# Analysis & PDE

Volume 9 No. 8 2016


msp

# Analysis & PDE

msp.org/apde





See inside back cover or msp.org/apde for submission instructions.

The subscription price for 2016 is US $235/year for the electronic version, and $430/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

APDE peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ANALYSIS AND PDE
Vol. 9, No. 8, 2016
dx.doi.org/10.2140/apde.2016.9.1773

# ESTIMATES FOR RADIAL SOLUTIONS OF THE HOMOGENEOUS LANDAU EQUATION WITH COULOMB POTENTIAL

MARIA PIA GUALDANI AND NESTOR GUILLEN

Motivated by the question of existence of global solutions, we obtain pointwise upper bounds for radially symmetric and monotone solutions to the homogeneous Landau equation with Coulomb potential. The estimates say that blow-up in the $L^\infty$ norm at some finite time $T$ occurs only if a certain quotient involving $f$ and its Newtonian potential concentrates near zero, which implies blow-up in more standard norms, such as the $L^{3/2}$ norm. This quotient is shown to be always less than a universal constant, suggesting that the problem of regularity for the Landau equation is in some sense critical.

The bounds are obtained using the comparison principle both for the Landau equation and for the associated mass function. In particular, the method provides long-time existence results for a modified version of the Landau equation with Coulomb potential, recently introduced by Krieger and Strain.

## 1. Introduction

This manuscript is concerned with the Cauchy problem for the homogeneous Landau equation. This equation takes the general form

$$\partial_t f(v,t) = Q(f,f), \quad f(v,0) = f_{\text{in}}(v), \quad v \in \mathbb{R}^3, \quad t > 0, \tag{1-1}$$

where $Q(f,f)$ is a quadratic operator known as the Landau collision operator:

$$Q(f,f) = \operatorname{div}\left( \int_{\mathbb{R}^3} A(v-y)(f(y)\nabla_v f(v) - f(v)\nabla_y f(y))\, dy \right). \tag{1-2}$$

The term $A(v)$ denotes a positive and symmetric matrix

$$A(v) := C_\gamma \left( \mathbb{I} - \frac{v \otimes v}{|v|^2} \right) \varphi(|v|), \quad v \neq 0, \quad C_\gamma > 0,$$

which acts as the projection operator onto the space orthogonal to the vector $v$. The function $\varphi(|v|)$ is a scalar-valued function determined from the original Boltzmann kernel describing how particles interact. If the interaction strength between particles at a distance $r$ is proportional to $r^{1-s}$, then

$$\varphi(|v|) := |v|^{\gamma+2}, \quad \gamma = \frac{s-5}{s-1}. \tag{1-3}$$

Note that $s = 2$ corresponds to the Coulomb potential, in which case we have $\gamma = -3$ [Villani 2002, Chapter 1, Section 1.4]. Any solution to (1-1)–(1-2) is an integrable and nonnegative scalar field

*MSC2010:* 35B65, 35K57, 35B44, 35K61, 35Q20.
*Keywords:* Landau equation, Coulomb potential, homogeneous solutions, upper bounds, barriers, regularity.

$f(v,t) : \mathbb{R}^3 \times [0,T] \to \mathbb{R}^+$. Equation (1-1) describes the evolution of a plasma in spatially homogeneous regimes, which means that the density function $f$ depends only on the velocity component $v$. Landau's original intent in deriving this approximation was to make sense of the Boltzmann collision operator, which always diverges when considering purely grazing collisions.

The Cauchy problem for (1-1)–(1-3) is very well understood for the case of hard potentials, which correspond to $\gamma \geq 0$ above. Desvillettes and Villani showed the existence of global classical solutions for hard potentials and studied its long-time behavior; see [Desvillettes and Villani 2000a; 2000b; Villani 2002] and references therein. In this case there is a unique global smooth solution, which converges exponentially to an equilibrium distribution, known as the Maxwellian function

$$\mathcal{M}(v) = \frac{1}{(2\pi)^{3/2}} e^{-|v|^2/2}.$$

Analyzing the soft potentials case, $\gamma < 0$, has proved to be more difficult. Using a probabilistic approach, [Wu 2014; Fournier and Guérin 2009; Alexandre et al. 2015] show uniqueness and existence of weak solutions for $\gamma \in [-2, 0]$. For $\gamma \in [-3, -2]$, existence is known for small-time or global in-time with smallness assumption on initial data [Alexandre et al. 2015; Arsen'ev and Peskov 1977]. Finally, for the Coulomb case $\gamma = -3$, Fournier [2010] showed the uniqueness of weak solutions as long as they remain in $L^\infty$.

Villani [1998] introduced the so called $H$-solutions, which enjoy (weak) a priori bounds in a weighted Sobolev space. However, the issue of their uniqueness and regularity (i.e., no finite-time breakdown occurs) has remained open, even for smooth initial data; see [Villani 2002, Chapters 1 and 5] for further discussion.

Guo [2002] employed a completely different approach based on perturbation theory for the existence of periodic solutions to the spatially inhomogeneous Landau equation in $\mathbb{R}^3$. He showed that if the initial data is sufficiently close to the unique equilibrium in a certain high Sobolev norm, then a unique global solution exists. Moreover, as remarked in [loc. cit.], this approach also extends to the case of potentials (1-3), where $\gamma$ might even take values below $-3$.

Due to the lack of a global well-posedness theory, several conjectures about possible finite-time blow-up for general initial data have been made throughout the years. Villani [2002] discussed the possibility that (1-1)–(1-3) could blow up for $\gamma = -3$. Note that for smooth solutions, (1-1)–(1-3) with $\gamma = -3$ can be rewritten as

$$\partial_t f = \operatorname{div}(A[f]\nabla f - f\nabla a[f]) = \operatorname{Tr}(A[f]D^2 f) + f^2, \tag{1-4}$$

where

$$A[f] := A(v) * f = \frac{1}{8\pi|v|}\left(\mathbb{I} - \frac{v \otimes v}{|v|^2}\right) * f, \quad \Delta a = -f.$$

Equation (1-4) can be thought of as a quasilinear nonlocal heat equation. Support for blow-up conjectures were given by the fact that (1-4) is reminiscent of the well studied semilinear heat equation

$$\partial_t f = \Delta f + f^2. \tag{1-5}$$

Blow-up for (1-5) is known to happen for every $L^p$ norm for $p > \frac{3}{2}$; see [Giga and Kohn 1985].

However, despite the apparent similarities, (1-4) behaves differently from (1-5). The Landau equation admits a richer class of equilibrium solution: every Maxwellian $\mathcal{M}$ solves $Q(\mathcal{M}, \mathcal{M}) = 0$, which holds, in particular, for those with arbitrarily large mass.

From a different perspective, Krieger and Strain [2012] considered a modified version of (1-4),

$$\partial_t f = a[f]\Delta f + \alpha f^2, \tag{1-6}$$

and showed global existence of smooth radial solutions starting from radial initial data when $\alpha < \frac{2}{3}$. This range for $\alpha$ later was expanded to any $\alpha < \frac{74}{75}$ by means of a nonlocal inequality obtained by Gressman, Krieger and Strain [Gressman et al. 2012]. Note that when $\alpha = 1$, the above equation can be written in divergence form,

$$\partial_t f = \operatorname{div}(a[f]\nabla f - f\nabla a[f]). \tag{1-7}$$

These results put in evidence how a nonlinear equation with a nonlocal diffusivity such as (1-7) behaves drastically differently from (and better than) (1-5).

Our main results in this manuscript are twofold. The first one gives necessary conditions for the finite-time blow-up of solutions to (1-4). The second (unconditional) result says that solutions to (1-7) do not blow up at all, and in fact become instantaneously smooth (even for initial data that might be initially unbounded). Both results deal only with radially symmetric, decreasing initial conditions; more precisely, we assume that

$$\begin{gathered} f_{\text{in}} \geq 0, \quad f_{\text{in}} \in L^\infty(\mathbb{R}^3), \\ \int_{\mathbb{R}^3} f_{\text{in}}\, dv = 1, \quad \int_{\mathbb{R}^3} f_{\text{in}}|v|^2\, dv = 3, \quad \int_{\mathbb{R}^3} f_{\text{in}} \log(f_{\text{in}})\, dv < \infty, \\ |v| \leq |w| \Rightarrow f_{\text{in}}(v) \geq f_{\text{in}}(w). \end{gathered} \tag{1-8}$$

The normalization of the initial data is standard and follows a standard change of variables. The main results are the following.

**Theorem 1.1.** *Let $f_{\text{in}}$ be as in (1-8). Then there exist $T_0 > 0$ and $f : \mathbb{R}^3 \times (0, T_0) \to \mathbb{R}_+$ such that $f$ is smooth and solves (1-4) for $t \in (0, T_0)$, with $f(\cdot, 0) = f_{\text{in}}$. Moreover, $T_0$ is maximal in the sense that either $T_0 = \infty$ or else the $L^{3/2}$ norm of $f$ accumulates near $v = 0$ as $t \to T_0^-$, in particular*

$$\lim_{t \to T_0^-} \|f(\cdot, t)\|_{L^p(B_1)} = \infty, \quad \forall p > \tfrac{3}{2}.$$

In fact, the above theorem is a consequence of the following sharper result.

**Theorem 1.2.** *There is a constant $\varepsilon_0 \geq \frac{1}{96}$ such that if $T_0 < \infty$, then*

$$\limsup_{r \to 0^+} \sup_{t \in (0, T_0)} \left\{ r^2 \frac{\int_{B_r} f(v, t)\, dv}{\int_{B_r} a[f](v, t)\, dv} \right\} \geq \varepsilon_0.$$

Neither of the above theorems are enough to guarantee long-time existence of classical solutions to (1-4). However, Theorem 1.2 suggests that (1-4) is in some sense "critical" for regularity. It can be

shown (see Proposition 5.6) that for any nonnegative $f \in L^1(\mathbb{R}^3)$,

$$r^2 \frac{\int_{B_r} f(v)\,dv}{\int_{B_r} a[f](v)\,dv} \leq 3, \quad \forall r > 0.$$

In particular, if the $\varepsilon_0$ in Theorem 1.2 could be shown to be at least 3 (or in general if the upper bound in the last inequality could be improved to something less than $\varepsilon_0$), it would immediately follow that solutions to the Landau equation (1-4) cannot blow up in finite time. It is not clear if this can be guaranteed for general $f$ without at least using some partial time regularization.

On the other hand, methods used in the proof of Theorem 1.1 and Theorem 1.2 yield long-time existence for the modified Landau equation (1-7) (again, in the radial case).

