geq \kappa$, we have

$$\|(8.86)\|_N + \|(8.87)\|_N \lesssim \langle s2^{2k}\rangle^{-10}\langle s^{-1}2^{-2k_{\max}}\rangle^{-1}2^{-\frac12\delta_1\kappa}2^{-\frac12\delta_1(k_{\max}-k_{\min})}\|P_{k_1}A\|_{S^1}\|P_{k_2}B\|_{S^1}.$$

On the other hand, when $k_{\max} - k_{\min} \leq \kappa$, we may apply Lemma 8.7 (in particular, (8.15) and (8.16)) and Remark 8.19, which implies

$$\|(8.86)\|_N + \|(8.87)\|_N \lesssim \langle s2^{2k}\rangle^{-10}\langle s^{-1}2^{-2k_{\max}}\rangle^{-1}2^{C\kappa}\varepsilon^{c\delta_1}\|P_{k_1}A\|_{\underline{S}^1}M.$$

Choosing $2^\kappa = \varepsilon^c$ for a sufficiently small $c > 0$, and performing a similar frequency envelope computation as in (8.84), we arrive at (4.59). □

Proof of Proposition 4.29. We first note that both $\mathbf{w}_0$ and $\mathbf{w}_x$ depend on $\partial_t B_1$, for which we control $\|\partial_t B_1\|_{S_c}$ and $\|\mathbf{P}^\perp\partial_t B_1\|_{Y_c}$. We may assume that

$$\|\partial_t B^{(1)}\|_{S_c[I]},\ \|\mathbf{P}^\perp\partial_t B^{(1)}\|_{Y_c[I]},\ \|B^{(2)}\|_{S^1_d[I]},\ \|B\|_{S^1_e[I]} \leq 1.$$

We can now extend $\partial_t B_1$ by zero outside I, and $B^{(2)}$ and B by free waves. Then the problem is reduced to the similar problem on the real line. We begin with the simpler $L^2\dot H^{-\frac12}$ bound. For that we use (4.53) and (4.58) to obtain

$$\|P_k\mathbf{w}_0\|_{L^2\dot H^{-\frac12}} + \|P_k\mathbf{w}_x\|_{N\cap\Box\underline{X}^1} \lesssim \langle s2^{2k'}\rangle^{-10}\langle s^{-1}2^{-2k_{\max}}\rangle^{-\delta_2}c_k d_k \tag{8.89}$$

and then conclude with (8.17) respectively (8.20).

It remains to prove the N bound. We define

$$\begin{aligned}\mathcal{I}(k',k_1,k_2,k,s) =\big(&-[\Delta^{-1}P_{k'}\mathbf{w}_0^2(P_{k_1}B^{(1)},P_{k_2}B^{(2)},s),\partial_t P_k B]\\ &+[\Box^{-1}P_{k'}\mathbf{P}_\ell\mathbf{w}_x^2(P_{k_1}B^{(1)},P_{k_2}B^{(2)},s),\partial^\ell P_k B]\big),\end{aligned}$$

so that $\mathrm{Diff}^\kappa_{\mathbf{P}A}B = \sum_{k',k_1,k_2,k:k'<k-\kappa}\mathcal{I}(k',k_1,k_2,k)$ on I. Introducing the shorthands

$$k_{\max} = \max\{k',k_1,k_2\}, \qquad k_{\min} = \min\{k',k_1,k_2\}$$

and

$$\alpha(k',k_1,k_2,s) = \langle s2^{2k'}\rangle^{-10}\langle s^{-1}2^{-2k_{\max}}\rangle^{-1}2^{-c\delta_1(k_{\max}-k_{\min})}$$

we claim that

$$\|\mathcal{I}(k',k_1,k_2,k,s)\|_N \lesssim \alpha(k',k_1,k_2,s)c_{k_1}d_{k_2}e_k. \tag{8.90}$$

This would conclude the proof of the proposition after summation with respect to k_1 and k_2.

We start with a simple observation, namely that we can easily dispense with the high modulations of $\partial_t B_1$ and B_2 using Lemma 8.2, combined with Hölder and Bernstein's inequalities and also (8.28) and (8.32). Thus from here on we assume that

$$P_{k_1}\partial_t B^{(1)} = P_{k_1}Q_{<k_1}\partial_t B^{(1)}, \qquad P_{k_2}\partial_t B^{(2)} = P_{k_2}Q_{<k_2}\partial_t B^{(2)}$$

In view of (8.85) and the identity

$$\mathbf{w}_0^2(A,B,s) = -2\mathbf{W}(\partial_t A,\partial_t^2 B,s) - 2\mathbf{W}(\partial_t A,\Box B,s),$$

we may expand

$$
\begin{aligned}
\mathcal{I}(k',k_1,k_2,k,s) = {} & 2[P_{k'}\Delta^{-1}\mathbf{W}(P_{k_1}\partial_t B^{(1)}, \Box P_{k_2}B^{(2)}, s), \partial_t P_k B] && (8.91)\\
& + 4[\Box^{-1}P_{k'}\mathbf{P}_\ell \mathbf{W}(P_{k_1}\mathbf{P}\partial_t B^{(1),m}, \partial_m P_{k_2}\partial_t B^{(2)}, s), \partial^\ell P_k B] && (8.92)\\
& + 4[\Box^{-1}P_{k'}\mathbf{P}_\ell \mathbf{W}(P_{k_1}\mathbf{P}^\perp\partial_t B^{(1),m}, \partial_m P_{k_2}\partial_t B^{(2)}, s), \partial^\ell P_k B] && (8.93)\\
& + 2[\Delta^{-1}P_{k'}\mathbf{W}(P_{k_1}\partial_t B^{(1)}, \partial_t P_{k_2}\partial_t B^{(2)}, s), \partial_t P_k B] && (8.94)\\
& - 2[\Box^{-1}P_{k'}\mathbf{P}_\ell \mathbf{W}(P_{k_1}\partial_t B^{(1),m}, \partial_x P_{k_2}\partial_t B^{(2)}_m, s), \partial^\ell P_k B]. && (8.95)\\
= {} & \mathcal{I}_{(1)} + \mathcal{I}_{(2)} + \mathcal{I}_{(3)} + \mathcal{I}_{(4)} + \mathcal{I}_{(5)} && (8.96)
\end{aligned}
$$

The first term is easily estimated in L^1L^2 using Lemma 8.2 and Holder and Bernstein's inequality by

$$
\begin{aligned}
\|\mathcal{I}_{(1)}\|_{L^1L^2} &\lesssim \|P_{k'}\Delta^{-1}\mathbf{W}(P_{k_1}\partial_t B^{(1)}, \Box P_{k_2}B^{(2)}, s)\|_{L^1L^\infty}\|\partial_t P_k B]\|_{L^\infty L^2}\\
&\lesssim \langle s2^{2k'}\rangle^{-10}\langle s^{-1}2^{-2k_{\max}}\rangle^{-1}2^{\frac{1}{2}(k_{min}-k_{max})}\|\partial_t P_{k_1}B^{(1)}\|_{L^2\dot{W}^{1,8}}\|\Box P_{k_2}B^{(2)}\|_{L^2\dot{H}^{-\frac{1}{2}}}e_k
\end{aligned}
$$

which suffices.

To continue, we use (8.25), (8.35) and the embedding $L^1L^2 \subseteq \Box Z^1$, we have

$$
\begin{aligned}
\|P_{k'}\mathbf{P}_\ell \mathbf{W}(P_{k_1}\mathbf{P}\partial_t B^{(1)}, \partial_x P_{k_2}\partial_t B^{(2)}, s)\|_{N\cap\Box Z^1} &\lesssim \alpha(k',k_1,k_2,s)c_{k_1}d_{k_2}\\
\|P_{k'}\mathbf{P}_\ell \mathbf{W}(P_{k_1}\mathbf{P}^\perp\partial_t B^{(1)}, \partial_x P_{k_2}\partial_t B^{(2)}, s)\|_{N\cap\Box Z^1} &\lesssim \alpha(k',k_1,k_2,s)c_{k_1}d_{k_2}
\end{aligned}
$$

This yields

$$
\begin{aligned}
\|\Box^{-1}P_{k'}\mathbf{P}_\ell \mathbf{W}(P_{k_1}\mathbf{P}\partial_t B^{(1)}, \partial_x P_{k_2}\partial_t B^{(2)}, s)\|_{S\cap Z^1} &\lesssim \alpha(k',k_1,k_2,s)c_{k_1}d_{k_2}\\
\|\Box^{-1}P_{k'}\mathbf{P}_\ell \mathbf{W}(P_{k_1}\mathbf{P}^\perp\partial_t B^{(1)}, \partial_x P_{k_2}\partial_t B^{(2)}, s)\|_{S\cap Z^1} &\lesssim \alpha(k',k_1,k_2,s)c_{k_1}d_{k_2}
\end{aligned}
$$

We use this directly for the next two terms $\mathcal{I}_{(2)}$ and $\mathcal{I}_{(3)}$, arguing in a bilinear fashion. The desired N bound for both is obtained using both (8.28) and (8.32) with $\kappa = 0$.

The final two terms are combined together in a trilinear null form,

$$
\mathcal{I}_{(4)} + \mathcal{I}_{(5)} = \mathrm{Diff}^\kappa_{\mathbf{P}\tilde{A}}B
$$

where

$$
\tilde{A}_0 = \Delta^{-1}P_{k'}\mathbf{W}(P_{k_1}\partial_t B^{(1)}, \partial_t P_{k_2}\partial_t B^{(2)}, s),
$$

and

$$
\mathbf{A}_x = \Box^{-1}P_{k'}\mathbf{P}_\ell \mathbf{W}(P_{k_1}\partial_t B^{(1),m}, \partial_x P_{k_2}\partial_t B^{(2)}_m, s)
$$

At this point we have placed ourselves in the same setting as in the proof of Proposition 4.27. Then the same argument applies, with the only difference that, due to Lemma 8.2, we obtain an additional factor of

$$
\langle s2^{2k'}\rangle^{-10}\langle s^{-1}2^{-2k_{\max}}\rangle^{-1}2^{-2k_{max}}2^{k_1+k_2}
$$

as needed. Here the factors 2^{k_1} and 2^{k_2} come from one time derivative on $B^{(1)}$, respectively $B^{(2)}$ at low modulation. Thus the N bound for $\mathcal{I}_{(4)} + \mathcal{I}_{(5)}$ follows. □

8.3.6. *Estimates for* $\mathrm{Rem}^3(A)B$ *and* $\mathrm{Rem}^3_s(A)B$. Finally, we sketch the proof of Proposition 4.19.

Proof of Proposition 4.19. By Holder's inequality and Bernstein, it suffices to show that the following nonlinear maps are Lipschitz and envelope preserving:

$$\mathrm{Str}^1 \ni A \to (\mathbf{DA}_0, \mathbf{DA}) \in L^{2-}\dot{H}^{\frac{1}{2}+} \cap L^{2+}\dot{H}^{\frac{1}{2}-}$$

$$\mathrm{Str}^1 \ni A \to \mathbf{A}_0 \in L^2\dot{H}^{\frac{3}{2}}$$

The same applies for the maps

$$\mathrm{Str}^1 \ni A \to \mathbf{DA}_{0,s} \in L^{2-}\dot{H}^{\frac{1}{2}+} \cap L^{2+}\dot{H}^{\frac{1}{2}-}$$

$$\mathrm{Str}^1 \ni A \to \mathbf{A}_{0;s} \in L^2\dot{H}^{\frac{3}{2}}$$

with the addition that now the output has to be also concentrated at frequency $k(s)$.

The $\mathbf{A}_0$ property is a consequence of (4.30) for the quadratic term, and (3.23) for the cubic part $\mathbf{A}^3_0$. Similarly, the $A_{0;s}$ property is a consequence of (4.53) for the quadratic term, and (3.36) for the cubic part $\mathbf{A}^3_{0;s}$.

The $\mathbf{DA}$ property follows from (a minor variation of) (4.36) for the quadratic part, and (3.18) for the cubic part $\mathbf{DA}^3$.

Finally, the $\mathbf{DA}_0$ property is a consequence of (a small variation of) (4.30) for the quadratic part and of (3.24) for the cubic part. Similarly, for $\mathbf{DA}^s_0$ we need (a small variation of) (4.53) and of (3.37). □

8.4. **Proof of the global-in-time dyadic estimates.** In this subsection, we prove the global-in-time dyadic estimates stated in Section 8.2.

8.4.1. *Preliminaries on orthogonality.* Let $\mathcal{O}$ be a translation-invariant bilinear operator on $\mathbb{R}^{1+4}$. Consider the expression

$$\iint u^{(0)}\mathcal{O}(u^{(1)}, u^{(2)})\, dtdx. \tag{8.97}$$

Our general strategy for proving the dyadic estimates stated in Section 8.2 will be as follows: (1) Decompose $u^{(i)}$ by frequency projection into various sets, (2) Estimate each such piece, and (3) Exploit vanishing (or orthogonality) properties of (8.97), which depend on the relative configuration of the frequency supports of $u^{(i)}$'s, to sum up. Some simple examples of orthogonality properties of (8.97) that we will use are as follows:

- (**Littlewood–Paley trichotomy**) If $u^{(i)} = P_{k_1}u^{(i)}$, then (8.97) vanishes unless the largest two numbers of k_0, k_1, k_2 are part by at most (say) 5. This property has already been used freely.
- (**Cube decomposition**) If $u^{(i)} = P_{k_i}P_{\mathcal{C}^i}u^{(i)}$ with $\mathcal{C}^i = \mathcal{C}_{k_{\min}}(0)$ (i.e., is a cube of dimension $2^{k_{\min}} \times \cdots 2^{k_{\min}}$) situated in $\{|\xi| \simeq 2^{k_i}\}$, then (8.97) vanishes unless $\mathcal{C}^0 + \mathcal{C}^1 + \mathcal{C}^2 \ni 0$.

  To obtain more useful statements, let $\mathcal{C}^{\max}$, $\mathcal{C}^{\mathrm{med}}$ and $\mathcal{C}^{\min}$ denote the re-indexing of the cubes $\mathcal{C}^0$, $\mathcal{C}^1$ and $\mathcal{C}^2$, which are situated at the annuli $\{|\xi| \simeq 2^{k_{\max}}\}$, $\{|\xi| \simeq 2^{k_{\mathrm{med}}}\}$ and $\{|\xi| \simeq 2^{k_{\min}}\}$, respectively. Then for every fixed $\mathcal{C}^{\min}$ and $\mathcal{C}^{\max}$ [resp. $\mathcal{C}^{\mathrm{med}}$], there are only $O(1)$-many cubes $\mathcal{C}^{\mathrm{med}}$ [resp. $\mathcal{C}^{\max}$] satisfying $\mathcal{C}^{\min} + \mathcal{C}^{\mathrm{med}} + \mathcal{C}^{\max} \ni 0$. Moreover, we have

  $$|\angle(\mathcal{C}^{\max}, -\mathcal{C}^{\mathrm{med}})| \lesssim 2^{k_{\max}-k_{\min}}.$$

  Geometrically, such cubes $\mathcal{C}^{\max}$ and $\mathcal{C}^{\mathrm{med}}$ are "nearly antipodal."

We will also exploit the relationship between modulation localization and angular restriction for (8.97). In the proofs below, we will only need the following simple statement. For a more complete discussion, see, e.g., [25].

Lemma 8.21 (Geometry of the cone)**.** *Consider integers* $k_0, k_1, k_2, j_0, j_1, j_2 \in \mathbb{Z}$ *be such that* $|k_{\mathrm{med}} - k_{\max}| \leq 5$. *For* $i = 0, 1, 2$, *let* $\omega_i \subseteq \mathbb{S}^3$ *be an angular cap of radius* $r_i < 2^{-5}$, $\pm_i \in \{+, -\}$, *and* $u^{(i)} \in \mathcal{S}(\mathbb{R}^{1+4})$ *have frequency support in the region* $\{|\xi| \simeq 2^{k_i},\ \frac{\xi}{|\xi|} \in \omega_i,\ |\tau - \pm_i|\xi|| \simeq 2^{j_i}\}$. *Suppose that* $j_{\max} \leq k_{\min}$, *and define* $\ell = \frac{1}{2}\min\{j_{\max} - k_{\min}, 0\}$.

Then the expression (8.97) *vanishes unless*

$$|\angle(\pm_i \omega_i, \pm_{i'} \omega_{i'})| \lesssim 2^{k_{\min} - \min\{k_i, k_{i'}\}} 2^{\ell} + \max\{r_i, r_{i'}\}$$

for every pair $i, i' \in \{0, 1, 2\}$ $(i \neq i')$.

Finally, we collect some often used estimates. For $k' \leq k$ and $\ell' < -5$, note that

$$2^{-\frac{5}{6}k} \|P_{\mathcal{C}_{k'}(\ell')} u_k\|_{L^2 L^6} + 2^{-k' - \frac{1}{2}k} 2^{-\frac{1}{2}\ell'} \|P_{\mathcal{C}_{k'}(\ell')} u_k\|_{L^2 L^\infty} \lesssim \|P_{\mathcal{C}_{k'}(\ell')} u_k\|_{S_k[\mathcal{C}_{k'}(\ell')]},$$

where, by (4.1), we have

$$\sum_{\mathcal{C} \in \{\mathcal{C}_{k'}(\ell')\}} \|P_{\mathcal{C}} u_k\|_{S_k[\mathcal{C}]}^2 \lesssim \|u_k\|_{S_k}^2 \simeq \|u_k\|_S^2.$$

Also note that, for any $j \leq k + 2\ell$, we have

$$\sum_{\omega} \|P_\ell^\omega Q_{<j} u_k\|_{L^\infty L^2}^2 \lesssim \|u_k\|_{S_k}^2 \simeq \|u_k\|_S^2,$$

by disposing $Q_{<j}$ (using boundedness on $L^\infty L^2$) and using $S_k^{ang} \supseteq S_k$.

8.4.2. *Bilinear estimates that do not involve any null forms.* We first prove Proposition 8.11, which does not involve any null forms.

Proof of Proposition 8.11. In this proof, we adopt the convention of writing $L^p L^{q+}$ for $L^p L^{\tilde{q}}$ with $\tilde{q}^{-1} = q^{-1} - \delta_0$. In particular, if (p, q) is a sharp Strichartz exponent with $\delta_0 \leq p^{-1} \leq \frac{1}{2} - \delta_0$, then $2^{(\frac{1}{p} + \frac{4}{q} - 2 - 4\delta_0)k} \mathrm{Str}_k^0 \subseteq L^p L^{q+}$.

To prove (8.17), we apply Hölder and Bernstein (on the lowest frequency factor), where we put u_{k_1} in $L^{\frac{9}{4}} L^{\frac{54}{11}+}$ and v_{k_2} in $L^{18} L^{\frac{27}{13}+}$. The proof of (8.18) is similar, except we put v_{k_2} in $L^9 L^{\frac{54}{23}+}$. The proofs of (8.20) and (8.21) are similar; for (8.20), we apply Hölder and Bernstein with u_{k_1} in $L^2 L^\infty$ and v_{k_2} in $L^\infty L^2$, and for (8.21) we put v_{k_2} in $L^{18} L^{\frac{27}{13}}$ instead.

It only remains to establish (8.19) and (8.22). First, (8.22) follows simply by applying Hölder and Bernstein (on the lowest frequency factor), where we put u_{k_1}, v_{k_2} in $L^2 L^6$. To prove (8.19), we divide into two cases. When $k \geq k_1 - 10$, the desired bound follows by Hölder, where we put both u_{k_1} and v_{k_2} in $L^2 L^\infty$. On the other hand, when $k < k_1 - 10$, we have $k = k_{\min}$ and $k_1 = k_2 + O(1)$ by Littlewood–Paley trichotomy. We decompose the inputs and the output by frequency projections to cubes of the form $\mathcal{C}_k(0)$, i.e.,

$$P_k \mathcal{O}(u_{k_1}, v'_{k_2}) = \sum_{\mathcal{C}, \mathcal{C}^1, \mathcal{C}^2} P_k P_{\mathcal{C}} \mathcal{O}(P_{\mathcal{C}^1} u_{k_1}, P_{\mathcal{C}^2} v'_{k_2}),$$

where $\mathcal{C}, \mathcal{C}^1, \mathcal{C}^2 \in \{\mathcal{C}_k(0)\}$. The summand on the RHS vanishes except when $-\mathcal{C} + \mathcal{C}^1 + \mathcal{C}^2 \ni 0$. For a pair $\mathcal{C}$ and $\mathcal{C}^1$ [resp. $\mathcal{C}^2$], there are only $O(1)$-many $\mathcal{C}^2$ [resp. $\mathcal{C}^1$] such that the preceding condition holds. Moreover, there are only $O(1)$-many $\mathcal{C}$ in the annulus $\{|\xi| \simeq 2^k\}$. Therefore, by Hölder and Cauchy–Schwarz (in $\mathcal{C}^1$ and $\mathcal{C}^2$), we have

$$\begin{aligned}
&2^{-2k}\|P_k\mathcal{O}(u_{k_1}, v'_{k_2})\|_{L^1L^\infty} \\
&\lesssim 2^{-2k}\left(\sum_{\mathcal{C}^1}\|P_{\mathcal{C}^1}u_{k_1}\|^2_{L^2L^\infty}\right)^{\frac{1}{2}}\left(\sum_{\mathcal{C}^2}\|P_{\mathcal{C}^2}v'_{k_2}\|^2_{L^2L^\infty}\right)^{\frac{1}{2}} \\
&\lesssim \|Du_{k_1}\|_S\|v'_{k_2}\|_S,
\end{aligned}$$

which completes the proof. □

8.4.3. *Bilinear null form estimates for the N-norm.* We now prove Proposition 8.12. We start with a lemma quantifying the gain from the null form $\mathcal{O}(\partial^\alpha(\cdot), \partial_\alpha(\cdot))$, which is a quick consequence of Lemmas 8.7 and 8.21.

Lemma 8.22. *Let k, k_1, k_2, j, j_1, j_2 satisfy $k_{\max} - k_{\mathrm{med}} \leq 5$, $j, j_1, j_2 \leq k_{\min} + C_0$, $j_1 = j + O(1)$ and $j_2 = j + O(1)$. Define $\ell = \min\{\frac{j-k_{\min}}{2}, 0\}$, and let $\mathcal{C}, \mathcal{C}^1, \mathcal{C}^2$ be rectangular boxes of the form $\mathcal{C}_{k_{\min}}(\ell)$. Then we have*

$$P_kQ_{<j}P_{\mathcal{C}}\mathcal{O}(\partial^\alpha Q_{<j_1}P_{\mathcal{C}^1}u_{k_1}, \partial_\alpha Q_{<j_2}P_{\mathcal{C}^2}v_{k_2}) = C2^{2\ell}P_{\mathcal{C}}\tilde{\mathcal{O}}(\nabla P_{\mathcal{C}^1}u_{k_1}, \nabla P_{\mathcal{C}^2}v_{k_2}) \tag{8.98}$$

for some universal constant C and a disposable operator $\tilde{O}$.

Proof. By disposability of $P_kQ_{<j}P_{\mathcal{C}}$, $P_{k_1}Q_{<j_1}P_{\mathcal{C}^1}$ and $P_{k_2}Q_{<j_2}P_{\mathcal{C}^2}$, we may harmlessly assume that (say) $j, j_1, j_2 < k_{\min} - 5$. Then we can decompose

$$P_kQ_{<j}P_{\mathcal{C}}\mathcal{O}(\partial^\alpha Q_{<j_1}P_{\mathcal{C}^1}u_{k_1}, \partial_\alpha Q_{<j_2}P_{\mathcal{C}^2}v_{k_2}) = \sum_{\pm,\pm_1,\pm_2} P_kQ^{\mp}_{<j}P_{\mathcal{C}}\mathcal{O}(\partial^\alpha Q^{\pm_1}_{<j_1}P_{\mathcal{C}^1}u_{k_1}, \partial_\alpha Q^{\pm_2}_{<j_2}P_{\mathcal{C}^2}v_{k_2}).$$

By Lemma 8.21, the summand on the RHS vanishes (and thus (8.98) holds trivially) unless $|\angle(\pm_1\mathcal{C}^1, \pm_2\mathcal{C}^2)| \lesssim 2^\ell$. In such a case, (8.98) follows from the decompositions (8.13) in Lemma 8.7 and the schematic identities

$$\begin{aligned}
\mathcal{N}_{0,\pm_1\pm_2}(Q^{\pm_1}_{<j_1}P_{\mathcal{C}^1}u_{k_1}, Q^{\pm_2}_{<j_2}P_{\mathcal{C}^2}v_{k_2}) &= C2^{k_1+k_2}2^{2\ell}\tilde{\mathcal{O}}(P_{\mathcal{C}^1}u_{k_1}, P_{\mathcal{C}^2}v_{k_2}), \\
\mathcal{R}_0(Q^{\pm_1}_{<j_1}P_{\mathcal{C}^1}u_{k_1}, Q^{\pm_2}_{<j_2}P_{\mathcal{C}^2}v_{k_2}) &= C2^j2^{-\min\{k_1,k_2\}}\tilde{\mathcal{O}}(\nabla P_{\mathcal{C}^1}u_{k_1}, \nabla P_{\mathcal{C}^2}v_{k_2}),
\end{aligned}$$

which in turn follow from Definition 8.3 (see also Remark 8.4) and (8.14), respectively. □

Proof of Proposition 8.12. Estimates (8.23) and (8.26) were proved in [17, Proposition 7.1]. Estimate (8.25) is a simple consequence of Hölder and Bernstein for u'_{k_1}, v_{k_2} or the output, depending on which has the lowest frequency. In the remainder of the proof, we prove (8.24) and (8.27) simultaneously.

Step 1: High modulation inputs/output. The goal of this step is to prove

$$\|P_k\mathcal{O}(\partial^\alpha u_{k_1}, \partial_\alpha v_{k_2}) - P_kQ_{<k_{\min}}\mathcal{O}(\partial^\alpha Q_{<k_{\min}}u_{k_1}, \partial_\alpha Q_{<k_{\min}}v_{k_2})\|_N \lesssim 2^{\frac{k_{\min}+k_{\max}}{2}}\|\nabla u_{k_1}\|_S\|\nabla v_{k_2}\|_S \tag{8.99}$$

Note that this step is vacuous for (8.27). Here we do not need the null form, and simply view $\mathcal{O}(\partial^\alpha u_{k_1}, \partial^\alpha v_{k_2})$ as $\tilde{\mathcal{O}}(\nabla u_{k_1}, \nabla v_{k_2})$ for some disposable $\tilde{\mathcal{O}}$.

We begin by reducing (8.99) into an atomic form. For $j, j_1, j_2 \geq k_{\min}$, we claim that

$$\left|\int Q_jw_k\tilde{\mathcal{O}}(Q_{<j_1}u'_{k_1}, Q_{<j_2}v'_{k_2})\, dtdx\right| \lesssim 2^{-\frac{1}{2}j}2^{k_{\min}}2^{\frac{1}{2}k_1}\|w_k\|_{X^{0,\frac{1}{2}}_\infty}\|u'_{k_1}\|_S\|v'_{k_2}\|_{L^\infty L^2}. \tag{8.100}$$

Once we prove (8.100), then by duality (recall that $N^* = L^\infty L^2 \cap X_\infty^{0,\frac{1}{2}}$) we would have

$$\begin{aligned}
\sum_{j \geq k_{\min}} \| P_k Q_j \mathcal{O}(\partial^\alpha u_{k_1}, \partial_\alpha v_{k_2}) \|_N \lesssim & 2^{\frac{1}{2} k_{\min}} 2^{\frac{1}{2} k_1} \| \nabla u_{k_1} \|_S \| \nabla v_{k_2} \|_{L^\infty L^2}, \\
\sum_{j \geq k_{\min}} \| P_k Q_{<k_{\min}} \mathcal{O}(\partial^\alpha Q_j u_{k_1}, \partial_\alpha v_{k_2}) \|_N \lesssim & 2^{\frac{1}{2} k_{\min}} 2^{\frac{1}{2} k_2} \| \nabla u_{k_1} \|_{X_\infty^{0,\frac{1}{2}}} \| \nabla v_{k_2} \|_S, \\
\sum_{j \geq k_{\min}} \| P_k Q_{<k_{\min}} \mathcal{O}(\partial^\alpha Q_{<k_{\min}} u_{k_1}, \partial_\alpha Q_j v_{k_2}) \|_N \lesssim & 2^{\frac{1}{2} k_{\min}} 2^{\frac{1}{2} k_1} \| \nabla u_{k_1} \|_S \| \nabla v_{k_2} \|_{X_\infty^{0,\frac{1}{2}}},
\end{aligned}$$

from which (8.99) would follow.

To prove (8.100), we decompose u', v', w by frequency projection to cubes of the form $\mathcal{C}_{k_{\min}}(0)$, i.e.,

$$\int Q_j w_k \tilde{\mathcal{O}}(Q_{<j_1} u'_{k_1}, Q_{<j_2} v'_{k_2}) \, dt dx = \sum_{\mathcal{C}^0, \mathcal{C}^1, \mathcal{C}^2} \int Q_j P_{\mathcal{C}^0} w_k \tilde{\mathcal{O}}(Q_{<j_1} P_{\mathcal{C}^1} u'_{k_1}, Q_{<j_2} P_{\mathcal{C}^1} v'_{k_2}) \, dt dx,$$

where $\mathcal{C}, \mathcal{C}^1, \mathcal{C}^2 \in \{\mathcal{C}_{k_{\min}}(0)\}$.

Let $\mathcal{C}^{\max}$, $\mathcal{C}^{\mathrm{med}}$ and $\mathcal{C}^{\min}$ denote the re-indexing of the boxes $\mathcal{C}^0, \mathcal{C}^1, \mathcal{C}^2$, which are situated at the frequency annuli $\{|\xi| \simeq 2^{k_{\max}}\}$, $\{|\xi| \simeq 2^{k_{\mathrm{med}}}\}$ and $\{|\xi| \simeq 2^{k_{\min}}\}$, respectively. The summand on the RHS vanishes unless $\mathcal{C}^{\max} + \mathcal{C}^{\mathrm{med}} + \mathcal{C}^{\min} \ni 0$. For a fixed pair $\mathcal{C}^{\min}$ and $\mathcal{C}^{\max}$ [resp. $\mathcal{C}^{\mathrm{med}}$], this happens only for $O(1)$-many $\mathcal{C}^{\mathrm{med}}$ [resp. $\mathcal{C}^{\max}$]. Moreover, note that each $\mathcal{C}^i$ lies within an angular sector of size $O(2^{k_{\min} - k_i})$; hence, $Q_{<j_i} P_{\mathcal{C}^i}$ is disposable ($i = 1, 2$). Thus, by Hölder, Cauchy–Schwarz (in $\mathcal{C}^{\max}$ and $\mathcal{C}^{\mathrm{med}}$) and the fact that there are only $O(1)$-many cubes $\mathcal{C}^{\min}$ situated in $\{|\xi| \simeq 2^{k_{\min}}\}$ (so any ℓ^r-sums over $\mathcal{C}^{\min}$ are equivalent), we have

$$\begin{aligned}
& |\sum_{\mathcal{C}^0, \mathcal{C}^1, \mathcal{C}^2} \int Q_j P_{\mathcal{C}^0} w_k \tilde{\mathcal{O}}(Q_{<j_1} P_{\mathcal{C}^1} u'_{k_1}, Q_{<j_2} P_{\mathcal{C}^2} v'_{k_2}) \, dt dx| \\
& \lesssim \| (\sum_{\mathcal{C}^0} \| Q_j P_{\mathcal{C}^0} w_k(t, \cdot) \|_{L^2}^2)^{\frac{1}{2}} \|_{L^2_t} \| (\sum_{\mathcal{C}^1} \| P_{\mathcal{C}^1} u'_{k_1}(t, \cdot) \|_{L^\infty}^2)^{\frac{1}{2}} \|_{L^2_t} \| (\sum_{\mathcal{C}^2} \| P_{\mathcal{C}^2} v'_{k_2}(t, \cdot) \|_{L^2}^2)^{\frac{1}{2}} \|_{L^\infty_t} \\
& \lesssim \| Q_j w_k \|_{L^2 L^2} (\sum_{\mathcal{C}^1} \| P_{\mathcal{C}^1} u'_{k_1} \|_{L^2 L^\infty}^2)^{\frac{1}{2}} \| v'_{k_2} \|_{L^\infty L^2} \\
& \lesssim 2^{-\frac{1}{2} j} 2^{k_{\min}} 2^{\frac{1}{2} k_1} \| w_k \|_{X_\infty^{0,\frac{1}{2}}} \| u'_{k_1} \|_S \| v'_{k_2} \|_{L^\infty L^2},
\end{aligned}$$

as desired.

Step 2: Proofs of (8.24) **and** (8.27)**.** For $j < k_{\min}$ and $\ell = \frac{j - k_{\min}}{2}$, we claim that

$$\| P_k Q_j \mathcal{O}(\partial^\alpha Q_{<j} u_{k_1}, \partial_\alpha Q_{<j} v_{k_2}) \|_N \lesssim 2^{-\frac{1}{2}(j - k_{\min})} 2^{\frac{5}{2} \ell} 2^{\frac{1}{2} k_{\min}} 2^{\frac{1}{2} k_1} \| \nabla u_{k_1} \|_S \| \nabla v_{k_2} \|_S, \tag{8.101}$$

$$\| P_k Q_{\leq j} \mathcal{O}(\partial^\alpha Q_j u_{k_1}, \partial_\alpha Q_{<j} v_{k_2}) \|_N \lesssim 2^{-\frac{1}{2}(j - k_{\min})} 2^{\frac{5}{2} \ell} 2^{\frac{1}{2} k_{\min}} 2^{\frac{1}{2} k_2} \| \nabla u_{k_1} \|_S \| \nabla v_{k_2} \|_S, \tag{8.102}$$

$$\| P_k Q_{\leq j} \mathcal{O}(\partial^\alpha Q_{\leq j} u_{k_1}, \partial_\alpha Q_j v_{k_2}) \|_N \lesssim 2^{-\frac{1}{2}(j - k_{\min})} 2^{\frac{5}{2} \ell} 2^{\frac{1}{2} k_{\min}} 2^{\frac{1}{2} k_1} \| \nabla u_{k_1} \|_S \| \nabla v_{k_2} \|_S. \tag{8.103}$$

Assuming that these estimates hold, we first conclude the proofs of (8.24) and (8.27). We start with (8.24). By Step 1, it suffices to estimate $P_k Q_{<k_{\min}} \mathcal{O}(\partial^\alpha Q_{<k_{\min}} u_{k_1}, Q_{<k_{\min}} v_{k_2})$. Decomposing the inputs and the output using $Q_{<k_{\min}} = \sum_{j < k_{\min}} Q_j$, and dividing cases according to which has dominant modulation (corresponding to j in the above estimates), (8.24) follows by summing (8.101)–(8.103) over j. To prove (8.27), observe simply that the modulation restrictions of the inputs and the output restricts the j-summation to $j < k_{\min}-\kappa$ in the preceding argument.

It remains to establish (8.101)–(8.103).

Step 2.1: Proof of (8.101). Here we provide a detailed proof of (8.101); similar arguments involving orthogonality and the null form gain will be used repeatedly in the remainder of this subsection.

We expand

$$P_k Q_j \mathcal{O}(\partial^\alpha Q_{<j} u_{k_1}, \partial_\alpha Q_{<j} v_{k_2}) = \sum_{\pm_0,\pm_1,\pm_2} \sum_{\mathcal{C}^0,\mathcal{C}^1,\mathcal{C}^2} P_k Q_j^{\mp_0} P_{-\mathcal{C}^0} \mathcal{O}(\partial^\alpha Q_{<j}^{\pm_1} P_{\mathcal{C}^1} u_{k_1}, \partial_\alpha Q_{<j}^{\pm_2} P_{\mathcal{C}^2} v_{k_2}),$$

where $\mathcal{C}^0, \mathcal{C}^1, \mathcal{C}^2 \in \{\mathcal{C}_{k_{\min}}(\ell)\}$. By duality, in order to estimate the summand on the RHS, it suffices to bound

$$\int P_k Q_j^{\pm_0} P_{\mathcal{C}^0} w \, \mathcal{O}(\partial^\alpha Q_{<j}^{\pm_1} P_{\mathcal{C}^1} u_{k_1}, \partial_\alpha Q_{<j}^{\pm_2} P_{\mathcal{C}^2} v_{k_2}) \, dtdx. \tag{8.104}$$

Let $\mathcal{C}^{\max}$, $\mathcal{C}^{\mathrm{med}}$ and $\mathcal{C}^{\min}$ denote the re-indexing of the boxes $-\mathcal{C}, \mathcal{C}^1, \mathcal{C}^2$, which are situated at the frequency annuli $\{|\xi| \simeq 2^{k_{\max}}\}$, $\{|\xi| \simeq 2^{k_{\mathrm{med}}}\}$ and $\{|\xi| \simeq 2^{k_{\min}}\}$, respectively.

Note that (8.104) vanishes unless $\mathcal{C}^0 + \mathcal{C}^1 + \mathcal{C}^2 \ni 0$. Combined with the geometry of the cone (Lemma 8.21) we see that: For a fixed $\mathcal{C}^{\max}$ [resp. $\mathcal{C}^{\mathrm{med}}$], (8.104) vanishes except for $O(1)$-many $\mathcal{C}^{\min}$ and $\mathcal{C}^{\mathrm{med}}$ [resp. $\mathcal{C}^{\max}$]. By Hölder, Cauchy–Schwarz (in $\mathcal{C}^{\max}$ and $\mathcal{C}^{\mathrm{med}}$) and Lemma 8.22, we obtain

$$\begin{aligned}
&\Big| \sum_{\pm_0,\pm_1,\pm_2} \sum_{\mathcal{C}^0,\mathcal{C}^1,\mathcal{C}^2} (8.104) \Big| \\
&\lesssim \sum_{\pm_0} 2^{2\ell} \| (\sum_{\mathcal{C}^0} \| P_k Q_j^{\pm_0} P_{\mathcal{C}^0} w(t,\cdot) \|_{L^2}^2)^{\frac{1}{2}} \|_{L^2_t} \\
&\quad \times \| (\sum_{\mathcal{C}^1} \| \nabla P_{\mathcal{C}^1} u_{k_1}(t,\cdot) \|_{L^\infty}^2)^{\frac{1}{2}} \|_{L^2_t} \| (\sum_{\mathcal{C}^2} \| \nabla P_{\mathcal{C}^2} v_{k_2}(t,\cdot) \|_{L^2}^2)^{\frac{1}{2}} \|_{L^\infty_t} \\
&\lesssim \sum_{\pm_0} 2^{2\ell} \| P_k Q_j^{\pm_0} w \|_{L^2 L^2} (\sum_{\mathcal{C}^1} \| \nabla P_{\mathcal{C}^1} u_{k_1} \|_{L^2 L^\infty}^2)^{\frac{1}{2}} \| \nabla v_{k_2} \|_{L^\infty L^2} \\
&\lesssim 2^{-\frac{1}{2} j} 2^{\frac{5}{2}\ell} 2^{k_{\min}} 2^{\frac{1}{2} k_1} \| w \|_{X^{0,\frac{1}{2}}_\infty} \| \nabla u_{k_1} \|_S \| \nabla v_{k_2} \|_{L^\infty L^2}.
\end{aligned}$$

By duality, (8.101) follows.

Steps 1.2 & 1.3: Proofs of (8.102) *&* (8.103). We now sketch the proofs of (8.102) and (8.103), which are very similar to Step 2.1. As before, we expand each modulation projection to the $\pm$-parts, and decompose the output, u, v by frequency projection to $-\mathcal{C}^0, \mathcal{C}^1, \mathcal{C}^2 \in \{\mathcal{C}_{k_{\min}}(\ell)\}$, respectively.

We proceed as in Step 1.1 but put the test function w in $L^\infty L^2$ and the input with the dominant modulation in $L^2 L^2$. Then we obtain

$$
\begin{aligned}
&|\sum_{\pm_0,\pm_1,\pm_2}\sum_{\mathcal{C}^0,\mathcal{C}^1,\mathcal{C}^2}\iint P_k Q^{\pm_0}_{\leq j} P_{\mathcal{C}^0}\mathcal{O}(\partial^\alpha Q^{\pm_1}_{j} P_{\mathcal{C}^1} u_{k_1}, \partial_\alpha Q^{\pm_2}_{<j} P_{\mathcal{C}^2} v_{k_2})| \\
&\lesssim 2^{-\frac{1}{2}j} 2^{\frac{5}{2}\ell} 2^{k_{\min}} 2^{\frac{1}{2}k_2} \|w\|_{L^\infty L^2} \|\nabla u_{k_1}\|_{X^{0,\frac{1}{2}}_\infty} \|\nabla v_{k_2}\|_S, \\
&|\sum_{\pm_0,\pm_1,\pm_2}\sum_{\mathcal{C}^0,\mathcal{C}^1,\mathcal{C}^2}\iint P_k Q^{\pm_0}_{\leq j} P_{\mathcal{C}^0}\mathcal{O}(\partial^\alpha Q^{\pm_1}_{\leq j} P_{\mathcal{C}^1} u_{k_1}, \partial_\alpha Q^{\pm_2}_{j} P_{\mathcal{C}^2} v_{k_2})| \\
&\lesssim 2^{-\frac{1}{2}j} 2^{\frac{5}{2}\ell} 2^{k_{\min}} 2^{\frac{1}{2}k_1} \|w\|_{L^\infty L^2} \|\nabla u_{k_1}\|_S \|\nabla v_{k_2}\|_{X^{0,\frac{1}{2}}_\infty}.
\end{aligned}
$$

By duality, (8.102) and (8.103) follow. ☐

8.4.4. *Bilinear estimates for the $X^{s,b,p}_r$-type norms.* Next, we prove Propositions 8.13, 8.14, 8.15 and 8.16.

Proof of Proposition 8.13. Estimates (8.28) and (8.29) were proved in [10, Eqns. (132) and (133)]; note that the slightly stronger S^1-norm is used on the RHS in [10, Eqns. (132) and (133)], but the proofs in fact lead to (8.28) and (8.29). Estimates (8.30) and (8.31) follow from slight modifications of the proofs of [10, Eqns. (134) and (140)] (the Z-norm in [10] is stronger than ours), as we outline below.

For (8.30), we first recall the definition of $\mathcal{H}^*$. For each $j < k_1 - C$, we introduce $\ell = \frac{1}{2}(j - k_1)$ and decompose

$$
P_k Q_{<j-C}\mathcal{N}(|D|^{-1} Q_j u_{k_1}, Q_{<j-C} v_{k_2}) = \sum_{\omega,\omega'} P_k Q_{<j-C}\mathcal{N}(|D|^{-1} P^\omega_\ell Q_j u_{k_1}, P^\ell_{\omega'} Q_{<j-C} v_{k_2}).
$$

By the geometry of the cone (Lemma 8.21), the summand vanishes unless $|\angle(\omega, \pm\omega')| \lesssim 2^\ell$ for some sign $\pm$. In this case, the null form $\mathcal{N}$ gains $2^{k_1+k_2} 2^\ell$ (cf. Definition 8.3), and hence we have

$$
\begin{aligned}
&\|P_k Q_{<j-C}\mathcal{N}(|D|^{-1} Q_j u_{k_1}, Q_{<j-C} v_{k_2})\|_{L^1 L^2} \\
&\lesssim \sum_{\omega,\omega':\min_\pm|\angle(\omega,\pm\omega')|\lesssim 2^\ell} 2^{k_2} 2^\ell \|P^\omega_\ell Q_j u_{k_1}\|_{L^1 L^\infty} \|P^{\omega'}_\ell Q_{<j-C} v_{k_2}\|_{L^\infty L^2} \\
&\lesssim 2^{k_2} 2^{(\frac{1}{2}-2b_0)\ell} \left(\sum_\omega (2^{(\frac{1}{2}+2b_0)\ell} \|P^\omega_\ell Q_{k+2\ell} u_{k_1}\|_{L^1 L^\infty})^2\right)^{\frac{1}{2}} \left(\sum_{\omega'} \|P^{\omega'}_\ell Q_{<j-C} v_{k_2}\|^2_{L^\infty L^2}\right)^{\frac{1}{2}} \\
&\lesssim 2^{(\frac{1}{2}-2b_0)\ell} \left(\sum_\omega (2^{(\frac{1}{2}+2b_0)\ell} \|P^\omega_\ell Q_{k+2\ell} u_{k_1}\|_{L^1 L^\infty})^2\right)^{\frac{1}{2}} \|D v_{k_2}\|_S.
\end{aligned}
$$

In the second inequality, we used Cauchy–Schwarz (or Schur's test) with the fact that the ω, ω' is essentially diagonal (i.e., for a fixed ω, there are only $O(1)$ many ω''s such that the sum is nonvanishing, and vice versa). Summing up in $j < k_1 - C$, then using the definition of the Z^1-norm, (8.30) follows.

Next, (8.31) is proved by essentially the same argument (with the same numerology) as above. Here we do not gain 2^ℓ from the null form $\mathcal{N}$, but rather from the extra factor $\Delta^{-\frac{1}{2}}\Box^{\frac{1}{2}}$

in the norm $\Delta^{-\frac{1}{2}}\Box^{\frac{1}{2}}Z^1$. Finally, (8.32) and (8.33) follow from the preceding proofs, once we observe that the modulation localization of u_{k_1} restricts the j-summation to $j < k_1 - \kappa$, which then leads to the small factor $2^{-(\frac{1}{2}-2b_0)\kappa}$. $\Box$

Proof of Proposition 8.14. In view of the embedding $N \cap \Box Z^1 \subseteq \Box Z^1_{p_0}$, (8.34) would follow once (8.35) is proved. Estimates (8.36) and (8.37) follow from (134) and (141) in [10], respectively. Moreover, when $k \geq k_1 - C$, (8.35) follows from (134) and (135) in [10]. In using the estimates from [10], we remind the reader that the Z-norm in [10] (which is equal to $\sum_k \|P_k Q_{<k} u\|_{X_\infty^{-\frac{1}{4},\frac{1}{4},1}}$) is stronger the Z-norm in this work. Moreover, although (134), (135) and (141) in [10] are stated with the S^1-norm on the RHS, an inspection of the proof reveals that only the S-norm is used.

It remains to establish (8.35) in the case $k < k_1 - C$. By Littlewood–Paley trichotomy, note that the LHS vanishes unless $k = k_{\min}$ and $k_1 = k_2 + O(1)$. By (8.36), we are only left to show that the $\Box Z^1$-norm of

$$P_k \mathcal{H}_k \mathcal{N}(u_{k_1}, v_{k_2}) = \sum_{j<k+C} P_k Q_j \mathcal{N}(Q_{<j-C} u_{k_1}, Q_{<j-C} v_{k_2}) \tag{8.105}$$

is bounded by $\lesssim 2^k \|Du_{k_1}\|_S \|Dv_{k_2}\|_S$.

Consider the summand of (8.105). We decompose the inputs and the output by frequency projections to rectangular boxes of the form $\mathcal{C}_k(\ell)$, where $\ell = \min\{\frac{j-k}{2}, 0\}$. Then we need to consider the expression

$$P_k Q_j P_{\mathcal{C}} \mathcal{N}(Q_{<j-C} P_{\mathcal{C}^1} u_{k_1}, Q_{<j-C} P_{\mathcal{C}^2} v_{k_2})$$

where $\mathcal{C}, \mathcal{C}^1, \mathcal{C}^2 \in \{\mathcal{C}_k(\ell)\}$. This expression is nonvanishing only when $-\mathcal{C} + \mathcal{C}^1 + \mathcal{C}^2 \ni 0$. In fact, combined with the geometry of the cone (Lemma 8.21), we see that for each fixed $\mathcal{C}^1$ [resp. $\mathcal{C}^2$], it is nonvanishing only for $O(1)$-many $\mathcal{C}$ and $\mathcal{C}^2$ [resp. $\mathcal{C}^1$]. The null form gains the factor $2^{k_1+k_2}2^{\ell}$. By Hölder and Cauchy–Schwarz (in $\mathcal{C}^1$ and $\mathcal{C}^2$), we have

$$\begin{aligned}
&\|P_k Q_j \mathcal{N}(Q_{<j-C} u_{k_1}, Q_{<j-C} v_{k_2})\|_{\Box Z^1} \\
&= 2^{-\frac{3}{2}k} 2^{-\frac{1}{2}j} \Big\| \sum_{\mathcal{C},\mathcal{C}^1,\mathcal{C}^2} P_k Q_j P_{\mathcal{C}} \mathcal{N}(Q_{<j-C} P_{\mathcal{C}^1} u_{k_1}, Q_{<j-C} P_{\mathcal{C}^2} v_{k_2}) \Big\|_{L^1 L^\infty} \\
&\lesssim 2^{-\frac{3}{2}k} 2^{-\frac{1}{2}j} 2^{k_1+k_2} 2^{\ell} \left(\sum_{\mathcal{C}^1} \|Q_{<j-C} P_{\mathcal{C}^1} u_{k_1}\|_{L^2 L^\infty}^2 \right)^{\frac{1}{2}} \left(\sum_{\mathcal{C}^2} \|Q_{<j-C} P_{\mathcal{C}^2} v_{k_2}\|_{L^2 L^\infty}^2 \right)^{\frac{1}{2}} \\
&\lesssim 2^{-\frac{1}{2}(k-j)} 2^k \|Du_{k_1}\|_S \|Dv_{k_2}\|_S.
\end{aligned}$$

Summing up in $j < k + C$, the desired estimate follows. $\Box$

Proof of Proposition 8.15. For all the estimates, the most difficult case is when $k_1 < k - 10$ (low-high interaction) and when u_{k_1} has the dominant modulation, i.e., the expression $P_k \mathcal{H}^*_{k_1} \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})$.

Step 1: Proof of (8.38)**,** (8.39) **and** (8.40)**.** We divide into three cases: (1) $k_1 \geq k - 10$; (2) $k_1 < k - 10$ but either the output or v_{k_2} has the dominant modulation; or (3) $k_1 < k - 10$ and u_{k_1} has the dominant modulation.

Step 1.1: $k_1 \geq k - 10$. In this case, all three bounds can be proved simultaneously. The idea is to apply Propositions 8.12 and 8.14. Indeed, by (8.35) and the fact that the LHS vanishes unless $k_1 = k_{\max} + O(1)$ (Littlewood–Paley trichotomy), we see that

$$\begin{aligned}\|P_k \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})\|_{\Box Z^1} &\lesssim 2^{k-k_1} \|P_k |D|^{-1} \mathcal{N}(u_{k_1}, v_{k_2})\|_{\Box Z^1} \\ &\lesssim 2^{-C\delta_1(k_{\max}-k_{\min})} \|D u_{k_1}\|_S \|D v_{k_2}\|_S.\end{aligned}$$

Combined with (8.23), it follows that

$$\|P_k \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})\|_{N \cap \Box Z^1} \lesssim 2^{-C\delta_1(k_{\max}-k_{\min})} \|D u_{k_1}\|_S \|D v_{k_2}\|_S.$$

By the chain of embeddings $N \cap \Box Z^1 \subseteq \Box Z^1_{p_0} \subseteq \Box \tilde{Z}^1_{p_0}$, the desired bounds follow.

Step 1.2: $k_1 < k - 10$, contribution of $1 - \mathcal{H}^*_{k_1}$. Note that, by Littlewood–Paley trichotomy, $P_k \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})$ vanishes unless $k_1 = k_{\min}$ and $k = k_{\max} + O(1)$. In Steps 1.2.a–1.2.c below, we estimate the $\Box Z^1$-norm of $P_k(1 - \mathcal{H}^*_{k_1}) \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2})$. Then in Step 1.2.d, we conclude the proof by interpolating with (8.28).

Step 1.2.a: High modulation inputs/output. The goal of this step is to prove

$$\|P_k \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2}) - P_k Q_{<k_1} \mathcal{N}(|D|^{-1} Q_{<k_1+C} u_{k_1}, Q_{<k_1} v_{k_2})\|_{\Box Z^1} \lesssim 2^{-\frac{1}{4}(k-k_1)} \|D u_{k_1}\|_S \|D v_{k_2}\|_S. \tag{8.106}$$

Here there is no need for null structure, so we simply write $\mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2}) = \mathcal{O}(u_{k_1}, D v_{k_2})$. We begin by proving

$$\|P_k Q_{\geq k_1} \mathcal{O}(u_{k_1}, D v_{k_2})\|_{\Box Z^1} \lesssim 2^{-b_0(k-k_1)} \||D|^{-\frac{1}{2}} u_{k_1}\|_{L^2 L^\infty} \|D v_{k_2}\|_S. \tag{8.107}$$

For $j \geq k_1$, we decompose

$$P_k Q_j P^\omega_{\frac{j-k}{2}} \mathcal{O}(u_{k_1}, D v_{k_2}) = \sum_{\omega'} P_k Q_j P^\omega_{\frac{j-k}{2}} \mathcal{O}(u_{k_1}, D P^{\omega'}_{\frac{j-k}{2}} v_{k_2}).$$

Since $\frac{j-k}{2} \geq k_1 - k$, for each fixed ω there are only $O(1)$-many ω' such that the summand on the RHS is (possibly) non-vanishing, and vice versa. Therefore, by Hölder, Bernstein and Cauchy–Schwarz, we have

$$\begin{aligned}&2^{(-\frac{3}{4}+b_0)(j-k)} 2^{-2k} \left(\sum_\omega \|P_k Q_j P^\omega_{\frac{j-k}{2}} \mathcal{O}(u_{k_1}, D P^{\omega'}_{\frac{j-k}{2}} v_{k_2})\|^2_{L^1 L^\infty} \right) \\ &\lesssim 2^{(-\frac{1}{2}+b_0)(j-k)} 2^{-\frac{1}{2}(k-k_1)} (2^{-\frac{1}{2}k_1} \|u_{k_1}\|_{L^2 L^\infty}) \left(\sum_{\omega'} (2^{\frac{1}{6}k_2} \|P^{\omega'}_{\frac{j-k}{2}} v_{k_2}\|_{L^2 L^6})^2 \right)^{\frac{1}{2}} \\ &\lesssim 2^{(-\frac{1}{2}+b_0)(j-k_1)} 2^{-b_0(k-k_1)} \||D|^{-\frac{1}{2}} u_{k_1}\|_{L^2 L^\infty} \|D v_{k_2}\|_S.\end{aligned}$$

Summing up in $j \geq k_1$, we obtain (8.107).

Next, we prove

$$\|P_k Q_{<k_1} \mathcal{O}(u_{k_1}, D Q_{\geq k_1} v_{k_2})\|_{\Box Z^1} \lesssim 2^{-b_0(k-k_1)} \||D|^{-\frac{1}{2}} u_{k_1}\|_{L^2 L^\infty} \|D v_{k_2}\|_S. \tag{8.108}$$

By (4.6) and (uniform-in-j) boundedness of Q_j on $L^1 L^2$, we have

$$\|P_k Q_{<k_1} f\|_{\Box Z^1} \lesssim 2^{-b_0(k-k_1)} \|f\|_{L^1 L^2}. \tag{8.109}$$

Therefore,

$$\begin{aligned}\|P_k Q_{<k_1}\mathcal{O}(u_{k_1}, DQ_j v_{k_2})\|_{\Box Z^1} \lesssim & 2^{-b_0(k-k_1)}\|P_k Q_{<k_1}\mathcal{O}(u_{k_1}, DQ_j v_{k_2})\|_{L^1L^2} \\ \lesssim & 2^{-\frac{1}{2}(j-k_1)}2^{-b_0(k-k_1)}(2^{-\frac{1}{2}k_1}\|u_{k_1}\|_{L^2L^\infty})\|DQ_j v_{k_2}\|_{X^{0,\frac{1}{2}}_\infty} \\ \lesssim & 2^{-\frac{1}{2}(j-k_1)}2^{-b_0(k-k_1)}\||D|^{-\frac{1}{2}}u_{k_1}\|_{L^2L^\infty}\|Dv_{k_2}\|_S.\end{aligned}$$

Then summing up in $j \geq k_1$, (8.108) follows.

To conclude the proof of (8.106), note that $\||D|^{-\frac{1}{2}}u_{k_1}\|_{L^2L^\infty} \lesssim \|Du_{k_1}\|_S$. Moreover, observe that

$$P_k Q_{<k_1}\mathcal{O}(Q_j u_{k_1}, DQ_{<k_1} v_{k_2})$$

vanishes unless $j < k_1 + 10$.

Step 1.2.b: Output has dominant modulation. Here we prove

$$\sum_{j<k_1} \|P_k Q_j \mathcal{N}(|D|^{-1}Q_{<j_1}u_{k_1}, Q_{<j_2}v_{k_2})\|_{\Box Z^1} \lesssim 2^{-b_0(k-k_1)}\|Du_{k_1}\|_S\|Dv_{k_2}\|_S, \tag{8.110}$$

where $j_1, j_2 = j + O(1)$.

Let $\ell = \frac{1}{2}(j-k_1)$. After decomposing $u_{k_1} = \sum_{\omega'} P^{\omega'}_\ell u_{k_1}$ and $v_{k_2} = \sum_{\omega''} P^{\omega''}_{\frac{j-k}{2}} v_{k_2}$, consider the expression

$$P_k Q_j P^{\omega}_{\frac{j-k}{2}}\mathcal{N}(|D|^{-1}Q_{<j_1}P^{\omega'}_\ell u_{k_1}, Q_{<j_2}P^{\omega''}_{\frac{j-k}{2}}v_{k_2}).$$

Using the geometry of the cone (Lemma 8.21), observe that for every fixed ω [resp. ω''], the preceding expression vanishes except for $O(1)$-many ω' and ω'' [resp. ω]. Moreover, for such a triple $\omega, \omega', \omega''$, the null form $\mathcal{N}$ gains a factor of 2^ℓ. By Hölder, Bernstein (for $P^{\omega}_{\frac{j-k}{2}}v_{k_2}$) and Cauchy–Schwarz (in ω, ω''), we have

$$\begin{aligned}&\|P_k Q_j \mathcal{N}(|D|^{-1}Q_{<j_1}u_{k_1}, Q_{<j_2}v_{k_2})\|_{\Box Z^1} \\ &\lesssim 2^{(-\frac{3}{4}+b_0)(j-k)}2^{-2k}\left(\sum_\omega \|P_k Q_j P^\omega_{\frac{j-k}{2}}\mathcal{N}(|D|^{-1}Q_{<j_1}u_{k_1}, Q_{<j_2}v_{k_2})\|^2_{L^1L^\infty}\right)^{\frac{1}{2}} \\ &\lesssim 2^{(-\frac{1}{2}+b_0)(j-k)}2^\ell 2^{-\frac{1}{2}(k-k_1)}\left(\sup_{\omega'} 2^{-\frac{1}{2}k_1}\|Q_{<j_1}P^{\omega'}_\ell u_{k_1}\|_{L^2L^\infty}\right)\left(\sum_\omega (2^{\frac{1}{6}k_2}\|Q_{<j_2}P^\omega_{\frac{j-k}{2}}v_{k_2}\|_{L^2L^6})^2\right)^{\frac{1}{2}} \\ &\lesssim 2^{-b_0(k_1-j)}2^{-b_0(k-k_1)}\|Du_{k_1}\|_S\|Dv_{k_2}\|_S.\end{aligned}$$

Summing up in $j < k_1$, (8.110) follows.

Step 1.2.c: v has dominant modulation. Next, we prove

$$\sum_{j<k_1} \|P_k Q_{<j_0}\mathcal{N}(|D|^{-1}Q_{<j_1}u_{k_1}, Q_j v_{k_2})\|_{\Box Z^1} \lesssim 2^{-b_0(k-k_1)}\|Du_{k_1}\|_S\|Dv_{k_2}\|_S, \tag{8.111}$$

where $j_0, j_1 = j + O(1)$. As before, let $\ell = \frac{j-k_1}{2}$. By (4.6) and (uniform-in-j) boundedness of Q_j on L^1L^2, we have

$$\|P_k Q_{<j} f\|_{\Box Z^1} \lesssim 2^{-b_0(k-j)}\|f\|_{L^1L^2}.$$

Hence it suffices to estimate the L^1L^2 norm of the output. This time, we decompose $u_{k_1} = \sum_\omega P_\ell^\omega u_{k_1}$ and $v_{k_2} = \sum_{\omega'} P_\ell^{\omega'} v_{k_2}$. By the geometry of the cone, for a fixed ω, the expression

$$P_k Q_{<j_0}\mathcal{N}(|D|^{-1}Q_{<j_1}P_\ell^\omega u_{k_1}, Q_j P_\ell^{\omega'} v_{k_1})$$

vanishes except for $O(1)$-many ω' and vice versa. Moreover, the null form $\mathcal{N}$ gains a factor of 2^ℓ. By Hölder and Cauchy–Schwarz (in ω, ω'), we have

$$\begin{aligned}
&2^{-b_0(k-j)}\|P_k Q_{<j_0}\mathcal{N}(|D|^{-1}Q_{<j_1}P_\ell^\omega u_{k_1}, Q_j P_\ell^{\omega'} v_{k_2})\|_{L^1L^2} \\
&\lesssim 2^{-b_0(k-j)}2^{\frac{3}{2}\ell}2^{\frac{1}{2}k_1}2^{-\frac{1}{2}j}\left(\sum_\omega (2^{-\frac{1}{2}k_1}2^{-\frac{1}{2}\ell}\|Q_{<j_1}P_\ell^\omega u_{k_1}\|_{L^2L^\infty})^2\right)^{\frac{1}{2}}\left(\sum_{\omega'}(2^{k_2}\|Q_j P_\ell^{\omega'} v_{k_2}\|_{X_\infty^{0,\frac{1}{2}}})^2\right)^{\frac{1}{2}} \\
&\lesssim 2^{(-\frac{1}{4}-b_0)(k_1-j)}2^{-b_0(k-k_1)}\|Du_{k_1}\|_S\|Dv_{k_2}\|_S.
\end{aligned}$$

Summing up in $j < k_1$, (8.111) is proved.

Step 1.2.d: Interpolation with (8.28). Combining (8.106), (8.110) and (8.111), we obtain

$$\|P_k(1-\mathcal{H}_{k_1}^*)\mathcal{N}(|D|^{-1}u_{k_1}, v_{k_2})\|_{\Box Z^1} \lesssim 2^{-b_0(k-k_1)}\|Du_{k_1}\|_S\|Dv_{k_2}\|_S.$$

On the other hand, (8.28) and the embedding $N \subseteq X_\infty^{0,-\frac{1}{2}}$ yields a similar bound for the $X_\infty^{0,-\frac{1}{2}}$-norm without the exponential gain. Nevertheless, since we have $\|f\|_{\Box Z^1_{p_0}} \lesssim \|f\|_{\Box Z^1}^{\theta_0}\|f\|_{X_\infty^{0,-\frac{1}{2}}}^{1-\theta_0}$ where $\theta_0 = 2(\frac{1}{p_0}-\frac{1}{2}) > 0$,

$$\|P_k(1-\mathcal{H}_{k_1}^*)\mathcal{N}(|D|^{-1}u_{k_1}, v_{k_2})\|_{\Box Z^1_{p_0}} \lesssim 2^{-\theta_0 b_0(k-k_1)}\|Du_{k_1}\|_S\|Dv_{k_2}\|_S.$$

Then the desired estimate for $\Box\tilde{Z}^1_{p_0}$ follows as well, thanks to the embedding $\Box Z^1_{p_0} \subseteq \Box\tilde{Z}^1_{p_0}$.

Step 1.3: $k_1 < k-10$, contribution of $\mathcal{H}_{k_1}^*$. This is the most difficult case. We consider

$$P_k\mathcal{H}_{k_1}^*\mathcal{N}(|D|^{-1}u_{k_1}, v_{k_2}) = \sum_{j<k_1+C} P_k Q_{<j-C}\mathcal{N}(|D|^{-1}Q_j u_{k_1}, Q_{<j-C}v_{k_2})$$

As before, by Littlewood–Paley trichotomy, this expression vanishes unless $k_1 = k_{\min}$ and $k = k_{\max} + O(1)$.

Recall that all three norms $\Box\tilde{Z}^1_{p_0}$, $\Box Z^1_{p_0}$ and $\Box Z^1$ are of the type $X_1^{s,b,p}$. To ensure the ℓ^2-summability in ω in the definition (4.3), we go through the L^pL^2 norm. More precisely, by Bernstein and L^2-orthogonality of $P^\omega_{\frac{j-k}{2}}$, note that

$$\|P_k Q_j f\|_{X_1^{s,b,p}} \lesssim 2^{sk}2^{\frac{5}{2}(\frac{1}{p}-\frac{1}{2})k}2^{bj}2^{\frac{3}{2}(\frac{1}{p}-\frac{1}{2})j}\|f\|_{L^pL^2}.$$

Since $b + \frac{3}{2}(\frac{1}{p}-\frac{1}{2}) > 0$ in all of these cases by (4.4), we have

$$\|P_k Q_{<j} f\|_{X_1^{s,b,p}} \lesssim 2^{sk}2^{\frac{5}{2}(\frac{1}{p}-\frac{1}{2})k}2^{bj}2^{\frac{3}{2}(\frac{1}{p}-\frac{1}{2})j}\|f\|_{L^pL^2}. \tag{8.112}$$

Hereafter, the proofs of the three bounds differ.

Step 1.3.a: Proof of (8.38). We decompose the inputs and the output by frequency projections to rectangular boxes of the form $\mathcal{C}_{k_1}(\ell)$. Then we need to consider the expression

$$P_k Q_{<j-C}P_{\mathcal{C}}\mathcal{N}(|D|^{-1}Q_j P_{\mathcal{C}^1}u_{k_1}, Q_{<j-C}P_{\mathcal{C}^2}v_{k_2})$$

where $\mathcal{C}, \mathcal{C}^1, \mathcal{C}^2 \in \{\mathcal{C}_{k_1}(\ell)\}$. Note that the above expression is nonvanishing only when $-\mathcal{C} + \mathcal{C}^1 + \mathcal{C}^2 \ni 0$. Moreover, by the geometry of the cone (Lemma 8.21), for each fixed $\mathcal{C}$ [resp. $\mathcal{C}^2$], this expression is nonvanishing only for $O(1)$-many $\mathcal{C}^1$ and $\mathcal{C}^2$ [resp. $\mathcal{C}$], and the null form gains the factor $2^{k_1+k_2}2^{\ell}$.