**Theorem 1.3.** *Let $f_{\text{in}}$ be as in (1-8) and such that for some $p > 6$,*

$$f_{\text{in}} \in L^p_{\text{weak}}(\mathbb{R}^3).$$

*Then there exists a function $f : \mathbb{R}^3 \times \mathbb{R}_+ \to \mathbb{R}$, smooth for positive times, with $f(\cdot, 0) = f_{\text{in}}$ which solves, for $t > 0$,*

$$\partial_t f = a[f]\Delta f + f^2.$$

We approach the analysis from the point of view of nonlinear parabolic equations. The nonlocal dependence of the coefficients on the solution prevents the equation from satisfying a comparison principle: if $v_0$ is a contact point of two functions $f$ and $g$, i.e., $f(v_0) = g(v_0)$ and everywhere else $f(v) \leq g(v)$, it does not follow that $Q(f, f)(v_0) \leq Q(g, g)(v_0)$. More precisely, for the case where $Q(f, f)$ corresponds to (1-2) one cannot expect an inequality such as

$$\operatorname{Tr}(A[f]D^2 f)(v_0) \leq \operatorname{Tr}(A[g]D^2 g)(v_0).$$

In fact, due to the nonlocality of $A$ one only has $A[f](v_0) \leq A[g](v_0)$. Equality $A[f](v_0) = A[g](v_0)$ holds only when $f \equiv g$ for every $v \in \mathbb{R}^3$. The maximum principle is not useful either, since at a maximum point for $f$ we only obtain $\partial_t f \leq -f\Delta a[f]$, which does not rule out growth of the maximum of $f$. The same observations apply to $Q(f, f)$ corresponding to (1-7).

On the other hand, if one could construct (using only properties of $f$ that are independent of $t$) a function $U(v)$ such that

$$\begin{aligned} \operatorname{Tr}(A[f]D^2 U) + fU \leq 0 \quad &\text{in } \mathbb{R}^3, \\ a[f]\Delta U + fU \leq 0 \quad &\text{in } \mathbb{R}^3, \end{aligned}$$

then the comparison principle (for linear parabolic equations) would guarantee that $f \leq cU$ for all times provided $f(t = 0) \leq cU$. Our main observation is that (under radial symmetry) the above can be made to work with $U(v) = |v|^{-\gamma}$, $\gamma \in (0, 1)$. From here higher local integrability of $f$ can be propagated, and from there higher regularity follows by standard elliptic regularization.

A previous attempt by the authors, also based on upper barrier arguments (but meant to cover any bounded, fast decaying initial data), was ultimately undone by a computational error. However, Theorems 1.1–1.3 show that the use of upper barriers to study (1-4) is fruitful at least for radially symmetric and

decreasing initial conditions. On the other hand, the authors in [Gualdani and Guillen ≥ 2016] show a local $L^\infty$-regularization estimate using the De Giorgi iteration method for $\gamma > -2$.

**Remark 1.4.** After the submission of this article, the authors learned of related work of Silvestre [2016] on the Boltzmann equation, covering the spatially inhomogeneous case. In that paper, a priori estimates rely on maximum principle arguments and make use of the regularity for parabolic integro-differential equations, particularly recent work of Schwab and Silvestre [2016].

***Outline.*** The rest of the paper is organized as follows. After a brief review in Section 2 on nonlinear parabolic theory that will be needed to construct local solutions to the nonlinear problems, in Section 3 we outline the symmetry properties of (1-4). Section 4 deals with short-time existence. In Section 5 we present a barrier argument that allows us to prove conditional non-blow-up results for the Landau equation and global well-posedness for the modified Landau equation in Section 6.

***Notation.*** Universal constants will be denoted by $c, c_0, c_1, C_0, C_1, C$. Vectors in $\mathbb{R}^3$ will be denoted by $v, w, x, y$ and so on. The inner product between $v$ and $w$ will be written $(v, w)$. $B_R(v_0)$ will denote the closed ball of radius $R$ centered at $v_0$; if $v_0 = 0$ we simply write $B_R$. The identity matrix will be denoted by $\mathbb{I}$, the trace of a matrix $X$ will be denoted $\mathrm{Tr}(X)$. The initial condition for the Cauchy problem will always be denoted by $f_{\mathrm{in}}$.

The letter $\Omega$ will denote a general compact subset of $\mathbb{R}^3$. $Q \subset \mathbb{R}^3 \times \mathbb{R}_+$ will be a space-time cylinder of parabolic diameter $R$ with $R > 0$ a general constant, unless otherwise specified. Finally, $\partial_p Q$ will denote the parabolic boundary of $Q$.

## 2. A rapid review of linear parabolic equations

We work with two bilinear operators, namely the one associated to (1-4),

$$Q_{\mathcal{L}}(g, f) := \mathrm{div}(A[g]\nabla f - f\nabla a[g]) = \mathrm{Tr}[A[g]D^2 f] + fg,$$

and the one associated to (1-7),

$$Q_{\mathcal{KS}}(g, f) := \mathrm{div}(a[g]\nabla f - f\nabla a[g]) = a[g]\Delta f + fg.$$

As is well known, through $Q_{\mathcal{L}}$ (and also $Q_{\mathcal{KS}}$), any $g : \mathbb{R}^3 \times \mathbb{R}_+ \to \mathbb{R}$ gives rise to a linear elliptic operator with variable coefficients as follows:

$$\begin{aligned}\phi \to Q_{\mathcal{L}}(g, \phi) &:= \mathrm{div}(A[g]\nabla\phi - \phi\nabla a[g]) = \mathrm{Tr}(A[g]D^2\phi) + \phi g,\\ \phi \to Q_{\mathcal{KS}}(g, \phi) &:= \mathrm{div}(a[g]\nabla\phi - \phi\nabla a[g]) = a[g]\Delta\phi + \phi g.\end{aligned}$$

Accordingly, given such a $g$ and initial data $f_{\mathrm{in}}$, one considers the linear Cauchy problem,

$$\begin{cases}\partial_t f = Q(g, f) & \text{in } \mathbb{R}^3 \times \mathbb{R}_+,\\ f(\cdot, 0) = f_{\mathrm{in}},\end{cases} \tag{2-1}$$

both for $Q = Q_{\mathcal{L}}$ and $Q = Q_{\mathcal{KS}}$.

**Remark.** Note that $Q_{\mathcal{L}}(g, f)$ and $Q_{\mathcal{KS}}(g, f)$ can both be expressed as a divergence, so any solution to (2-1) preserves its mass over time, i.e.,

$$\|f(\cdot, t)\|_{L^1(\mathbb{R}^3)} = \|f_{\text{in}}(\cdot)\|_{L^1(\mathbb{R}^3)} =: M_{\text{in}}, \quad \forall t > 0.$$

**Lemma 2.1** (see [Ladyženskaja et al. 1968, Theorem 5.1, page 320]). *Let $f_{\text{in}} : \mathbb{R}^3 \to \mathbb{R}$ and $g : \mathbb{R}^3 \times \mathbb{R}_+ \to \mathbb{R}$ be nonnegative functions such that for some $\beta \in (0, 1)$ we have*

$$\begin{aligned} f_{\text{in}} &\in L^1(\mathbb{R}^3) \cap C^{2+\beta}(\mathbb{R}^3), \\ A[g], \nabla a[g] &\in C^{\beta, \beta/2}(\mathbb{R}^3 \times \mathbb{R}_+). \end{aligned} \tag{2-2}$$

*Then for every $\delta > 0$, there exists a unique $f : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$ with $f \in C^{2+\beta, 1+\beta/2}(\mathbb{R}^3 \times \mathbb{R}_+)$ which is a classical solution of*

$$\begin{cases} \partial_t f = \delta \Delta f + Q(g, f) & \text{in } \mathbb{R}^3 \times \mathbb{R}_+, \\ f(\cdot, 0) = f_{\text{in}}, \end{cases} \tag{2-3}$$

*where $Q(\cdot, \cdot)$ denotes either $Q = Q_{\mathcal{L}}$ or $Q = Q_{\mathcal{KS}}$.*

Next we summarize in three theorems several classical local regularity estimates for parabolic equations of the form

$$\partial_t f = \operatorname{div}(B \nabla f + f b),$$

where $f : Q \to \mathbb{R}$ and $Q = B_R(v_0) \times (t_0 - R^2, t_0) \subset \mathbb{R}^d \times \mathbb{R}_+$ is the parabolic cylinder of radius $R$ centered at some points $x_0, t_0$. The first two theorems are, respectively, a local Hölder estimate (from De Giorgi–Nash–Moser theory) and an $L^\infty$ estimate for $f$ in terms of its boundary data (Stampacchia estimate); see [Ladyženskaja et al. 1968, Chapter III, Theorem 10.1, page 204 and Chapter IV, Theorem 10.1, page 351 of] as well as [Lieberman 1996, Chapter VI, Theorem 6.29, page 131] for the respective proofs. The main point of these theorems is that they do not require any regularity assumption on the diffusion matrix $B$ (beyond ellipticity and boundedness).

**Theorem 2.2** (De Giorgi–Nash–Moser estimate). *Suppose $f$ is a weak solution of the equation*

$$\partial_t f = \operatorname{div}(B \nabla f + f b),$$

*where $b$ is a vector field and $B$ is a symmetric matrix such that*

$$\lambda \mathbb{I} \le B(v, t) \le \Lambda \mathbb{I} \quad \text{a.e. in } Q.$$

*Then there is some $\alpha \in (0, 1)$ and $C > 0$ such that the estimate*

$$[f]_{C^{\alpha, \alpha/2}(Q_{1/2})} \le C\big(\|f\|_{L^\infty(Q)} + R^2 \|b\|_{L^\infty(Q)}\big) \tag{2-4}$$

*holds, where $Q_{1/2} := B_{R/2}(x_0) \times (t_0 - (R/2)^2, t_0)$ and $\alpha$ and $C$ are determined by $\lambda$, $\Lambda$, $R$ and $d$.*

**Theorem 2.3** (Stampacchia estimate). *If $f$ is a weak solution of*

$$\partial_t f \le \operatorname{div}(B \nabla f + f b),$$

*with $B$ and $b$ as in the previous theorem, then there exists a constant $C > 0$ such that*

$$\|f\|_{L^\infty(Q)} \le C(\|f\|_{L^2(Q)} + \|b\|_{L^\infty(Q)}). \tag{2-5}$$

*As before, $C$ is determined by $\lambda$, $\Lambda$, $d$ and $R$.*

The last theorem recalls interior classical regularity estimates when the coefficients are Hölder continuous in time and space. See [Ladyženskaja et al. 1968, Chapter IV] or also [Lieberman 1996, Chapter III, Theorem 6.17] for a proof.

**Theorem 2.4** (Schauder estimates). *If $B, b \in C^{\beta,\beta/2}(Q)$, then there is a finite $C$ such that*

$$[D^2 f]_{C^{\beta,\beta/2}(Q_{1/2})} + [\partial_t f]_{C^{\beta,\beta/2}(Q_{1/2})} \le C\big(\lambda, \Lambda, R, \|B\|_{C^{\beta,\beta/2}(Q)}, \|b\|_{C^{\beta,\beta/2}(Q)}, \|f\|_{L^\infty(Q)}\big).$$

## 3. Radial symmetry

This section is devoted to some technical lemmas. The proofs of the first two propositions are rather technical and can be found in the Appendix.

**Proposition 3.1.** *Suppose $f_{\text{in}}$ and $g(\cdot, t)$ are both radially symmetric, and let $Q(\cdot,\cdot)$ denote either $Q_{\mathcal{L}}$ or $Q_{\mathcal{KS}}$. Then any solution of the linear Cauchy problem*

$$\partial_t f = Q(g, f), \quad f(v, 0) = f_{\text{in}}(v),$$

*is radially symmetric for all $t$. Furthermore, if $f_{\text{in}}$ and $g$ are radially decreasing, then so is $f$.*

Let $h : \mathbb{R}^3 \to \mathbb{R}_+$. Define

$$A^*[h](v) := (A[h](v)\hat{v}, \hat{v}), \quad v \neq 0, \quad \hat{v} := v|v|^{-1}. \tag{3-1}$$

There are two useful expressions for $A^*[h]$ and $a[h]$ when $h$ is radially symmetric.

**Proposition 3.2.** *Let $h \in L^1(\mathbb{R}^3)$ be radially symmetric and nonnegative. Then*

$$A^*[h](v) = \frac{1}{12\pi|v|^3}\int_{B_{|v|}} h(w)|w|^2\,dw + \frac{1}{12\pi}\int_{B^c_{|v|}} \frac{h(w)}{|w|}\,dw, \tag{3-2}$$

$$a[h](v) = \frac{1}{4\pi|v|}\int_{B_{|v|}} h(w)\,dw + \frac{1}{4\pi}\int_{B^c_{|v|}} \frac{h(w)}{|w|}\,dw. \tag{3-3}$$

The second formula above is simply the classical formula for the Newtonian potential in the case of radial symmetry; the formula for $A^*[h]$ is new and the proof can be found in the Appendix.

**Lemma 3.3.** *Let $h \in L^1(\mathbb{R}^3)$ be a nonnegative, spherically symmetric function.*

(1) *If $h$ is monotone decreasing with $|v|$, and*

$$\int_{B_{R_1}\setminus B_{R_0}} h\,dv \ge \theta > 0$$

*for some $\delta > 0$ and $0 < R_0 < R_1$, then*

$$A[h](v) \geq \frac{\theta R_0^2}{12\pi(1+R_1^3)} \frac{1}{1+|v|^3}\mathbb{I}. \tag{3-4}$$

(2) *If $h$ is bounded, i.e., if $\|h\|_{L^\infty(\mathbb{R}^3)} = h(0) < +\infty$, then*

$$A[h](v) \leq a[h]\mathbb{I} \leq 2\left(\frac{\|h\|_{L^\infty(\mathbb{R}^3)} + \|h\|_{L^1(\mathbb{R}^3)}}{1+|v|}\right)\mathbb{I}, \quad \forall v \in \mathbb{R}^3. \tag{3-5}$$

*Proof.* (1) Let $A^*[h]$ be as in (3-2). If $|v| \geq R_1$, then

$$\begin{aligned} A^*[h](v) &\geq \frac{1}{12\pi|v|^3}\int_{B_{R_1}} h(w)|w|^2\,dw \geq \frac{1}{12\pi|v|^3}\int_{B_{R_1}\setminus B_{R_0}} h(w)|w|^2\,dw \\ &\geq \frac{R_0^2}{12\pi|v|^3}\int_{B_{R_1}\setminus B_{R_0}} h(w,t)\,dw \geq \frac{\theta R_0^2}{12\pi|v|^3}. \end{aligned}$$

Note that Proposition 3.2 guarantees that $A^*[h]$ is radially decreasing. Thus,

$$A^*[h](v) \geq \frac{\theta R_0^2}{12\pi R_1^3}, \quad \forall v \in B_{R_1}.$$

Combining both estimates, we conclude that

$$A^*[h](v) \geq \frac{\theta R_0^2}{12\pi(1+R_1^3)} \frac{1}{1+|v|^3}.$$

(2) If $h \in L^\infty$, then we may use (3-3) to obtain the estimate

$$\begin{aligned} [h] \leq a[h](v)\mathbb{I} &\leq \left(\frac{h(0)}{4\pi|v|}\int_{B_{|v|}} dw + \frac{1}{4\pi}\int_{B_{|v|}^c} h(w)\,dw\right)\mathbb{I} \\ &\leq (\|h\|_{L^\infty(\mathbb{R}^3)} + \|h\|_{L^1(\mathbb{R}^3)})\mathbb{I}, \quad \text{if } |v| \leq 1, \end{aligned}$$

and

$$A[h] \leq a[h](v)\mathbb{I} \leq \left(\frac{\|h\|_{L^1(\mathbb{R})}}{2\pi|v|}\right)\mathbb{I} \leq \left(\frac{\|h\|_{L^1(\mathbb{R})}}{1+|v|}\right)\mathbb{I}, \quad \text{if } |v| \geq 1. \qquad \square$$

**Proposition 3.4.** *Let $h$ be a positive and radially symmetric and decreasing function. For any $\gamma \in (0,1)$, define $U_\gamma(v)$ as*

$$U_\gamma(v) := |v|^{-\gamma}.$$

*Then for $Q = Q_{\mathcal{L}}$ or $Q = Q_{\mathcal{KS}}$,*

$$Q(h, U_\gamma) \leq U_\gamma\left(-\tfrac{1}{3}\gamma(1-\gamma)a[h]|v|^{-2} + h\right).$$

*Proof.* As $U_\gamma$ is radial

$$\nabla U_\gamma(v) = U_\gamma'(v)\frac{v}{|v|}, \quad D^2U_\gamma(v) = U_\gamma''(v)\frac{v}{|v|}\otimes\frac{v}{|v|} + U_\gamma'(v)\frac{1}{|v|}\left(\mathbb{I} - \frac{v}{|v|}\otimes\frac{v}{|v|}\right).$$

Thus, in the case $Q = Q_{\mathcal{L}}$,

$$Q(h, U_\gamma) = \mathrm{Tr}(A[h]D^2U_\gamma) + hU_\gamma = A^*[h]U_\gamma'' + \frac{a[h] - A^*[h]}{|v|}U_\gamma' + hU_\gamma.$$

In particular, since $U_\gamma' = -\gamma r^{-1}U_\gamma$, $U_\gamma'' = \gamma(\gamma+1)|v|^{-2}U_\gamma$, it follows that

$$Q_{\mathcal{L}}(h, U_\gamma) = U_\gamma\big(\gamma(\gamma+1)A^*[h]|v|^{-2} - \gamma(a[h] - A^*[h])|v|^{-2} + h\big).$$

The thesis follows by noticing that $A^*[h] \le \frac{1}{3}a[h]$.

For the case $Q = Q_{\mathcal{KS}}$, an analogous computation shows that

$$\begin{aligned} Q(h, U_\gamma) &= U_\gamma\big(-\gamma(1-\gamma)a[h]|v|^{-2} + h\big) \\ &\le U_\gamma\big(-\tfrac{1}{3}\gamma(1-\gamma)a[h]|v|^{-2} + h\big), \end{aligned}$$

where in the last inequality we use $\gamma \in (0, 1)$ and $a[h] \ge 0$. □

## 4. Short-time existence

In this section, $Q$ denotes either $Q_{\mathcal{L}}$ or $Q_{\mathcal{KS}}$. For some nontrivial interval of existence $[0, T)$, a smooth solution to

$$\begin{cases} \partial_t f = Q(f, f) & \text{in } \mathbb{R}^3 \times [0, T), \\ f(\cdot, 0) = f_{\mathrm{in}}, \end{cases}$$

will be obtained by taking the limit of a sequence of functions $\{f_k\}_{k\ge 0}$ constructed recursively (as explained further below). The interval of existence $[0, T)$ is maximal in the sense that either $T = \infty$ or else the $L^\infty$ norm of $f(\cdot, t)$ blows up as $t$ approaches $T$, so the classical solution cannot be extended to a longer time interval.

**Remark 4.1.** As mentioned in the introduction, existence and uniqueness of bounded weak solutions to (1-4) have been obtained, respectively, by Arsen'ev and Peskov [1977] and by Fournier [2010]. It is likely (but not at all obvious) that the method used in [Fournier 2010] will carry over to the case of the isotropic equation (1-7). Thus, for the sake of completeness, we provide in this section a detailed proof of existence (but not uniqueness) of a classical solution for the nonlinear problem that covers the isotropic equation. For completely classical solutions this is certainly new for the isotropic equation (1-7) with $\alpha = 1$, although the methods used in the proof — a priori estimates for linear equations, which yield compactness for a sequence of approximate solutions to the nonlinear problem — are fairly well known, but still somewhat different from the approach used in [Krieger and Strain 2012] for the case $\alpha < \frac{3}{4}$. Uniqueness for classical solutions of (1-4) is contained in Fournier's [2010] result, since classical solutions are in particular weak solutions, and as it was just mentioned above, it is likely that this result can be expanded to cover (1-7).

For technical reasons we first assume that $f_{\mathrm{in}}$ satisfies (1-8) and for some $c > 0$,

$$f_{\mathrm{in}} \in C^{2+\beta}(\mathbb{R}^3), \quad \|f_{\mathrm{in}}\|_{C^{2+\beta}(B_1(v))} \le \frac{c}{1+|v|^5}, \quad \forall v \in \mathbb{R}^3. \tag{4-1}$$

The inequality yields a rate of decay for the second derivatives of $f_{\text{in}}$ which somewhat simplifies the existence proof. The assumptions (4-1) are auxiliary, and will be removed (by an approximation argument) in the proof of Theorem 4.14 at the end of this section.

Fix $\delta > 0$. A sequence $\{f_k^\delta\}_{k\geq 0}$ will be constructed recursively, so that for every $k$,

$$f_k^\delta \in L^\infty(\mathbb{R}_+, L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)) \cap C^{2+\beta, 1+\beta/2}(\mathbb{R}^3 \times \mathbb{R}_+) \tag{4-2}$$

for some $\alpha \in (0, 1)$ independent of $k$. The construction is done as follows: First, we set $f_0(v, t) := f_{\text{in}}(v)$ for all $v$ and $t > 0$. Next, given $f_{k-1}^\delta \in L^\infty(\mathbb{R}_+, L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)) \cap C^{2+\beta, 1+\beta/2}(\mathbb{R}^3 \times \mathbb{R}_+)$, define $f_k^\delta$ as the unique classical solution to the linear Cauchy problem

$$\begin{cases} \partial_t f = \delta \Delta f + Q(f_{k-1}^\delta, f) & \text{in } \mathbb{R}^3 \times \mathbb{R}_+, \\ f(\cdot, 0) = f_{\text{in}}. \end{cases} \tag{4-3}$$

The fact that the sequence $f_k^\delta$ is well defined and satisfies (4-2) follows by repeatedly applying Lemma 2.1, making use of the fact that for every $k \geq 1$, $\beta' \in (0, 1)$,

$$f_k^\delta \text{ satisfies (4-2) and solves (4-3)} \Rightarrow A[f_k^\delta], \nabla a[f_k^\delta] \in C^{\beta', \beta'/2}(\mathbb{R}^3 \times [0, \infty)). \tag{4-4}$$

That this is so is essentially a consequence of the fact that $A[f_k^\delta]$ and $\nabla a[f_k^\delta]$ are convolutions of $f_k^\delta$ with relatively nice kernels; we do not write out the explicit proof of the above fact here, as the proof is essentially the same as that of Lemma 4.7, where a quantified version of the assertion (4-4) is proved. Thus, we have entirely constructed the sequence $\{f_k^\delta\}_{k\geq 0}$, each $f_k^\delta$ being also radially symmetric and monotone, thanks to Proposition 3.1 and (1-8).