For exponents $p_1, p_2, q_1, q_2 \geq 2$ such that $p_1^{-1} + p_2^{-1} = p^{-1}$ and $q_1^{-1} + q_2^{-1} = 2^{-1}$, proceeding carefully to exploit spatial orthogonality in L^2, we have

$$
\begin{aligned}
&\|P_k Q_{<j-C} \mathcal{N}(|D|^{-1} Q_j u_{k_1}, Q_{<j-C} v_{k_2})\|_{L^p L^2} \\
&= \| \sum_{\mathcal{C}, \mathcal{C}^1, \mathcal{C}^2} P_k Q_{<j-C} P_{\mathcal{C}} \mathcal{N}(|D|^{-1} Q_j P_{\mathcal{C}^1} u_{k_1}, Q_{<j-C} P_{\mathcal{C}^2} v_{k_2})\|_{L^p L^2} \\
&\lesssim \|(\sum_{\mathcal{C}} \| \sum_{\mathcal{C}^1, \mathcal{C}^2} P_k Q_{<j-C} P_{\mathcal{C}} \mathcal{N}(|D|^{-1} Q_j P_{\mathcal{C}^1} u_{k_1}, Q_{<j-C} P_{\mathcal{C}^2} v_{k_2})(t, \cdot)\|_{L^2}^2)^{\frac{1}{2}} \|_{L^p_t} \\
&\lesssim 2^{\ell} 2^{k_2} \| \sup_{\mathcal{C}_1} \|Q_j P_{\mathcal{C}^1} u_{k_1}(t, \cdot)\|_{L^{q_1}} \|_{L^{p_1}_t} \| (\sum_{\mathcal{C}^2} \|Q_{<j-C} P_{\mathcal{C}^2} v_{k_2}(t, \cdot)\|_{L^{q_2}}^2)^{\frac{1}{2}} \|_{L^{p_2}_t} \\
&\lesssim 2^{\ell} 2^{k_2} \|Q_j u_{k_1}\|_{L^{p_1} L^{q_1}} \left(\sum_{\mathcal{C}^2} \|Q_{<j-C} P_{\mathcal{C}^2} v_{k_2}\|_{L^{p_2} L^{q_2}}^2 \right)^{\frac{1}{2}}. \qquad (8.113)
\end{aligned}
$$

We now apply (8.112) and (8.113) with

$$
(s, b, p, p_1, q_1, p_2, q_2) = (\frac{5}{4} - \frac{3}{p_0} + (\frac{1}{4} - b_0)\theta_0, -\frac{1}{4} - (\frac{1}{4} - b_0)\theta_0, p_0, 2, 2, \frac{2p_0}{2 - p_0}, \infty),
$$

where $\theta_0 = 2(\frac{1}{p_0} - \frac{1}{2})$. We then obtain

$$
\begin{aligned}
&\|P_k Q_{<j-C} \mathcal{N}(|D|^{-1} Q_j u_{k_1}, Q_{<j-C} v_{k_2})\|_{\Box \tilde{Z}^1_{p_0}} \\
&\lesssim 2^{-(1-\frac{1}{p_0})k} 2^k 2^{(-\frac{1}{4} - (\frac{1}{4} - b_0)\theta_0)(j-k)} 2^{-\frac{3}{2}(1-\frac{1}{p_0})(j-k)} 2^{\frac{3}{4}(j-k)} 2^{\ell} \|Q_j u_{k_1}\|_{L^2 L^2} \left(\sum_{\mathcal{C}^2} \|P_{\mathcal{C}^2} Q_{<j-C} v_{k_2}\|_{L^{p_2} L^\infty}^2 \right)^{\frac{1}{2}} \\
&\lesssim 2^{(-\frac{3}{4} + \frac{1}{2}(1-\frac{1}{p_0}) + (\frac{1}{4} - b_0)\theta_0)(k_1 - j)} 2^{(-\frac{1}{2}(1-\frac{1}{p_0}) + (\frac{1}{4} - b_0)\theta_0)(k - k_1)} \|Q_j u_{k_1}\|_{X^{1, \frac{1}{2}}_\infty} \left(\sum_{\mathcal{C}^2} \|D P_{\mathcal{C}^2} v_{k_2}\|_{S_{k_2}[\mathcal{C}_{k_1}(\ell)]}^2 \right)^{\frac{1}{2}}.
\end{aligned}
$$

On the last line, we used

$$
\|Q_{<j-C} P_{\mathcal{C}^2} v_{k_2}\|_{L^{p_2} L^\infty} \lesssim 2^{(\frac{3}{2} - \theta_0)\ell} 2^{(2-\theta_0)(k_1 - k_2)} 2^{(2 - \frac{1}{2}\theta_0)k_2} \|P_{\mathcal{C}^2} v_{k_2}\|_{S_{k_2}[\mathcal{C}_{k_1}(\ell)]},
$$

which follows from interpolation. By (4.4), the factors in front of $(k_1 - j)$ and $(k - k_1)$ are both negative. Summing up in $j < k_1 + C$, we obtain (8.38).

Step 1.3.b: Proof of (8.39). As in the proof of (8.111) (Step 1.2.c), we decompose $u_{k_1} = \sum_\omega P^\omega_\ell u_{k_1}$ and $v_{k_2} = \sum_{\omega'} P^{\omega'}_\ell v_{k_2}$, where $\ell = \frac{j-k_1}{2}$. By the geometry of the cone (Lemma 8.21), the null form gain, Hölder, Cauchy–Schwarz (in ω, ω') and Bernstein (for u_{k_1}), we have

$$
\begin{aligned}
&\|P_k Q_{<j} \mathcal{N}(|D|^{-1} Q_j P^\omega_\ell u_{k_1}, Q_{<j-C} P^{\omega'}_\ell v_{k_2})\|_{L^p L^2} \\
&\lesssim 2^{(1+3(1-\frac{1}{p}))\ell} 2^{4(1-\frac{1}{p})k_1} 2^{k_2} \left(\sum_\omega \|P^\omega_\ell Q_j u_{k_1}\|_{L^p L^{p'}}^2 \right)^{\frac{1}{2}} \left(\sum_{\omega'} \|P^{\omega'}_\ell Q_{<j-C} v_{k_2}\|_{L^\infty L^2}^2 \right)^{\frac{1}{2}}. \qquad (8.114)
\end{aligned}
$$

Applying (8.112) and (8.114) with $(s, b, p) = (\frac{3}{2} - \frac{3}{p_0} + (\frac{1}{4} - b_0)\theta_0, -\frac{1}{2} - (\frac{1}{4} - b_0)\theta_0, p_0)$, where $\theta_0 = 2(\frac{1}{p_0} - \frac{1}{2})$, we obtain

$$\begin{aligned}
&\|P_k Q_{<j} \mathcal{N}(|D|^{-1} Q_j P_\ell^\omega u_{k_1}, Q_{<j} P_\ell^{\omega'} v_{k_2})\|_{\Box Z^1_{p_0}} \\
&\lesssim 2^{-(1-\frac{1}{p_0})k} 2^{(\frac{1}{4} - (\frac{1}{4} - b_0)\theta_0)(j-k)} 2^{-\frac{3}{2}(1-\frac{1}{p_0})(j-k)} \|P_k Q_{<j} \mathcal{N}(|D|^{-1} Q_j P_\ell^\omega u_{k_1}, Q_{<j} P_\ell^{\omega'} v_{k_2})\|_{L^{p_0} L^2} \\
&\lesssim 2^{(-\frac{1}{4} + (\frac{1}{4} - b_0)\theta_0 + \frac{1}{2}(1-\frac{1}{p_0}))(k-k_1)} \|Q_j u_{k_1}\|_{X_1^{\frac{9}{4} - \frac{3}{p_0} + (\frac{1}{4} - b_0)\theta_0, \frac{3}{4} - (\frac{1}{4} - b_0)\theta_0, p_0}} \|D v_{k_2}\|_S.
\end{aligned}$$

By our choices of b_0 and p_0, the overall factor in front of $(k - k_1)$ is negative. Summing up in $j < k_1 + C$, we obtain the desired conclusion.

Step 1.3.c: Proof of (8.40). We again decompose $u_{k_1} = \sum_\omega P_\ell^\omega u_{k_1}$ and $v_{k_2} = \sum_{\omega'} P_\ell^{\omega'} v_{k_2}$, where $\ell = \frac{j-k_1}{2}$. We use (8.112) with $(s, b, p) = (-\frac{5}{4} - b_0, -\frac{3}{4} + b_0, 1)$. By the geometry of the cone (Lemma 8.21), the null form gain, Hölder and Cauchy–Schwarz (in ω, ω'), we have

$$\begin{aligned}
&2^{b_0(j-k)} \|P_k Q_{<j} \mathcal{N}(|D|^{-1} Q_j P_\ell^\omega u_{k_1}, Q_{<j-C} P_\ell^{\omega'} v_{k_2})\|_{L^1 L^2} \\
&\lesssim 2^{b_0(j-k)} 2^\ell 2^{k_2} \left(\sum_\omega \|Q_j P_\ell^\omega u_{k_1}\|^2_{L^{p_0} L^{p_0'}} \right)^{\frac{1}{2}} \left(\sum_{\omega'} \|Q_{<j-C} P_\ell^{\omega'} v_{k_2}\|^2_{L^{p_0'} L^{q_0}} \right)^{\frac{1}{2}} \\
&\lesssim 2^{(b_0 + (\frac{1}{4} - b_0)\theta_0)(k_1 - j)} 2^{-b_0(k-k_1)} 2^{3(1-\frac{1}{p_0})(k-k_1)} \|u_{k_1}\|_{X_\infty^{3(1-\frac{1}{p_0}) - \frac{1}{2} + (\frac{1}{4} - b_0)\theta_0, \frac{1}{2} - (\frac{1}{4} - b_0)\theta_0, p_0}} \|D v_{k_2}\|_S,
\end{aligned}$$

where $q_0^{-1} = 2^{-1} - (p_0')^{-1}$ and $\theta_0 = 2(\frac{1}{p_0} - \frac{1}{2})$. By our choices of p_0 and b_0, the overall factors in front of $(k_1 - j)$ and $(k - k_1)$ are both negative. Summing up in $j < k_1$, the proof is complete.

Step 2: Proof of (8.41)**.** As in Step 1, we divide into three cases.

Step 2.1: $k_1 \geq k - 10$. In view of the embedding $N \cap L^2 \dot{H}^{-\frac{1}{2}} \subseteq X^{-\frac{1}{2} + b_1, -b_1}$ (for any $0 < b_1 < \frac{1}{2}$), the desired bound follows from (8.17) and (8.23).

Step 2.2: $k_1 < k - 10$, *contribution of* $1 - \mathcal{H}^*_{k_1}$. Consider the expression

$$P_k (1 - \mathcal{H}^*_{k_1}) \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2}).$$

Interpolating the N-norm bound (8.28) (recall that $N \subseteq X_\infty^{0, -\frac{1}{2}}$) with an $L^2 \dot{H}^{-\frac{1}{2}}$-norm bound (which is a minor modification of (8.17)), the desired estimate for this expression follows for $0 < b_1 < \frac{1}{2}$.

Step 2.3: $k_1 < k - 10$, *contribution of* $\mathcal{H}^*_{k_1}$. Finally, we estimate

$$P_k \mathcal{H}^*_{k_1} \mathcal{N}(|D|^{-1} u_{k_1}, v_{k_2}) = \sum_{j < k_1 + C} P_k Q_{<j-C} \mathcal{N}(|D|^{-1} Q_j u_{k_1}, Q_{<j-C} v_{k_2}).$$

By (8.114), we have

$$\begin{aligned}
&2^{(\frac{1}{p_0} - 1)k} \|P_k Q_{<j-C} \mathcal{N}(|D|^{-1} Q_j u_{k_1}, Q_{<j-C} v_{k_2})\|_{L^{p_0} L^2} \\
&\lesssim 2^{(\frac{1}{2} + \frac{3}{2}(1-\frac{1}{p_0}))(j-k_1)} 2^{(\frac{1}{p_0} - 1)k} 2^{4(1-\frac{1}{p_0})k_1} 2^{-3(1-\frac{1}{p_0})k_1} 2^{(-\frac{1}{2} + (\frac{1}{4} - b_0)\theta_0)(j-k_1)} \|u_{k_1}\|_{Z^1_{p_0}} \|D v_{k_2}\|_{L^\infty L^2} \\
&\lesssim 2^{(-\frac{3}{2}(1-\frac{1}{p_0}) - (\frac{1}{4} - b_0)\theta_0)(k_1 - j)} 2^{-(1-\frac{1}{p_0})(k-k_1)} \|u_{k_1}\|_{Z^1_{p_0}} \|D v_{k_2}\|_S.
\end{aligned}$$

Summing up in $j < k_1 + C$ and using the embedding $2^{(1-\frac{1}{p_0})k} P_k Q_{<k} L^{p_0} L^2 \subseteq X^{-\frac{1}{2}+b_1,-b_1}$, which holds by Bernstein since $b_1 < \frac{1}{p_0} - \frac{1}{2}$, the proof of (8.41) is complete. ☐

Proof of Proposition 8.16. As in Proposition 8.15, we divide the proof into two cases: $k_1 \geq k-10$ and $k_1 < k-10$.

Step 1: $k_1 \geq k-10$**.** In this case, by (8.20), (8.25) and the embeddings $L^1 L^2 \subseteq \Box Z^1_{p_0} \cap \Box Z^1$ and $L^1 L^2 \cap L^2 \dot{H}^{-\frac{1}{2}} \subseteq X^{-\frac{1}{2}+b_1,-b_1}$, the three bounds follow simultaneously.

Step 2: $k_1 < k-10$**.** We begin with (8.42) and (8.44). By Hölder and Bernstein, we have

$$2^{(\frac{1}{p_0}-1)k} \| P_k \mathcal{O}(u_{k_1}, v'_{k_2}) \|_{L^{p_0} L^2} \lesssim 2^{-(1-\frac{1}{p_0})(k-k_1)} \| u_{k_1} \|_{L^{p_0} \dot{W}^{2-\frac{3}{p_0}, p'_0}} \| v'_{k_2} \|_{L^\infty L^2}$$

By (8.112) with $(s,b,p) = (\frac{3}{2} - \frac{3}{p_0}, -\frac{1}{2}, p_0)$, (8.42) follows. Moreover, by the $L^2 \dot{H}^{-\frac{1}{2}}$-norm estimate (8.17) and the embedding $P_k Q_{<k} L^{p_0} L^2 \subseteq X^{-\frac{1}{2}+b_1,-b_1}$, (8.44) follows as well.

It remains to prove (8.43). Applying (8.107) (from Step 1.2.a of the proof of Proposition 8.15) with $D v_{k_2} = v'_{k_2}$ and the embedding $2^{-\frac{3}{2}k_1} P_{k_1} Y \subseteq L^2 L^\infty$, we have

$$\| P_k Q_{\geq k_1} \mathcal{O}(u_{k_1}, v'_{k_2}) \|_{\Box Z^1} \lesssim 2^{-b_0(k-k_1)} \| D u_{k_1} \|_Y \| v'_{k_2} \|_S.$$

On the other hand, by (8.109) and Hölder, we have

$$\begin{aligned}
\| P_k Q_{<k_1} \mathcal{O}(u_{k_1}, v'_{k_2}) \|_{\Box Z^1} \lesssim & 2^{-b_0(k-k_1)} \| P_k \mathcal{O}(u_{k_1}, v'_{k_2}) \|_{L^1 L^2} \\
\lesssim & 2^{-b_0(k-k_1)} 2^{3(1-\frac{1}{p_0})(k-k_1)} \| D u_{k_1} \|_Y (2^{(\frac{3}{p_0}-3)k_2} \| v'_{k_2} \|_{L^{p'_0} L^{q_0}}) \\
\lesssim & 2^{-b_0(k-k_1)} 2^{3(1-\frac{1}{p_0})(k-k_1)} \| D u_{k_1} \|_Y \| v'_{k_2} \|_S,
\end{aligned}$$

where $q_0^{-1} = 2^{-1} - (p'_0)^{-1}$. By our choice of p_0, the overall factor in front of $(k-k_1)$ is negative; hence, (8.43) follows. ☐

8.4.5. *Trilinear null form estimates.*

Proofs of Propositions 8.17 and 8.18. Estimate (8.45) would follow from Lemma 8.10 and the core estimates (8.46), (8.47) and (8.48), combined with Lemma 8.21 and (4.1).

Estimates (8.46), (8.47) and (8.48) can be established by repeating the proofs of (136), (137) and (138) in [10] with the following modifications:

- Thanks to the frequency localization of the inputs and the output to rectangular boxes of the type $\mathcal{C}_k(\ell)$, the bilinear operators $\mathcal{O}$ and $\mathcal{O}'$ can be safely disposed.
- Moreover, for any disposable multilinear operator $\mathcal{M}$ and rectangular boxes $\mathcal{C}, \mathcal{C}'$ of the type $\mathcal{C}_k(\ell)$ situated in the annuli $\{|\xi| \simeq 2^{k_1}\}$ and $\{|\xi| \simeq 2^{k_2}\}$, respectively, note that (by Lemma 8.7)

$$\begin{aligned}
&\mathcal{M}(\partial^\alpha Q^{\pm}_{<j-C} P_{\mathcal{C}} u_{k_1}, \partial_\alpha Q^{\pm'}_{<j-C} P_{\mathcal{C}'} v_{k_2}, \cdots) \\
&= C 2^{k_1+k_2} \max\{ |\angle(\pm \mathcal{C}, \pm' \mathcal{C}')|^2, 2^{j-\min\{k_1,k_2\}} \} \tilde{\mathcal{M}}(P_{\mathcal{C}} u_{k_1}, P_{\mathcal{C}'} v_{k_2}, \cdots)
\end{aligned}$$

  for some disposable $\tilde{\mathcal{M}}$, which suffices for the proofs in [10].

We also note that although (136)–(138) in [10] are stated with the factor $2^{\delta(k-\min\{k_i\})}$ on the RHS, an inspection of the proofs reveals that the actual gain is $2^{\delta(k-k_1)}$, as claimed in (8.46)–(8.48). We omit the straightforward details. ☐

9. The paradifferential wave equation

Sections 9, 10 and 11 are devoted to the proofs of Theorem 4.24 and Proposition 4.25. In this section, we first reduce the task of proving these results to that of constructing an appropriate parametrix (Section 9.1). Parametrix construction, in turn, is reduced to constructing a renormalization operator that roughly conjugates $\Box + \mathrm{Diff}^{\kappa}_{\mathbf{P}A}$ to $\Box$. Sections 10 and 11 are devoted proofs of the desired properties of the renormalization operator.

9.1. **Reduction to parametrix construction.** We start with a quick reduction of the problem (4.90). After peeling off perturbative terms using commutator estimates (which will be sketched in more detail below), we are led to consideration of the frequency localized problem

$$\begin{cases} \Box u_k + 2[P_{<k-\kappa}\mathbf{P}_\alpha A, \partial^\alpha u_k] = f_k, \\ (u_k, \partial_t u_k)(0) = (u_{0,k}, u_{1,k}), \end{cases} \tag{9.1}$$

for each $k \in \mathbb{Z}$. By scaling, we may normalize $k = 0$.

Our goal is to construct a parametrix to (9.1). We summarize the main properties of the parametrix in this case, as well as the precise hypotheses on A_α that we need, in the following theorem.

Theorem 9.1 (Parametrix construction). *Let A_α be a $\mathfrak{g}$-valued 1-form on $I \times \mathbb{R}^4$ such that*

$$\|A\|_{S^1[I]} + \|\Box A\|_{\ell^1 X^{-\frac{1}{2}+b_1, -b_1}[I]} \leq M \tag{9.2}$$

for some $M > 0$ and $b_1 > \frac{1}{4}$. Let $\varepsilon > 0$. Assume that $\kappa > \kappa_1(\varepsilon, M)$ and

$$\|A\|_{DS^1[I]} + \|\Box A\|_{\ell^1 L^2 \dot{H}^{-\frac{1}{2}}} < \delta_p(\varepsilon, M, \kappa_1), \tag{9.3}$$

for some functions $\kappa_1(\varepsilon, M) \gg 1$, $0 < \delta_p(\varepsilon, M, \kappa_1) \ll 1$ independent of A_α. Moreover, assume that there exists $\tilde{A}_\alpha$ such that

$$\|\tilde{A}\|_{\underline{S}^1[I]} + \|(D\tilde{A}_0, D\mathbf{P}^\perp \tilde{A})\|_{Y[I]} \leq M, \tag{9.4}$$

$$\|\tilde{A}\|_{DS^1[I]} + \|(\tilde{A}_0, \mathbf{P}^\perp \tilde{A})\|_{L^2 \dot{H}^{\frac{3}{2}}[I]} < \delta_p(\varepsilon, M, \kappa_1), \tag{9.5}$$

and

$$\|\Delta A_0 - \mathbf{O}(\tilde{A}^\ell, \partial_0 \tilde{A}_\ell)\|_{\ell^1(\Delta L^1 L^\infty \cap L^2 \dot{H}^{-\frac{1}{2}})[I]} < \delta_p^2(\varepsilon, M, \kappa_1), \tag{9.6}$$

$$\|\Box \mathbf{P} A - \mathbf{P}\mathbf{O}(\tilde{A}^\ell, \partial_x \tilde{A}_\ell)) - \mathbf{P}\mathbf{O}'(\tilde{A}^\alpha, \partial_\alpha \tilde{A})\|_{\ell^1(L^1 L^2 \cap L^2 \dot{H}^{-\frac{1}{2}})[I]} < \delta_p^2(\varepsilon, M, \kappa_1), \tag{9.7}$$

where $\mathbf{O}(\cdot,\cdot)$ and $\mathbf{O}'(\cdot,\cdot)$ are disposable bilinear operators on $\mathbb{R}^4$. Then the following statements hold.

(1) Given any $(u_0, u_1) \in \dot{H}^1 \times L^2$ and $f \in N \cap L^2 \dot{H}^{-\frac{1}{2}}$ such that u_0, u_1, f are all frequency-localized in $\{C^{-1} \leq |\xi| \leq C\}$, there exists a $\mathfrak{g}$-valued function $u(t)$ on I which obeys

$$\|u\|_{S^1[I]} \lesssim_M \|(u_0, u_1)\|_{\dot{H}^1 \times L^2} + \|f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]}, \tag{9.8}$$

$$\|\Box u + 2[P_{<-\kappa}\mathbf{P}_\alpha A, \partial^\alpha u] - f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]} \leq \varepsilon \left(\|(u_0, u_1)\|_{\dot{H}^1 \times L^2} + \|f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]} \right), \tag{9.9}$$

$$\|u[0] - (u_0, u_1)\|_{\dot{H}^1 \times L^2} \leq \varepsilon \left(\|(u_0, u_1)\|_{\dot{H}^1 \times L^2} + \|f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]} \right). \tag{9.10}$$

Moreover, u is frequency-localized in $\{(2C)^{-1} \leq |\xi| \leq 2C\}$.

(2) Assume furthermore that

$$\|A_x\|_{\ell^\infty S^1[I]} + \|A_0\|_{\ell^\infty L^2 \dot{H}^{\frac{3}{2}}[I]} < \delta_o(M) \tag{9.11}$$

for some $\delta_o(M) \ll 1$ independent of A_α. Then the approximate solution u constructed above obeys (9.8) *with a universal constant, i.e.,*

$$\|u\|_{S^1[I]} \lesssim \|(u_0, u_1)\|_{\dot{H}^1 \times L^2} + \|f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]}. \tag{9.12}$$

In the remainder of this subsection, we sketch the proofs of Theorem 4.24 and Proposition 4.25 assuming Theorem 9.1. Then in the rest of this section, as well as in Sections 10 and 11, our goal will be to establish Theorem 9.1.

Lemma 9.2. *a) Let $A_{t,x}$ and $\tilde{A}_{t,x}$ be $\mathfrak{g}$-valued 1-forms on $I \times \mathbb{R}^4$, which satisfy* (9.2), (9.3), (9.4), (9.5), (9.6) *and* (9.7). *Then for $\varepsilon > 0$ sufficiently small (depending on M) and κ sufficiently large (depending on ε, M), given any $(u_0, u_1) \in \dot{H}^1 \times L^2$ and $f \in N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]$, there exists a unique solution $u \in S^1[I]$ to the IVP*

$$\begin{cases} (\Box + \mathrm{Diff}^\kappa_{\mathbf{P}A}) u = f, \\ u[0] = (u_0, u_1). \end{cases} \tag{9.13}$$

which obeys

$$\|u\|_{S^1[I]} \lesssim_M \|(u_0, u_1)\|_{\dot{H}^1 \times L^2} + \|f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]}. \tag{9.14}$$

b) If, in addition, $\|A\|_{\ell^\infty S^1[I]}$ obeys (9.11), *then the solution u constructed above obeys* (9.14) *with a universal constant, i.e.,*

$$\|u\|_{S^1[I]} \lesssim \|(u_0, u_1)\|_{\dot{H}^1 \times L^2} + \|f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]}. \tag{9.15}$$

Proof. Let u_k be the function given by (the rescaled) Theorem 9.1 which is determined by the initial data $(P_k u_0, P_k u_1, P_k f)$. We set

$$u_{app} = \sum_{k'} u_{k'}.$$

We claim that u is a good approximate solution to (9.13) in the sense that in any subinterval $J \subset I$ we have

$$\|u_{app}\|_{S^1[J]} \lesssim_M \|(u_0, u_1)\|_{\dot{H}^1 \times L^2} + \|f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[J]}, \tag{9.16}$$

$$\|u_{app}[0] - (u_0, u_1)\|_{\dot{H}^1 \times L^2} \lesssim \epsilon (\|(u_0, u_1)\|_{\dot{H}^1 \times L^2} + \|f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[J]}), \tag{9.17}$$

respectively

$$\begin{aligned} \|(\Box + \mathrm{Diff}^\kappa_{\mathbf{P}A}) u_{app} - f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[J]} \lesssim_M & \Big(\epsilon + 2^{-\delta_2 \kappa} + 2^{C\kappa} (\|\mathbf{P}A\|_{\ell^\infty DS^1[I]} + \|A_0\|_{\ell^\infty L^2 \dot{H}^{\frac{3}{2}}[J]} \Big) \\ & \Big(\|(u_0, u_1)\|_{\dot{H}^1 \times L^2} + \|f\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[J]} \Big) \end{aligned} \tag{9.18}$$

Assume that we have these bounds. Then the solution u to (9.13) is obtained as follows:

(i) We choose κ large enough so that $2^{-\delta_2} \ll_M 1$.

(ii) We divide the interval I into subintervals J_j so that

$$2^{C\kappa} \|\mathbf{P}A\|_{DS^1[I]} + \|A_0\|_{L^2 \dot{H}^{\frac{3}{2}}[I_j]} \ll_M 1$$

(iii) Within the interval J_1 we now have small errors for the approximate solution u_{app}; hence we can obtain an exact solution by reiterating.
(iv) We successively repeat the previous step on each of the subintervals I_j.

It remains to prove the bounds (9.16), (9.17) and (9.18). The first two follow directly from (9.8) and (9.9) for u_k after summation in k. We now consider (9.18), where we write

$$(\Box + \mathrm{Diff}^{\kappa}_{\mathbf{P}A})u - f = \sum_k (\Box u_k + 2[P_{<k-\kappa}\mathbf{P}A_\alpha, \partial^\alpha u_k] - P_k f) + \sum_k g_k$$

where

$$g_k = 2[P_{<k-\kappa}\mathbf{P}A_\alpha, \partial^\alpha u_k] - \sum_{k'} [P_{-k'-\kappa}\mathbf{P}A_\alpha, \partial^\alpha P_{k'} u_k]$$

The first sum is estimated directly via (9.9), so it remains to estimate g_k. We split

$$g_k = g_k^1 + g_k^2$$

where

$$g_k^1 = \sum_{k'=k+O(1)} P_{k'}[P_{-k'-\kappa}\mathbf{P}A_\alpha, \partial^\alpha P_{k'} u_k] - [P_{-k'-\kappa}\mathbf{P}A_\alpha, \partial^\alpha P_{k'} u_k]$$

and

$$g_k^2 = \sum_{k'=k+O(1)} [P_{[-k'-\kappa, k-\kappa)}\mathbf{P}A_\alpha, \partial^\alpha P_{k'} u_k]$$

Here g_k^1 has a commutator structure, so we can estimate it as in Proposition 4.30, yielding a $2^{-\delta_2 \kappa}$ factor. For the expression g_k^2, on the other hand, we can apply Proposition 4.20 to split it into a small part and a large part but which uses only divisible norms. Thus (9.18) follows, and the proof of the Lemma is concluded.

b) The same iterative construction applies, but no we no longer need to subdivide the interval as (9.11) insures that the divisible norms in (9.18) are actually small.

□

Proof of Theorem 4.24 assuming Theorem 9.1. We prove the theorem by repeatedly applying the lemma in successive intervals. To achieve this, we begin by choosing ϵ and κ depending only on M so that Lemma 9.2 holds. It remains to insure that we can divide the interval I into subintervals J_j where the conditions (9.2), (9.3), (9.4), (9.5), (9.6) and (9.7) hold.

We choose $\tilde{A} = A$. We carefully observe that we cannot use Theorem 5.1 here, as Theorem 4.24 is used in the proof of Theorem 5.1. However, we can use the weaker result in Proposition 5.4, which immediately gives. (9.2) and (9.4) from Theorem 5.1.

The remaining bounds are for divisible norms, so it suffices to establish them with a large constant depending on M; then we gain smallness by subdividing. Indeed, for (9.3) and (9.5) this still follows from Proposition 5.4.

For (9.6) we choose $\mathbf{O}(A, \partial_0 A) = [A, \partial_0 A]$. Then we can use (3.23) and (4.37). Finally for (9.7) we choose in addition $\mathbf{O}(A_\alpha, \partial^\alpha A) = -2[A_\alpha, \partial^\alpha A]$. Then by Theorem 9.1 we have

$$\Box A - \mathbf{O}(A, \partial_x A) - \mathbf{O}(A_\alpha, \partial^\alpha A) = R(A) + \mathrm{Rem}^3(A)A$$

and it suffices to use (3.21) and (4.74).

To conclude, we note that the second part of the lemma is proved as □

Proof of Proposition 4.25 assuming Theorem 9.1. We divide

$$A_{t,x} = A^{pert}_{t,x} + A^{nonpert}_{t,x}$$

where

$$A^{pert}_{t,x} = \sum_{k \in K} P_k A_{t,x}$$

with $|K| = O_{\delta_o(M)^{-1}M}(1)$ and

$$\|A^{nonpert}\|_{\ell^\infty S^1[I]} < \delta_o(M).$$

By Proposition 4.23, it follows that the contribution of any finite number of dyadic pieces of $A_{t,x}$ in $\mathrm{Diff}^\kappa_{\mathbf{P}A}$ is perturbative. More precisely, for A^{pert}, we have

$$\|\mathrm{Diff}^\kappa_{\mathbf{P}A^{pert}} B\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]} \lesssim_{|K|,M} \|B\|_{S^1[I]}. \tag{9.19}$$

Thus B solves also

$$(\Box + \mathrm{Diff}^\kappa_{\mathbf{P}A^{nonpert}}) B = \tilde{G},$$

where

$$\|\tilde{G}\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]} \lesssim_M \|G\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]} + \|B\|_{S^1[I]}$$

We now claim that Theorem 9.1 and thus Lemma 9.2 apply for $A^{nonpert}$. If that were true, then the conclusion of the proposition is achieved by subdividing the interval I into finitely many subintervals J_j, depending only on M, so that

(i) Lemma 9.2 applies in J_j
(ii) The size of the inhomogeneous term $\|\tilde{G}\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}}[I]}$ is small in J_j.

Indeed, to verify the hypothesis of Theorem 9.1 with A replaced by $A^{nonpert}$ it suffices to leave $\tilde{A} = A$, unchanged, but instead replace the operators $\mathbf{O}$ and $\mathbf{O}'$ by $(1 - \sum_{k \in K} P_k)\mathbf{O}$, respectively $(1 - \sum_{k \in K} P_k)\mathbf{O}'$, which are still disposable. □

9.2. **Extension and spacetime Fourier projections.** As in [11], our parametrix will be constructed by conjugating the usual Fourier representation formula for the $\pm$-half-wave equations by a *renormalization operator* $Op(Ad(O_\pm)_{<0})$; see (9.50). The renormalization operator is designed so that it cancels the most dangerous part of the paradifferential term $2[\mathbf{P}A_{\alpha,<-\kappa}, \partial^\alpha P_0 u]$ (Theorem 9.9), and furthermore enjoys nice mapping properties in functions spaces we use (Theorem 9.6).

9.2.1. *Extension to a global-in-time wave.* As in [11], our parametrix construction for (9.1) involves fine spacetime Fourier localization of $\mathbf{P}A$, which necessitates extension of $\mathbf{P}A$ outside I. Here we specify the extension procedure, and collect some of its properties that will be used later.

We extend $\mathbf{P}A$ by homogeneous waves outside I. By (9.2), this extension (still denoted by $\mathbf{P}A$) obeys the global-in-time bound

$$\|\mathbf{P}A\|_{S^1} + \|\Box \mathbf{P}A\|_{\ell^1 X^{-\frac{1}{2}+b_1,-b_1}} \lesssim M. \tag{9.20}$$

By Proposition 4.10, for any $p \geq 2$ note that

$$\|\chi^k_I P_k \mathbf{P}A\|_{L^p L^\infty} \lesssim \|P_k \mathbf{P}A\|_{L^p L^\infty[I]}. \tag{9.21}$$

Moreover, by (9.3), we have

$$\sum_k \|P_k \Box \mathbf{P} A\|_{L^2 \dot{H}^{-\frac{1}{2}}} = \|\Box \mathbf{P} A\|_{\ell^1 L^2 \dot{H}^{-\frac{1}{2}}[I]} < \delta_p. \tag{9.22}$$

Next, we specify the extension of A_0, and also of the relations (9.6) and (9.7) outside I. We first extend $\tilde{A}$ by homogeneous wave outside I and $\tilde{A}_0$ by zero outside I. These extensions (still denoted by $\tilde{A}$ and $\tilde{A}_0$, respectively) satisfy the global-in-time bound

$$\|\tilde{A}\|_{\underline{S}^1} + \|D\tilde{A}_0\|_Y \lesssim M. \tag{9.23}$$

In addition, we introduce the extension $\tilde{G}$ of $\mathbf{P}^\perp \tilde{A}$ by zero outside I. It obeys

$$\|D\tilde{G}\|_Y \lesssim M. \tag{9.24}$$

We emphasize that, in general, $\mathbf{P}^\perp \tilde{A}$ does *not* coincide with $\tilde{G}$ outside I.