**Remark 4.2.** Note that, for the purpose of iteration in $k$, the coefficients $A[f_{\text{in}}]$ and $\nabla a[f_{\text{in}}]$ (which are independent of time) are Hölder continuous in space thanks to (4-1).

Once we have constructed the sequence $\{f_k^\delta\}_k$, we focus on showing that it converges locally uniformly in $\mathbb{R}^3 \times [0, T_*^\delta)$ ($\delta$ fixed, $k \to \infty$) to some function $f^\delta$ in $\mathbb{R}^3 \times [0, T_*^\delta)$, where $f^\delta$ is a classical solution of

$$\partial_t f^\delta = \delta \Delta f^\delta + Q(f^\delta, f^\delta), \quad f^\delta = f_{\text{in}}.$$

The proof of this fact will take most of this section, and is achieved in Theorem 4.12. The selection of $T_*^\delta$ will guarantee that either $T_*^\delta = \infty$ or else $\|f^\delta(\cdot, t)\|_\infty$ blows up as $t \to T_*^\delta$. Then, we take the limit $\delta \to 0$ along a subsequence, making sure $f^\delta$ and its derivatives converge locally uniformly to a solution of the original nonlinear problem. This is done in Theorem 4.14, where the auxiliary assumption (4-1) is also removed.

We start by using a differential inequality argument to control the $L^\infty$ norm of the $f_k^\delta$ uniformly in $k$ and $\delta$ for at least some time interval depending only on $\|f_{\text{in}}\|_{L^\infty(\mathbb{R}^3)}$.

**Lemma 4.3.** *Let $\{f_k^\delta\}_k$ be the sequence defined above. Then for every $k \in \mathbb{N}$ we have*

$$f_k^\delta(0, t) \leq \frac{f_{\text{in}}(0)}{1 - f_{\text{in}}(0)t}, \quad \forall t \in \left[0, \frac{1}{f_{\text{in}}(0)}\right).$$

*Proof.* Since $f_{\text{in}}(0) > 0$, it is immediate that the estimate holds for $k = 0$. Arguing by induction, suppose that

$$f_{k-1}^{\delta}(0,t) \le \frac{f_{\text{in}}(0)}{1 - f_{\text{in}}(0)t}, \quad \forall t \in \Big[0, \frac{1}{f_{\text{in}}(0)}\Big).$$

Let us prove the corresponding inequality for $f_k^{\delta}$. By virtue of $f_k^{\delta}$ being smooth, radially symmetric and monotone decaying, it follows that $f_k^{\delta}(0,t) \ge f_k^{\delta}(v,t)$ for all $v$ and $t$ and $D^2 f_k^{\delta}(0,t) \le 0$ for all $t$. Plugging this information into the equation solved by $f_k^{\delta}$, we obtain

$$\begin{aligned}\partial_t f_k^{\delta}(0,t) &= 2^{-k}\Delta f_k^{\delta}(0,t) + \operatorname{Tr}(A[f_{k-1}^{\delta}](0,t) D^2 f_k^{\delta}(0,t)) + f_{k-1}^{\delta}(0,t) f_k^{\delta}(0,t)\\ &\le f_{k-1}^{\delta}(0,t) f_k^{\delta}(0,t).\end{aligned}$$

Then we may integrate the differential inequality

$$\partial_t f_k^{\delta}(0,t) \le f_{k-1}^{\delta}(0,t) f_k^{\delta}(0,t)$$

in time, and it follows that

$$f_k^{\delta}(0,t) \le f_{\text{in}}(0) e^{\int_0^t f_{k-1}^{\delta}(0,s)\,ds} \le f_{\text{in}}(0) e^{\int_0^t f_{\text{in}}(0)/(1 - f_{\text{in}}(0)s)\,ds}, \quad \forall t \in \Big[0, \frac{1}{f_{\text{in}}(0)}\Big),$$

where the last inequality was due to the inductive hypothesis. Since

$$\int_0^t \frac{f_{\text{in}}(0)}{1 - f_{\text{in}}(0)s}\,ds = -\log(1 - f_{\text{in}}(0)t),$$

it follows, as desired, that

$$f_k^{\delta}(0,t) \le \frac{f_{\text{in}}(0)}{1 - f_{\text{in}}(0)t}, \quad \forall t \in \Big[0, \frac{1}{f_{\text{in}}(0)}\Big). \qquad \square$$

Continuing with our analysis of the sequence $\{f_k^{\delta}\}_k$, we introduce a quantity that will play a crucial role in what follows: for every $T > 0$, $\delta > 0$, let

$$M(f_{\text{in}}, T, \delta) := \sup_k \|f_k^{\delta}\|_{L^{\infty}(\mathbb{R}^3 \times [0,T])} = \sup_k \sup_{0 \le t \le T} f_k^{\delta}(0,t). \tag{4-5}$$

Lemma 4.3 shows that $M(f_{\text{in}}, T, \delta) < \infty$ for at least every $T < f_{\text{in}}(0)^{-1}$ and any $\delta > 0$. For the rest of this section, we will be concerned only with those $T$ such that

$$M(f_{\text{in}}, T, \delta) < \infty. \tag{4-6}$$

**Remark 4.4.** In the following series of lemmas and propositions, culminating with Theorem 4.12, we use a series of estimates that depend on $f_{\text{in}}$, $T$, $\delta$ and the function $M(f_{\text{in}}, T, \delta)$. For the sake of brevity, throughout this section we write $C(f_{\text{in}}, T, \delta)$, $C_0(f_{\text{in}}, T, \delta)$, $C_1(f_{\text{in}}, T, \delta)$, $C'(f_{\text{in}}, T, \delta)$ (as well as $c(f_{\text{in}}, T, \delta)$ et cetera) to denote constants that depend solely on $f_{\text{in}}$, $T$, $\delta$ and $M(f_{\text{in}}, T, \delta)$, with the understanding that the constants may change from one line to the next.

The next proposition says that we can control the $L^{\infty}$ norm of the coefficients of (4-3) uniformly in $k$ and $\delta$, as long as (4-6) holds.

**Proposition 4.5.** *Let $\delta, k$ be arbitrary and $M(f_{\text{in}}, T, \delta)$ as in (4-5). For any $t \le T$ and $v \in \mathbb{R}^3$ we have the pointwise bounds*

$$A[f_k^\delta](v,t) \le a[f_k^\delta](v,t)\mathbb{I} \le \frac{2(M(f_{\text{in}},T,\delta)+1))}{1+|v|}\,\mathbb{I}, \tag{4-7}$$

$$|\nabla a[f_k^\delta](v,t)| \le \frac{M(f_{\text{in}},T,\delta)+1}{1+|v|^2}. \tag{4-8}$$

*Proof.* The bound (4-7) follows immediately from (3-2) in Lemma 3.3 applied to $h = f_k^\delta$. On the other hand, from Newton's formula (3-3) one sees immediately that

$$\nabla a[f_k^\delta] = -\frac{v}{4\pi|v|^3}\int_{B_{|v|}} f_k^\delta(w,t)\,dw. \tag{4-9}$$

Therefore,

$$|\nabla a[f_k^\delta](v,t)| = \frac{1}{4\pi|v|^2}\int_{B_{|v|}} f_k^\delta(w,t)\,dw.$$

Using the fact that $\|f_k^\delta(\cdot,t)\|_{L^1} = 1$ yields

$$|\nabla a[f_k^\delta](v,t)| \le \frac{1}{4\pi|v|^2}, \quad \forall (v,t),$$

while

$$\begin{aligned}|\nabla a[f_k^\delta](v,t)| &\le \frac{1}{4\pi|v|^2}\frac{4\pi}{3}|v|^3\|f_k^\delta(\cdot,t)\|_{L^\infty}\\ &\le \frac{1}{3}M(f_{\text{in}},T,\delta), \quad \forall (v,t)\in B_1(0)\times[0,T].\end{aligned}$$

Using that $4\pi|v|^2 \ge 1+|v|^2$ if $|v|\ge 1$, we combine the previous inequalities to obtain the bound

$$|\nabla a[f_k^\delta](v,t)| \le \frac{M(f_{\text{in}},T,\delta)+1}{1+|v|^2}, \quad \forall (v,t)\in\mathbb{R}^3\times[0,T],$$

which proves (4-8). □

For the purpose of controlling the size of $f_k^\delta(v,t)$ for large $v$, it is necessary to bound the second moment of $f_k^\delta$, in a manner which is uniform in $k$.

**Proposition 4.6.** *Let $T > 0$ and $\delta \in \left(0, \frac{1}{10}\right)$. For any $k \in \mathbb{N}$, $f_k^\delta$ satisfies the bound*

$$\int_{\mathbb{R}^3} f_k^\delta(v,t)|v|^2\,dv \le 3 + 10(1+M(f_{\text{in}},T,\delta))T, \quad \forall t\in[0,T]. \tag{4-10}$$

*Proof.* Let $\phi(v)$ be a smooth function with compact support. Using the equation solved by $f_k^\delta$, and integrating by parts, we obtain for every $t > 0$

$$\frac{d}{dt}\int f_k^\delta(v,t)\phi(v)\,dv = \int f_k^\delta\left(\delta\Delta\phi + \mathrm{Tr}(B[f_{k-1}^\delta]D^2\phi) + 2(\nabla a[f_{k-1}^\delta],\nabla\phi)\right)dv.$$

Above, $B[f_{k-1}^\delta]$ denotes $a[f_{k-1}^\delta]\mathbb{I}$ or $A[f_{k-1}^\delta]$ depending on whether $Q = Q_{\mathcal{KS}}$ or $Q = Q_{\mathcal{L}}$. Integrating in time, it follows that

$$\int f_k^\delta(v,t_2)\phi(v)\,dv - \int f_k^\delta(v,t_1)\phi(v)\,dv = \int_{t_1}^{t_2}\int f_k^\delta\big(\delta\Delta\phi + \operatorname{Tr}(B[f_{k-1}^\delta]D^2\phi) + 2(\nabla a[f_{k-1}^\delta],\nabla\phi)\big)\,dv\,dt,$$

for all $0 \le t_1 < t_2$. Next, we apply this identity to the sequence $\phi_j(v) = |v|^2\eta_j(v)$, where $\eta_j \in C_c^\infty(\mathbb{R}^3)$, and $\eta_j(v) \to 1$ locally uniformly. Due to the integrability of $f_k^\delta$ and the bounds (4-7)–(4-8), we have enough decay at infinity to pass to the limit $j \to \infty$ in the integral and conclude that the identity also holds for the function $\phi(v) = |v|^2$. Therefore, given $0 \le t_1 < t_2$, we have the identity

$$\int f_k^\delta(v,t_2)|v|^2\,dv - \int f_k^\delta(v,t_1)|v|^2\,dv = \int_{t_1}^{t_2}\int f_k^\delta\big(\delta 6 + 2\operatorname{Tr}(B[f_{k-1}^\delta]) + 4(\nabla a[f_{k-1}^\delta],v)\big)\,dv\,dt.$$

Now, the bounds (3-2)–(3-3) guarantee that in $\mathbb{R}^3\times[0,T]$ we have

$$\operatorname{Tr}(B[f_{k-1}^\delta]) \le 2M(f_{\text{in}},T,\delta)+2,$$
$$|(\nabla a[f_{k-1}^\delta],v)| \le \frac{(M(f_{\text{in}},T,\delta)+1)|v|}{1+|v|^2} \le M(f_{\text{in}},T,\delta)+1.$$

Therefore, as long as $t\in[0,T]$,

$$\left|\int_{t_1}^{t_2}\int f_k^\delta\big(\delta 6 + 2\operatorname{Tr}(B[f_{k-1}^\delta]) + 4(\nabla a[f_{k-1}^\delta],v)\big)\,dv\,dt\right| \le \int_{t_1}^{t_2}\int f_k^\delta(\delta 6 + 8M(f_{\text{in}},T,\delta)+8)\,dv\,dt$$
$$\le (6\delta + 8 + 8M(f_{\text{in}},T,\delta))(t_2-t_1).$$

Taking $t_1 = 0$ it follows that for $\delta \in \big(0,\frac{1}{10}\big)$,

$$\int f_k^\delta(v,t_2)|v|^2\,dv \le \int f_{\text{in}}|v|^2\,dv + 10(1+M(f_{\text{in}},T,\delta))T, \quad \forall t\in[0,T].$$

Since $\int f_{\text{in}}|v|^2\,dv = 3$ by assumption (1-8), this proves the proposition. □

Next, we show how $f_{k-1}^\delta \in L^\infty(\mathbb{R}_+, L^1(\mathbb{R}^3)\cap L^\infty(\mathbb{R}^3))\cap C^{\alpha,\alpha/2}(\mathbb{R}^3\times\mathbb{R}_+)$ implies Hölder continuity of the coefficients appearing in $Q(f_{k-1}^\delta, f)$, emphasizing that the estimate is uniform in $k$ for $\delta > 0$ fixed whenever $T$ is such that (4-6) holds.

**Lemma 4.7.** *Let $\delta \in \big(0,\frac{1}{10}\big)$ and $T>0$ be such that (4-6) holds. Then there is an absolute constant $C>0$ such that for any $\alpha\in(0,1)$ we have, for every $k\ge 1$, the bound*

$$[A[f_k^\delta]]_{C^{\alpha,\alpha/2}(\mathbb{R}^3\times[0,T])} \le C\big([f_k^\delta]_{C^{\alpha,\alpha/2}(\mathbb{R}^3\times[0,T])} + M(f_{\text{in}},T,\delta)+1\big),$$
$$[\nabla a[f_k^\delta]]_{C^{\alpha,\alpha/2}(\mathbb{R}^3\times[0,T])} \le C\big([f_k^\delta]_{C^{\alpha,\alpha/2}(\mathbb{R}^3\times[0,T])} + M(f_{\text{in}},T,\delta)+1\big).$$

*Proof.* Let $\eta\in C^\infty(\mathbb{R}^3)$ be an even function such that $\eta\equiv 1$ in $B_1(0)$ and $\eta\equiv 0$ outside $B_2$. Let us write

$$A[f_k^\delta] = A_1[f_k^\delta] + A_2[f_k^\delta].$$

Each $A_i$ $(i = 1, 2)$ is given by convolutions $A_i[f_k^\delta] = K_i * f_k^\delta$ with the respective kernels

$$K_1(v) := \frac{1}{8\pi|v|}\left(\mathbb{I} - \frac{v\otimes v}{|v|^2}\right)\eta(v), \quad K_2(v) := \frac{1}{8\pi|v|}\left(\mathbb{I} - \frac{v\otimes v}{|v|^2}\right)(1-\eta(v)).$$

Let us show that $A_1$, $A_2$ are Hölder continuous in $v$ ant $t$. We make use of the fact that there is a constant $C(\eta)$ such that

$$\int_{\mathbb{R}^3} |K_1(v)|\,dv + \sup_v |K_2(v)| + \sum_{i=1}^3 \sup_v |\partial_i K_2(v)| + \sum_{i,j=1}^3 \sup_v |\partial_{ij} K_2(v)| \le C(\eta),$$

where the matrix norm used is the standard $L^2$ norm $|A| = \mathrm{Tr}(AA^*)^{1/2}$. For $A_1$ it is straightforward that

$$\begin{aligned}|A_1(v_1,t_1) - A_1(v_2,t_2)| &\le \int_{B_2} |K_1(w)||f_k^\delta(v_1-w,t_1) - f_k^\delta(v_2-w,t_2)|\,dw\\ &\le \left(\int_{B_2}|K_1(w)|\,dw\right)\sup_{w\in B_2(0)} |f_k^\delta(v_1-w,t_1) - f_k^\delta(v_2-w,t_2)|,\end{aligned}$$

the above holding for any $(v_i, t_i)$, so that

$$[A_1]_{C^{\alpha,\alpha/2}} \le C(\eta)[f_k^\delta]_{C^{\alpha,\alpha/2}}.$$

Next we deal with $A_2$, which in fact will be Lipschitz continuous. Fix $e\in\mathbb{S}^2$ and set $K_{2,e}(v) := (K_2(v)e, e)$. Using the equation for $f_k^\delta$ and integration by parts,

$$\begin{aligned}&\partial_t(A_2[f_k^\delta](v)e, e)\\ &\quad= \int_{B_1^c} K_{2,e}(w-v)\partial_t f_k^\delta\,dw\\ &\quad= -\int_{B_1^c}(\nabla_w K_{2,e}(w-v), (A[f_{k-1}^\delta] + \delta\mathbb{I})\nabla_w f_k^\delta)\,dw + \int_{B_1^c} f_k^\delta(\nabla_w a[f_{k-1}^\delta], \nabla_w K_{2,e}(w-v))\,dw.\end{aligned}$$

Integrating by parts once again,

$$\begin{aligned}&-\int_{B_1^c}(\nabla_w K_{2,e}(w-v), (A[f_{k-1}^\delta] + \delta\mathbb{I})\nabla_w f_k^\delta)\,dw\\ &\quad= \int_{B_1^c}\mathrm{div}_w((A[f_{k-1}^\delta] + \delta\mathbb{I})\cdot\nabla_w K_{2,e}(w-v))f_k^\delta\,dw\\ &\quad= \int_{B_1^c} f_k^\delta\,\mathrm{Tr}(A[f_{k-1}^\delta]D_w^2 K_{2,e}(w-v))\,dw + \int_{B_1^c} f_k^\delta\nabla_w a[f_{k-1}^\delta]\cdot\nabla_w K_{2,e}(w-v)\,dw\\ &\qquad\qquad + \delta\int_{B_1^c} f_k^\delta\Delta_w K_{2,e}(w-v)\,dw.\end{aligned}$$

Gathering all of the above, it follows that

$$\begin{aligned}\partial_t(A_2[f_k^\delta]e, e) = &\int_{B_1^c} f_k^\delta\,\mathrm{Tr}(A[f_{k-1}^\delta]D_w^2 K_{2,e}(w-v))\,dw\\ &+ 2\int_{B_1^c} f_k^\delta(\nabla_w a[f_{k-1}^\delta], \nabla_w K_{2,e}(w-v))\,dw + \delta\int_{B_1^c} f_k^\delta\Delta_w K_{2,e}(w-v)\,dw.\end{aligned}$$

Therefore, we have the bound

$$\begin{aligned}
&|\partial_t(A_2[f_k^\delta](v)e, e)| \\
&\quad \le \|D^2K_{2,e}\|_{L^\infty}\|A[f_{k-1}^\delta]\|_{L^\infty}\|f\|_{L^1} + 2\|\nabla K_{2,e}\|_{L^\infty}\|\nabla a[f_{k-1}^\delta]\|_{L^\infty}\|f\|_{L^1} + \delta\|\Delta K_{2,e}\|_{L^\infty}\|f_k^\delta\|_{L^1} \\
&\quad \le \|K_{2,e}\|_{C^2}(\|A[f_{k-1}^\delta]\|_{L^\infty} + \|\nabla a[f_{k-1}^\delta]\|_{L^\infty} + \delta) \\
&\quad \le \|K_{2,e}\|_{C^2}(3M(f_{\text{in}}, T, \delta) + 4),
\end{aligned}$$

where we used (4-7)–(4-8) and $\delta \in \left(0, \frac{1}{10}\right)$ in the last inequality. Since $\|K_{2,e}\| \le C(\eta)$ for all $e$,

$$|\partial_t(A_2[f_k^\delta](v)e, e)| \le 4C(\eta)(M(f_{\text{in}}, T, \delta) + 1).$$

This immediately implies a Lipschitz bound in time for $A_2$, namely

$$|A_2(v, t_1) - A_2(v, t_2)| \le 12\|K_2\|_{C^2}(M(f_{\text{in}}, T, \delta) + 1)|t_1 - t_2|, \quad \forall v \in \mathbb{R}^3,\ t_1, t_2 \ge 0.$$

For the spatial regularity, from the definition of $A_2$ and the triangle inequality it follows that

$$\begin{aligned}
|A_2(v_1, t) - A_2(v_2, t)| &\le \int |K_2(w - v_1) - K_2(w - v_2)| f_k^\delta(w, t)\, dw \\
&\le C(\eta)|v_1 - v_2| \int f_k^\delta(w, t)\, dw \quad \forall v_1, v_2 \in \mathbb{R}^3,\ t \ge 0.
\end{aligned}$$

Then, thanks to $\|f_k^\delta(\cdot, t)\|_{L^1} = 1$, it follows that

$$|A_2(v_1, t) - A_2(v_2, t)| \le C(\eta)|v_1 - v_2|, \quad \forall v_1, v_2 \in \mathbb{R}^3,\ t > 0.$$

Finally, we combine the estimates in time and space to see that

$$\begin{aligned}
|A_2(v_1, t_1) - A_2(v_2, t_2)| &\le |A_2(v_1, t_1) - A_2(v_2, t_1)| + |A_2(v_2, t_1) - A_2(v_2, t_2)| \\
&\le 15C(\eta)(M(f_{\text{in}}, T, \delta) + 1)(|v_1 - v_2| + |t_1 - t_2|), \quad \forall (v_i, t_i),\ i = 1, 2.
\end{aligned}$$

Since $|v_1 - v_2| + |t_1 - t_2| \le |v_1 - v_2|^\alpha + |t_1 - t_2|^{\alpha/2}$ when $|v_1 - v_2|, |t_1 - t_2| \le 1$, we conclude that

$$[A_2]_{C^{\alpha,\alpha/2}(\mathbb{R}^3 \times [0,T])} \le 15C(\eta)(M(f_{\text{in}}, T, \delta) + 1).$$

The proof of Hölder regularity for $\nabla a[f_k^\delta](v, t)$ can be done in an entirely analogous manner, writing the kernel as the sum of integrable and $C^2$ parts. One may also make a slightly different argument, using the fact that since $f_k^\delta$ is spherically symmetric, we have the identity (4-9), which yields a similar bound. □

For the purposes of the proof of existence of solutions, we require several parabolic estimates that are local in space but uniform up to $t = 0$. Notice these are different to the interior estimates stated in Section 2, namely Theorems 2.2, 2.3 and 2.4, which will be of chief importance in later sections. The parabolic estimates hold in a space-time cylinder, which starts at time $t = 0$, and are in terms of norms of the initial data. They guarantee in particular that under the auxiliary assumptions (4-1) on $f_{\text{in}}$ the functions $f_k^\delta$ have spatial decay on their second derivatives.