Define $\tilde{R}_0$ and $\mathbf{P}\tilde{R}$ as

$$\begin{aligned} \tilde{R}_0(t) =& \Delta A_0(t) - \mathbf{O}(\tilde{A}^\ell(t), \partial_t \tilde{A}_\ell(t)) \quad \text{ for } t \in I, \\ \mathbf{P}\tilde{R}(t) =& \Box \mathbf{P} A(t) - \mathbf{P}\mathbf{O}(\tilde{A}^\ell(t), \partial_x \tilde{A}_\ell(t)) + \mathbf{P}\mathbf{O}'(\tilde{A}_\alpha, \partial^\alpha \tilde{A}) \quad \text{ for } t \in I, \end{aligned}$$

and 0 for $t \notin I$. By the hypotheses (9.6) and (9.7), we have

$$\|\tilde{R}_0\|_{\ell^1(\Delta L^1 L^\infty \cap L^2 \dot{H}^{-\frac{1}{2}})} < \delta_p^2, \tag{9.25}$$

$$\|\mathbf{P}\tilde{R}\|_{\ell^1(L^1 L^2 \cap L^2 \dot{H}^{-\frac{1}{2}})} < \delta_p^2. \tag{9.26}$$

We extend A_0 outside I by solving the equation

$$\Delta A_0 = \mathbf{O}(\chi_I \tilde{A}^\ell, \partial_t \tilde{A}_\ell) + \chi_I \tilde{R}_0. \tag{9.27}$$

By (8.17), (8.19), (9.5), (9.23) and (9.25), it follows that

$$\|DA_0\|_{\ell^1 Y} \lesssim M^2, \tag{9.28}$$

$$\|\Delta A_0\|_{\ell^1 L^2 \dot{H}^{-\frac{1}{2}}} \lesssim \delta_p^2. \tag{9.29}$$

Moreover, observe that the extension $\mathbf{P}A$ obeys the equation

$$\begin{aligned} \Box \mathbf{P} A =& \mathbf{P}\mathbf{O}(\chi_I \tilde{A}^\ell, \partial_x \tilde{A}_\ell) + \mathbf{P}\mathbf{O}'(\mathbf{P}_\ell \tilde{A}, \chi_I \partial^\ell \tilde{A}) \\ & - \mathbf{P}\mathbf{O}'(\tilde{A}_0, \chi_I \partial_t \tilde{A}) + \mathbf{P}\mathbf{O}'(\tilde{G}_\ell, \chi_I \partial^\ell \tilde{A}) + \chi_I \mathbf{P}\tilde{R}. \end{aligned} \tag{9.30}$$

9.2.2. *Spacetime Fourier projections.* Here we introduce the spacetime Fourier projections needed for definition of the renormalization operator. We denote by $(\tau, \xi) \in \mathbb{R} \times \mathbb{R}^4$ the Fourier variables for the input, and by $(\sigma, \eta) \in \mathbb{R} \times \mathbb{R}^4$ the Fourier variables for the symbol, which will be constructed from $\mathbf{P}A$. We remind the reader that our sign convention is such that the characteristic cone for a $\pm$-wave is $\{\tau \pm |\xi| = 0\}$.

Consider the following (overlapping) decomposition of $\mathbb{R}^{1+4}$, which is symmetric and homogeneous with respect to the origin:

$$\begin{aligned} D^{\omega,\pm}_{cone} =& \{\mathrm{sgn}(\sigma)(\sigma \pm \eta\cdot\omega) > \tfrac{1}{16}|\eta|^{-1}(|\eta_\perp|^2 + |\sigma \pm \eta\cdot\omega|^2)\} \\ &\cap \{\mathrm{sgn}(\sigma)(\sigma \pm \eta\cdot\omega) < \tfrac{4}{5}|\sigma|^{-1}(|\eta_\perp|^2 + |\sigma \pm \eta\cdot\omega|^2)\}, \\ D^{\omega,\pm}_{null} =& \{|\sigma \pm \eta\cdot\omega| < \tfrac{1}{8}|\eta|^{-1}(|\eta_\perp|^2 + |\sigma \pm \eta\cdot\omega|^2)\}, \\ D^{\omega,\pm}_{out} =& \{\mathrm{sgn}(\sigma)(\sigma \pm \eta\cdot\omega) < -\tfrac{1}{16}|\eta|^{-1}(|\eta_\perp|^2 + |\sigma \pm \eta\cdot\omega|^2)\} \\ &\cup \{\mathrm{sgn}(\sigma)(\sigma \pm \eta\cdot\omega) > \tfrac{2}{3}|\sigma|^{-1}(|\eta_\perp|^2 + |\sigma \pm \eta\cdot\omega|^2)\}. \end{aligned}$$

where $\eta_\perp = \eta - (\eta\cdot\omega)\omega$. See Figure 1 below for a plot of these domains.

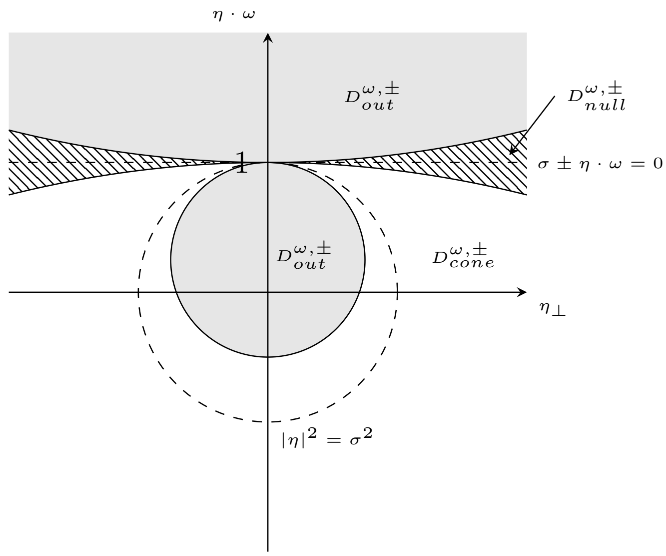

FIGURE 1. Caricature of $D^{\omega,\pm}_{cone}$, $D^{\omega,\pm}_{med}$ and $D^{\omega,\pm}_{out}$ in the hyperplane $\{\sigma = 1\}$ with $\pm = -$. Note that the actual domains are defined to be slightly overlapping.

We construct a smooth partition of unity adapted to the decomposition $D^{\omega,\pm}_{cone} \cup D^{\omega,\pm}_{null} \cup D^{\omega,\pm}_{out} = \mathbb{R}^{1+4}$ as follows. We begin with the preliminary definitions

$$\begin{aligned} \widetilde{\Pi}^{\omega,\pm}_{in}(\sigma,\eta) =& m_{>1}\left(\frac{4}{5}\frac{\sigma(\sigma \pm \eta\cdot\omega)}{(|\eta|^2 - (\eta\cdot\omega)^2) + |\sigma \pm \eta\cdot\omega|^2}\right), \\ \widetilde{\Pi}^{\omega,\pm}_{med}(\sigma,\eta) =& m_{>1}\left(8\frac{\mathrm{sgn}(\sigma)|\eta|(\sigma \pm \eta\cdot\omega)}{(|\eta|^2 - (\eta\cdot\omega)^2) + |\sigma \pm \eta\cdot\omega|^2}\right), \\ \widetilde{\Pi}^{\omega,\pm}_{out}(\sigma,\eta) =& m_{>1}\left(-8\frac{\mathrm{sgn}(\sigma)|\eta|(\sigma \pm \eta\cdot\omega)}{(|\eta|^2 - (\eta\cdot\omega)^2) + |\sigma \pm \eta\cdot\omega|^2}\right), \end{aligned}$$

where $m_{>1}(z) : \mathbb{R} \to [0,1]$ is a smooth cutoff to the region $\{z > 1\}$ (i.e., equals 1 there), which vanishes outside $\{z > \frac{5}{6}\}$. Then we define the symbols $\Pi^{\omega,\pm}_{cone}(\sigma,\eta)$, $\Pi^{\omega,\pm}_{null}(\sigma,\eta)$, $\Pi^{\omega,\pm}_{out}(\sigma,\eta)$ as follows:

$$\Pi^{\omega,\pm}_{cone}(\sigma,\eta) = \widetilde{\Pi}^{\omega,\pm}_{med}(\sigma,\eta) - \widetilde{\Pi}^{\omega,\pm}_{in}(\sigma,\eta), \tag{9.31}$$

$$\Pi^{\omega,\pm}_{null}(\sigma,\eta) = 1 - \widetilde{\Pi}^{\omega,\pm}_{med}(\sigma,\eta) - \widetilde{\Pi}^{\omega,\pm}_{out}(\sigma,\eta), \tag{9.32}$$

$$\Pi^{\omega,\pm}_{out}(\sigma,\eta) = \widetilde{\Pi}^{\omega,\pm}_{out}(\sigma,\eta) + \widetilde{\Pi}^{\omega,\pm}_{in}(\sigma,\eta). \tag{9.33}$$

Observe that $1 = \Pi^{\omega,\pm}_{cone} + \Pi^{\omega,\pm}_{null} + \Pi^{\omega,\pm}_{out}$, and $\operatorname{supp}\Pi^{\omega,\pm}_{*} \subseteq D^{\omega,\pm}_{*}$ for $* \in \{cone, null, out\}$. Moreover, by symmetry, $\Pi^{\omega,\pm}_{*}$ preserves the real-valued property.

We also make use of a dyadic angular decomposition with respect to ω. Given $\theta > 0$, we define the symbol

$$\Pi^{\omega,\pm}_{>\theta}(\sigma,\eta) = m_{>1}\left(\frac{|\angle(\omega, -\mathrm{sgn}(\sigma)|\eta)}{\theta}\right)$$

Furthermore, we define

$$\Pi^{\omega,\pm}_{\leq\theta}(\sigma,\eta) = 1 - \Pi^{\omega,\pm}_{>\theta}(\sigma,\eta), \qquad \Pi^{\omega,\pm}_{\theta}(\sigma,\eta) = (\Pi^{\omega,\pm}_{>\theta} - \Pi^{\omega,\pm}_{>\theta/2})(\sigma,\eta).$$

Since these symbols are real-valued and odd, the corresponding multipliers (which we simply denote by $\Pi^{\omega,\pm}_{>\theta}$, $\Pi^{\omega,\pm}_{\geq\theta}$ and $\Pi^{\omega,\pm}_{\theta}$, respectively) preserve the real-valued property.

The regularity of the symbols $\Pi^{\omega,\pm}_{cone}$, $\Pi^{\omega,\pm}_{null}$ and $\Pi^{\omega,\pm}_{out}$ degenerate as $|\eta_\perp| \to 0$; however, they are well-behaved when composed with $\Pi^{\omega,\pm}_{\theta} P_h$. The following lemma will play a basic role for our construction.

Lemma 9.3. *For any fixed $\pm$, $\omega \in \mathbb{S}^3$, $n \in \mathbb{N}$, $h \in 2^{\mathbb{R}}$ and $* \in \{cone, null, out\}$, the multiplier*[10] *$\theta^n \partial^{(n)}_{\xi}(\Pi^{\omega,\pm}_{*}\Pi^{\omega,\pm}_{\theta} P_h)$ is disposable.*

Proof. In this proof, we take $h = 0$ by scaling, and fix $\pm = +$. Let $* \in \{cone, null\}$.

We begin with some elementary reductions. First, since $1 = \Pi^{\omega,\pm}_{cone} + \Pi^{\omega,\pm}_{null} + \Pi^{\omega,\pm}_{out}$, and $\theta^n \partial^{(n)}_{\xi}\Pi^{\omega,\pm}_{\theta} P_0$ is disposable, it suffices to prove the lemma for just $\Pi^{\omega,\pm}_{cone}$ and $\Pi^{\omega,\pm}_{null}$. In this case, note that the symbol $\Pi^{\omega,\pm}_{*}\Pi^{\omega,\pm}_{\theta} m_h(\eta)$ (where m_h is the symbol of P_h) is compactly supported. Furthermore, the lemma is obvious if $\theta \gtrsim 1$, since then the symbol is smooth in ξ, σ, η on the unit scale. Therefore, we may assume that $\theta \ll 1$.

We now consider the case $n = 0$, when there is no ξ-differentiation. We fix $\omega \in \mathbb{S}^3$. To ease our computation, we introduce the null coordinate system $(\underline{\upsilon}, \upsilon, \tilde{\eta}_\perp)$, where

$$\underline{\upsilon} = \sigma - \eta\cdot\omega, \quad \upsilon = \sigma + \eta\cdot\omega,$$

and $\tilde{\eta}_\perp \in \mathbb{R}^3$ are the coordinates for the constant $\underline{\upsilon}, \upsilon$-spaces. Observe that

$$\frac{\sigma + \eta\cdot\omega}{|\eta_\perp|^2 + |\sigma + \eta\cdot\omega|^2} = \frac{\upsilon}{|\tilde{\eta}_\perp|^2 + \upsilon^2} \simeq 1, \quad |\eta_\perp| = |\tilde{\eta}_\perp| \simeq \theta, \quad |\upsilon| \simeq \theta^2, \quad |\underline{\upsilon}| \simeq 1 \tag{9.34}$$

on the support of $\Pi^{\omega,\pm}_{*}\Pi^{\omega,\pm}_{\theta} m_0$. Moreover, $\sigma = \sigma(\underline{\upsilon}, \upsilon, \tilde{\eta}_\perp)$ and $|\eta| = |\eta|(\underline{\upsilon}, \upsilon, \tilde{\eta}_\perp)$ are comparable to 1, and are also smooth on the unit scale on the support of $\Pi^{\omega,\pm}_{*}\Pi^{\omega,\pm}_{\theta} m_0$. Recalling the definition of $\Pi^{\omega,\pm}_{*}$, it can be computed from (9.34) that

$$|\partial^{\alpha}_{\underline{\upsilon}}\partial^{\beta}_{\upsilon}\partial^{\gamma}_{\tilde{\eta}_\perp}\Pi^{\omega,\pm}_{*}| \lesssim \theta^{-2|\beta|-|\gamma|} \quad \text{on } \operatorname{supp}\Pi^{\omega,\pm}_{*}\Pi^{\omega,\pm}_{\theta} m_0.$$

On the other hand,

$$|\partial^{\alpha}_{\underline{\upsilon}}\partial^{\beta}_{\upsilon}\partial^{\gamma}_{\tilde{\eta}_\perp}(\Pi^{\omega,\pm}_{\theta} m_0)| \lesssim \theta^{-|\gamma|} \quad \text{on } \operatorname{supp}\Pi^{\omega,\pm}_{*}\Pi^{\omega,\pm}_{\theta} m_0,$$

[10] We quantize $(\sigma,\eta) \mapsto (D_t, D_x)$.

so it follows that

$$|\partial_{\underline{\upsilon}}^{\alpha}\partial_{\upsilon}^{\beta}\partial_{\tilde{\eta}_\perp}^{\gamma}(\Pi_*^{\omega,\pm}\Pi_\theta^{\omega,\pm})| \lesssim \theta^{-2|\beta|-|\gamma|}. \tag{9.35}$$

Furthermore, from (9.34) we have

$$|\mathrm{supp}\, \Pi_*^{\omega,\pm}\Pi_\theta^{\omega,\pm} m_0| \lesssim \theta^5. \tag{9.36}$$

From these bounds, we see that the multiplier $\Pi_*^{\omega,\pm}\Pi_\theta^{\omega,\pm} P_0$ has a kernel with a universal bound on the mass, and thus is disposable.

Finally, we sketch the proof in the case $n \geq 1$. We first claim that

$$|\partial_\xi^{(n)}(\Pi_*^{\omega,\pm}\Pi_\theta^{\omega,\pm} m_0)| \lesssim \theta^{-n}. \tag{9.37}$$

Clearly $|\partial_\xi^{(n)}\Pi_\theta^{\omega,\pm}| \lesssim_n \theta^{-n}$, so it suffices to verify that $|\partial_\xi^{(n)}\Pi_*^{\omega,\pm}| \lesssim_n \theta^{-n}$ on the support of $\Pi_*^{\omega,\pm}\Pi_\theta^{\omega,\pm} m_0$. Note that

$$|\partial_\xi^{\alpha}(\eta\cdot\omega)| \lesssim_{|\alpha|} \begin{cases} \theta & |\alpha| = 1 \\ 1 & |\alpha| \geq 2 \end{cases} \qquad \text{on } \mathrm{supp}\, \Pi_*^{\omega,\pm}\Pi_\theta^{\omega,\pm} m_0. \tag{9.38}$$

Then recalling the definition of $\Pi_*^{\omega,\pm}$ and using the chain rule, the claim (9.37) follows. We remark that a differentiation in $\sigma + \eta\cdot\omega$ loses θ^{-2}, but we gain back a factor of θ through the chain rule and (9.38).

Next, we fix $\omega \in \mathbb{S}^3$ and start differentiating in $(\underline{\upsilon}, \upsilon, \tilde{\eta}_\perp)$. Using the chain rule, (9.38) and (9.34), it can be proved that

$$|\partial_{\underline{\upsilon}}^{\alpha}\partial_{\upsilon}^{\beta}\partial_{\tilde{\eta}_\perp}^{\gamma}\partial_\xi^{(n)}(\Pi_*^{\omega,\pm}\Pi_\theta^{\omega,\pm})| \lesssim \theta^{-2|\beta|-|\gamma|}\theta^{-n}. \tag{9.39}$$

We omit the details. Combined with (9.36), we see that $\theta^n\partial_\xi^{(n)}\Pi_*^{\omega,\pm}\Pi_\theta^{\omega,\pm} P_0$ is disposable. $\square$

As a corollary of the proof of Lemma 9.3, we obtain the following disposability statement.

Corollary 9.4. *For any fixed $\pm$, $\omega \in \mathbb{S}^3$, $h, k \in 2^{\mathbb{R}}$ and $* \in \{cone, null, out\}$, the translation-invariant bilinear operator on $\mathbb{R}^{1+4}$ with symbol*

$$\Pi_*^{|\xi|^{-1}\xi,\pm}\Pi_{2^\ell}^{|\xi|^{-1}\xi,\pm} P_h(\sigma,\eta) P_k P_\ell^\omega(\xi)$$

is disposable.

Clearly, the same corollary holds with any of the continuous Littlewood-Paley projections P_h, P_k replaced by the discrete analogue.

We also record a lemma which describes how the operator $\Box$ acts in the presence of $\Pi_{cone}^{\omega,\pm}\Pi_\theta^{\omega,\pm} P_h$.

Lemma 9.5. *For any fixed $\pm$, $\omega \in \mathbb{S}^3$, $n \in \mathbb{N}$ and $h \in 2^{\mathbb{R}}$, the multiplier*

$$(2^{-2h}\theta^{-2}\Box)\theta^n\partial_\xi^{(n)}(\Pi_{cone}^{\omega,\pm}\Pi_\theta^{\omega,\pm} P_h) \tag{9.40}$$

is disposable.

Proof. We set $h = 0$ by scaling. The symbol of $\Box$ is $-\sigma^2 + |\eta|^2$. For a fixed ω, we introduce the null coordinate system $(\underline{\upsilon}, \upsilon, \eta_\perp)$ as before. Then observe that

$$|\partial_{\underline{\upsilon}}^{\alpha}\partial_{\upsilon}^{\beta}\partial_{\tilde{\eta}_\perp}^{\gamma}(-\sigma^2 + |\eta|^2)| = |\partial_{\underline{\upsilon}}^{\alpha}\partial_{\upsilon}^{\beta}\partial_{\tilde{\eta}_\perp}^{\gamma}(-\underline{\upsilon}\upsilon + |\tilde{\eta}_\perp|^2)| \lesssim \theta^2\theta^{-2|\beta|-|\gamma|}$$

on the support of $\Pi_{cone}^{\omega,\pm}\Pi_\theta^{\omega,\pm} P_0$. The lemma follows by combining this bound with the proof of Lemma 9.3. $\square$

9.3. **Pseudodifferential renormalization operator.** In this subsection we define the pseudodifferential renormalization operator, and describe its main properties.

9.3.1. *Definition of the pseudodifferential renormalization operator.* As mentioned before, the aim for our renormalization operator is not to remove all of $\mathbf{P}A$, but only the most harmful (nonperturbative) part of it. This part is defined as

$$A^{main,\pm}_{j,<h} = \Pi^{\omega,\pm}_{\geq|\eta|^\delta} \Pi^{\omega,\pm}_{cone} P_{<h}(\mathbf{P}A)_j. \tag{9.41}$$

Precisely, given a direction ω, it selects the region which is both near the cone in a parabolic fashion near the direction ω, but also away from ω, on an angular scale that is slowly decreasing as the frequency η of A approaches 0. We emphasize that this decomposition depends on ω, which is what will make our renormalization operator a pseudodifferential operator.

To account for the fact that our gauge group is noncommutative, and also to better take advantage of previous work in this area, we divide the construction of the renormalization operator in two steps. The first step is microlocal but linear, and mirrors the renormalization construction in the (MKG) case, see [10] and also [17]. Precisely, we define the intermediate symbol

$$\Psi_{\pm,<h} = - L^\omega_\mp \Delta^{-1}_{\omega^\perp} A^{main,\pm}_{j,<h} \omega^j. \tag{9.42}$$

Here the operator $L^\omega_\mp \Delta^{-1}_{\omega^\perp}$ is chosen as a good approximate inverse for $L^\omega_\pm$, within the frequency localization region for $A^{main,\pm}_{j,<h}$. In effect this frequency localization region is chosen exactly so that this property holds within. This is based on the decomposition

$$-L^\omega_\pm L^\omega_\mp + \Delta_{\omega^\perp} = \Box.$$

which gives

$$L^\omega_\pm L^\omega_\mp \Delta^{-1}_{\omega^\perp} = 1 - \Box \Delta^{-1}_{\omega^\perp}$$

Given $A^{main,\pm}_{j,<h}$ and $\Psi_{\pm,<h}$ as above, we define their Littlewood-Paley pieces as

$$A^{main,\pm}_{j,h} = \frac{d}{dh} A^{main,\pm}_{j,<h}, \quad \Psi_{\pm,h} = \frac{d}{dh} \Psi_{\pm,<h}.$$

Now we come to the second step in the construction of the renormalization operator. This step is nonlinear but local, and is based on the construction of the renormalization operator in [23] for the corresponding wave map problem. Precisely, we solve the ODE

$$\frac{d}{dh} O_{<h,\pm} O^{-1}_{<h,\pm} = \Psi_{\pm,h} \tag{9.43}$$

$$\lim_{h\to-\infty} \|\partial_x O_{<h,\pm}(t,x,\xi)\|_{L^\infty} = 0.$$

Thus our renormalization is achieved via the paradifferential operator

$$Ad(O_\pm)_{<0}$$

where the localization to small frequencies is so that this operator preserves the unit dyadic frequency shell.

The parameter $\delta > 0$ is a universal constant, which is chosen below so that the parametrix construction go through. In particular, we take $0 < \delta < \frac{1}{100}$. Logically, it is fixed at the end of Section 10.

9.3.2. *Properties of the pseudodifferential renormalization operator.* Now we state the key properties satisfied by the renormalization operator $Ad(O_{\pm})_{<0}$ that we just defined; see Theorems 9.6 and 9.9. Proofs of these results are the subjects of Sections 10 and 11, respectively.

Theorem 9.6 (Mapping properties of the pseudodifferential renormalization operator)**.** *Let A be a Lie algebra-valued spatial 1-form on $I \times \mathbb{R}^4$ such that $A = P_{<-\kappa} A$ and*

$$\|\mathbf{P}A\|_{S^1[I]} \leq M_0.$$

for some $\kappa, M_0 > 0$. Let $\Psi_{\pm,<h}$, $\Psi_{\pm,h}$ and $O_{<h,\pm}$ be defined on $\mathbb{R}^{1+4}$ as above from the homogeneous-wave extension of $\mathbf{P}A$. Let Z be any of the spaces L^2_x, N or N^*.

(1) For $\kappa > 20$, the following bounds hold:
- *(Boundedness)*

$$\|Op(Ad(O_{\pm})_{<0})(t,x,D)P_0\|_{Z\to Z} \lesssim_{M_0} 1, \tag{9.44}$$

- *(Dispersive estimates)*

$$\|Op(Ad(O_{\pm})_{<0})(t,x,D)P_0\|_{S_0^\sharp \to S_0} \lesssim_{M_0} 1. \tag{9.45}$$

(2) For any $\varepsilon > 0$, there exist $\kappa_0(\varepsilon, M_0) \gg 1$ (independent of A_x) such that if $\kappa > \kappa_0(\varepsilon, M_0)$, then
- *(Derivative bounds)*

$$\|[\partial_t, Op(Ad(O_{\pm})_{<0})(t,x,D)]P_0\|_{Z\to Z} \lesssim \varepsilon, \tag{9.46}$$

- *(Approximate unitarity)*

$$\|(Op(Ad(O_{\pm})_{<0})(t,x,D)Op(Ad(O_{\pm}^{-1})_{<0})(D,s,y) - I)P_0\|_{Z\to Z} \lesssim \varepsilon, \tag{9.47}$$

where the implicit constants are universal.

(3) There exists $0 < \delta_o(M_0) \ll 1$ (independent of A_x) such that if, in addition to the above hypothesis,

$$\|\mathbf{P}A_x\|_{\ell^\infty S^1[I]} < \delta_o(M_0), \tag{9.48}$$

then (9.44) and (9.45) hold with universal constants. That is, for $\kappa > 20$ we have
- *(Boundedness with a universal constant)*

$$\|Op(Ad(O_{\pm})_{<0})(t,x,D)P_0\|_{Z\to Z} \lesssim 1, \tag{9.44$'$}$$

- *(Dispersive estimates with a universal constant)*

$$\|Op(Ad(O_{\pm})_{<0})(t,x,D)P_0\|_{S_0^\sharp \to S_0} \lesssim 1. \tag{9.45$'$}$$

Here the frequency localization operator P_0 can easily be replaced by a more general localization to $\{|\xi| \simeq 1\}$.

Remark 9.7. As we will see in the proof below, $\kappa_0(\varepsilon, M_0) \simeq_\varepsilon \log M_0$ and $\delta_o(M_0) \ll_{M_0} 1$.

Remark 9.8. Note that the symbol of each of the above PDOs is independent of $\tau = \xi_0$, and thus it defines a PDO on $\mathbb{R}^4$ for each fixed t. By the mapping property $Z \to Z$ with $Z = L^2_x$, we mean that the PDO maps $L^2_x \to L^2_x$ for each fixed t, with a constant uniform in t.

Theorem 9.9 (Renormalization error)**.** *Let A_α be a $\mathfrak{g}$-valued 1-form on $I \times \mathbb{R}^4$ such that $A_\alpha = P_{<-\kappa} A_\alpha$ and $\|\mathbf{P} A_x\|_{S^1[I]} \leq M$ for some $\kappa, M > 0$. Let $\varepsilon > 0$. Assume that $\kappa > \kappa_1(\varepsilon, M)$ and (9.3)–(9.7) hold for some functions $\kappa_1(\varepsilon, M) \gg 1$ and $0 < \delta_p(\varepsilon, M, \kappa_1) \ll 1$ independent of A_α (to be specified below). Let $\Psi_{\pm,<h}$, $\Psi_{\pm,h}$ and $O_{<h,\pm}$ be defined as above from the homogeneous-wave extension of $\mathbf{P} A_x$. Then we have*

$$\|(\Box^p_{\mathbf{P}A} Op(Ad(O_\pm)_{<0}) - Op(Ad(O_\pm)_{<0})\Box) P_0\|_{S^\sharp_{0,\pm}[I] \to N_{0,\pm}[I]} < \varepsilon. \tag{9.49}$$

Remark 9.10. As we will see later, $\kappa_1(\varepsilon, M) \simeq_\varepsilon \log M$ and $\delta_p(\varepsilon, M, \kappa_1) \ll_{M,\kappa_1} \varepsilon$.

9.4. **Definition of the parametrix and proof of Theorem 9.1.** Our parametrix is given by:

$$\begin{aligned} u(t) = \sum_\pm \Big(&\frac{1}{2} Op(Ad(O_\pm)_{<0})(t,x,D) e^{\pm it|D|} Op(Ad(O^1_\pm)_{<0})(D,0,y)(u_0 \pm i|D|^{-1} u_1) \\ &+ Op(Ad(O_\pm)_{<0})(t,x,D) \frac{1}{|D|} K^\pm Op(Ad(O^{-1}_\pm)_{<0})(D,s,y) f \Big) \end{aligned} \tag{9.50}$$

where

$$K^\pm g(t) = \int_0^t e^{\pm i(t-s)|D|} g(s)\, ds.$$

With this definition, the proof of Theorem 9.1 starting from Theorems 9.6, 9.9 is essentially identical to the corresponding proof in [17], and is omitted.

10. Mapping properties of the renormalization operator

10.1. **Fixed-time pointwise bounds for the symbols Ψ and O.** Here we state fixed-time pointwise bounds for Ψ and O. We borrow these estimates from [11], while carefully noting dependence of constants on the frequency envelope of $A = A_x$ in S^1. The bounds below are stated using continuous Littlewood-Paley projections P_h, but we note that the same bounds hold for discrete Littlewood-Paley projections as well.

We begin with pointwise bounds for the $\mathfrak{g}$-valued symbol $\Psi_{h,\pm}(t, x, \xi)$.

Lemma 10.1. *The following bounds hold.*

(1) For $m \geq 0$ and $0 \leq n < \delta^{-1}$, we have

$$|\partial^{(n)}_\xi \partial^{(m-1)}_x \nabla \Psi^{(\theta)}_{\pm,h}(t,x,\xi)| \lesssim 2^{mh} \theta^{\frac{1}{2}-n} \|A_h\|_{S^1}. \tag{10.1}$$

When $m = 0$, we interpret the expression on the LHS as $\partial^n_\xi \Psi^{(\theta)}_{\pm,h}$.

(2) Let $\langle t-s, x-y\rangle^2 = 1 + |t-s|^2 + |x-y|^2$. We have

$$|\Psi_{\pm,h}(t,x,\xi) - \Psi_{\pm,h}(s,y,\xi)| \lesssim \min\{2^h \langle t-s, x-y\rangle, 1\} \|A_h\|_{S^1}. \tag{10.2}$$

(3) Finally, for $1 \leq n < \delta^{-1}$ we have

$$|\partial^{(n)}_\xi (\Psi_{\pm,h}(t,x,\xi) - \Psi_{\pm,h}(s,y,\xi))| \lesssim \min\{2^h \langle t-s, x-y\rangle, 1\} 2^{-(n-\frac{1}{2})\delta h} \|A_h\|_{S^1}. \tag{10.3}$$

For a proof, we refer to [11, Section 7.3]. As a corollary of (10.1) we have

$$|\nabla \Psi_{\pm,h}| \lesssim 2^h \|A_h\|_{S^1} \tag{10.4}$$

Next, we consider the $\mathbf{G}$-valued symbol $O_{<h,\pm}$.

Lemma 10.2. *Let c_h be an admissible frequency envelope for A in S^1. Then the following bounds hold.*

(1) For $0 \le n < \delta^{-1}$, we have

$$|\partial_\xi^{(n)}(O_{<h,\pm})_{;t,x}(t,x,\xi)| \lesssim_{\|A\|_{S^1}} 2^{(1-n\delta)h} c_h \tag{10.5}$$

(2) We have

$$d(O_{<h,\pm}(t,x,\xi)O^{-1}_{<h,\pm}(s,y,\xi), Id) \lesssim_{\|A\|_{S^1}} \log(1+2^h\langle t-s, x-y\rangle)c_h. \tag{10.6}$$

(3) Finally, for $1 \le n < \delta^{-1}$, we have

$$\begin{aligned}&|\partial_\xi^{(n-1)}(O_{<h,\pm}(t,x,\xi)O^{-1}_{<h,\pm}(s,y,\xi))_{;\xi}| \\ &\quad\lesssim_{\|A\|_{S^1}} \min\{2^h\langle t-s, x-y\rangle, 1\}^{1-(n-\frac{1}{2})\delta}(1+\langle t-s,x-y\rangle)^{(n-\frac{1}{2})\delta} c_h.\end{aligned} \tag{10.7}$$

For a proof, we refer to [11, Section 7.7].

10.2. **Decomposability calculus.** To handle symbol multiplications, we use the decomposability calculus introduced in [22, 9], which allows us to roughly regard these operations as multiplication by a function in L^pL^q. In the present work, we need an interval-localized version in order to exploit small divisible norms.

Given $\theta \in 2^{-\mathbb{N}}$, consider a covering of the unit sphere $\mathbb{S}^3 = \{\omega \in \mathbb{R}^4 : |\xi| = 1\}$ by solid angular caps of the form $\{\omega \in \mathbb{S}^3 : |\phi - \omega| < \theta\}$ with uniformly finite overlaps. We index these caps by their centers $\phi \in \mathbb{S}^3$, and denote by $\{(m_\theta^\phi)^2(\omega)\}$ the associated nonnegative smooth partition of unity on $\mathbb{S}^3$.

Let I be an interval. Consider a $\mathrm{End}(\mathfrak{g})$-valued symbol $c(t,x,\xi)$ on $I_t \times \mathbb{R}^4_x \times \mathbb{R}^4_\xi$, which is zero homogeneous in ξ, i.e., depends only on the angular variable $\omega = \frac{\xi}{|\xi|}$. We say that $c(t,x,\xi)$ is *decomposable in $L^qL^r[I]$* if $c = \sum_\theta c^{(\theta)}$, $\theta \in 2^{-\mathbb{N}}$ and

$$\sum_\theta \|c^{(\theta)}\|_{D_\theta L^q L^r[I]} < \infty, \tag{10.8}$$

where

$$\|c^{(\theta)}\|_{D_\theta L^q L^r[I]} = \|\Big(\sum_{n=0}^{40}\sum_\phi \sup_\omega \big(m_\theta^\phi(\omega)\|\theta^n\partial_\xi^{(n)}c^{(\theta)}\|_{L^r_x}\big)^2\Big)^{\frac{1}{2}}\|_{L^q_t[I]}. \tag{10.9}$$

We define $\|c\|_{DL^qL^r[I]}$ to be the infimum of (10.8) over all possible decompositions $c = \sum_\theta c^{(\theta)}$. In what follows, we will use the convention of omitting $[I]$ when $I = \mathbb{R}$.

In the following lemma, we collect some basic properties of the symbol class $DL^qL^r[I]$.

Lemma 10.3. *(1) For any two intervals such that $I \subset I'$, we have*

$$\|c\|_{DL^qL^r[I]} \le \|c\|_{DL^qL^r[I']}.$$

(2) For any symbols $c \in DL^{q_1}L^{r_1}[I]$ and $d \in DL^{q_2}L^{r_2}[I]$, its product obeys the Hölder-type bound

$$\|cd\|_{DL^q_tL^r_x[I]} \lesssim \|c\|_{DL^{q_1}L^{r_1}[I]}\|d\|_{DL^{q_2}L^{r_2}[I]}$$

where $1 \le q_1, q_2, q, r_1, r_2, r \le \infty$, $\frac{1}{q_1} + \frac{1}{q_2} = \frac{1}{q}$ and $\frac{1}{r_1} + \frac{1}{r_2} = \frac{1}{r}$.

(3) Let $a(t,x,\xi)$ be a $\mathrm{End}(\mathfrak{g})$-valued smooth symbol on $I \times \mathbb{R}^4_x \times \mathbb{R}^4_\xi$ whose left quantization $Op(a)$ satisfies the fixed-time bound

$$\sup_{t\in I} \|Op(a)(t,x,D)\|_{L^2\to L^2} \leq C_a.$$

Then for any symbol $c \in DL^qL^r$, we have the spacetime bound

$$\|Op(ac)(t,x,D)\|_{L^{q_1}L^2[I]\to L^{q_2}L^{r_2}[I]} \lesssim C_a\|c\|_{DL^qL^r[I]}$$

where $1 \leq q_1, q_2, q, r_2, r \leq \infty$, $\frac{1}{q_1}+\frac{1}{q}=\frac{1}{q_2}$ and $\frac{1}{2}+\frac{1}{r}=\frac{1}{r_2}$. An analogous statement holds in the case of right quantization.

The proof is essentially the same as the global-in-time versions in [9, Chapter 10] and [10, Lemma 7.1]; we omit the details.

10.3. **Decomposability bounds for A, Ψ and O.** Here we collect some decomposability bounds for A, Ψ and O that we will use in our proof of Theorems 9.6 and 9.9. As before, we state the bounds using continuous Littlewood-Paley projections P_h, but note that the same bounds hold for discrete Littlewood-Paley projections as well. For simplicity of notation, we will usually write $\|G\|_{DL^qL^r} = \|ad(G)\|_{DL^qL^r}$ for a $\mathfrak{g}$-valued symbol G, respectively $\|O\|_{DL^qL^r} = \|Ad(O)\|_{DL^qL^r}$ for a $\mathbf{G}$-valued symbol O.

For any $\theta > 0$, $h \in \mathbb{R}$ and $* \in \{cone, null, out\}$, recall the definition

$$A^{(\theta)}_{\alpha,h,*,\pm} = P_h \Pi^{\omega,\pm}_* \Pi^{\omega,\pm}_\theta (\mathbf{P}A)_\alpha.$$

As before, we will often omit the subscript x for simplicity, and write $A^{(\theta)}_{h,*,\pm} = A^{(\theta)}_{x,h,*,\pm}$ etc.

These symbols obey the following global-in-time decomposability bounds:

Lemma 10.4. *For $q \geq 2$ and $* \in \{cone, null, out\}$, we have*

$$\|A^{(\theta)}_{h,*,\pm}\cdot\omega\|_{DL^qL^\infty} \lesssim 2^{(1-\frac{1}{q})h}\theta^{\frac{5}{2}-\frac{2}{q}}\|A_h\|_{S^1}, \tag{10.10}$$

$$\|A^{(\theta)}_{0,h,*,\pm}\|_{DL^qL^\infty} \lesssim 2^{(1-\frac{1}{q})h}\theta^{\frac{5}{2}-\frac{2}{q}}\|A_{0,h}\|_{Y^1}. \tag{10.11}$$

Furthermore, for $* = cone$ we have

$$\|\Box A^{(\theta)}_{h,cone,\pm}\cdot\omega\|_{DL^qL^\infty} \lesssim 2^{(3-\frac{1}{p})h}\theta^{\frac{9}{2}-\frac{2}{q}}\|A_h\|_{S^1}, \tag{10.12}$$

$$\|\Delta^{-1}_{\omega^\perp}\Box A^{(\theta)}_{h,cone,\pm}\cdot\omega\|_{DL^qL^\infty} \lesssim 2^{(1-\frac{1}{p})h}\theta^{\frac{5}{2}-\frac{2}{q}}\|A_h\|_{S^1}. \tag{10.13}$$

Proof. The symbols $(\theta\partial_\omega)^n(\Pi^{\omega,\pm}_*\Pi^{\omega,\pm}_\theta)$ are smooth, homogeneous and uniformly bounded, and the corresponding multipliers are disposable for fixed Ω. Then the bounds (10.10) and (10.11) follow by Bernstein's inequality using the Strichartz component of the S^1 norm, respectively the $L^2\dot{H}^{\frac{1}{2}}$ component of the ∇Y^1 norm.

For the bounds (10.12) and (10.13) we need in addition to consider the size of the symbol of $\Box$, respectively $\Delta^{-1}_{\omega^\perp}$ within the support of $P_h\Pi^{\omega,\pm}_{cone}\Pi^{\omega,\pm}_\theta$. This is $\theta^2 2^{2h}$, respectively $\theta^{-2}2^{-2h}$. Precisely, we have the representations

$$\Box P_h \Pi^{\omega,\pm}_{cone}\Pi^{\omega,\pm}_\theta = \theta^2 2^{2h}\mathbf{O}\Pi^{\omega,\pm}_{cone}\Pi^{\omega,\pm}_\theta, \qquad \Delta^{-1}_{\omega^\perp}P_h\Pi^{\omega,\pm}_*\Pi^{\omega,\pm}_\theta = \theta^{-2}2^{-2h}\mathbf{O}\Pi^{\omega,\pm}_{cone}\Pi^{\omega,\pm}_\theta$$

with $\mathbf{O}$ disposable, see e.g. Lemma 9.5. Then (10.12) and (10.13) immediately follow from (10.10). □

Next, we consider the phase $\Psi_{\pm}$, which was defined in (9.42). Given $\theta > 0$ and $h \in \mathbb{R}$, let

$$\Psi^{(\theta)}_{h,\pm} = P_h \Pi^{\omega,\pm}_{\theta} \Psi_{\pm}.$$

We have the following global-in-time decomposability bounds.

Lemma 10.5. *For $q, r \geq 2$ and $\frac{2}{q} + \frac{3}{r} \leq \frac{3}{2}$, we have*

$$\|(\Psi^{(\theta)}_{h,\pm}, 2^{-h}\nabla\Psi^{(\theta)}_{h,\pm})\|_{DL^q L^r} \lesssim 2^{-(\frac{1}{q}+\frac{4}{r})h} \theta^{\frac{1}{2}-\frac{2}{q}-\frac{3}{r}} \|A_h\|_{S^1}. \tag{10.14}$$

In addition, suppose that $\theta \lesssim 2^a$ for some $a \in -\mathbb{N}$. Then for $q, r \geq 2$, we also have

$$\|Q_{h+2a}(\Psi^{(\theta)}_{h,\pm}, 2^{-h}\nabla\Psi^{(\theta)}_{h,\pm})\|_{DL^q L^r} \lesssim 2^{-(\frac{1}{q}+\frac{4}{r})h} 2^{-\frac{2}{q}a} \theta^{\frac{1}{2}-\frac{3}{r}} \|A_h\|_{S^1}. \tag{10.15}$$

Furthermore,

$$\|\Box\Psi^{(\theta)}_{h,\pm}\|_{DL^2 L^\infty} \lesssim \theta^{\frac{3}{2}} 2^{\frac{3}{2}h} \|A_h\|_{S^1}. \tag{10.16}$$

Proof. Observing that within the support of $P_h \Pi^{\omega,\pm}_{cone} \Pi^{\omega,\pm}_{\theta}$ the symbol $L^{\mp}\Delta^{-1}_{\omega^\perp}$ has the form $2^{-h}\theta^{-2}\mathbf{O}$ with $\mathbf{O}$ disposable and depending smoothly on ω on the θ scale, the first bound (10.14) is again a direct consequence of the Strichartz bounds in the S^1 norm for A.

For (10.15) it suffices to prove the case $p = q = 2$ and then use Bernstein's inequality. But in this case it suffices to use the $X^{1,\frac{1}{2}}_{\infty}$ component of the S^1 norm at fixed modulation.

For the last bound (10.16) it suffices to combine the $L^2 L^\infty$ case of (10.14) with Lemma 9.5. ∎

We now consider the $\mathbf{G}$-valued symbol $O_{<h,\pm}$, which was defined in (9.43). It obeys the following global-in-time decomposability bounds.

Lemma 10.6. *Let c_h be an admissible frequency envelope for A in S^1. Then for any $q > 4$, we have*

$$\|(O_{<h,\pm;x}, O_{<h,\pm;t})\|_{DL^q L^\infty} \lesssim_{\|A\|_{S^1}} 2^{(1-\frac{1}{q})h} c_h. \tag{10.17}$$

When $q = 2$, an analogous bound with a slight loss holds:

$$\|(O_{<h,\pm;x}, O_{<h,\pm;t})\|_{DL^2 L^\infty} \lesssim_{\|A\|_{S^1}} 2^{\frac{1}{2}(1-\delta)h} c_h. \tag{10.18}$$

Proof. These bounds are a consequence of the $\Psi^{(\theta)}_{h,\pm}$ bounds in the previous lemma. The proof is similar to the proof of the similar result in [11, Lemma 7.9] and is omitted. We note that the constraint $q > 4$ in the first bound is to prevent losses in the θ summation in (10.14). ∎

Finally, we consider *interval-localized* decomposability bounds, which will be needed to exploit divisibility (i.e., the hypothesis (9.3)) to gain smallness.

Lemma 10.7. *Let $|I| \geq 2^{-h-\kappa}$, where $h \in \mathbb{R}$ and $\kappa \geq 0$. For $q \geq 2$, we have*

$$\|\Psi^{(\theta)}_{h}\|_{DL^q L^\infty[I]} \lesssim 2^{C\kappa}\theta^{-C} 2^{-h} \|A_h\|_{L^q L^\infty[I]}, \tag{10.19}$$

$$\|\Delta^{-1}_{\omega^\perp}\Box(\omega \cdot A^{(\theta)}_{h,cone,\pm})\|_{DL^q L^\infty[I]} \lesssim 2^{C\kappa}\theta^{-C} \|A_h\|_{L^q L^\infty[I]}. \tag{10.20}$$

$$\|\omega \cdot A^{(\theta)}_{h}\|_{DL^q L^\infty[I]} \lesssim 2^{C\kappa}\theta^{-C} \|A_h\|_{L^q L^\infty[I]}. \tag{10.21}$$

$$\|\omega \cdot A^{(\theta)}_{0,h}\|_{DL^q L^\infty[I]} \lesssim 2^{C\kappa}\theta^{-C} \|A_{0,h}\|_{L^q L^\infty[I]}. \tag{10.22}$$

Proof. We will prove (10.19), and leave the similar cases of (10.20), (10.21), (10.22) to the reader.

By scaling, we set $h = 0$. By the definition of the class $DL^qL^\infty[I]$, we have

$$\begin{aligned}\|\Psi_0^{(\theta)}\|_{DL^qL^\infty[I]} \lesssim & \theta^{-2}\Big(\sum_{n=0}^{40}\sum_{\phi}\sup_{\omega}\|m_\theta^\phi(\omega)\theta^n\partial_\xi^{(n)}\Pi_\theta^\omega\Pi_{cone}^\omega P_0(\omega\cdot\mathbf{P}A)\|_{L^qL^\infty[I]}^2\Big)^{\frac{1}{2}}\\ \lesssim & \theta^{-C}\sum_{n=0}^{40}\|\theta^n\partial_\xi^{(n)}\Pi_\theta^\omega\Pi_{cone}^\omega P_0(\omega\cdot\mathbf{P}A)\|_{L^qL^\infty[I]}.\end{aligned}$$

Fix $n \in [1,40]$ and $\omega \in \mathbb{S}^3$. From the proof of Lemma 9.3, we see that the projection $\theta^{n'}\partial_\xi^{(n')}\Pi_\theta^\omega\Pi_{cone}^\omega P_0$, when viewed as a Fourier multiplier in (σ,η), has a symbol which is supported in a spacetime cube of radius $\lesssim 1$, and its derivatives (up to 40, say) are bounded by θ^{-C} for some large universal constant C. Moreover, we have $|\theta^{n''}\partial_\xi^{(n'')}\omega| \lesssim_{n''} 1$. Denoting by χ_I^0 a generalized cutoff adapted at the unit scale as in (4.22), we have

$$\|\theta^n\partial_\xi^{(n)}\Pi_\theta^\omega\Pi_{cone}^\omega P_0(\omega\cdot\mathbf{P}A)\|_{L^qL^\infty[I]} \lesssim \theta^{-C}\|\chi_I^0 P_0 A\|_{L^qL^\infty}$$

Recall that A is extended outside I by homogeneous waves. By Proposition 4.10, the last expression is bounded by

$$\lesssim 2^{C\kappa}\theta^{-C}\|P_0A\|_{L^qL^\infty[I]},$$

which proves (10.19). $\square$

10.4. **Collection of symbol bounds.** Before we continue, we introduce the quantity M_σ, which collects various symbol bounds that we have so far.

We fix large enough N and a small universal constant $\delta_\sigma > 0$. Then we let $M_\sigma > 0$ be the minimal constant such that:

- The following pointwise bounds hold for all $0 \le n \le \delta^{-1}$ and $0 \le m \le N$:

$$\begin{aligned}|\partial_\xi^{(n)}\partial_x^{(m-1)}\nabla\Psi_{\pm,h}^{(\theta)}| \le & 2^{mh}\theta^{\frac{1}{2}-n}M_\sigma,\\ |\Psi_{\pm,h}(t,x,\xi)-\Psi_{\pm,h}(s,y,\xi)| \le & \min\{2^h\langle t-s,x-y\rangle,1\}M_\sigma,\\ |\partial_\xi^{(n)}(\Psi_{\pm,h}(t,x,\xi)-\Psi_{\pm,h}(s,y,\xi))| \le & \min\{2^h\langle t-s,x-y\rangle,1\}2^{-(n-\frac{1}{2})\delta h}M_\sigma,\\ |\partial_\xi^{(n)}(O_{<h,\pm})_{;t,x}(t,x,\xi)| \le & 2^{(1-n\delta)h}M_\sigma,\\ d(O_{<h,\pm}(t,x,\xi)O_{<h,\pm}^{-1}(s,y,\xi),Id) \le & \log(1+2^h\langle t-s,x-y\rangle)M_\sigma,\\ |\partial_\xi^{(n-1)}(O_{<h,\pm}(t,x,\xi)O_{<h,\pm}^{-1}(s,y,\xi))_{;\xi}| \le & \min\{2^h\langle t-s,x-y\rangle,1\}^{1-(n-\frac{1}{2})\delta}\\ & \times(1+\langle t-s,x-y\rangle)^{(n-\frac{1}{2})\delta}M_\sigma.\end{aligned}$$

- The following decomposability bounds hold for all $* \in \{cone, null, out\}$, $q, r \geq 2$ and $\frac{2}{q} + \frac{3}{r} \leq \frac{3}{2}$:

$$\begin{aligned}
\|A^{(\theta)}_{h,*,\pm} \cdot \omega\|_{DL^q L^\infty} \leq& 2^{(1-\frac{1}{q})h} \theta^{\frac{5}{2}-\frac{2}{q}} M_\sigma, \\
\|A^{(\theta)}_{0,h,*,\pm}\|_{DL^q L^\infty} \leq& 2^{(1-\frac{1}{q})h} \theta^{\frac{5}{2}-\frac{2}{q}} M_\sigma, \\
\|\Box A^{(\theta)}_{h,cone,\pm} \cdot \omega\|_{DL^q L^\infty} \leq& 2^{(3-\frac{1}{p})h} \theta^{\frac{9}{2}-\frac{2}{q}} M_\sigma, \\
\|\Delta^{-1}_{\omega^\perp} \Box A^{(\theta)}_{h,cone,\pm} \cdot \omega\|_{DL^q L^\infty} \leq& 2^{(1-\frac{1}{p})h} \theta^{\frac{5}{2}-\frac{2}{q}} M_\sigma, \\
\|(\Psi^{(\theta)}_{h,\pm}, 2^{-h} \nabla \Psi^{(\theta)}_{h,\pm})\|_{DL^q L^r} \leq& 2^{-(\frac{1}{q}+\frac{4}{r})h} \theta^{\frac{1}{2}-\frac{2}{q}-\frac{3}{r}} M_\sigma, \\
\|Q_{h+2a}(\Psi^{(\theta)}_{h,\pm}, 2^{-h} \nabla \Psi^{(\theta)}_{h,\pm})\|_{DL^q L^r} \leq& 2^{-(\frac{1}{q}+\frac{4}{r})h} 2^{-\frac{2}{q}a} \theta^{\frac{1}{2}-\frac{3}{r}} M_\sigma, \qquad (\theta \lesssim 2^a \lesssim 1) \\
\|\Box \Psi^{(\theta)}_{h,\pm}\|_{DL^2 L^\infty} \leq& \theta^{\frac{3}{2}} 2^{\frac{3}{2}h} M_\sigma, \\
\|(O_{<h,\pm;x}, O_{<h,\pm;t})\|_{DL^q L^\infty} \leq& 2^{(1-\frac{1}{q})h} M_\sigma, \qquad (q \geq 4 + \delta_\sigma) \\
\|(O_{<h,\pm;x}, O_{<h,\pm;t})\|_{DL^2 L^\infty} \leq& 2^{\frac{1}{2}(1-\delta)h} M_\sigma.
\end{aligned}$$

By the preceding results, there exists a M_σ such that

$$M_\sigma \lesssim_M \|A\|_{\ell^\infty S^1} + \|A_0\|_{\ell^\infty Y^1}. \tag{10.23}$$

In particular, note that all of the above symbol bounds are small if $\|A\|_{\ell^\infty S^1}$ and $\|A_0\|_{\ell^\infty Y^1}$ are.

10.5. **Oscillatory integral bounds.** Given a smooth function a, let

$$K^a_{<0}(t,x;s,y) = \int Ad(O_{<h,\pm})_{<0}(t,x,\xi) a(\xi) e^{\pm i(t-s)|\xi|} e^{i\xi\cdot(x-y)} Ad(O^{-1}_{<h,\pm})_{<0}(\xi,y,s) \frac{d\xi}{(2\pi)^4}.$$

Lemma 10.8. *For a sufficiently small universal constant $\delta > 0$, the following bounds hold for the kernel $K^a_{<0}(t,x;s,y)$.*

(1) Assume that a is a smooth bump function on the unit scale. Then

$$|K^a_{<0}(t,x;s,y)| \lesssim_{M_\sigma} \langle t-s\rangle^{-\frac{3}{2}} \langle |t-s| - |x-y|\rangle^{-100}. \tag{10.24}$$

(2) Let $a = a_{\mathcal{C}}$ be a smooth bump function on a radially oriented rectangular box $\mathcal{C}$ of size $2^k \times (2^{k+\ell})^3$, where $k, \ell \leq 0$. Then

$$|K^a_{<0}(t,x;s,y)| \lesssim_{M_\sigma} 2^{4k+3\ell} \langle 2^{2(k+\ell)}(t-s)\rangle^{-\frac{3}{2}} \langle 2^k(|t-s| - |x-y|)\rangle^{-100}. \tag{10.25}$$

(3) Let $a = a_{\mathcal{C}}$ be a smooth bump function on a radially oriented rectangular box $\mathcal{C}$ of size $1 \times (2^\ell)^3$, where $\ell \leq 0$. Let $\omega \in \mathbb{S}^3$ be at angle $\simeq 2^\ell$ from $\mathcal{C}$. Then for $t - s = (x-y)\cdot\omega + O(1)$,

$$|K^a_{<0}(t,x;s,y)| \lesssim_{M_\sigma} 2^{3\ell} \langle 2^{2\ell}(t-s)\rangle^{-100} \langle 2^\ell (x'-y')\rangle^{-100} \tag{10.26}$$

where $x' = x - (x\cdot\omega)\omega$ and $y' = y - (y\cdot\omega)\omega$.

This lemma is proved as in [11, Section 8.1] by stationary phase, using the symbol bounds in Lemmas 10.1 and 10.2.

10.6. **Fixed-time L^2 bounds.** The goal of this subsection is to prove (9.44), (9.46), (9.47) and (9.44′) for $Z = L^2$. The common key ingredient is the following fixed-time L^2 estimate:

Proposition 10.9. *For $\delta > 0$ sufficiently small, there exists $\delta_{(0)} > 0$ such that the following statement holds. the following statement holds. Let $h + 10 \leq k \leq 0$. Then for every fixed t, we have*

$$\| \big(Op(Ad(O_{<h,\pm})_{<k})(x, D)Op(Ad(O^{-1}_{<h,\pm})_{<k})(D, y) - 1\big) P_0\|_{L^2\to L^2} \lesssim_{M_\sigma} 2^{\delta_{(0)}h} + 2^{-10(k-h)}. \tag{10.27}$$

Lemma 10.10. *There exists $\delta_{(0)} > 0$ such that the following statement holds. Let $h \leq 0$ and $a(\xi)$ be a smooth bump function adapted to $\{|\xi| \lesssim 1\}$. Then for every fixed t, we have*

$$\|Op(Ad(O_{<h,\pm}))(x, D)a(D)Op(Ad(O^{-1}_{<h,\pm}))(D, y) - a(D)\|_{L^2\to L^2} \lesssim_{M_\sigma} 2^{\delta_{(0)}h}. \tag{10.28}$$

Proof. For simplicity of notation, we omit $\pm$ in $O_{<h,\pm}$, $O^{-1}_{<h,\pm}$ and $\Psi_{\pm,h}$. Following the hypothesis, we fix $t \in \mathbb{R}$.

The idea is to derive a kernel estimate as in Lemma 10.8, but taking into account the frequency gap. The kernel of the End($\mathfrak{g}$)-valued operator in (10.28) is given by

$$K_{<h}(x, y) = \int \big(Ad(O_{<h}(x, \xi)O^{-1}_{<h}(y, \xi)) - 1\big)\, a(\xi)e^{i(x-y)\cdot\xi} \frac{d\xi}{(2\pi)^4}. \tag{10.29}$$

We obtain two different estimates depending on whether $|x-y| \lesssim 2^{-\delta_{(0)}h}$ or $|x-y| \gtrsim 2^{-\delta_{(0)}h}$.

Case 1: $|x - y| \lesssim 2^{-\delta_{(0)}h}$. In this case, we use the fundamental theorem of calculus and simply bound

$$\begin{aligned}|K_{<h}(x, y)| &\lesssim \int \int_{-\infty}^{h} \left|\frac{d}{d\ell}\big(Ad(O_{<\ell}(x, \xi)O^{-1}_{<\ell}(y, \xi))\big)\right| |a(\xi)|\, d\ell\, d\xi \\ &\lesssim \sup_{|\xi|\lesssim 1} \int_{-\infty}^{h} \left|\frac{d}{d\ell}\big(Ad(O_{<\ell}(x, \xi)O^{-1}_{<\ell}(y, \xi))\big)\right| d\ell\end{aligned}$$

By the algebraic property

$$O[u, v]O^{-1} = [OuO^{-1}, OvO^{-1}], \quad O \in \mathbf{G},\, u, v \in \mathfrak{g}$$

we have

$$ad(u)Ad(O) = Ad(O)ad(Ad(O^{-1})u), \quad Ad(O^{-1})ad(u) = ad(Ad(O^{-1})u)Ad(O^{-1}).$$

Therefore,

$$\begin{aligned}&\frac{d}{d\ell}\big(Ad(O_{<\ell}(x, \xi)O^{-1}_{<\ell}(y, \xi))\big)\\ &= ad(\Psi_\ell)Ad(O_{<\ell})(x, \xi)Ad(O^{-1}_{<\ell})(y, \xi) - Ad(O_{<\ell})(x, \xi)Ad(O^{-1}_{<\ell})ad(\Psi_\ell)(y, \xi)\\ &= Ad(O_{<\ell})(x, \xi)ad(Ad(O^{-1}_{<\ell})\Psi_\ell(x, \xi) - Ad(O^{-1}_{<\ell})\Psi_\ell(y, \xi))Ad(O^{-1}_{<\ell})(y, \xi).\end{aligned}$$

Then using the fact that the norm on End($\mathfrak{g}$) is invariant under $Ad(O)$ for any $O \in \mathbf{G}$, we have

$$\left|\frac{d}{d\ell}\big(Ad(O_{<\ell}(x, \xi)O^{-1}_{<\ell}(y, \xi))\big)\right| = \left|Ad(O^{-1}_{<\ell})\Psi_\ell(x, \xi) - Ad(O^{-1}_{<\ell})\Psi_\ell(y, \xi)\right|.$$

By the symbol bounds (10.5) and (10.4), we have $|\partial_x(Ad(O_{<\ell}^{-1})\Psi_\ell)| \lesssim_{M_\sigma} 2^{\ell}$. Thus, by the mean value theorem,

$$\left|\frac{d}{d\ell}\left(Ad(O_{<\ell}(x,\xi)O_{<\ell}^{-1}(y,\xi))\right)\right| \lesssim_{M_\sigma} 2^{\ell}2^{-\delta_{(0)}h}.$$

Integrating in ℓ, we arrive at

$$|K_{<h}(x,y)| \lesssim_{M_\sigma} 2^{(1-\delta_{(0)})h}. \tag{10.30}$$

Case 2: $|x-y| \gtrsim 2^{-\delta_{(0)}h}$. Here, the idea is to repeatedly integrate by parts in ξ. Since

$$\partial_\xi Ad(O_{<h}(x,\xi)O_{<h}^{-1}(y,\xi)) = ad((O_{<h}(x,\xi)O_{<h}^{-1}(y,\xi))_{;\xi})Ad(O_{<h}(x,\xi)O_{<h}^{-1}(y,\xi)),$$

the symbol bound (10.5) implies

$$|\partial_\xi^{(n)} Ad(O_{<h}(x,\xi)O_{<h}^{-1}(y,\xi))| \lesssim_{n,M_\sigma} 2^{\delta|n-\frac{1}{2}|h}.$$

Therefore, integrating by parts in ξ for N-times in (10.29), we obtain

$$|K_{<h}(x,y)| \lesssim_{\delta,N,M_\sigma} \frac{1}{|x-y|^{(1-\delta)N+\frac{1}{2}\delta}} \quad \text{for } |x-y| \gtrsim 2^{-\delta_{(0)}h},\ 0 \leq N < \delta^{-1}.$$

Finally, combining Cases 1 and 2, we obtain

$$\sup_x \int |K_{<h}(x,y)|\,dy + \sup_y \int |K_{<h}(x,y)|\,dx \lesssim_{M_\sigma} 2^{(1-5\delta_{(0)})h} \lesssim 2^{\delta_{(0)}h}$$

provided that $\delta_{(0)}$ is small enough. Bound (10.28) now follows. □

Corollary 10.11. *For any $k \in \mathbb{R}$ we have*

$$\|Op(Ad(O_{<h,\pm}))(x,D)P_0\|_{L^2\to L^2} \lesssim_{M_\sigma} 1, \tag{10.31}$$

$$\|Op(Ad(O_{<h,\pm})_{<k})(x,D)P_0\|_{L^2\to L^2} \lesssim_{M_\sigma} 1. \tag{10.32}$$

Proof. The first bound follows by a TT^*-argument from Lemma 10.10. Next, note that $Ad(O_{<h,\pm})_{<k}(x,\xi)$ is simply a smooth average of translates of $Ad(O_{<h,\pm})(x,\xi)$ in x. Therefore, the second bound follows from the first by translation invariance of L^2. □

Next, we borrow a lemma from [11], which handles $Ad(O_{<h,\pm})_k$ when k is large compared to h.

Lemma 10.12. *Let $t \in \mathbb{R}$, $h \leq 0$ and $k \geq h+10$. Then we have*

$$\|Op(Ad(O_{<h,\pm})_k)(t,x,D)P_0\|_{L^2\to L^2} \lesssim_{M_\sigma} 2^{-10(k-h)} \tag{10.33}$$

Furthermore, for $1 \leq q \leq p \leq \infty$, $h \leq 0$ and $k \geq h+10$, we have

$$\|Op(Ad(O_{<h,\pm})_k)(t,x,D)P_0\|_{L^pL^2\to L^qL^2} \lesssim_{M_\sigma} 2^{(\frac{1}{p}-\frac{1}{q})h}2^{-10(k-h)}. \tag{10.34}$$

Same estimates hold for the right quantization $Op(Ad(O_{<h,\pm})_k(D,s,y)$.

Remark 10.13. The specific factor 10 in the gain $2^{-10(k-h)}$ is not of any significance, but it is important to note that this number is much bigger than 1; see the proof of Proposition 10.14 below.

For the proof, we refer to [11, Proof of Lemma 8.4] or [17, Proof of Lemma 9.11].

Proof of Proposition 10.9. Due to the frequency localization of the symbols in (10.27), we can harmlessly insert a multiplier $a(D)$ whose symbol is a smooth bump function $a(\xi)$ adapted to $\{|\xi| \lesssim 1\}$, and then discard P_0 to replace (10.27) by

$$\|Op(Ad(O_{<h,\pm})_{<k})(x,D)a(D)Op(Ad(O^{-1}_{<h,\pm})_{<k})(D,y) - a(D)\|_{L^2\to L^2} \lesssim_{M_\sigma} 2^{\delta_{(0)}h} + 2^{-10(k-h)}.$$

Now it suffices to combine the last two Lemmas. □

Proof of (9.44), (9.46), (9.47) *and* (9.44′) *in the case* $Z = L^2$. By a TT^* argument, the bounds (9.44) and (9.44′) are immediate consequences of (10.27). Also from (10.27) we obtain the estimate (9.47) with a constant $2^{-\delta_{(0)}\kappa}$, which is less than ϵ if κ is chosen large enough depending only on M_0.

Finally, for (9.46) we compute

$$\partial_t (Ad(O))_{<0} = (ad(O_{;t})Ad(O))_{<0}$$

therefore it suffices to combine the decomposability bound (10.17) for $O_{;t}$ with $q = \infty$ with (10.31). The former bound yields a $2^{-\kappa}$ factor which again yields ϵ smallness if κ is large enough. □

10.7. **Spacetime L^2L^2 bounds.** Next, we establish (9.44), (9.46), (9.47) and (9.44′) when $Z = N$ or N^*. As we will see below, (9.44), (9.46) and (9.44′) follow from the arguments in [11]. In the bulk of this subsection, we focus on the task of establishing (9.47).

To state the key estimates, it is convenient to set up some notation. We introduce the compound $\mathbf{G}$-valued symbol

$$\mathbf{O}_{<h,\pm}(t,x,s,y,\xi) = O_{<h,\pm}(t,x,\xi)O^{-1}_{<h,\pm}(s,y,\xi).$$

The quantization of $Ad(\mathbf{O}_{<h,\pm})$, which is a $\mathrm{End}(\mathfrak{g})$-valued compound symbol, takes the form

$$Op(Ad(\mathbf{O}_{<h,\pm}))(t,x,D,y,s) = Op(Ad(O_{<h,\pm}))(t,x,D)Op(Ad(O^{-1}_{<h,\pm}))(D,y,s).$$

Given a compound $\mathrm{End}(\mathfrak{g})$-valued symbol $a(t,x,s,y,\xi)$, we define the double spacetime frequency projection

$$(a)_{\ll k}(t,x,s,y,\xi) = S^{t,x}_{<k}S^{s,y}_{<k}a(t,x,s,y,\xi).$$

Therefore, according to our conventions,

$$Ad(\mathbf{O}_{<h,\pm})_{\ll k}(t,x,s,y,\xi) = Ad(O_{<h,\pm})_{<k}(t,x,\xi)Ad(O^{-1}_{<h,\pm})_{<k}(s,y,\xi).$$

Proposition 10.14. *For $\delta > 0$ sufficiently small, there exists $\delta_{(1)}$ such that the following bounds hold for any $h < -20$:*

$$\|\left(Op(Ad(\mathbf{O}_{<h,\pm})_{\ll 0})(t,x,D,t,y) - 1\right)P_0\|_{N^*\to X^{0,\frac{1}{2}}_\infty} \lesssim_{M_\sigma} 2^{\delta_{(1)}h}. \quad (10.35)$$

Before we begin the proof, we state a lemma for passing to a double spacetime frequency localization of $Ad(\mathbf{O}_{<h,\pm})$, which is used several times in our argument below.

Lemma 10.15. *For $2 \leq q \leq \infty$ and $h + 10 \leq k \leq 0$, we have*

$$\|\left(Op(Ad(\mathbf{O}_{<h,\pm})_{\ll 0}) - Op(Ad(\mathbf{O}_{<h,\pm})_{\ll k})\right)P_0\|_{L^pL^2\to L^qL^2} \lesssim_{M_\sigma} 2^{(\frac{1}{p}-\frac{1}{q})h}2^{10(h-k)}. \quad (10.36)$$

This lemma is a straightforward consequence of Lemma 10.12; we omit the proof.

Proof of (10.35). We follow [17, Proof of Proposition 9.13]. For simplicity, we omit $\pm$ in $O_{<h,\pm}$, $\mathbf{O}_{<h,\pm}$ etc.

Step 1: High modulation input. For any $j \in \mathbb{Z}$ and $j' \geq j-5$, we claim that

$$\|Q_j(Op(Ad(\mathbf{O}_{<h})_{\ll 0}) - 1)P_0 Q_{j'}\|_{N^* \to X^{0,\frac{1}{2}}_\infty} \lesssim_{M_\sigma} 2^{\delta_{(0)} h} 2^{\frac{1}{2}(j-j')}. \tag{10.37}$$

Step 2: Low modulation input, $\frac{1}{2}h \leq j$. Here, we take care of the easy case $\frac{1}{2}h \leq j$. Under this assumption, we claim that

$$\|Q_j(Op(Ad(\mathbf{O}_{<h})_{\ll 0}) - 1)P_0 Q_{<j-5}\|_{N^* \to X^{0,\frac{1}{2}}_\infty} \lesssim_{M_\sigma} 2^{4h}. \tag{10.38}$$

Note that

$$Q_j(Op(Ad(\mathbf{O}_{<h})_{\ll j-5}) - 1)P_0 Q_{<j-5} = 0.$$

Thus, using the $L^\infty L^2$ portion of N^*, it suffices to prove

$$\|Q_j(Op(Ad(\mathbf{O}_{<h})_{\ll 0} - Ad(\mathbf{O}_{<h})_{\ll j-5})P_0 Q_{<j-5}\|_{N^* \to X^{0,\frac{1}{2}}_\infty} \lesssim_{M_\sigma} 2^{4h}.$$

Since Q_j and $Q_{<j-5}$ are disposable in L^2L^2 and $L^\infty L^2$, respectively, this estimate follows from Lemma 10.15.

Step 3: Low modulation input, $j < \frac{1}{2}h$, main decomposition. The goal of Steps 3–6 is to establish

$$\|Q_j(Op(Ad(\mathbf{O}_{<h})_{\ll 0}) - Ad(\mathbf{O}_{<j+\tilde{\delta}h})_{\ll 0})P_0 Q_{<j-5}\|_{N^* \to X^{0,\frac{1}{2}}_\infty} \lesssim_{M_\sigma} 2^{\delta_{(0)} h}. \tag{10.39}$$

provided that $j + \tilde{\delta}h \leq h$.