**Lemma 4.8** (Hölder estimate for regular initial data). *There exists some $\alpha \in (0,1)$ and constant $c$, which only depends on $\delta$, $f_{\mathrm{in}}$, $T$ and $[f_{\mathrm{in}}]_{C^{2+\beta}(\mathbb{R}^3)}$, such that for any $v \in \mathbb{R}^3$ and $k \geq 1$,*

$$[f_k^\delta]_{C^{\alpha,\alpha/2}(B_1(v)\times[0,T])} \leq c(\delta, M(f_{\mathrm{in}}, T, \delta), [f_{\mathrm{in}}]_{C^1(B_2(v))}). \tag{4-11}$$

(Schauder estimate up to the initial time). *Let $\beta \in (0,1)$. Then for any $v \in \mathbb{R}^3$, $k \geq 1$,*

$$[f_k^\delta]_{C^{2+\alpha,1+\alpha/2}(B_1(v)\times[0,T])} \leq C(\|f_k^\delta\|_{L^\infty(B_2(v)\times[0,T])} + [f_{\mathrm{in}}]_{C^{2+\beta}(B_2(v))}), \tag{4-12}$$

*where $C = C(f_{\mathrm{in}}, T, \delta)$.*

*Proof.* For the proof of the first estimate we refer to [Ladyženskaja et al. 1968, Theorem 10.1, page 204]. Note that the constant does not depend in any way on the regularity of the coefficients in the equation solved by $f_k^\delta$, and depends only on the ellipticity constants and the regularity of $f_{\mathrm{in}}$. The second estimate follows from [Ladyženskaja et al. 1968, Theorem 10.1, page 351], noting that the space-time Hölder norm of the coefficients $A[f_{k-1}^\delta]$, $\nabla a[f_{k-1}^\delta]$ is bounded by a constant $C(f_{\mathrm{in}}, T, \delta)$, thanks to Lemma 4.7 and the first estimate (4-11) applied to $f_{k-1}^\delta$ (when $k > 1$; $f_0^\delta \equiv f_{\mathrm{in}}$ for $k = 1$, which is regular in space and constant in time). □

Next we show that the diffusion matrices $A[f_k^\delta] + \delta\mathbb{I}$ are Hölder continuous in a manner which is uniform in $k$ (but possibly depending on $\delta$). In this case, standard estimates for linear parabolic equations yield Hölder bounds on the second-order spatial derivatives and first-order temporal derivatives for $f_k^\delta$, these being uniform in $k$. Particularly, since we are assuming a spatial decay for the second derivatives of $f_{\mathrm{in}}$ (see (2-2)), the same holds for $f_k^\delta$.

**Proposition 4.9.** *Let $\delta \in \left(0, \frac{1}{10}\right)$ and $0 < T < \infty$ be such that (4-6) holds. Then there is a $C$ depending only on $f_{\mathrm{in}}$, $\delta$, $T$, $M(f_{\mathrm{in}}, T, \delta)$ such that*

$$\|D^2 f_k^\delta\|_{C^\alpha(B_1(v)\times[0,T])} \leq C(1+|v|^5)^{-1}, \quad \forall v \in \mathbb{R}^3. \tag{4-13}$$

*Proof.* We first show that $f_k^\delta(v,t)$ decays as $(1+|v|^5)^{-1}$ for $v$ large. Fixing $v \in \mathbb{R}^3$, the spherical symmetry and radial monotonicity of $f_k^\delta$ implies that

$$\frac{7}{6}\pi|v|^3 f_k^\delta(v,t) \leq \int_{B_{|v|}\setminus B_{|v|/2}} f_k^\delta(w,t)\,dw \leq \frac{4}{|v|^2}\int_{\mathbb{R}^3} f_k^\delta(w,t)|w|^2\,dw.$$

Using the second moment bound (4-10), we arrive at the estimate

$$f_k^\delta(v,t) \leq \frac{4}{\pi|v|^5}\big(3 + 10(1+M(f_{\mathrm{in}}, T, \delta))T\big)$$

for all $|v| \geq 1$ and $t \in [0,T]$. Since $f_k^\delta(v,t) \leq M(f_{\mathrm{in}}, T, \delta)$ as long as $t \leq T$, we conclude that

$$f_k^\delta(v,t) \leq \frac{C'(f_{\mathrm{in}}, T, \delta)}{1+|v|^5}, \quad \forall (v,t) \in \mathbb{R}^3 \times [0,T], \tag{4-14}$$

with $C'(f_{\mathrm{in}}, T, \delta) := \max\big\{M, \frac{4}{\pi}\big(3 + 10(1+M(f_{\mathrm{in}}, T, \delta))T\big)\big\}$. The bound follows, combining the initial bound (4-1), the decay estimate(4-14) and the estimate (4-12) from Lemma 4.8. □

So far we have shown the existence of the sequence $\{f_k^\delta\}$, and proven several uniform estimates which are uniform in $k$ for times $T < T_*^\delta$. Moving towards obtaining a limit from this sequence, we prove an iterative estimate on the size of the functions $\{f_k^\delta - f_{k-1}^\delta\}_k$ in $\mathbb{R}^3 \times [0, T]$, for $\delta > 0$ fixed and $T$ such that $C(f_{\text{in}}, T, \delta) < \infty$.

**Lemma 4.10.** *Let $\delta \in \left(0, \frac{1}{10}\right)$ and $T > 0$ be such that (4-6) holds, and let $w_k^\delta := f_{k-1}^\delta - f_k^\delta$ for each $k \geq 1$. There is a number $0 < T_0 < T$, $T_0 = T_0(f_{\text{in}}, T, \delta)$ with the following properties*:

(1) *For each $k \geq 2$,*

$$\|w_k(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[0,T_0])} \leq \tfrac{1}{4}\|w_{k-1}(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[0,T_0])}.$$

(2) *For each $k \geq 2$ and $l = 1, \ldots, l_0$, we have*

$$\|w_k^\delta(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[t_{l-1},t_l])} \leq \tfrac{1}{4}\|w_{k-1}^\delta(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[t_{l-1},t_l])} + 2\|w_k^\delta(t_{l-1})\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)}.$$

*Here $l_0 \in \mathbb{N}$ is the largest such that $(l_0 - 1)T_0 \leq T$, and $t_l := \min\{lT_0, T\}$.*

*Proof.* We drop the superscript $\delta$ for convenience. Using the equations for $f_{k-1}$ and $f_k$ we get that $w_k = f_{k-1} - f_k$ satisfies

$$\begin{cases} \partial_t w_k = \delta\Delta w_k + \operatorname{Tr}(A[f_{k-2}]D^2 w_k) + f_{k-2}w_k + \operatorname{Tr}(A[w_{k-1}]D^2 f_k) + f_k w_{k-1}, & \text{for } t > 0, \\ w_k = 0, & \text{for } t = 0. \end{cases} \tag{4-15}$$

*Step 1.* According to Proposition 4.9, there is a positive constant $C(f_{\text{in}}, T, \delta)$ such that

$$|D^2 f_k^\delta(v,t)| \leq C(f_{\text{in}}, T, \delta)(1 + |v|^5)^{-1}, \quad \forall v \in \mathbb{R}^3,\ t \in [0, T]. \tag{4-16}$$

The estimate (4-16) and the estimate (3-5) applied to $w_{k-1}$ imply the inequality

$$|\operatorname{Tr}(A[w_{k-1}]D^2 f_k(v,t))| \leq C(f_{\text{in}}, T, \delta)\left(\frac{\|w_{k-1}(\cdot,t)\|_{L^\infty(\mathbb{R}^3)} + \|w_{k-1}(\cdot,t)\|_{L^1(\mathbb{R}^3)}}{1 + |v|^5}\right),$$

which holds for any $(v,t) \in \mathbb{R}^3 \times [0,T]$. On the other hand, $\langle v\rangle^{-4} \in L^1(\mathbb{R}^3)$. Therefore,

$$\|w_k(t)\|_{L^1} = \int_{\mathbb{R}^3} |w_k(v,t)|\langle v\rangle^4\langle v\rangle^{-4}\,dv \leq \|w_k(t)\langle v\rangle^4\|_{L^\infty}\|\langle v\rangle^{-4}\|_{L^1(\mathbb{R}^3)}.$$

Substituting this in the last estimate, we arrive at the bound,

$$|\operatorname{Tr}(A[w_{k-1}]D^2 f_k(v,t))| \leq C(f_{\text{in}}, T, \delta)\|w_{k-1}(t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)}(1 + |v|^5)^{-1}.$$

*Step 2.* Consider the function $h_0(v) := \langle v\rangle^{-4} = (1 + |v|^2)^{-2}$. We have

$$\begin{aligned} Dh_0(v) &= -4(1 + |v|^2)^{-3}v, \\ D^2 h_0(v) &= -4(1 + |v|^2)^{-3}\mathbb{I} + 24(1 + |v|^2)^{-4}v \otimes v. \end{aligned}$$

In particular,

$$\begin{aligned} \Delta h_0 &= 12(|v|^2 - 1)\langle v\rangle^{-8}, \\ \operatorname{Tr}(A[f_{k-2}]D^2 h_0) &= -4\langle v\rangle^{-6}a[f_{k-2}] + 24\langle v\rangle^{-8}(A[f_{k-2}]v, v). \end{aligned}$$

Using the inequalities $||v|^2-1|$, $|v|^2 \le \langle v\rangle^2$, the above leads to

$$|\delta \Delta h_0| \le 12\delta \langle v\rangle^{-6},$$
$$|\operatorname{Tr}(A[f_{k-2}]D^2 h_0)| \le 4\langle v\rangle^{-6} a[f_{k-2}] + 24\langle v\rangle^{-6} a[f_{k-2}].$$

Then, recalling that $\delta \in \left(0, \frac{1}{10}\right) \Rightarrow 12\delta < \frac{3}{2}$, we combine the above inequalities into one,

$$|\delta\Delta h_0 + \operatorname{Tr}(A[f_{k-2}]D^2 h_0)| \le 28(1+a[f_{k-2}])\langle v\rangle^{-6} \le 56(1+C(f_{\mathrm{in}},T,\delta))h_0,$$

where we have used (4-7) to bound $a[f_{k-2}]$.