At the level of $\mathrm{End}(\mathfrak{g})$-valued compound symbols, we expand

$$Ad(\mathbf{O}_{<h}) - Ad(\mathbf{O}_{<j+\tilde{\delta}h}) = \mathcal{L} + \mathcal{Q} + \mathcal{C},$$

where

$$\begin{aligned}
\mathcal{L} &= \int_{j+\tilde{\delta}h \leq \ell \leq h} \mathcal{L}_{\ell, <j+\tilde{\delta}h} \, d\ell \\
\mathcal{Q} &= \int_{j+\tilde{\delta}h \leq \ell' \leq \ell \leq h} \mathcal{Q}_{\ell,\ell', <j+\tilde{\delta}h} \, d\ell' \, d\ell \\
\mathcal{C} &= \int_{j+\tilde{\delta}h \leq \ell'' \leq \ell' \leq \ell \leq h} \mathcal{C}_{\ell,\ell',\ell'', <\ell''} \, d\ell'' \, d\ell' \, d\ell
\end{aligned}$$

and the integrands $\mathcal{L}_{\ell,<k}$, $\mathcal{Q}_{\ell,\ell',<k}$ and $\mathcal{C}_{\ell,\ell',\ell'',<k}$ are defined recursively as

$$\begin{aligned}
\mathcal{L}_{\ell,<k}(t,x,s,y,\xi) =& ad(\Psi_\ell)(t,x,\xi) Ad(\mathbf{O}_{<k})(t,x,s,y,\xi) \\
&- Ad(\mathbf{O}_{<k})(t,x,s,y,\xi) ad(\Psi_\ell)(s,y,\xi) \\
\mathcal{Q}_{\ell,\ell',<k}(t,x,s,y,\xi) =& ad(\Psi_\ell)(t,x,\xi) \mathcal{L}_{\ell',<k}(t,x,s,y,\xi) \\
&- \mathcal{L}_{\ell',<k}(t,x,s,y,\xi) ad(\Psi_\ell)(s,y,\xi) \\
\mathcal{C}_{\ell,\ell',\ell'',<k}(t,x,s,y,\xi) =& ad(\Psi_\ell)(t,x,\xi) \mathcal{Q}_{\ell',\ell'',<k}(t,x,s,y,\xi) \\
&- \mathcal{Q}_{\ell',\ell'',<k}(t,x,s,y,\xi) ad(\Psi_\ell)(s,y,\xi)
\end{aligned}$$

The three terms $\mathcal{L}_{\ell,<k}$, $\mathcal{Q}_{\ell,\ell',<k}$ and $\mathcal{C}_{\ell,\ell',\ell'',<k}$ are successively considered in the next three steps.

Step 4: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{L}$. Our goal here is to prove

$$\|Q_j \mathcal{L}_{\ll 0} P_0 Q_{<j-5}\|_{N^* \to X_\infty^{0,\frac{1}{2}}} \lesssim_{M_\sigma} 2^{\delta_{(0)} h}. \tag{10.40}$$

We introduce

$$\begin{aligned}\mathcal{L}_{\ell,<k,\ll k'} =& ad(\Psi_\ell)(t,x,\xi) Ad(\mathbf{O}_{<k})_{\ll k'}(t,x,s,y,\xi) \\ &- Ad(\mathbf{O}_{<k})_{\ll k'}(t,x,s,y,\xi) ad(\Psi_\ell)(s,y,\xi) \\ \mathcal{L}_{\ell,<-\infty} =& ad(\Psi_\ell)(t,x,\xi) - ad(\Psi_\ell)(s,y,\xi)\end{aligned}$$

and decompose

$$\begin{aligned}\mathcal{L} =& \int_{j+\tilde{\delta}h \le \ell \le h} \left(\mathcal{L}_{\ell,<j+\tilde{\delta}h} - \mathcal{L}_{\ell,<j+\tilde{\delta}h,\ll j-5}\right) d\ell \\ &+ \int_{j-10\tilde{\delta}h \le \ell \le h} \mathcal{L}_{\ell,<j+\tilde{\delta}h,\ll j-5}\, d\ell \\ &+ \int_{j+\tilde{\delta}h \le \ell \le j-10\tilde{\delta}h} \left(\mathcal{L}_{\ell,<j+\tilde{\delta}h,\ll j-5} - \mathcal{L}_{\ell,<-\infty}\right) d\ell \\ &+ \int_{j+\tilde{\delta}h \le \ell \le j-10\tilde{\delta}h} \mathcal{L}_{\ell,<-\infty}\, d\ell \\ =:& \mathcal{L}_{(1)} + \mathcal{L}_{(2)} + \mathcal{L}_{(3)} + \mathcal{L}_{(4)}.\end{aligned}$$

Step 4.1: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{L}_{(1)}$. For this term we can add a double frequency localization $\ll C$ on $\mathcal{L}_{\ell,<j+\tilde{\delta}h}$ and then harmlessly discard the double $\ll 0$ localization in (10.40). Then it suffices to prove that for $\ell > j + \delta m$ we have

$$\|Q_j Op\left(\mathcal{L}_{\ell,<j+\tilde{\delta}h,\ll C} - \mathcal{L}_{\ell,<j+\tilde{\delta}h,\ll j-5}\right) P_0 Q_{<j-5}\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{6}[\ell-(j+\tilde{\delta}h)]} 2^{(10+\frac{1}{2})\tilde{\delta}h},$$

and then integrate with respect to ℓ. But this is a consequence of the decomposability bound (10.14) with $q = 6$ and $r = \infty$, together with the bound (10.34) with $p = 6$ and $q = 2$.

Step 4.2: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{L}_{(2)}$. Here as well as in the next two cases the $\ll 0$ localization in ℓ has no effect and is discarded. The two terms in $\mathcal{L}_{\ell,<j+\tilde{\delta}h,\ll j-5}$ are similar; we restrict our attention to the first one. Consider now the operator

$$Q_j Op(ad(\Psi_\ell) Ad(\mathbf{O}_{<j+\tilde{\delta}h})_{\ll j-5}) Q_{<j-5} = \sum_\theta Q_j Op(ad(\Psi_\ell^{(\theta)}) Ad(\mathbf{O}_{<j+\tilde{\delta}h})_{\ll j-5}) Q_{<j-5}$$

The important observation here is that, because of the geometry of the cone, the frequency localizations for both $Ad(\mathbf{O}_{<j+tdh})_{\ll j-5})$ and $\Psi_\ell^{(\theta)}$ force a large angle $\theta > 2^{\frac{1}{2}(j-\ell)}$, or else the above operator vanishes.

Given this bound for θ, we can now use the decomposability bound (10.14) with $q = 2$ and $r = \infty$ combined with (10.34) with $p = \infty$ and $q = \infty$ to obtain

$$\|Op(ad(\Psi_\ell^{(\theta)}) Ad(\mathbf{O}_{<j+\tilde{\delta}h})_{\ll j-5}) P_0\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} 2^{\frac{1}{2}(j-\ell)} \theta^{-\frac{1}{2}}$$

which after θ summation in the range $\theta > 2^{\frac{1}{2}(j-\ell)}$ yields

$$\|Q_j Op(\mathcal{L}_{(2)}) P_0 Q_{<j-5}\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} 2^{\frac{5}{2}\tilde{\delta}h}.$$

which suffices.

Step 4.3: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{L}_{(3)}$. Here we have the same angle constraint as above but this levels off for $\ell < j$, namely $\theta > 2^{-\frac{1}{2}(\ell-j)_+}$. However, we can now replace (10.32) with (10.27) to obtain

$$\|Op(ad(\Psi_\ell^{(\theta)})(Ad(\mathbf{O}_{<j+\tilde{\delta}h})_{\ll j-5} - I))P_0\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} 2^{-\frac{1}{2}(\ell-j)} \theta^{-\frac{1}{2}} (2^{\delta_{(0)}(j+\tilde{\delta}h)} + 2^{10\tilde{\delta}h})$$

which after θ and ℓ summation yields

$$\|Q_j Op(\mathcal{L}_{(3)}) P_0 Q_{<j-5}\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} (2^{(\delta_{(0)} - \frac{1}{4}\tilde{\delta})h} + 2^{9\tilde{\delta}h}).$$

This suffices provided that $\tilde{\delta}$ is small enough $\tilde{\delta} < \delta_{(0)}$.

Step 4.4: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{L}_{(4)}$. Here we have the same range $j - \tilde{\delta}h < \ell < j + 10\tilde{\delta}h$ for ℓ. We also have the same constraint on the angle $\theta > 2^{-\frac{1}{2}(\ell-j)_+}$ but this is no longer relevant in this case, as we will gain in frequency, and this can override any angular losses.

This time we are able to take advantage of the difference structure for Ψ. Precisely, it suffices to show that for a localized at frequency 1 we have

$$\|Op(ad(\Psi_\ell^{(\theta)}))(t,x,D)a(D) - a(D)Op(ad(\Psi_\ell^{(\theta)}))(t,x,D)\|_{L^\infty L^2 \to L^q L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{q}\ell} 2^{\ell} \theta^{-C} \quad (10.41)$$

But this was already proved in [17], (9.40).

Step 5: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{Q}$. We proceed in the same manner as in the case of $\mathcal{L}$. Defining the symbols

$$\begin{aligned}
\mathcal{Q}_{\ell,\ell',<k,\ll k'} =& ad(\Psi_\ell)(t,x,\xi)\mathcal{L}_{\ell',<k,\ll k'}(t,x,s,y,\xi) \\
&- \mathcal{L}_{\ell',<k,\ll k'}(t,x,s,y,\xi) ad(\Psi_\ell)(s,y,\xi) \\
\mathcal{Q}_{\ell,\ell',<-\infty} =& ad(\Psi_\ell)(t,x,\xi)\mathcal{L}_{\ell',<-\infty}(t,x,s,y,\xi) \\
&- \mathcal{L}_{\ell',<-\infty}(t,x,s,y,\xi) ad(\Psi_\ell)(s,y,\xi)
\end{aligned}$$

we decompose $\mathcal{Q}$ as follows:

$$\begin{aligned}
\mathcal{Q} =& \int_{j+\tilde{\delta}h \le \ell' \le \ell \le h} \left(\mathcal{Q}_{\ell,\ell',<j+\tilde{\delta}h} - \mathcal{Q}_{\ell,\ell',<j+\tilde{\delta}h,\ll j-10}\right) d\ell' d\ell \\
&+ \int_{\substack{j+\tilde{\delta}h \le \ell' \le \ell \le h \\ j-10\tilde{\delta}h \le \ell}} \mathcal{Q}_{\ell,\ell',<j+\tilde{\delta}h,\ll j-10} \, d\ell' d\ell \\
&+ \int_{j+\tilde{\delta}h \le \ell' \le \ell \le j-10\tilde{\delta}h} \left(\mathcal{Q}_{\ell,\ell',<j+\tilde{\delta}h,\ll j-10} - \mathcal{Q}_{\ell,\ell',<-\infty}\right) d\ell' d\ell \\
&+ \int_{j+\tilde{\delta}h \le \ell' \le \ell \le j-10\tilde{\delta}h} \mathcal{Q}_{\ell,\ell',<-\infty} \, d\ell' d\ell \\
=:& \mathcal{Q}_{(1)} + \mathcal{Q}_{(2)} + \mathcal{Q}_{(3)} + \mathcal{Q}_{(4)}
\end{aligned}$$

Then we consider each term separately.

Step 5.1: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{Q}_{(1)}$. Proceeding as in Step 4.1, we have

$$\mathcal{Q}_{\ll 1} = \int_{j+\tilde{\delta}h \leq \ell' \leq \ell \leq h} \left(\mathcal{Q}_{\ell,\ell',<j+\tilde{\delta}h,\ll C} - \mathcal{Q}_{\ell,\ell',<j+\tilde{\delta}h,\ll j-5}\right)_{\ll 0} \, d\ell' d\ell$$

and we can again harmlessly discard the outer $\ll 0$. Applying the decomposability bound (10.14) with $q = 6$ for Ψ_ℓ and with $q = \infty$ for $\Psi_{\ell'}$ and $r = \infty$, together with the bound (10.34) with $p = \infty$ and $q = 3$, we obtain

$$\|\mathcal{Q}_{\ell,\ell',<j+\tilde{\delta}h,\ll C} - \mathcal{Q}_{\ell,\ell',<j+\tilde{\delta}h,\ll j-5}\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{6}[\ell-(j+\tilde{\delta}h)]} 2^{(10+\frac{1}{2})\tilde{\delta}h}.$$

Summing up with respect to ℓ and ℓ' we obtain

$$\|Op(\mathcal{Q}_{(1)})P_0\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{10\tilde{\delta}h}.$$

which suffices.

Step 5.2: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{Q}_{(2)}$. Here and also for $\mathcal{Q}_{(3)}$ and $\mathcal{Q}_{(4)}$ we can remove the outer frequency localization $\ll 0$ which does nothing. The expression $\mathcal{Q}_{(2)}$ contains four terms depending on whether Ψ_ℓ and $\Psi_{\ell'}$ act on the left or on the right. We consider one of them, for which we need to bound the operator

$$Q_j Op(ad(\Psi_\ell) Ad(\mathcal{O}_{<j+\tilde{\delta}h})_{\ll j-5} ad(\Psi_{\ell'})) Q_{<j-5} P_0$$

We decompose with respect to angles into

$$\sum_{\theta,\theta'} Q_j Op(ad(\Psi_\ell^{(\theta)}) Ad(\mathcal{O}_{<j+\tilde{\delta}h})_{\ll j-5} ad(\Psi_{\ell'}^{(\theta')})) Q_{<j-5} P_0$$

and consider the nontrivial scenarios. This is as in Step 5.2 but now we have two angles, which must satisfy non-exclusively

$$\text{either } \theta > 2^{\frac{1}{2}(j-\ell)}, \qquad \text{or } \theta' > 2^{\frac{1}{2}(j-\ell')}.$$

We can now use the decomposability bound (10.14) with $q = 3$ and $r = \infty$ for the large[11] angle respectively $q = 6$ and $r = \infty$ for the other angle combined with (10.34) with $p = \infty$ and $q = \infty$ to obtain either

$$\|Op(ad(\Psi_\ell^{(\theta)}) Ad(\mathcal{O}_{<j+\tilde{\delta}h})_{\ll j-5} ad(\Psi_{\ell'}^{(\theta')})) P_0\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} 2^{\frac{1}{3}(j-\ell)} \theta^{-\frac{1}{6}} 2^{\frac{1}{6}(j-\ell')} \theta'^{\frac{1}{6}}$$

or the same bound with the pairs (l,θ) and (l',θ') reversed. Summing with respect to ℓ, ℓ', and also with respect to θ, θ' subject to the constraints above, we obtain

$$\|Q_j Op(\mathcal{Q}_{(2)}) P_0 Q_{<j-5}\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} 2^{\frac{5}{3}\tilde{\delta}h}.$$

which suffices.

Step 5.3: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{Q}_{(3)}$. We repeat the angle localization analysis in the previous step, but as in Step 4.3, we again replace (10.32) with (10.27). The outcome is similar to the one in Step 4.3; details are omitted.

Step 5.4: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{Q}_{(4)}$. Again we apply the same angle localization analysis as in the previous two steps. However, as in Step 4.4, we

[11] i.e. which satisfies the bound on the previous line

also need to exploit the difference between one of the two Ψ's and its adjoint. Consider one such term, e.g.

$$ad(\Psi_\ell^{(\theta)})(t,x,\xi)[ad(\Psi_{\ell'}^{(\theta')})(t,x,\xi) - ad(\Psi_{\ell'}^{(\theta')})(\xi,y,s)]$$

For this it suffices to apply the disposability bound (10.14) for $\Psi_\ell^{(\theta)}$ combined with (10.41). The choice of the exponents is no longer important. We obtain

$$\|Op(\mathcal{Q}_{(4)})P_0\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} 2^{(1-C\tilde{\delta})j}.$$

Step 6: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{C}$. This repeats the analysis for $\mathcal{L}$ and $\mathcal{Q}$, but we no longer need to keep track of angular separation. Denoting

$$\begin{aligned}\mathcal{C}_{\ell,\ell',\ell'',<k,\ll k'} =& ad(\Psi_\ell)(t,x,\xi)\mathcal{Q}_{\ell',\ell'',<k,\ll k'}(t,x,s,y,\xi)\\ &- \mathcal{Q}_{\ell',\ell'',<k,\ll k'}(t,x,s,y,\xi)ad(\Psi_\ell)(s,y,\xi)\\ \mathcal{C}_{\ell,\ell',\ell'',<-\infty} =& ad(\Psi_\ell)(t,x,\xi)\mathcal{Q}_{\ell',\ell'',<-\infty}(t,x,s,y,\xi)\\ &- \mathcal{Q}_{\ell',\ell'',<-\infty}(t,x,s,y,\xi)ad(\Psi_\ell)(s,y,\xi)\end{aligned}$$

we decompose $\mathcal{C}$ as

$$\begin{aligned}\mathcal{C} =& \int_{j+\tilde{\delta}h \le \ell'' \le \ell' \le \ell \le h} (\mathcal{C}_{\ell,\ell',\ell'',<\ell''} - \mathcal{C}_{\ell,\ell',\ell'',<\ell'',\ll -5})\, d\ell'' d\ell' d\ell\\ &+ \int_{\substack{j+\tilde{\delta}h \le \ell'' \le \ell' \le \ell \le h\\ j-10\tilde{\delta}h \le \ell}} \mathcal{C}_{\ell,\ell',\ell'',<\ell'',\ll -5}\, d\ell'' d\ell' d\ell\\ &+ \int_{j+\tilde{\delta}h \le \ell'' \le \ell' \le \ell \le j-10\tilde{\delta}h} (\mathcal{C}_{\ell,\ell',\ell'',<\ell'',\ll j-5} - \mathcal{C}_{\ell,\ell',\ell'',<-\infty})\, d\ell'' d\ell' d\ell\\ &+ \int_{j+\tilde{\delta}h \le \ell'' \le \ell' \le \ell \le j-10\tilde{\delta}h} \mathcal{C}_{\ell,\ell',\ell'',<-\infty}\, d\ell'' d\ell' d\ell\\ =:& \mathcal{C}_{(1)} + \mathcal{C}_{(2)} + \mathcal{C}_{(3)} + \mathcal{C}_{(4)}\end{aligned}$$

and consider each of the terms separately.

Step 6.1: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{C}_{(1)}$. The same argument as in Steps 4.1 and 5.1 yields the bound

$$\begin{aligned}&\|Op(ad(\Psi_\ell)ad(\Psi_{\ell'})ad(\Psi_{\ell''})(Ad((\mathbf{O}_{<\ell''}) - Ad((\mathbf{O}_{<\ell''})_{\ll -5}))_{\ll 0}\|_{L^\infty L^2 \to L^2}\\ &\qquad \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} 2^{\frac{1}{6}(j+\tilde{\delta}h-\ell)} 2^{\frac{1}{6}(j+\tilde{\delta}h-\ell')} 2^{\frac{1}{6}(j+\tilde{\delta}h-\ell')} 2^{10\ell''} 2^{\frac{1}{2}\tilde{\delta}h}\end{aligned}$$

as well as for any of the other choices of left/right quantizations for the Ψ's. Integration over $j + \tilde{\delta}h < \ell'' < \ell' < \ell < \frac{m}{2}$ is now harmless.

Step 6.2: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{C}_{(2)}$. Applying the decomposability bound (10.14) with $q = 6$ for each of the three Ψ's in the $\mathcal{C}_2$ integrand, as well as the L^2 bound for $Op(Ad((\mathbf{O}_{<\ell''})_{\ll -5})$ yields the bound

$$\|Op(ad(\Psi_\ell)ad(\Psi_{\ell'})ad(\Psi_{\ell''})Ad((\mathbf{O}_{<j+\tilde{\delta}h}))_{\ll -5}\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} 2^{\frac{1}{6}(j-\ell)} 2^{\frac{1}{6}(j-\ell')} 2^{\frac{1}{6}(j-\ell')}$$

which suffices after integration in $\ell > j - 10\tilde{\delta}h$ and $\ell', \ell'' > j + \tilde{\delta}h$.

Step 6.3: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{C}_{(3)}$. This is the same argument as in the previous step, but using (10.27) instead of (10.32).

Step 6.4: Low modulation input, $j < \frac{1}{2}h$, contribution of $\mathcal{C}_{(4)}$. Here we are concerned with symbols of the form

$$ad(\Psi_\ell)(t,x,\xi)ad(\Psi_{\ell'})(t,x,\xi)[ad(\Psi_{\ell''}(t,x,\xi) - ad(\Psi_{\ell''}(\xi,y,s)]$$

where one or both of $ad(\Psi_\ell)$ and $ad(\Psi_{\ell'})$ may be switched to the right and in the right quantization. Here we use again the decomposability bound (10.14) with $q = 6$ for Ψ_ℓ and $ad(\Psi_{\ell'}$, respectively (10.41) for the $\Psi_{\ell''}$ difference.

Step 7: Low modulation input, $j < \frac{1}{2}h$, low frequency O. To complete the proof of the estimate (10.35) it remains to show that

$$\|Q_j Op(Ad(\mathbf{O}_{<j+\tilde{\delta}h})_{\ll 0}(t,x,D,y,s) - 1)P_0 Q_{<j-5}\|_{N^* \to X_\infty^{0,\frac{1}{2}}} \lesssim_{M_\sigma} 2^{\delta_{(1)}h} \tag{10.42}$$

If $j + \tilde{\delta}h \leq h$ this is combined with the bound (10.39), which is the main outcome of Steps 3-6. Else, this is used by itself, simply observing that we can harmlessly replace $j + \tilde{\delta}h$ by h.

The above bound is identical to

$$\|Q_j Op(Ad(\mathbf{O}_{<j+\tilde{\delta}h})_{\ll 0} - Ad(\mathbf{O}_{<j+\tilde{\delta}h})_{\ll j-5})(t,x,D,y,s)P_0 Q_{<j-5}\|_{N^* \to X_\infty^{0,\frac{1}{2}}} \lesssim_{M_\sigma} 2^{\delta_{(1)}h}$$

which in turn would follow from

$$\|Op(Ad(\mathbf{O}_{<j+\tilde{\delta}h})_{\ll 0} - Ad(\mathbf{O}_{<j+\tilde{\delta}h})_{\ll j-5})(t,x,D,y,s))P_0\|_{L^\infty L^2 \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} 2^{\delta_{(1)}h}$$

But this is a direct consequence of the bound (10.34). ☐

Proof of (9.47) *in the case $Z = N$ or N^*.* For the estimate (9.47) with $Z = N^*$ we combine the $L^\infty L^2$ bound given by (10.27) with (10.35). If on the other hand $Z = N$, then the same bound follows by duality. ☐

It remains to prove (9.44), (9.46) and (9.44′) when $Z = N$ or N^*. For this purpose, we recall the following result from [11]:

Lemma 10.16. *For $\ell \leq k' \pm O(1)$, we have*

$$\|Q_\ell Op(Ad(O_{<h,\pm})_{k'})(t,x,D)Q_{<0}P_0\|_{N^* \to X_\infty^{0,\frac{1}{2}}} \lesssim_{M_\sigma} 2^{\delta_1(\ell-k')}, \tag{10.43}$$

$$\|Q_\ell Op(Ad(O^{-1}_{<h,\pm})_{k'})(D,y,s)Q_{<0}P_0\|_{N^* \to X_\infty^{0,\frac{1}{2}}} \lesssim_{M_\sigma} 2^{\delta_1(\ell-k')}. \tag{10.44}$$

In particular, summing over all (ℓ, k') with $\ell \leq k$ and $k \leq k' + O(1)$, we have

$$\|Q_{<k}(Op(Ad(O_{<h,\pm})_{<0}) - Op(Ad(O_{<h,\pm})_{<k-C}))(t,x,D)Q_{<0}P_0\|_{N^* \to X_\infty^{0,\frac{1}{2}}} \lesssim_{M_\sigma} 1, \tag{10.45}$$

$$\|Q_{<k}(Op(Ad(O^{-1}_{<h,\pm})_{<0}) - Op(Ad(O^{-1}_{<h,\pm})_{<k-C}))(D,y,s)Q_{<0}P_0\|_{N^* \to X_\infty^{0,\frac{1}{2}}} \lesssim_{M_\sigma} 1. \tag{10.46}$$

Proof. The proof of this lemma is similar to that of Proposition 10.14, but simpler in the sense the frequency gap need not be exploited. It can be proved with exactly the same arguments as in [11, Proof of Proposition 8.5] (there, $M_\sigma \lesssim \epsilon$). Because of this, we will merely indicate here how to modify the preceding proof of (10.35) to obtain (10.43). We leave the details, as well as the entire case of (10.44), to the reader.

As before, we omit $\pm$ in the symbols. Throughout the proof of (10.35), we replace $Ad(\mathbf{O}_{<h})_{\ll k}(t,x,s,y,\xi) - 1$ by $Ad(O_{<h})_{<k}(t,x,\xi)$. The main decomposition (Step 4) now takes the form

$$\begin{aligned} Ad(O_{<h})(t,x,\xi) - Ad(O_{<j+\tilde{\delta}h}) = \mathcal{L}' + \mathcal{Q}' + \mathcal{C}' = & \int_{j+\tilde{\delta}h\leq \ell\leq h} \mathcal{L}'_{\ell,<j+\tilde{\delta}h}\, d\ell \\ & + \int_{j+\tilde{\delta}h\leq \ell'\leq \ell\leq h} \mathcal{Q}'_{\ell,\ell',<j+\tilde{\delta}h}\, d\ell' d\ell \\ & + \int_{j+\tilde{\delta}h\leq \ell''\leq \ell'\leq \ell\leq h} \mathcal{C}'_{\ell,\ell',\ell'',<j+\tilde{\delta}h}\, d\ell'' d\ell'\, d\ell \end{aligned}$$

where

$$\begin{aligned} \mathcal{L}'_{\ell,<k}(t,x,\xi) =& ad(\Psi_\ell) Ad(O_{<k})(t,x,\xi), \\ \mathcal{Q}'_{\ell,<k}(t,x,\xi) =& ad(\Psi_\ell)\mathcal{L}'_{\ell',<k}(t,x,\xi) = ad(\Psi_\ell) ad(\Psi_{\ell'}) Ad(O_{<k})(t,x,\xi), \\ \mathcal{C}'_{\ell,<k}(t,x,\xi) =& ad(\Psi_\ell)\mathcal{Q}'_{\ell',\ell'',<k}(t,x,\xi) = ad(\Psi_\ell) ad(\Psi_{\ell'}) ad(\Psi_{\ell''}) Ad(O_{<k})(t,x,\xi). \end{aligned}$$

For the expansion of $\mathcal{L}$, $\mathcal{Q}$ and $\mathcal{C}$ in Steps 5, 6 and 7, we replace $\mathcal{L}_{\ell,<k,\ll k'}$, $\mathcal{L}_{\ell,<-\infty}$, $\mathcal{Q}_{\ell,\ell',<k,\ll k'}$, $\mathcal{Q}_{\ell,\ell',<-\infty}$, $\mathcal{C}_{\ell,\ell',\ell'',<k,\ll k'}$ and $\mathcal{C}_{\ell,\ell',\ell'',<-\infty}$ by, respectively,

$$\begin{aligned} \mathcal{L}'_{\ell,<k,<k'} =& ad(\Psi_\ell) Ad(O_{<k})_{<k'}(t,x,\xi), \\ \mathcal{L}'_{\ell,<-\infty} =& ad(\Psi_\ell)(t,x,\xi), \\ \mathcal{Q}'_{\ell,\ell',<k,<k'} =& ad(\Psi_\ell)\mathcal{L}'_{\ell',<k,<k'}(t,x,\xi) = ad(\Psi_\ell) ad(\Psi_{\ell'}) Ad(O_{<k})_{<k'}(t,x,\xi), \\ \mathcal{Q}'_{\ell,\ell',<-\infty} =& ad(\Psi_\ell)\mathcal{L}'_{\ell',<-\infty}(t,x,\xi) = ad(\Psi_\ell) ad(\Psi_{\ell'})(t,x,\xi), \\ \mathcal{C}'_{\ell,\ell',\ell''<k,<k'} =& ad(\Psi_\ell)\mathcal{Q}'_{\ell',\ell'',<k,<k'}(t,x,\xi) = ad(\Psi_\ell) ad(\Psi_{\ell'}) ad(\Psi_{\ell''}) Ad(O_{<k})_{<k'}(t,x,\xi), \\ \mathcal{C}'_{\ell,\ell',\ell'',<-\infty} =& ad(\Psi_\ell)\mathcal{Q}'_{\ell',\ell'',<-\infty}(t,x,\xi) = ad(\Psi_\ell) ad(\Psi_{\ell'}) ad(\Psi_{\ell''})(t,x,\xi). \end{aligned}$$

Accordingly, we replace the use of (10.27) and (10.36) by (10.32) and (10.34), respectively, which results in loss of the smallness factor $2^{\delta_{(1)}h}$ in (10.43) compared to (10.35). ☐

Proof of (9.44), (9.46) *and* (9.44′). in the case $Z = N$ or N^*] It suffices to consider the $Z = N^*$; then the case $Z = N$ follows by duality. The $L^\infty L^2$ bound follows from the $Z = L^2$ case, so for (9.44) and (9.44′) it remains to establish that

$$\| Q_j Op(Ad(O_{<h,\pm})_{<0}) P_0 \|_{N^* \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j}$$

By Lemma 10.16 this reduces to

$$\| Q_j Op(Ad(O_{<h,\pm})_{<j-5}) P_0 \|_{N^* \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j}$$

Now due to the frequency localization for $Op(Ad(O_{<h,\pm})_{<j-5}$ we can insert a (slight enlargement of) Q_j on the right, in which case we can simply use again the $Z = L^2$ case.

Similarly, in the case of (9.44′) it suffices to show that

$$\| Q_j [\partial_t, Op(Ad(O_{<h,\pm})_{<0})] Q_{<j} P_0 \|_{N^* \to L^2} \lesssim_{M_\sigma} 2^{-\frac{1}{2}j} 2^h$$

We split into two cases. If $j \leq \frac{3}{4}h$ then we write

$$\partial_t Ad(O_{<h,\pm}) = ad(O_{<h,\pm;t}) Ad(O_{<h,\pm})_{<0}).$$

and then we can easily combine the decomposability bound (10.18) with the L^2 boundedness of $Op(Ad(O_{<h,\pm})_{<0})$. Else we have

$$Q_j[\partial_t, Op(Ad(O_{<h,\pm})_{<0})]Q_{<j}P_0 = Q_j[\partial_t, Op(Ad(O_{<h,\pm})_{[j-5,0]})]Q_{<j}P_0$$

Now we discard Q_j, $Q_{<j-5}$ and ∂_t and use directly (10.34) with $p = \infty$ and $q = 2$.

□

10.8. **Dispersive estimates.** Finally, we sketch the proofs of (9.45) and (9.45′). As in [11], we exactly follow the argument in [10, Section 11]. In the case of (9.45), we replace the use of the oscillatory integral estimates (108), (110) and (111) in [10] by (10.24), (10.25) and (10.26), respectively, the fixed-time L^2 bound (114) in [10] by (10.32), (118) in [10] by (10.45) etc. In case of (9.45′), observe that all the constants in these bounds are *universal* under the smallness assumption (9.48) for a suitable choice of $\delta_o(M)$, as we may take $M_\sigma \lesssim 1$.

There is one exception to the above strategy, namely the square function bound

$$\|Op(Ad(O_\pm)_{<0}(t,x,D)\|_{S_0^\sharp \to L_x^{\frac{10}{3}} L_t^2} \lesssim_{M_\sigma} 1. \tag{10.47}$$

This is due to the fact that the square function norm was not part of the S_0 norm in [10, 11], and was added only here. The same approach as in [11] allows us, via a TT^* type argument, to reduce the problem to an estimate of the form

$$\left\| \int \chi_{-l}(t-s)\mathbf{S}(t,s)B(s)ds \right\|_{L_x^{\frac{10}{3}} L_t^2} \lesssim_{M_\sigma} \|B\|_{L_x^{\frac{10}{7}} L_t^2}$$

where

$$\mathbf{S}(t,s) = Op(Ad(O_\pm)_{<0}(t,x,D)e^{\pm i(t-s)|D|}Op(Ad(O_\pm)_{<0}(D,s,y)$$

and the bump function χ_{-l} corresponds to the modulation scale 2^l in $S_0^\sharp$. It is easily seen that the bump function is disposable and can be harmlessly discarded. Hence in order to prove (10.47) it remains to show that

$$\| \int \mathbf{S}(t,s)B(s)ds\|_{L_x^{\frac{10}{3}} L_t^2} \lesssim_{M_\sigma} \|B\|_{L_x^{\frac{10}{7}} L_t^2} \tag{10.48}$$

To prove this we use Stein's analytic interpolation theorem. We consider the analytic family of operators

$$T_z B(t) = e^{z^2} \int (t-s)^z \mathbf{S}(t,s)B(s)ds$$

for z in the strip

$$-1 \le \mathrm{Im} z \le \frac{3}{2}$$

Then it suffices to establish the uniform bounds

$$\|T_z\|_{L^2 \to L^2} \lesssim_{M_\sigma} 1, \qquad \mathrm{Re} z = -1 \tag{10.49}$$

respectively

$$\|T_z\|_{L_x^1 L_t^2 \to L_x^\infty L_t^2} \lesssim_{M_\sigma} 1, \qquad \mathrm{Re} z = \frac{3}{2} \tag{10.50}$$

For (10.49) we can use the bound (10.31) to discard the L^2 bounded operators

$$Op(Ad(O_\pm)_{<0}(t,x,D)e^{\pm it|D|}, \qquad e^{\mp is|D|}Op(Ad(O_\pm)_{<0}(D,s,y).$$

Then we are left with the time convolutions with the kernels $e^{z^2}t^z$. But these are easily seen to be multipliers with uniformly bounded symbols.

For (10.50), on the other hand, we consider the kernel $K_z(t,x,s,y)$ of T_z. This is given by

$$K_z(t,x,s,y) = e^{z^2}(t-s)^z K^a_{<0}(t,x,s,y)$$

with a a smooth bump function on the unit scale. Hence by (10.24) we have the kernel bound

$$|K_z(t,x,s,y)| \lesssim_{M_\sigma} \langle |t-s| - |x-y| \rangle^{-100}, \qquad \mathrm{Re} z = \frac{3}{2}$$

Fixing x and y we have the obvious bound

$$\|K_z(\cdot, x, \cdot, y)\|_{L^2 \to L^2} \lesssim_{M_\sigma} 1.$$

Then (10.50) easily follows.