*Step 3.* Next, let

$$H_0(v,t) := RA^{-1}(e^{At}-1)h_0(v),$$

for $A, R > 0$ to be determined. It is immediate that

$$\partial_t H_0 = AH_0 + Rh_0.$$

The last inequality in Step 2 implies that

$$|\delta\Delta H_0 + \operatorname{Tr}(A[f_{k-2}]D^2 H_0)| + f_{k-2}H_0 \le 60(1+C(f_{\mathrm{in}},T,\delta))H_0.$$

The estimates from Step 1, the definition of $h_0(v)$ and (4-14) yield

$$\operatorname{Tr}(A[w_{k-1}]D^2 f_k) + f_k w_{k-1} \le C_0 \|w_{k-1}(t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)} h_0(v),$$

with $C_0 = C_0(f_{\mathrm{in}},T,\delta)$. In light of this, for any $T_0 \in (0,T)$, we choose $A$ and $R$ to be

$$A = 60(1+C(f_{\mathrm{in}},T,\delta)),$$
$$R = C_0 \sup_{0\le t\le T_0} \|w_{k-1}(t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)},$$

in which case we have, for any $(v,t) \in \mathbb{R}^3 \times [0,T_0]$,

$$\begin{aligned}\partial_t H_0 &\ge 60(1+C(f_{\mathrm{in}},T,\delta))H_0 + C_0\big(\|w_{k-1}(\cdot,t)\|_{L^\infty(\mathbb{R}^3)} + \|w_{k-1}(\cdot,t)\|_{L^1(\mathbb{R}^3)}\big)h_0 \\ &\ge \delta\Delta H_0 + \operatorname{Tr}(A[f_{k-2}]D^2 H_0) + f_{k-2}H_0 + (\operatorname{Tr}(A[w_{k-1}]D^2 f_k) + f_k w_{k-1}).\end{aligned}$$

This means that $H_0$ is a supersolution of (4-15), the parabolic equation solved by $w_k$. Furthermore, $H_0(\cdot,0) = w_k(\cdot,0) = 0$. Then, thanks to the comparison principle,

$$w_k \le H_0 \quad \text{in } \mathbb{R}^3 \times [0,T_0].$$

The same argument applied to $-w_k$ yields

$$\eta_k \le H_0 \quad \text{in } \mathbb{R}^3 \times [0,T_0].$$

We have shown that there are constants $C_0(f_{\mathrm{in}},T,\delta)$ and $C_1(f_{\mathrm{in}},T,\delta)$ such that

$$|w_k(v,t)| \le C_0 \|w_{k-1}(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[0,T_0])}(e^{C_1(f_{\mathrm{in}},T,\delta)t}-1)\langle v\rangle^{-4} \quad \text{in } \mathbb{R}^3\times[0,T_0].$$

In particular, there is a $T_0$, depending only on $T$ and $C_0(f_{\text{in}}, T, \delta)$, such that

$$T_0 \in (0, T) \quad \text{and} \quad C_0\big(e^{C_1(f_{\text{in}},T,\delta)T_0} - 1\big) \le \tfrac{1}{4}.$$

This results in the estimate

$$\|w_k(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[0,T_0])} \le \tfrac{1}{4}\|w_{k-1}(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[0,T_0])},$$

and the first part of the lemma is proved.

*Step 4.* Fix $k \ge 2$. Assume for now that $2T_0 < T$ — the same $T_0$ as in Step 3 — and define the function $H_1 : \mathbb{R}^3 \times [T_0, \infty) \to \mathbb{R}$ by

$$H_1(v,t) := RA^{-1}(e^{A(t-T_0)} - 1)h_0(v) + \|w_k(T_0)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)}h_0(v),$$

where $A$ and $R$ are to be determined. A straightforward computation yields

$$\begin{aligned}\partial_t H_1 &= Re^{A(t-t_0)}h_0(v)\\ &= A\big(RA^{-1}(e^{A(t-T_0)} - 1)h_0(v) + \|w_k(T_0)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)}h_0(v)\big) + Rh_0(v) - \|w_k(T_0)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)}h_0(v)\\ &= AH_1 + Rh_0(v) - \|w_k(T_0)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)}h_0(v).\end{aligned}$$

As in the previous step, we have

$$\begin{aligned}\delta\Delta H_1 + \operatorname{Tr}(A[f_{k-2}]D^2H_1) + f_{k-2}H_1 + \operatorname{Tr}(A[w_{k-1}]D^2f_k) + f_k w_{k-1} + 2^{-k}\Delta f_k&\\ \le 60(1 + C(f_{\text{in}}, T, \delta))H_1 + h_0C_0\|w_{k-1}(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[T_0,2T_0])}&\\ = AH_1 + h_0(R - A\|w_k(v,T_0)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)}) = \partial_t H_1&\end{aligned}$$

by choosing

$$\begin{aligned}A =& 60(1 + C(f_{\text{in}}, T, \delta)),\\ R =& C_0\|w_{k-1}(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[T_0,2T_0])} + 60(1 + C(f_{\text{in}}, T, \delta))\|w_k(v,T_0)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)}.\end{aligned}$$

Likewise, $H_1(\cdot, T_0) \ge w_k(\cdot, T_0)$. Then, just as before, the comparison principle says that $H_1(\cdot, t) \ge w_k(\cdot, t)$ for $t \in [T_0, 2T_0]$,

$$\begin{aligned}|w_k(v,t)| \le C_0\big(e^{C_1(f_{\text{in}},T,\delta)t} - 1\big)\|w_{k-1}(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[T_0,2T_0])}\langle v\rangle^{-4}&\\ + \|w_k(v,T_0)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)}\big(e^{C_1(f_{\text{in}},T,\delta)t} - 1\big)\langle v\rangle^{-4} + \|w_k(v,T_0)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3)}\langle v\rangle^{-4}.&\end{aligned}$$

Hence for $t \in [T_0, 2T_0]$ we get

$$\|w_k(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[T_0,2T_0])} \le \tfrac{1}{4}\|w_{k-1}(v,t)\langle v\rangle^4\|_{L^\infty(\mathbb{R}^3\times[T_0,2T_0])} + 2\|w_k(v,T_0)(1 + |v|^2)^2\|_{L^\infty(\mathbb{R}^3)}.$$

This yields the second estimate in the case $l = 2$. The above argument can be repeated to obtain a further estimate in the interval $[2T_0, 3T_0]$, and so on. After a finite number of iterations we will reach some $l_0 \in \mathbb{N}$ such that $(l_0 - 1)T_0 \le T$ and $l_0T_0 > T$. In that case we repeat the above argument on the interval $[(l_0 - 1)T_0, T]$, yielding the respective bound and completing the proof of the second estimate. □

The next lemma shows that if $\delta \in \left(0, \frac{1}{10}\right)$ and $T$ is a time for which (4-6) holds, the sequence $f_k^\delta$ converges uniformly in $\mathbb{R}^3 \times [0, T]$ to a continuous limit $f^\delta$.

**Lemma 4.11.** *Let $\{f_k^\delta\}_k$, $\delta \in \left(0, \frac{1}{10}\right)$, and $T > 0$ be such that (4-6) holds. Then there is a continuous function $f^\delta : \mathbb{R}^3 \times [0, T] \to \mathbb{R}$ such that*

$$\lim_k \| f^\delta - f_k^\delta \|_{L^\infty(\mathbb{R}^3 \times [0,T])} = 0,$$
$$\lim_k \| f^\delta - f_k^\delta \|_{L^\infty(0,T;L^1(\mathbb{R}^3))} = 0.$$

*Proof.* Let $T_0 > 0$ and $l_0$ and $t_l$ be as in Lemma 4.10. Define, for $l = 0, 1, \dots, l_0$ and $k \in \mathbb{N}$,

$$E_{k,l} := \| w_k(v,t) \langle v \rangle^4 \|_{L^\infty(\mathbb{R}^3 \times [t_{l-1}, t_l])}.$$

Then Lemma 4.10 says that the recursive relations

$$E_{k,1} \le \tfrac{1}{4} E_{k-1,1},$$
$$E_{k,l} \le 4 E_{k,l-1} + \tfrac{1}{2} E_{k-1,l}$$

hold for $k \ge 2$ and $l = 0, \dots, l_0$. We claim that these recurrence relations guarantee the summability in $k$ of the sequence $\{E_{k,l}\}_k$ for any fixed $l = 1, \dots, l_0$. The first recurrence relation implies that $E_{k,1}$ decays geometrically, thus we immediately have

$$\sum_{k=3}^{\infty} E_{k,1} < \infty.$$

Next, suppose that for some $1 < l < l_0$ we have

$$\sum_{k=3}^{\infty} E_{k,l} < \infty.$$

Taking the sum for $k$ from 3 to $N$ of the second recursive relation, we get

$$\sum_{k=3}^{N} \frac{1}{2} E_{k,l+1} \le 4 \sum_{k=3}^{N} E_{k,l} + \frac{1}{2} E_{2,l+1}.$$

We can then pass to the limit $N \to +\infty$, and use the summability for $E_{k,l}$ to obtain

$$\sum_{k=3}^{N} \frac{1}{2} E_{k,l+1} < +\infty.$$

Combining the summability of the sequences $\{E_{k,l}\}_k$ for every $l \le l_0$, we conclude that

$$\sum_k \| (f_k(v,t) - f_{k-1}(v,t)) \langle v \rangle^4 \|_{L^\infty(\mathbb{R}^3 \times [0,T])} < \infty.$$

Since $\langle v\rangle \geq 1$ for all $v$, and $\langle v\rangle^{-4} = (1+|v|^2)^{-2} \in L^1(\mathbb{R}^3)$, this implies that

$$\sum_k \|f_k - f_{k-1}\|_{L^\infty(\mathbb{R}^3\times[0,T])} < \infty,$$
$$\sum_k \|f_k - f_{k-1}\|_{L^\infty(0,T;L^1(\mathbb{R}^3))} < \infty.$$

This summability implies $\{f_k\}$ is a Cauchy sequence in each norm, proving the lemma. □