11. Renormalization error bounds

Without loss of generality, we fix the sign $\pm = +$. In this section, unless we specify otherwise, $Op(\cdot)$ denotes the left quantization. For the sake of simplicity, we also adopt the convention of simply writing A_x for $\mathbf{P}_x A$.

11.1. Preliminaries.

We collect here some technical tools for proving the renormalization error bound.

We begin with a tool that allows us to split $Op(ab)$ into $Op(a)Op(b)$. The idea of the proof is based on the heuristic identity $Op(ab) - Op(a)Op(b) \approx Op(-i\partial_\xi a \cdot \partial_x b)$ for left-quantized pseudodifferential operators (cf. [10, Lemma 7.2] and [11, Lemma 7.2]).

Lemma 11.1 (Composition via pseudodifferential calculus). *Let $a(t,x,\xi)$ and $b(t,x,\xi)$ be $\mathrm{End}(\mathfrak{g})$-valued symbols on $I_t \times \mathbb{R}^4_x \times \mathbb{R}^4_\xi$ with bounded derivatives, such that $a(t,x,\xi)$ is homogeneous of degree 0 in ξ and $b(t,x,\xi) = P^x_{<h_\theta - 10} b(t,x,\xi)$ for some $0 < \theta < 1$ and $2^{h_\theta} = \theta$. Then we have*

$$\|(Op(a)Op(b) - Op(ab))P_0\|_{L^q L^2[I] \to L^r L^2[I]} \lesssim \|\theta \partial_\xi a\|_{D_\theta L^{p_2} L^\infty[I]} \|Op(\theta^{-1}\partial_x b)P_0\|_{L^q L^2[I] \to L^{p_1} L^2[I]}, \tag{11.1}$$

where $r^{-1} = p_1^{-1} + p_2^{-1}$.

Proof. For simplicity, in this proof we only present formal computation, which can be justified using the qualitative assumptions on a and b.

Let us fix $t \in I$. Thanks to the frequency localization condition $b(x,\xi) = P^x_{<h_\theta - 10} b(x,\xi)$, we may write

$$(Op(a)Op(b) - Op(ab))P_0 = \sum_\phi Op(a^\phi_\theta) Op(b^\phi_\theta) - Op(a^\phi_\theta b^\phi_\theta)$$

where

$$a^\phi_\theta(x,\xi) = a(x,\xi)(m^\phi_\theta)^2(\xi)\tilde{m}^2_0(\xi), \quad b^\phi_\theta(x,\xi) = b(x,\xi)\tilde{m}^\phi_\theta(\xi) m_0(\xi).$$

Here ϕ runs over caps of radius $\simeq \theta$ on $\mathbb{S}^3$ with uniformly finite overlaps, $(m^\phi_\theta)^2(\xi) = (m^\phi_\theta)^2(\xi/|\xi|)$ are the associated smooth partition of unity on $\mathbb{S}^3$ and $m_0(\xi)$ is the symbol for P_0. The functions $\tilde{m}^\phi_\theta(\xi) = \tilde{m}^\phi_\theta(\xi/|\xi|)$ and $\tilde{m}^2_0(\xi)$ are smooth cutoffs to the supports of m^ϕ_θ and m_0, respectively, which can be inserted thanks to the frequency localization condition $b(x,\xi) = P^x_{<h_\theta - 10} b(x,\xi)$.

For each ϕ, we claim that

$$\|Op(a_\theta^\phi)Op(b_\theta^\phi) - Op(a_\theta^\phi b_\theta^\phi)\|_{L^2\to L^2} \lesssim \big(\sum_{n=1}^{20} \sup_\omega m_\theta^\phi(\omega)\|\theta^n \partial_\xi^{(n)} a(\cdot,\omega)\|_{L^\infty}\big)\|Op(\theta^{-1}\partial_x b_\theta^\phi)\|_{L^2\to L^2} \tag{11.2}$$

Assuming the claim, the proof can be completed as follows. Let us restore the dependence of the symbols on t. By the definition of $D_\theta L^q L^r$, we have

$$\|\Big(\sum_\phi \big(\sum_{n=1}^{20} \sup_\omega m_\theta^\phi(\omega)\|\theta^n \partial_\xi^{(n)} a(t,\cdot,\omega)\|_{L^\infty}\big)^2\Big)^{\frac{1}{2}}\|_{L_t^{p_2}[I]} \lesssim \|\theta\partial_\xi a\|_{D_\theta L^{p_2} L^\infty[I]}$$

On the other hand, by L^2-almost orthogonality of $\tilde{m}_\theta^\phi(\xi)$ and Hölder in t, we have

$$\|\Big(\sum_\phi \|Op(\theta^{-1}\partial_x b_\theta^\phi)\|_{L^2\to L^2}^2\Big)^{\frac{1}{2}}\|_{L_t^{p_0}[I]} \lesssim \|Op(\theta^{-1}\partial_x b)P_0\|_{L^q L^2 \to L^{p_1} L^2[I]}$$

where $r^{-1} + p_0^{-1} = p_1^{-1}$. Therefore, by Cauchy–Schwarz in ϕ and Hölder in t, (11.1) would follow.

We now turn to the proof of (11.2). For simplicity of notation, we use the shorthands $a = a_\theta^\phi$ and $b = b_\theta^\phi$ for now. Then the kernel of $Op(a)Op(b) - Op(ab)$ can be computed as follows:

$$\begin{aligned}
K(x,y) &= \int e^{i(x-z)\cdot\xi} e^{i(z-y)\cdot\eta}(a(x,\xi) - a(x,\eta))b(z,\eta)\, dz \frac{d\xi}{(2\pi)^4}\frac{d\eta}{(2\pi)^4} \\
&= \int_0^1 \int e^{i(x-z)\cdot\xi} e^{i(z-y)\cdot\eta}(\xi - \eta)\cdot(\partial_\xi a)(x, s\xi + (1-s)\eta) b(z,\eta)\, dz \frac{d\xi}{(2\pi)^4}\frac{d\eta}{(2\pi)^4}\, ds \\
&= -i \int_0^1 \int e^{i(x-z)\cdot\xi} e^{i(z-y)\cdot\eta}(\partial_\xi a)(x, s\xi + (1-s)\eta)(\partial_x b)(z,\eta)\, dz \frac{d\xi}{(2\pi)^4}\frac{d\eta}{(2\pi)^4}\, ds.
\end{aligned}$$

Expanding $\partial_\xi a(x,\cdot) = \int e^{-i(\cdot)\cdot\Xi}(\partial_\xi a)^\vee(x,\Xi)\, d\Xi$ and making the change of variables $\tilde{z} = z - (1-s)\Xi$, we further compute

$$\begin{aligned}
K(x,y) &= -i \int_0^1 \int e^{i(x-s\Xi-z)\cdot\xi} e^{i(z-(1-s)\Xi-y)\cdot\eta}(\partial_\xi a)^\vee(x,\Xi)(\partial_x b)(z,\eta)\, d\Xi\, dz \frac{d\xi}{(2\pi)^4}\frac{d\eta}{(2\pi)^4}\, ds \\
&= -i \int_0^1 \int e^{i(x-\Xi-\tilde{z})\cdot\xi} e^{i(\tilde{z}-y)\cdot\eta}(\partial_\xi a)^\vee(x,\Xi)(\partial_x b)(\tilde{z} + (1-s)\Xi, \eta)\, d\Xi\, d\tilde{z} \frac{d\xi}{(2\pi)^4}\frac{d\eta}{(2\pi)^4}\, ds \\
&= -i \int_0^1 \int (\partial_\xi a)^\vee(x,\Xi)\left(\int e^{i(x-s\Xi-y)\cdot\eta}(\partial_x b)(x - s\Xi, \eta)\, \frac{d\eta}{(2\pi)^4}\right) d\Xi\, ds
\end{aligned}$$

On the last line, observe that the η-integral inside the parentheses is precisely the kernel of $Op(\partial_x b)(x - s\Xi, D)$. By translation invariance, we have

$$\theta^{-1}\|(\partial_x b)(x - s\Xi, D)\|_{L^2\to L^2} = \|(\theta^{-1}\partial_x b)(x, D)P_0\|_{L^2\to L^2}$$

On the other hand, returning to the full notation $a_\theta^\phi = a$ and rotating the axes so that $\phi = (1,0,0,0)$, note that $a_\theta^\phi(x,\cdot)$ is supported on a rectangle of dimension $\simeq 1\times\theta\times\theta\times\theta$, and smooth on the corresponding scale. Integrating by parts in ξ to obtain rapid decay in Ξ (of the form $\langle \Xi^1 \rangle^{-N} \langle \theta \Xi' \rangle^{-N}$, where $\Xi' = (\Xi^2, \Xi^3, \Xi^4)$), we may estimate

$$\theta \int \|(\partial_\xi a^\phi_\theta)^\vee(\cdot, \Xi)\|_{L^\infty} d\Xi \leq \int \| \int e^{i\Xi\cdot\xi} \theta \partial_\xi a(\cdot, \xi) (m^\phi_\theta)^2(\xi) \tilde{m}^2_0(\xi) \, \frac{d\xi}{(2\pi)^4} \|_{L^\infty} \, d\Xi$$
$$\lesssim \theta^{-3} \sum_{n=1}^{20} \int \|\theta^n \partial^{(n)}_\xi a(\cdot, \xi)\|_{L^\infty} m^\phi_\theta(\xi) \tilde{m}_0(\xi) \, d\xi.$$

Passing to the polar coordinates $\xi = \lambda\omega$ (where $\lambda = |\xi|$), integrating out λ and using Hölder in ω (which cancels the factor θ^{-3}), we arrive at

$$\theta \int \|(\partial_\xi a^\phi_\theta)^\vee(\cdot, \Xi)\|_{L^\infty} d\Xi \lesssim \sum_{n=1}^{20} \sup_\omega m^\phi_\theta(\omega) \|\theta^n \partial^{(n)}_\xi a(\cdot, \omega)\|_{L^\infty},$$

which proves (11.2). □

Remark 11.2. As it is evident from the proof, we in fact have the simpler bound

$$\|(Op(a)Op(b) - Op(ab))P_0\|_{L^q L^2[I] \to L^r L^2[I]} \lesssim \|a\|_{D_\theta L^{p_2} L^\infty[I]} \|Op(\theta^{-1}\partial_x b) P_0\|_{L^q L^2[I] \to L^{p_1} L^2[I]}, \tag{11.1'}$$

In other words, control of the $D_\theta L^{p_2} L^\infty$-norm already encodes the fact that a is smooth in ξ on the scale θ.

In practice, Lemma 11.1 can be only be applied when we know that the symbol on the right (b in Lemma 11.1) is smooth in x on the scale θ^{-1}. Fortunately, when $b = Ad(O)$, the remainder can be controlled using decomposability bounds for Ψ. We therefore have the following useful composition lemma.

Lemma 11.3 (Composition lemma)**.** *Let $G = G(t, x, \xi)$ be a smooth $\mathfrak{g}$-valued symbol on $I \times \mathbb{R}^4 \times \mathbb{R}^4$, which is homogeneous of degree 0 in ξ and admits a decomposition of the form $G = \sum_{\theta \in 2^{-\mathbb{N}}} G^{(\theta)}$, where*

$$\|G^{(\theta)}\|_{D_\theta L^2 L^\infty[I]} \leq \theta^\alpha B$$

for some $B > 0$ and $\alpha > \frac{1}{2} + \delta$. Then for every $\ell \leq 0$ we have

$$\|Op(ad(G)Ad(O_{<\ell}))P_0 - Op(ad(G))Op(Ad(O_{<\ell}))P_0\|_{N^*[I] \to N[I]} \lesssim_M B. \tag{11.3}$$

Proof. Let us assume that $\ell > h_\theta - 20$, as the alternative case is easier.

We decompose the expression on the LHS of (11.3) into $\sum_{\theta \in 2^{-\mathbb{N}}} D^{(\theta)}$, where

$$D^{(\theta)} = Op(ad(G^{(\theta)})Ad(O_{<\ell}))P_0 - Op(ad(G^{(\theta)}))Op(Ad(O_{<\ell}))P_0.$$

In order to reduce to the case when Lemma 11.1 is applicable, we introduce $h_\theta = \log_2 \theta$ and further decompose $D^{(\theta)}$ as follows:

$$\begin{aligned} D^{(\theta)} = &\int_{h_\theta - 20}^{\ell} Op(ad(G^{(\theta)})ad(\Psi_h)Ad(O_{<h}))P_0 \, dh \\ &- \int_{h_\theta - 20}^{\ell} Op(ad(G^{(\theta)}))Op(ad(\Psi_h)Ad(O_{<h}))P_0 \, dh \\ &+ Op(ad(G^{(\theta)})Ad(O_{<h_\theta - 20})_{\geq h_\theta - 10})P_0 - Op(ad(G^{(\theta)}))Op(Ad(O_{<h_\theta-20})_{h_\theta - 10})P_0 \\ &+ Op(ad(G^{(\theta)})Ad(O_{<h_\theta - 20})_{<h_\theta - 10})P_0 - Op(ad(G^{(\theta)}))Op(Ad(O_{<h_\theta-20})_{<h_\theta - 10})P_0. \end{aligned}$$

We claim that

$$\|D^{(\theta)}\|_{L^\infty L^2[I]\to L^1 L^2[I]} \lesssim \theta^{\alpha-\frac{1}{2}} B \tag{11.4}$$

Assuming (11.4), the proof can be completed by simply summing up in $\theta \in 2^{-\mathbb{N}}$, which is possible since $\alpha > \frac{1}{2} + \delta$.

For the first term in the above splitting of $D^{(\theta)}$, we have

$$\begin{aligned}
&\int_{h_\theta - 20}^{\ell} \|Op(ad(G^{(\theta)})ad(\Psi_h)Ad(O_{<h}))P_0\|_{L^\infty L^2[I]\to L^1 L^2[I]}\, dh \\
\lesssim_M &\int_{h_\theta - 20}^{\ell} \|G^{(\theta)}\|_{D_\theta L^2 L^\infty[I]} \|\Psi_h\|_{DL^2 L^\infty[I]}\, dh \\
\lesssim_M &\int_{h_\theta - 20}^{\ell} \theta^\alpha 2^{(-\frac{1}{2}-\delta)h} B \lesssim_M \theta^{\alpha - \frac{1}{2} - \delta} B.
\end{aligned}$$

The second term can be handled similarly. For the third term, we use the $DL^2 L^\infty$ bound for $G^{(\theta)}$ and apply Lemma 10.12 to $Ad(O_{<h_\theta - 20})_{\geq h_\theta - 10})$, which leads to the acceptable bounds

$$\begin{aligned}
\|Op(ad(G^{(\theta)})Ad(O_{<h_\theta-20})_{\geq h_\theta - 10})P_0\|_{L^\infty L^2[I]\to L^1 L^2[I]} &\lesssim_M \theta^\alpha B, \\
\|Op(ad(G^{(\theta)}))Op(Ad(O_{<h_\theta-20})_{\geq h_\theta - 10})P_0\|_{L^\infty L^2[I]\to L^1 L^2[I]} &\lesssim_M \theta^\alpha B.
\end{aligned}$$

Finally, for the last term we use Lemma 11.1 (in fact, (11.1′)). ☐

11.2. **Decomposition of the error.** Let

$$E = \Box_A^{p,\kappa} Op(Ad(O)_{<0}) - Op(Ad(O)_{<0})\Box$$

We may decompose

$$E = E_1 + \cdots + E_6$$

where

$$\begin{aligned}
E_1 &= 2iOp\left(\left(ad(\omega \cdot A_{x,<-\kappa} + A_{0,<-\kappa} + L_+^\omega \Psi)Ad(O)\right)_{<0}\right)|D_x| \\
E_2 &= 2iOp\left(\left(ad(\omega \cdot O_{;x} + O_{;t} - L_+^\omega \Psi)Ad(O)\right)_{<0}\right)|D_x|, \\
E_3 &= 2Op\left(ad(A_{\alpha,<-\kappa})\left(ad(O^{;\alpha})Ad(O)\right)_{<0}\right) + Op\left(\left(ad(O_{;\alpha})ad(O^{;\alpha})Ad(O)\right)_{<0}\right), \\
E_4 &= Op\left(\left(ad(\partial^\alpha O_{;\alpha})Ad(O)\right)_{<0}\right), \\
E_5 &= -2iOp\left(ad(A_{0,<-\kappa})Ad(O)_{<0}\right)(D_t + |D_x|) - 2iOp\left(\left(ad(O_{<-\kappa;t})Ad(O)\right)_{<0}\right)(D_t + |D_x|), \\
E_6 &= -2iOp\left([S_{<0}, ad(\omega \cdot A_{x,<-\kappa} + A_{0,<-\kappa})]Ad(O)\right)|D_x|.
\end{aligned}$$

In the remainder of this section, we estimate each error term in order.

11.3. **Estimate for E_1.** Here, our goal is to prove

$$\|E_1 P_0\|_{S_0^\sharp[I]\to N[I]} \leq \varepsilon \tag{11.5}$$

with κ_1 large enough and δ_p sufficiently small.

11.3.1. *Preliminary reduction.* For this term, we may simply work with $I = \mathbb{R}$ by extending the input by homogeneous waves outside I. The desired smallness comes from κ and bounds for $\Box A_x$ and ΔA_0 on I, which controls the size of the symbol of E_1 through our extension of A_α as in Section 9.2

We first dispose the symbol regularization $(\cdot)_{<0}$ by translation invariance, and also throw away $|D_x|$ using P_0. Using (9.42) and the identity $L_+^\omega L_-^\omega \Delta_{\omega^\perp}^{1} = -\Delta_{\omega^\perp}^{-1}\Box + 1$, (11.5) reduces to showing

$$\| \int_{-\infty}^{-\kappa} Op\left(ad(G_h)Ad(O)\right) P_0 \, dh\|_{S_0^\sharp \to N} \ll \varepsilon,$$

where

$$G_h = \omega \cdot A_{x,h} - \omega \cdot A_{x,h,cone}^{(\geq|\eta|^\delta)} + \Delta_{\omega^\perp}^{-1}\Box(\omega \cdot A_{x,h,cone}^{(\geq|\eta|^\delta)}) + A_{0,h}.$$

Note that each angular component $G_h^{(\theta)} = \Pi_\theta^{\omega,+} G_h$ obeys

$$\|G_h^{(\theta)}\|_{DL^2L^\infty} \lesssim 2^{\frac{1}{2}h}\theta^{\frac{3}{2}}(\|A_{x,h}\|_{S^1} + \|A_{0,h}\|_{Y^1}).$$

Therefore, by Lemma 11.3, we have

$$\| \int_{-\infty}^{-\kappa} \left(Op(ad(G_h)Ad(O)) - Op(ad(G_h))Op(Ad(O))\right) P_0 \, dh\|_{N^* \to N} \lesssim_M 2^{-\frac{1}{2}\kappa},$$

which is acceptable. By Lemma 10.12 applied to $Op(Ad(O)_{\geq 0})$, we also have

$$\begin{aligned}\| \int_{-\infty}^{-\kappa} Op(ad(G_h))Op(Ad(O)_{\geq 0})P_0 \, dh\|_{N^* \to N} &\lesssim_M \int_{-\infty}^{-\kappa} 2^{\frac{1}{2}h}\|Op(Ad(O)_{\geq 0})P_0\|_{L^\infty L^2 \to L^2L^2} \, dh \\ &\lesssim_M 2^{-\frac{1}{2}\kappa}.\end{aligned}$$

Thus it suffices to show that

$$\| \int_{-\infty}^{-\kappa} Op(ad(G_h))Op(Ad(O)_{<0})P_0 \, dh\|_{S_0^\sharp \to N} \ll \varepsilon.$$

By (9.45), we have $Op(Ad(O)_{<0})P_0 : S_0^\sharp \to S_0$. Thus, in order to prove (11.5), we are left to establish

$$\| \int_{-\infty}^{-\kappa} Op(ad(G_h))P_0 \, dh\|_{S_0 \to N} \ll \varepsilon. \tag{11.6}$$

where we abuse the notation a bit and denote by P_0 a frequency projection to a slightly enlarged region of the form $\{|\xi| \simeq 1\}$.

At this point it is convenient to observe that the contribution of $\tilde{R}_0$ to A_0 in (9.27) is easy to estimate in L^1L^∞ and can be harmlessly discarded. Thus from here on we assume that

$$\tilde{R}_0 = 0. \tag{11.7}$$

In order to proceed, we split

$$G_h = G_{h,cone} + G_{h,null} + G_{h,out},$$

where

$$\begin{aligned}G_{h,cone} &= \omega \cdot A_{x,h,cone}^{(<|\eta|^\delta)} + \Delta_{\omega^\perp}^{-1}\Box(\omega \cdot A_{x,h,cone}^{(\geq|\eta|^\delta)}) + A_{0,h,cone}, \\ G_{h,null} &= \omega \cdot A_{x,h,null} + A_{0,h,null}, \\ G_{h,out} &= \omega \cdot A_{x,h,out} + A_{0,h,out}.\end{aligned}$$

11.3.2. *Estimate for $G_{h,cone}$.* We claim that

$$\| \int_{-\infty}^{-\kappa} Op(ad(G_{h,cone}))P_0 \, dh\|_{N^* \to N} \ll \varepsilon. \tag{11.8}$$

Let $G^{(\theta)}_{h,cone} = \Pi^{\omega,\pm}_{\theta} G_{h,cone}$ and consider the expression $Op(ad(G^{(\theta)}_{h,cone}))P_0$. By the Fourier support property of $G^{(\theta)}_{h,cone}$ (more precisely, the mismatch between its modulation $\lesssim 2^h \theta^2$ and the angle θ), it is impossible that both the input and the output have modulation $\ll 2^h \theta^2$. Using the $L^2 L^2$ norm for the input or the output (whichever that has modulation $\gtrsim 2^h\theta^2$), we may estimate

$$\begin{aligned}
&\|Op(G_{h,cone})P_0\|_{N^* \to N} \\
&\quad \lesssim \sum_{\theta < 1} 2^{-\frac{1}{2}h} \theta^{-1} \|G^{(\theta)}_{h,cone}\|_{DL^2 L^\infty} \\
&\quad \lesssim 2^{\frac{\delta}{2}h} \|A_{x,h}\|_{S^1} + \sum_{\theta<1} 2^{-\frac{1}{2}h} \theta^{-\frac{1}{2}} \|Q_{<h+2\log_2 \theta + C} \Box A_x\|_{L^2 L^2} + \sum_{\theta<1} 2^{-\frac{1}{2}h} \theta^{\frac{1}{2}} \|\Delta A_{0,h}\|_{L^2 L^2}.
\end{aligned}$$

We now treat each term separately.

Case 1: Contribution of small angle interaction. The term $2^{\frac{\delta}{2}h}\|A_{x,h}\|_{S^1}$ is acceptable since it is integrable in $-\infty < h < -\kappa$, and we gain a small factor $2^{-\frac{\delta}{2}\kappa}$ as a result.

Case 2: Contribution of $\Box A_x$. For the second term, we split the θ-summation into $\theta < 2^{-\kappa}$ and $\theta \geq 2^{-\kappa}$. In the former case, note that

$$\|Q_{<h+2\log_2\theta+C} \Box A_x\|_{L^2L^2} \lesssim \theta^{2b_1} \|\Box A_{x,h}\|_{X^{-\frac{1}{2}+b_1, -b_1}}.$$

Since $b_1 > 1/4$, we may estimate

$$\sum_{\theta < 2^{-\kappa}} 2^{-\frac{1}{2}h} \theta^{-\frac{1}{2}} \|Q_{<h+2\log_2\theta+C} \Box A_x\|_{L^2L^2} \lesssim 2^{-(2b_1 - \frac{1}{2})\kappa} \|\Box A_{x,h}\|_{X^{-\frac{1}{2}+b_1, -b_1}}.$$

The last line is acceptable, since it is integrable in $-\infty < h < -\kappa$, and it is small thanks to $2^{-(2b-\frac{1}{2})\kappa}$. In the case $\theta \geq 2^{-\kappa}$, we estimate

$$\sum_{\theta \geq 2^{-\kappa}} 2^{-\frac{1}{2}h} \theta^{-\frac{1}{2}} \|Q_{<h+2\log_2\theta+C} \Box A_x\|_{L^2L^2} \lesssim 2^{\frac{1}{2}\kappa} \|\Box A_{x,h}\|_{L^2 \dot{H}^{-\frac{1}{2}}}.$$

After integration in h, this is acceptable thanks to (9.22).

Case 3: Contribution of A_0. In this case, we simply sum up in $\theta < 1$ and observe that

$$\sum_{\theta<1} 2^{-\frac{1}{2}h} \theta^{\frac{1}{2}} \|\Delta A_{0,h}\|_{L^2L^2} \lesssim \|\Delta A_{0,h}\|_{L^2 \dot{H}^{-\frac{1}{2}}}.$$

After integration in h, this term is then acceptable by (9.29).

11.3.3. *Estimate for $G_{h,out}$.* We claim that

$$\| \int_{-\infty}^{-\kappa} Op(ad(G_{h,out}))P_0 \, dh\|_{N^* \to N} \ll \varepsilon. \tag{11.9}$$

As in the case of $G_{h,cone}$, the idea is again to make use of the mismatch between modulation of $G_{h,out}$ and the angle θ. Let $G^{(\theta)}_{h,out} = \Pi^{\omega,\pm}_{\theta} G_{h,out}$, and consider the expression

$Op(ad(G^{(\theta)}_{h,out}))P_0$. By definition, $G^{(\theta)}_{h,out}$ has modulation $\gtrsim 2^h\theta^2$. Thus, we decompose $G^{(\theta)}_{h,out} = \sum_{a:2^a\gtrsim\theta} Q_{h+2a}G^{(\theta)}_{h,out}$. By the Fourier support property of the symbol $Q_{h+2a}G^{(\theta)}_{h,out}$ (more precisely, the mismatch between the angle θ and the modulation 2^{h+2a}), it is impossible that both the input and the output have modulation $\ll 2^{h+2a}$. Using the L^2L^2 norm for the input or the output, we have

$$\begin{aligned}
&\|Op(ad(G_{h,out}))P_0\|_{N^*\to N}\\
&\lesssim \sum_a \sum_{\theta<\min\{C2^a,1\}} 2^{-\frac{1}{2}(h+2a)}\|Q_{h+2a}G^{(\theta)}_{h,out}\|_{DL^2L^\infty}\\
&\lesssim \sum_a \sum_{\theta<\min\{C2^a,1\}} \left(2^{-\frac{1}{2}(h+2a)}2^{2h}\theta^{\frac{5}{2}}\|Q_{h+2a}A_{x,h}\|_{L^2L^2} + 2^{-\frac{1}{2}(h+2a)}2^{2h}\theta^{\frac{3}{2}}\|A_{0,h}\|_{L^2L^2}\right)\\
&\lesssim \sum_a \left(2^{\frac{5}{2}a_-}2^{-3a}2^{-\frac{1}{2}h}\|Q_{h+2a}\Box A_{x,h}\|_{L^2L^2} + 2^{\frac{3}{2}a_-}2^{-a}2^{-\frac{1}{2}h}\|\Delta A_{0,h}\|_{L^2L^2}\right).
\end{aligned}$$

We split the a-summation into $a<-\kappa$ and $a>-\kappa$. In the former case, the sum is bounded by

$$2^{-(2b_1-\frac{1}{2})\kappa}\|\Box A_{x,h}\|_{X^{b_1-\frac{1}{2},-b_1}} + 2^{-\frac{1}{2}\kappa}\|\Delta A_{0,h}\|_{L^2\dot{H}^{-\frac{1}{2}}},$$

which is integrable in h and small thanks to $2^{-(2b_1-\frac{1}{2})\kappa}$; therefore it is acceptable. When $a>-\kappa$, the sum is bounded by

$$2^{\frac{1}{2}\kappa}\|\Box A_{x,h}\|_{L^2\dot{H}^{\frac{1}{2}}} + \|\Delta A_{0,h}\|_{L^2\dot{H}^{-\frac{1}{2}}}.$$

After integrating in h, this term is therefore acceptable by (9.22) and (9.29).

11.3.4. *Estimate for $G_{h,null}$.* We claim that

$$\|\int_{-\infty}^{-\kappa} Op(ad(G_{h,null}))P_0\,dh\|_{S_0\to N} \ll \varepsilon. \tag{11.10}$$

Let $G^{(\theta)}_{h,null} = \Pi^{\omega,\pm}_\theta G_{h,null}$. Note that $G^{(\theta)}_{h,null}$ has modulation $\simeq 2^h\theta^2$. Hence if either the input or the output have modulation $\geq 2^{-C}2^h\theta^2$, the same argument as in the case of $G_{h,cone}$ applies. Writing $\theta = 2^\ell$, it remains to prove

$$\|\sum_{\ell\in-\mathbb{N}}\int_{-\infty}^{-\kappa} Q_{<h+2\ell-C}Op(ad(\omega\cdot A^{(2^\ell)}_{x,h,null} + A^{(2^\ell)}_{0,h,null}))P_0Q_{<h+2\ell-C}\,dh\|_{S_0\to N} \ll \varepsilon. \tag{11.11}$$

Our next simplification is to observe that we can harmlessly replace the symbols $A^{(2^\ell)}_{x,h,null}$ and $A^{(2^\ell)}_{0,h,null}$ with the functions $Q_{h+2\ell}A_{x,h}$ respectively $Q_{h+2\ell}A_{x,h}$. This is because the difference of the two is localized still at modulation $2^{h+2\ell}$, but also at distance $2^{h+2\ell}$ from the null plane $\{\sigma+\omega\cdot\eta=0\}$. This would force either the input or the output modulation in (11.11) to be $\geq 2^{-C}2^{h+2\ell}$, and again the same argument as in the case of $G_{h,cone}$ applies. Thus with $j=h+2\ell$ we have reduced the problem to estimating

$$\|\sum_{j<h}\int_{-\infty}^{-\kappa} Q_{<j-C}ad(Q_jA_{\alpha,h})\partial^\alpha P_0Q_{<j-C}\,dh\|_{S_0\to N} \ll \varepsilon. \tag{11.12}$$

respectively

$$\| \sum_{j<h} \int_{-\infty}^{-\kappa} Q_{<j-C} ad(Q_j A_{0,h})(D_0 + |D_x|) P_0 Q_{<j-C}\, dh \|_{S_0 \to N} \ll \varepsilon. \tag{11.13}$$

The second bound is straightforward since $(D_0 + |D_x|)P_0 Q_{<0} : S_0 \to L^2$ and $A_0 \in L^2 \dot{H}^{\frac{3}{2}}$.

Thus it remains to consider (11.12). From here on, we assume that A is determined by the expressions (9.27) and (9.30) in terms of $\tilde{A}$. By (11.7) we have already set $\tilde{R}_0 = 0$. It is equally easy to see that we can set $\tilde{R}_x = 0$. Indeed, by (4.6) and (8.32) we have

$$\|Q_{<j-C} ad(\Box^{-1} P_h R_\ell) \partial^\ell P_0 Q_{<j-C}\|_{S_0 \to N} \lesssim 2^{\delta_1 (j-h)} \|\Box^{-1} P_h R_\ell\|_{Z^1} \lesssim 2^{\delta_1(j-h)} \|P_h R_\ell\|_{L^1 L^2},$$

where $R = \chi_I \mathbf{P} \tilde{R}$. Now the summability in $j < h$ and the smallness is assured due to (9.26).

Once we have dispensed with the error terms, we are left with $A_{t,x}$ given by

$$A_0 = \Delta^{-1} \mathbf{O}(\chi_I \tilde{A}^\ell, \partial_t \tilde{A}_\ell) \tag{11.14}$$

$$A = \Box^{-1} \mathbf{P}(\mathbf{O}(\chi_I \tilde{A}^\ell, \partial_x \tilde{A}_\ell) + \mathbf{O}'(\mathbf{P}_\ell \tilde{A}, \chi_I \partial^\ell \tilde{A}) - \mathbf{O}'(\tilde{A}_0, \chi_I \partial_t \tilde{A}) + \mathbf{O}'(\tilde{G}_\ell, \chi_I \partial^\ell \tilde{A})). \tag{11.15}$$

We consider the contributions of each of these terms in (11.12).

1. The contribution of $A_0 = \Delta^{-1}\mathbf{O}(\chi_I \tilde{A}^\ell, \partial_t \tilde{A}_\ell)$ and $A_x = \Box^{-1}\mathbf{PO}(\chi_I \tilde{A}^\ell, \partial_x \tilde{A}_\ell)$. This is the main component, which we have to treat in a trilinear fashion. In particular we have to insure that we gain smallness. For this we use a trilinear Littlewood-Paley decomposition to set

$$A = \sum_{k,k_1,k_2} A(k,k_1,k_2) = \sum_{k,k_1,k_2} \mathcal{H} A(k,k_1,k_2) + \sum (1-\mathcal{H}^*) A(k,k_1,k_2)$$

where

$$\mathcal{H}A(k,k_1,k_2) := \mathcal{H} P_k \mathbf{P} A(P_{k_1} \chi_I \tilde{A}^\ell, P_{k_2} \partial_t \tilde{A}_\ell)$$
$$(1-\mathcal{H})A(k,k_1,k_2) := (1-\mathcal{H}) P_k \mathbf{P} A(P_{k_1} \chi_I \tilde{A}^\ell, P_{k_2} \partial_t \tilde{A}_\ell)$$

For the terms in the first sum we use the trilinear estimate (8.45), which gives

$$\|Q_{<j-C} ad(Q_j \mathcal{H} A_\alpha(k,k_1,k_2)) \partial^\alpha P_0 Q_{<j-C}\|_{S_0 \to L^1 L^2} \lesssim 2^{-\delta_1 |k_{max} - k_{min}|} 2^{\delta_1 (j-k)} \|P_{k_1} \tilde{A}\|_{S^1} \|P_{k_2} \tilde{A}\|_{S^1}$$

For the A_x terms in the second sum we first use (8.23) and (8.35), (8.36) to obtain

$$\|(1-\mathcal{H}) A_x(k,k_1,k_2)\|_{Z^1} \lesssim 2^{-\delta_1 |k_{max} - k_{min}|} \|P_{k_1} \tilde{A}\|_{S^1} \|P_{k_2} \tilde{A}\|_{S^1}$$

and then use (8.32) to conclude that

$$\|Q_{<j-C} ad(Q_j (1-\mathcal{H}) A_\ell(k,k_1,k_2)) \partial^\ell P_0 Q_{<j-C}\|_{S_0 \to N} \lesssim 2^{-\delta_1 |k_{min} - k_{max}|} 2^{\delta_1 (j-k)} \|P_{k_1} \tilde{A}\|_{S^1} \|P_{k_2} \tilde{A}\|_{S^1}$$

Similarly, for the A_0 terms in the second sum we use (8.37) and then (8.33) to obtain

$$\|Q_{<j-C} ad(Q_j (1-\mathcal{H}) A_0(k,k_1,k_2)) \partial^0 P_0 Q_{<j-C}\|_{S_0 \to N} \lesssim 2^{-\delta_1 |k_{min} - k_{max}|} 2^{\delta_1 (j-k)} \|P_{k_1} \tilde{A}\|_{S^1} \|P_{k_2} \tilde{A}\|_{S^1}$$

Adding the last three bounds, we obtain

$$\|Q_{<j-C} ad(Q_j A_\alpha(k,k_1,k_2)) \partial^\alpha P_0 Q_{<j-C}\|_{S_0 \to N} \lesssim 2^{-\delta_1 |k_{max} - k_{min}|} 2^{\delta_1 (j-k)} \|P_{k_1} \tilde{A}\|_{S^1} \|P_{k_2} \tilde{A}\|_{S^1}.$$

This gives both summability in k, k_1, k_2 and smallness provided we exclude the range of indices $j, k_1, k_2 \in [k - \kappa', k + \kappa']$ with $\kappa' \gg 1$.

On the other hand, in the range excluded above the operator $P_k Q_j$ is disposable while both $\Box$ and Δ are elliptic, i.e. of size 2^{2k}. Then we can estimate

$$\|Q_j A(k,k_1,k_2)\|_{L^1 L^\infty} \lesssim 2^{C\kappa'} \|P_{k_1}\tilde{A}\|_{DS^1} \|P_{k_2}\tilde{A}\|_{DS^1}$$

therefore we gain smallness from the divisible norm, see (9.5).

2. The contribution of $A_x = \Box^{-1}\mathbf{PO}'(\mathbf{P}_\ell \tilde{A}, \chi_I \partial^\ell \tilde{A})$. This is a milder contribution, which we can deal with in a bilinear fashion. Decomposing again

$$A_x = \sum_{k,k_1,k_2} A(k,k_1,k_2)$$

we use (8.40) to obtain

$$\|A_x(k,k_1,k_2)\|_{Z^1} \lesssim 2^{-\delta_1|k_{\max}-k_{min}|} \|P_{k_1}\tilde{A}\|_{\underline{S}^1} \|P_{k_1}\tilde{A}\|_{S^1}$$

Then by (8.32) it follows that

$$\|Q_{<j-C} ad(Q_j \mathcal{H} A_x(k,k_1,k_2)) \partial^\alpha P_0 Q_{<j-C}\|_{S_0 \to L^1 L^2} \lesssim 2^{-\delta_1|k_{max}-k_{min}|} 2^{\delta_1(j-k)} \|P_{k_1}\tilde{A}\|_{\underline{S}^1} \|P_{k_1}\tilde{A}\|_{S^1} \tag{11.16}$$

Again this is suitable outside the range $j, k_1, k_2 \in [k-\kappa', k+\kappa']$ with $\kappa' \gg 1$, whereas in this range we can use divisible norms as in the previous step.

3. The contribution of $\mathbf{PO}'(\tilde{A}_0, \chi_I \partial_t \tilde{A}) + \mathbf{PO}'(\tilde{G}_\ell, \chi_I \partial^\ell \tilde{A})$. These two terms are similar, as we have the same bounds available for $\tilde{A}_0$ and $\tilde{G}_l$. We will discuss $\tilde{A}_0$. Setting

$$A_x = \Box^{-1}\mathbf{PO}'(\tilde{A}_0, \chi_I \partial_t \tilde{A}), \qquad A_0 = 0,$$

we decompose as before

$$A_x = \sum A_x(k,k_1,k_2)$$

We can estimate the terms in the sum using (8.43) to get

$$\|A_x(k,k_1,k_2)\|_{Z^1} \lesssim 2^{-\delta_1|k_{\max}-k_{min}|} \|P_{k_1}\tilde{A}_0\|_{Y^1} \|P_{k_1}\tilde{A}\|_{S^1}$$

Then (11.16) follows again from (8.32), and we conclude as in Step 2.

11.4. **Estimate for E_2.** Our next goal is to estimate the error term E_2, which arises from the multilinear error between $O_{;\alpha}$ and $\partial_\alpha \Psi$. For this purpose, we rely crucially on interval localization of decomposable norms (Lemma 10.7).

11.4.1. *Expansion of $O_{;\alpha}$.* We will prove that

$$\|E_2 P_0\|_{N^*[I] \to N[I]} \leq \varepsilon \tag{11.17}$$

provided that κ_1 is large enough, and δ_p is sufficiently small.

As usual, we may dispose the symbol regularization $(\cdot)_{<0}$ by translation invariance. Also disposing $|D_x|$ using P_0, it suffices to prove

$$\|Op\left(ad(\omega \cdot (O_{;x} - \partial_x \Psi) + (O_{;t} - \partial_t \Psi)) Ad(O)\right) P_0\|_{N^*[I] \to N[I]} \ll \varepsilon. \tag{11.18}$$

Recall that $\partial_h O_{<h;\alpha} = \Psi_{h,\alpha} + [\Psi_h, O_{<h;\alpha}]$. Therefore,

$$\partial_h \left(ad(O_{<h;\alpha}) Ad(O_{<h})\right) = ad(\partial_\alpha \Psi_h) Ad(O_{<h}) + ad(\Psi_h) Ad(O_{<h;\alpha}) Ad(O_{<h}).$$

Repeatedly applying the fundamental theorem of calculus and this equation, we obtain the expansion

$$
\begin{aligned}
&ad(O_{;\alpha})Ad(O) \\
&= \int_{-\infty}^{-\kappa} ad(\partial_\alpha \Psi_{h_1}) Ad(O_{<h_1})\, dh_1 && (11.19)\\
&\quad + \int_{-\infty}^{-\kappa}\int_{-\infty}^{h_1} ad(\Psi_{h_1}) ad(\partial_\alpha \Psi_{h_2}) Ad(O_{<h_2})\, dh_2\, dh_1 && (11.20)\\
&\quad + \cdots \\
&\quad + \int_{-\infty}^{-\kappa}\int_{-\infty}^{h_1}\cdots\int_{-\infty}^{h_5} ad(\Psi_{h_1}) ad(\Psi_{h_2})\cdots ad(\partial_\alpha \Psi_{h_6}) Ad(O_{<h_6}) dh_6 \cdots dh_2\, dh_1. && (11.21)
\end{aligned}
$$

On the other hand,

$$\partial_h \left(ad(\partial_\alpha \Psi_{<h}) Ad(O_{<h})\right) = ad(\partial_\alpha \Psi_h) Ad(O_{<h}) + ad(\partial_\alpha \Psi_{<h}) ad(\Psi_h) Ad(O_{<h}),$$

so we have

$$
\begin{aligned}
ad(\partial_\alpha \Psi) Ad(O) = &\int_{-\infty}^{-\kappa} ad(\partial_\alpha \Psi_{h_1}) Ad(O_{<h_1})\, dh_1 && (11.22)\\
&+ \int_{-\infty}^{-\kappa}\int_{-\infty}^{h_1} ad(\partial_\alpha \Psi_{h_2}) ad(\Psi_{h_1}) Ad(O_{<h_1})\, dh_2\, dh_1. && (11.23)
\end{aligned}
$$

Observe that (11.19) and (11.22) coincide. Thus, we only need to consider the contribution of (11.20)–(11.21) and (11.23) in (11.18).

11.4.2. *Estimate for quadratic expressions.* We begin with the contribution of the quadratic terms in Ψ, namely (11.20) and (11.23), which are most delicate. We claim that

$$\| \int_{-\infty}^{-\kappa}\int_{-\infty}^{h_1} Op\left(ad(\Psi_{h_1}) ad(L_+^\omega \Psi_{h_2}) Ad(O_{<h_2})\right) P_0\, dh_2\, dh_1 \|_{N^*[I]\to N[I]} \leq \varepsilon, \qquad (11.24)$$

$$\| \int_{-\infty}^{-\kappa}\int_{-\infty}^{h_1} Op\left(ad(L_+^\omega \Psi_{h_2}) ad(\Psi_{h_1}) Ad(O_{<h_1})\right) P_0\, dh_2\, dh_1 \|_{N^*[I]\to N[I]} \leq \varepsilon, \qquad (11.25)$$

provided that κ_1 is large enough and δ_p is sufficiently small. In what follows, we will focus on establishing (11.24), as the proof for the other claim is analogous.

By (9.42) and the identity $L_+^\omega L_-^\omega \Delta_{\omega^\perp}^{1} = -\Delta_{\omega^\perp}^{-1}\Box + 1$, (11.24) would follow once we establish

$$\| \int_{-\infty}^{-\kappa}\int_{-\infty}^{h_1} Op\left(ad(\Psi_{h_1}) ad(\omega\cdot A_{h_2}^{main}) Ad(O_{<h_2})\right) P_0\, dh_2\, dh_1 \|_{N^*[I]\to N[I]} \ll \varepsilon, \qquad (11.26)$$

$$\| \int_{-\infty}^{-\kappa}\int_{-\infty}^{h_1} Op\left(ad(\Psi_{h_1}) ad(\Delta_{\omega^\perp}^{-1}\Box(\omega\cdot A_{h_2}^{main})) Ad(O_{<h_2})\right) P_0\, dh_2\, dh_1 \|_{N^*[I]\to N[I]} \ll \varepsilon. \qquad (11.27)$$

In Lemma 10.4 and Lemma 10.7, note that $\omega\cdot A_h^{main,(\theta)} (= \omega\cdot A_{x,h,cone,+}^{(\theta)})$ and $\Delta_{\omega^\perp}^{-1}\Box(\omega\cdot A_h^{main,(\theta)})$ obey the same bounds. Therefore, (11.26) and (11.27) are proved in exactly the same way. In what follows, we only consider (11.26).

Our first task is to remove $Ad(O_{<h_2})$. For $\theta\in 2^{-\mathbb{N}}$, define

$$G^{(\theta)} = ad(\Psi_{h_1}^{(\theta)}) ad(\omega\cdot A_{h_2}^{main,(<\theta)}) + ad(\Psi_{h_1}^{(\leq\theta)}) ad(\omega\cdot A_{h_2}^{main,(\theta)}).$$

so that $G := ad(\Psi_{h_1})ad(\omega \cdot A^{main}_{h_2}) = \sum_{\theta \in 2^{-\mathbb{N}}} G^{(\theta)}$. Note that

$$\|G^{(\theta)}\|_{DL^2L^\infty} \lesssim_M 2^{\frac{1}{2}h_1} 2^{\frac{1}{2}(h_2-h_1)} \theta^{\frac{3}{2}},$$

by Lemma 10.4 and Lemma 10.5. Applying Lemma 11.3, then integrating $-\infty < h_2 < h_1 < -\kappa$, it follows that

$$\| \int_{-\infty}^{-\kappa} \int_{-\infty}^{h_1} (Op(ad(G)Ad(O_{<h_2})) - Op(ad(G))Op(Ad(O_{<h_2})))\, P_0 \, dh_2 \, dh_1 \|_{N^*[I] \to N[I]} \lesssim 2^{-\frac{1}{2}\kappa}$$

which is acceptable. On the other hand, using the DL^2L^∞ bound for G and Lemma 10.12, we have

$$\begin{aligned}
&\| \int_{-\infty}^{-\kappa} \int_{-\infty}^{h_1} Op(ad(G))Op(Ad(O_{<h_2})_{\geq 0}) P_0 \, dh_2 \, dh_1 \|_{N^*[I] \to N[I]} \\
&\lesssim_M \int_{-\infty}^{-\kappa} \int_{-\infty}^{h_1} 2^{\frac{1}{2}h_1} 2^{\frac{1}{2}(h_2-h_1)} \|Op(Ad(O_{<h_2})_{\geq 0}) P_0\|_{L^\infty L^2[I] \to L^2 L^2[I]} \, dh_2 \, dh_1 \\
&\lesssim_M 2^{-\frac{1}{2}\kappa}
\end{aligned}$$

so we may replace $Op(Ad(O_{<h_2}))$ by $Op(Ad(O_{<h_2}))_{<0}$. Finally, by (9.44) we have

$$Op(Ad(O_{<h_2})_{<0}) P_0 : N^*[I] \to N^*[I],$$

so we are left to prove

$$\| \int_{-\infty}^{0} \int_{-\infty}^{h_1} Op\left(ad(\Psi_{h_1})ad(\omega \cdot A^{main}_{h_2})\right) \, dh_2 \, dh_1 \|_{N^*[I] \to N[I]} \ll \varepsilon. \tag{11.28}$$

In order to place ourselves in a context where we can apply Lemma 10.7, we begin by dispensing with the case of short intervals

$$|I| \leq 2^{-h_2 - C\kappa}$$

For very short intervals $|I| \leq 2^{-h_1 - C\kappa}$ we have the bound

$$\| \int_{-\infty}^{0} \int_{-\infty}^{h_1} Op\left(ad(\Psi_{h_1})ad(\omega \cdot A^{main}_{h_2})\right) \, dh_2 \, dh_1 \|_{L^\infty L^2 \to L^1 L^2} \lesssim_M 2^{h_2} |I|,$$

which is a consequence of fixed time decomposability bounds, namely (10.10) with $q = \infty$ and (10.14) with $q = \infty$ and $r = \infty$, combined with Holder's inequality in time. This suffices for the integration with respect to h_1 and h_2 in this range.

For merely short intervals $2^{-h_1 - C\kappa} \leq |I| \leq 2^{-h_2 - C\kappa}$ we are allowed to use spacetime decomposabilty bounds but only for Ψ_{h_1}. In this case we apply (10.10) with $q = \infty$ and (10.14) with $q = 6$ and $r = \infty$, combined with Holder's inequality in time, to obtain

$$\| \int_{-\infty}^{0} \int_{-\infty}^{h_1} Op\left(ad(\Psi_{h_1})ad(\omega \cdot A^{main}_{h_2})\right) \, dh_2 \, dh_1 \|_{L^\infty L^2 \to L^1 L^2} \lesssim_M 2^{-\frac{1}{6}h_1} 2^{h_2} |I|^{\frac{5}{6}}$$

This again suffices for the integration with respect to h_1 and h_2 in this range.

For large intervals, on the other hand, we will use Lemma 10.7. We begin by decomposing $\Psi_{h_1} = \sum_{\theta_1} \Psi^{(\theta_1)}_{h_1}$ and $A^{main}_{h_2} = \sum_{\theta_2} A^{main,(\theta_2)}_{h_2}$. First, we consider the case $2^{h_1}\theta_1^2 \geq 2^{-2\kappa} 2^{h_2} \theta_2^2$.

For fixed h_1, h_2 and θ_2, we use interval localized decomposability calculus to estimate

$$\begin{aligned}
&\sum_{\theta_1 \geq 2^{-\kappa} 2^{\frac{1}{2}(h_2-h_1)}\theta_2} \|Op\left(ad(\Psi_{h_1}^{(\theta_1)}) ad(\omega \cdot A_{h_2}^{main,(\theta_2)})\right)\|_{L^\infty L^2[I] \to L^1 L^2[I]} \\
&\lesssim \sum_{\theta_1 \geq 2^{-\kappa} 2^{\frac{1}{2}(h_2-h_1)}\theta_2} \|\Psi_{h_1}^{(\theta_1)}\|_{DL^2 L^\infty[I]} \|\omega \cdot A_{h_2}^{main,(\theta_2)}\|_{DL^2 L^\infty[I]} \\
&\lesssim 2^{\kappa} 2^{\frac{1}{4}(h_2-h_1)} \theta_2 \|A_{h_1}\|_{S^1} \left(2^{-\frac{1}{2}h_2} \theta_2^{-\frac{3}{2}} \|\omega \cdot A_{h_2}^{main,(\theta_2)}\|_{DL^2 L^\infty[I]}\right).
\end{aligned}$$

Summing up in $\theta_2 < 2^{-2\kappa}$, we see that

$$\begin{aligned}
&\sum_{\theta_2 < 2^{-2\kappa}} \sum_{\theta_1 \geq 2^{-\kappa} 2^{\frac{1}{2}(h_2-h_1)}\theta_2} \|Op\left(ad(\Psi_{h_1}^{(\theta_1)}) ad(\omega \cdot A_{h_2}^{main,(\theta_2)}) Ad(O_{<h_2})|\xi|\right)\|_{L^\infty L^2[I] \to L^1 L^2[I]} \\
&\lesssim 2^{-\kappa} 2^{\frac{1}{4}(h_2-h_1)} \|A_{h_1}\|_{S^1} \|A_{h_2}\|_{S^1},
\end{aligned}$$

which is acceptable. On the other hand, in the large angle case $\theta_2 \geq 2^{-2\kappa}$, we use Lemma 10.7 to bound

$$2^{-\frac{1}{2}h_2} \theta_2^{-\frac{3}{2}} \|\omega \cdot A_{h_2}^{main,(\theta_2)}\|_{DL^2 L^\infty[I]} \lesssim 2^{C\kappa} \|A_{h_2}\|_{DS^1[I]}.$$

When $2^{h_1}\theta_1^2 < 2^{-2\kappa} 2^{h_2} \theta_2^2$, we extend the input to $\mathbb{R} \times \mathbb{R}^4$ by zero outside I and use modulation localization. Here we do not apply Lemma 10.7, but rather gain smallness from $-\kappa$. In this case, observe that it is impossible for the input, the output and $\Psi_{h_1}^{(\theta_1)}$ to all have modulation $\ll 2^{h_2}\theta_2^2 =: j_2$. Therefore, we split into three cases:

Case 1. (High modulation input) We estimate

$$\begin{aligned}
&\sum_{\theta_2} \sum_{\theta_1 < 2^{-\kappa/2} 2^{\frac{1}{2}(h_2-h_1)}\theta_2} \|Op\left(ad(\Psi_{h_1}^{(\theta_1)}) ad(\omega \cdot A_{h_2}^{main,(\theta_2)})\right) Q_{\geq j_2 - C}\|_{X_0^{\frac{1}{2},\infty} \to L^1 L^2} \\
&\lesssim \sum_{\theta_2} \sum_{\theta_1 < 2^{-\kappa/2} 2^{\frac{1}{2}(h_2-h_1)}\theta_2} 2^{-\frac{1}{2}h_2} \theta_2^{-1} \|\Psi_{h_1}^{(\theta_1)}\|_{DL^6 L^\infty} \|\omega \cdot A_{h_2}^{main,(\theta_2)}\|_{DL^3 L^\infty} \\
&\lesssim \sum_{\theta_2} \sum_{\theta_1 < 2^{-\kappa/2} 2^{\frac{1}{2}(h_2-h_1)}\theta_2} 2^{\frac{1}{6}(h_2-h_1)} \theta_1^{\frac{1}{6}} \theta_2^{\frac{5}{6}} \|A_{x,h_1}\|_{S^1} \|A_{x,h_2}\|_{S^1} \\
&\lesssim 2^{-\frac{1}{6}\kappa} 2^{\frac{1}{4}(h_2-h_1)} \|A_{x,h_1}\|_{S^1} \|A_{x,h_2}\|_{S^1},
\end{aligned}$$

which is acceptable.

Case 2. (High modulation output) When the output has modulation $\geq 2^{j_2 - C}$, then we have exactly the same bound for $L^\infty L^2 \to X_0^{-\frac{1}{2},1}$ (we use boundedness of $Q_{<j_2-C}$ on $L^\infty L^2$).

Case 3. (High modulation for Ψ_{h_1}) By boundedness of $Q_{<j_2-C}$ on $L^\infty L^2$ and $L^1 L^2$, it suffices to have the following estimate:

$$\begin{aligned}
&\sum_{\theta_2} \sum_{\theta_1 < 2^{-\kappa/2} 2^{\frac{1}{2}(h_2-h_1)}\theta_2} \|Op\left(ad(Q_{\geq j_2-C}\Psi_{h_1}^{(\theta_1)}) ad(\omega \cdot A_{h_2}^{main,(\theta_2)})\right)\|_{L^\infty L^2 \to L^1 L^2} \\
&\lesssim \sum_{\theta_2} \sum_{\theta_1 < 2^{-\kappa/2} 2^{\frac{1}{2}(h_2-h_1)}\theta_2} \|Q_{\geq j_2-C}\Psi_{h_1}^{(\theta_1)}\|_{DL^2 L^\infty} \|\omega \cdot A_{h_2}^{main,(\theta_2)}\|_{DL^2 L^\infty} \\
&\lesssim \sum_{\theta_2} \sum_{\theta_1 < 2^{-\kappa/2} 2^{\frac{1}{2}(h_2-h_1)}\theta_2} \theta_1^{\frac{1}{2}} \theta_2^{\frac{1}{2}} \|A_{x,h_1}\|_{S^1} \|A_{x,h_2}\|_{S^1} \\
&\lesssim 2^{-\frac{1}{2}\kappa} 2^{\frac{1}{4}(h_2-h_1)} \|A_{x,h_1}\|_{S^1} \|A_{x,h_2}\|_{S^1}.
\end{aligned}$$

Here, we have use (10.15) for $\sum_{j \geq j_2 - C} Q_j \Psi_{h_1}^{(\theta_1)}$.

11.4.3. *Estimate for higher order expressions.* The contribution of the cubic, quartic and quintic terms in Ψ in the expansion of $O_{;\alpha}$ are treated in a similar manner as in the quadratic case; therefore, we omit the proof. The only remaining case is the contribution of (11.21). For this term, we claim that

$$\| \int_{-\infty}^{-\kappa} \int_{-\infty}^{h_1} \cdots \int_{-\infty}^{h_5} Op\left(ad(\Psi_{h_1}) \cdots ad(\Psi_{h_5}) ad(O_{<h_6;\alpha}) Ad(O_{<h_6})\right) \, dh_6 \cdots \, dh_2 \, dh_1 \|_{N^*[I] \to N[I]} \leq \varepsilon,$$

for κ_1 large enough and δ_p in (9.3) adequately small.

As in the case of the quadratic part, we start with very short intervals and move up the line. If $|I| < 2^{-h_1 - C\kappa}$ then we only apply fixed time decomposability estimates, namely (10.14) with $q = \infty$ and $r = \infty$ and (10.17) also with $q = \infty$, together with Hölder in time, to obtain

$$\|Op\left(ad(\Psi_{h_1}) \cdots ad(\Psi_{h_5}) ad(O_{<h_6;\alpha}) Ad(O_{<h_6})\right)\|_{L^\infty L^2 \to L^1 L^2} \lesssim_M 2^{h_6} |I|,$$

which suffices for the h integration.

If $2^{-h_1 - C\kappa} \leq |I| < 2^{-h_2 - C\kappa}$ then we switch to (10.14) with $q = 6$ and $r = \infty$ for Ψ_{h_1}, to obtain

$$\|Op\left(ad(\Psi_{h_1}) \cdots ad(\Psi_{h_5}) ad(O_{<h_6;\alpha}) Ad(O_{<h_6})\right)\|_{L^\infty L^2 \to L^1 L^2} \lesssim_M 2^{-\frac{1}{6}h_1} 2^{h_6} |I|^{\frac{5}{6}},$$

which again suffices for the h integration.

Repeating this procedure for increasingly large I we eventually arrive at the last case $|I| > 2^{-h_6 - C\kappa}$. There by Lemma 10.3 and boundedness of $Ad(O_{<h_6})$ on L^2, we have

$$\begin{aligned}
&\|Op\left(ad(\Psi_{h_1}) \cdots ad(\Psi_{h_5}) ad(O_{<h_6;\alpha}) Ad(O_{<h_2})\right)\|_{L^\infty L^2[I] \to L^1 L^2[I]} \\
&\lesssim \|\Psi_{h_1}\|_{DL^6 L^\infty[I]} \cdots \|\Psi_{h_5}\|_{DL^6 L^\infty[I]} \|O_{<h_6;\alpha}\|_{DL^6 L^\infty[I]}.
\end{aligned}$$

Using Lemma 10.5 for $\Psi_h^{(\theta)}$ with $\theta < 2^{-\kappa}$ and Lemma 10.7 for the rest, we have

$$\|\Psi_h\|_{DL^6 L^\infty[I]} \leq 2^{-\frac{1}{6}h} \left(2^{-\kappa} \|A_{x,h}\|_{S^1[I]} + C 2^{C\kappa} \|A_{x,h}\|_{DS^1[I]}\right).$$

This bound provides us with the desired smallness. By the previous estimate and (10.17), the h-integrals converge as well, which proves our claim.

11.5. **Estimates for E_3, ... E_6.** We finally handle the error terms E_3, ..., E_6, for which we gain smallness from the frequency gap κ.

11.5.1. *The estimate for E_3.* It suffices to show that

$$\|E_3 P_0\|_{L^\infty L^2 \to L^1 L^2} \lesssim_M 2^{-\frac{1}{2}\kappa}$$

But this is a consequence of the L^2 boundedness for $Op(Ad(O))$, combined with the $L^2 L^\infty$ decomposability estimates for A_α and $O_{;\alpha}$ in Lemmas 10.4,10.6.

11.5.2. *The estimate for E_4.* We expand with respect to h,

$$ad(\partial^\alpha O_{;\alpha}) Ad(O) = \int_{-\infty}^{-\kappa} \partial^\alpha (ad(O_{<h;\alpha}) ad(\Psi_h)) Ad(O_{<h}) ad(\Box \Psi_h) Ad(O_{<h})\, dh$$

For the first term we simply use two $L^2 L^\infty$ decomposability estimates as in the case of E_3. For the second term, in view of the bound (10.16), we can apply Lemma 11.3 to discard the $Ad(O_{<h})$ factor. Then it suffices to show that

$$\| \int_{-\infty}^{-\kappa} Op(ad(\Box \Psi_h)) P_0 \, dh\|_{S_0 \to N} \lesssim_M 2^h$$

After expanding Ψ_h in θ, we note that, due to the frequency localization of $\Psi_h^{(\theta)}$, either the input or the output has modulation $\gtrsim 2^h \theta^2$. We assume the former, as the other case is similar. Then we only need to prove the bound

$$\| \int_{-\infty}^{-\kappa} Op(ad(\Box \Psi_h^{(\theta)})) P_0 \, dh\|_{L^2 \to L^1 L^2} \lesssim_M \theta 2^{\frac{3}{2}h}$$

which is an immediate consequence of the decomposability bound (10.16) for $\Box \Psi_h^{(\theta)}$.

11.5.3. *The estimate for E_5.* It suffices to show that

$$\|E_3 P_0\|_{S_0^\sharp \to L^1 L^2} \lesssim_M 2^{-\frac{1}{2}\kappa}$$

Since $(D_t + |D_x|) P_0 : S_0^\sharp \to L^2$, this follows from the L^2 boundedness for $Op(Ad(O))$, combined with the $L^2 L^\infty$ decomposability estimates for A_α in Lemma 10.4,

11.5.4. *The estimate for E_6.* In view of the $L^2 L^\infty$ decomposability estimates for A_α in Lemma 10.4 and Lemma 11.3, we can discard the $Ad(O)$ factor. In addition, as in Proposition 4.30, we can express the commutator $[S_0, A_h]$ in the form

$$[S_0, A_h] f = 2^h \mathcal{O}(A_h, f)$$

Then we have reduced our problem to proving

$$\| \int_{-\infty}^{-\kappa} 2^h Op(ad(\omega \cdot \nabla A_{x,h})) P_0 \, dh\|_{S_0 \to N} \ll \varepsilon,$$

$$\| \int_{-\infty}^{-\kappa} 2^h Op(ad(A_{0,h})) P_0 \, dh\|_{S_0 \to N} \ll \varepsilon.$$

But then these follow, with the $2^{-\delta_1 \kappa}$ gain, from (8.23) and (8.25), thanks to the extra derivative (i.e. the 2^h factor).

References

[1] C. Gavrus, *Global well-posedness for the massive Maxwell-Klein-Gordon equation with small critical Sobolev data*, preprint (2016), arXiv:1610.03581.

[2] C. Gavrus and S.-J. Oh, *Global well-posedness of high dimensional Maxwell-Dirac for small critical data*, Mem. Amer. Math. Soc. (to appear), arXiv:1604.07900.

[3] C. E. Kenig and F. Merle, *Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case*, Invent. Math. **166** (2006), no. 3, 645–675, doi.

[4] ———, *Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation*, Acta Math. **201** (2008), no. 2, 147–212, doi.

[5] S. Klainerman and M. Machedon, *On the Maxwell-Klein-Gordon equation with finite energy*, Duke Math. J. **74** (1994), no. 1, 19–44, doi.

[6] ———, *Finite energy solutions of the Yang-Mills equations in* $\mathbb{R}^{3+1}$, Ann. of Math. (2) **142** (1995), no. 1, 39–119, doi.

[7] J. Krieger and J. Lührmann, *Concentration compactness for the critical Maxwell-Klein-Gordon equation*, Ann. PDE **1** (2015), no. 1, Art. 5, 208.

[8] J. Krieger and W. Schlag, *Concentration compactness for critical wave maps*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2012, doi.

[9] J. Krieger and J. Sterbenz, *Global regularity for the Yang-Mills equations on high dimensional Minkowski space*, Mem. Amer. Math. Soc. **223** (2013), no. 1047, vi+99, doi.

[10] J. Krieger, J. Sterbenz, and D. Tataru, *Global well-posedness for the Maxwell-Klein-Gordon equation in* $4+1$ *dimensions: small energy*, Duke Math. J. **164** (2015), no. 6, 973–1040, doi.

[11] J. Krieger and D. Tataru, *Global well-posedness for the Yang-Mills equation in* $4+1$ *dimensions. Small energy*, Ann. of Math. (2) **185** (2017), no. 3, 831–893.

[12] M. Machedon and J. Sterbenz, *Almost optimal local well-posedness for the* $(3+1)$*-dimensional Maxwell-Klein-Gordon equations*, J. Amer. Math. Soc. **17** (2004), no. 2, 297–359, doi.

[13] S.-J. Oh, *Gauge choice for the Yang-Mills equations using the Yang-Mills heat flow and local well-posedness in* H^1, J. Hyperbolic Differ. Equ. **11** (2014), no. 1, 1–108, doi.

[14] ———, *Finite energy global well-posedness of the Yang-Mills equations on* $\mathbb{R}^{1+3}$*: an approach using the Yang-Mills heat flow*, Duke Math. J. **164** (2015), no. 9, 1669–1732, doi.

[15] S.-J. Oh and D. Tataru, *Global well-posedness and scattering of the* $(4+1)$*-dimensional Maxwell-Klein-Gordon equation*, Invent. Math. **205** (2016), no. 3, 781–877, arXiv:1503.01562, doi.

[16] ———, *Local well-posedness of the* $(4+1)$*-dimensional Maxwell-Klein-Gordon equation at energy regularity*, Ann. PDE **2** (2016), no. 1, Art. 2, 70, arXiv:1503.01560, doi.

[17] ———, *Energy dispersed solutions for the (4+1)-dimensional Maxwell-Klein-Gordon equation*, Amer. J. Math. (2017), arXiv:1503.01561.

[18] ———, *The Yang-Mills heat flow and the caloric gauge*, preprint (2017), arXiv:1709.08599.

[19] ———, *The hyperbolic Yang-Mills equation for connections in an arbitrary topological class*, preprint (2017), arXiv:1709.08604.

[20] ———, *The Threshold Conjecture for the energy critical hyperbolic Yang-Mills equation*, preprint (2017), arXiv:1709.08606.

[21] ———, *The Threshold Theorem for the* $(4+1)$*-dimensional Yang-Mills equation: An overview of the proof*, preprint (2017), arXiv:1709.09088.

[22] I. Rodnianski and T. Tao, *Global regularity for the Maxwell-Klein-Gordon equation with small critical Sobolev norm in high dimensions*, Comm. Math. Phys. **251** (2004), no. 2, 377–426, doi.

[23] J. Sterbenz and D. Tataru, *Energy dispersed large data wave maps in* $2+1$ *dimensions*, Comm. Math. Phys. **298** (2010), no. 1, 139–230, doi.

[24] ———, *Regularity of wave-maps in dimension* $2+1$, Comm. Math. Phys. **298** (2010), no. 1, 231–264, doi.

[25] T. Tao, *Global regularity of wave maps. II. Small energy in two dimensions*, Comm. Math. Phys. **224** (2001), no. 2, 443–544, doi.

[26] ———, *Geometric renormalization of large energy wave maps*, Journées "Équations aux Dérivées Partielles", École Polytech., Palaiseau, 2004, pp. Exp. No. XI, 32.

[27] ———, *Global regularity of wave maps III. Large energy from βR^{1+2} to hyperbolic spaces*, preprint (2008), `arXiv:0805.4666`.
[28] ———, *Global regularity of wave maps IV. Absence of stationary or self-similar solutions in the energy class*, preprint (2008), `arXiv:0806.3592`.
[29] ———, *Global regularity of wave maps V. Large data local wellposedness and perturbation theory in the energy class*, preprint (2008), `arXiv:0808.0368`.
[30] ———, *Global regularity of wave maps VI. Abstract theory of minimal-energy blowup solutions*, preprint (2009), `arXiv:0906.2833`.
[31] ———, *Global regularity of wave maps VII. Control of delocalised or dispersed solutions*, preprint (2009), `arXiv:0908.0776`.
[32] D. Tataru, *On global existence and scattering for the wave maps equation*, Amer. J. Math. **123** (2001), no. 1, 37–77, `link`.

KIAS, Seoul, Korea 02455
Email address: `sjoh@kias.re.kr`

Department of Mathematics, UC Berkeley, Berkeley, CA, 94720
Email address: `tataru@math.berkeley.edu`