**Theorem 4.12.** *For each $\delta \in \left(0, \frac{1}{10}\right)$, there is a time $T_*^\delta = T_*^\delta(f_{\text{in}})$ with $0 < T_*^\delta \leq \infty$ and a function $f^\delta$ in $C^{2,1}_{\text{loc}}(\mathbb{R}^3 \times [0, T_*))$ such that*

$$\begin{cases} \partial_t f^\delta = \delta \Delta f^\delta + Q(f^\delta, f^\delta) & \text{in } \mathbb{R}^3 \times [0, T_*^\delta), \\ f^\delta(\cdot, 0) = f_{\text{in}}. \end{cases}$$

*Moreover, either $T_*^\delta = \infty$ or*

$$\limsup_{T \to T_*^{\delta -}} \|f^\delta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^3))} = \infty.$$

*Proof. Step 1.* Let

$$T_*^\delta := \sup\{T > 0 \mid M(f_{\text{in}}, T, \delta) < \infty\}.$$

By Lemma 4.3 we have $T_*^\delta \geq (2f_{\text{in}}(0))^{-1}$, thus $T_*^\delta > 0$. It may certainly be that $T_*^\delta = \infty$. Now, we may apply Lemma 4.11 to $f_k^\delta$ and any fixed $T < T_*^\delta$, resulting in a continuous function $f^\delta : \mathbb{R}^3 \times [0, T_*^\delta) \to \mathbb{R}$ such that

$$f_k^\delta \to f^\delta \text{ uniformly in } \mathbb{R}^3 \times [0, T), \quad \forall T < T_*^\delta.$$

On the other hand, we have the estimates from Lemma 4.8, which guarantee, by the Arzelà–Ascoli theorem, that for any subsequence $k_n \to \infty$ there is a subsequence $k_n'$ such that $\partial_t f_{k_n'}^\delta$ and $D^2 f_{k_n'}^\delta$ converge locally uniformly in $\mathbb{R}^3 \times [0, T_*)$ as $n \to \infty$. Since $f_k^\delta \to f$ locally uniformly and $\{k_n\}$ was arbitrary, it follows that (i) $f^\delta \in C^{2,1}_{\text{loc}}(\mathbb{R}^3 \times [0, T_*))$, and (ii) the sequences $D^2 f_k^\delta$ and $\partial_t f_k^\delta$ converge locally uniformly to $D^2 f^\delta$ and $\partial_t f^\delta$ as $k \to \infty$, respectively.

*Step 2.* Let us show the matrices $\{A[f_k^\delta]\}_k$ converge locally uniformly in $\mathbb{R}^3 \times [0, T_*^\delta)$ to $A[f^\delta]$. Indeed, let $t \in [0, T_*^\delta)$ and apply the estimate (3-5) to $g = |f_k(\cdot, t) - f_k^\delta(\cdot, t)|$ (which is a nonnegative, bounded, spherically symmetric function), which leads to the bound

$$|A[f_k^\delta](v,t) - A[f^\delta](v,t)| \leq 2\big(\|f_k(\cdot,t) - f_k^\delta(\cdot,t)\|_{L^\infty(\mathbb{R}^3)} + \|f_k(\cdot,t) - f_k^\delta(\cdot,t)\|_{L^1(\mathbb{R}^3)}\big)$$

for all $v$ and $t < T_*^\delta$. Then Lemma 4.11 shows that $A[f_k^\delta]$ converges uniformly to $A[f^\delta]$ uniformly in $\mathbb{R}^3 \times [0, T]$ for every $T < T_*^\delta$.

*Step 3.* Thanks to the local uniform convergence of $f_k^\delta$, $D^2 f_k^\delta$, $\partial_t f_k^\delta$ and $A[f_k^\delta]$ proved in the previous two steps, we can pass to the limit in the equation for $f_k^\delta$ and conclude that

$$\partial_t f^\delta = \delta \Delta f^\delta + Q(f^\delta, f^\delta) \quad \text{in } \mathbb{R}^3 \times [0, T_*^\delta).$$

*Step 4.* We show here that if $T_*^\delta$ is finite, then the $L^\infty$ norm of $f^\delta(\cdot, t)$ goes to infinity as $t$ approaches $T_*^\delta$. Arguing by contradiction, suppose that $T_*^\delta$ is finite, and

$$\limsup_{T \to T_*^\delta} f^\delta(0, t) < +\infty.$$

Since $f^\delta$ is continuous and bounded for any $t < T_*^\delta$, then $f^\delta$ is bounded for any $t \le T_*^\delta$ and in particular,

$$f^\delta(0, T_*^\delta - \varepsilon) \le C, \quad \varepsilon > 0.$$

The uniform convergence of $f_k^\delta \to f^\delta$ for all $t < T_*^\delta$ shows that for any small enough $\varepsilon > 0$ there is some $k_0$ such that

$$f_k^\delta(0, T_*^\delta - \varepsilon) < 2C, \quad \forall k > k_0. \tag{4-17}$$

Since $\sup_k f_k^\delta(0, T_*^\delta - \varepsilon) < +\infty$, we have that (4-17) implies

$$f_k^\delta(0, T_*^\delta - \varepsilon) < \widetilde{C}, \quad \forall k \ge 1.$$

Then the differential inequality argument from Lemma 4.3, applied with starting time shifted to $T_*^\delta - \varepsilon$, proves that

$$f_k^\delta(0, T_*^\delta - \varepsilon + t) \le \frac{\widetilde{C}}{1 - \widetilde{C} t}, \quad \forall k \ge 1,\ 0 < t < \frac{1}{\widetilde{C}}.$$

Taking now $t = 1/(2\widetilde{C})$ and $\varepsilon = 1/(4\widetilde{C})$ yields

$$f_k^\delta(0, T_*^\delta + \varepsilon) < 2\widetilde{C},$$

which contradicts the maximality of $T_*^\delta$ and the theorem is proved. □

Next, we show that as long as $f^\delta(v, t)$ is bounded in a time interval $[0, T]$, the mass of $f^\delta(v, t)$ cannot escape to infinity nor concentrate at the origin. The bound is independent of $\delta$. A consequence of this result is a local lower bound for $A[f^\delta]$ along radial directions.

**Proposition 4.13.** *Let $\delta \in \left(0, \frac{1}{10}\right)$, let $f^\delta$ be a function given by Theorem 4.14, let $T < T_*^\delta$ and let $M > 0$ be such that*

$$\| f^\delta \|_{L^\infty \times [0,T]} < M.$$

*Then there are radii $r(f_{\text{in}}, T, M)$ and $R(f_{\text{in}}, T, M)$ such that $0 < r < R < \infty$ and*

$$\int_{B_R \setminus B_r} f^\delta(v, t)\, dv \ge \frac{1}{2}, \quad \forall t \in [0, T]. \tag{4-18}$$

*As a consequence, there is a positive constant $c_0 = c_0(f_{\text{in}}, T, M)$ such that*

$$A^*[f^\delta](v, t) \ge \frac{c_0}{1 + |v|^3}, \quad \forall v \in \mathbb{R}^3,\ t \in [0, T],\ k \in \mathbb{N}, \tag{4-19}$$

*where $A^*[\,\cdot\,]$ is as defined in (3-2).*

*Proof.* Given $R > 0$, the mass of $f^\delta$ outside $B_R(0)$ may be estimated via its second moment

$$\int_{B_R^c} f^\delta \, dv \le \int_{B_R^c} f^\delta \frac{|v|^2}{R^2} \, dv \le \frac{1}{R^2} \int_{\mathbb{R}^3} f^\delta(v,t)|v|^2 \, dv.$$

Moreover, for any $r$, $R$ with $R > r > 0$ there is the obvious lower bound

$$\begin{aligned} \int_{B_R \setminus B_r} f^\delta(v,t) \, dv &= 1 - \int_{B_R^c} f^\delta(v,t) \, dv - \int_{B_r} f^\delta(v,t) \, dv \\ &\ge 1 - \frac{1}{R^2} \int_{\mathbb{R}^3} f^\delta(v,t)|v|^2 \, dv - \frac{4\pi}{3} r^3 M, \end{aligned} \tag{4-20}$$

using the fact that $\|f^\delta(\cdot,t)\|_{L^1} = 1$. Following exactly the same steps as in the proof of Proposition 4.6, one can show

$$\int_{\mathbb{R}^3} f^\delta(v,t)|v|^2 \, dv \le 3 + 10(1+M)T, \quad \forall t \in [0,T]. \tag{4-21}$$

Hence (4-18) follows from (4-20) and (4-21) by choosing

$$\begin{aligned} R &:= 2(3 + 10(1+M)T)^{1/2}, \\ r &:= (8\pi M)^{-1/3}. \end{aligned}$$

Finally, (4-19) follows from (4-18), the selection of $R$ and $r$ above and Lemma 3.3. □

**Theorem 4.14.** *Given $f_{\text{in}}$ as in (1-8), there is a time $T_*$ and a function $f \in C^{2;1}_{\text{loc}}(\mathbb{R}^3 \times (0, T_*))$ with initial data $f_{\text{in}}$, which solves (1-4) or (1-7). Moreover, either $T_* = \infty$ or*

$$\limsup_{t \to T_*^-} \|f\|_{L^\infty(\mathbb{R}^3 \times [0,t])} = \infty.$$

*The initial data is achieved in the sense that for any $\phi \in C_c^\infty(\mathbb{R}^3)$ and any $t \in (0, T_*)$ we have*

$$\int_{\mathbb{R}^3} f(v,t)\phi(v) \, dv - \int_{\mathbb{R}^3} f_{\text{in}}(v)\phi(v) \, dv = -\int_0^t \int (B[f]\nabla f - f\nabla a[f], \nabla\phi) \, dv \, dt.$$

*Here $B[f]$ denotes $A[f]$ or $a[f]\mathbb{I}$ depending on whether we are dealing with (1-4) or (1-7).*

*Proof. Step 1.* Let us assume first that $f_{\text{in}}$ satisfies the additional assumptions (4-1); this assumption will be removed in the final step. For each $n \in \mathbb{N}$, let $f_n := f^{\delta_n}$ and $T_n := T_*^{\delta_n}$ correspond to $f^\delta$ with $\delta = 10^{-n}$, as constructed in Theorem 4.12. Then each $f_n$ is a spherically symmetric, monotone solution to

$$\partial_t f_n = \frac{1}{10^n} \Delta f_n + Q(f_n, f_n) \quad \text{in } \mathbb{R}^3 \times [0, T_n), \qquad f_n(v,0) = f_{\text{in}}(v).$$

Moreover, for each $n$, we have that either $T_n = \infty$ or else $\|f_n(\cdot,t)\|_\infty \to \infty$ as $t \to T_n$.

We define $T_*$ by

$$T_* := \inf\{T \mid \liminf_n M(f_{\text{in}}, T, 10^{-n}) = \infty\}, \tag{4-22}$$

with the understanding that $T_* = \infty$ if the set above is empty. As before, it is not difficult to see that $T_* \ge (2f_{\text{in}}(0))^{-1}$. See Remark 4.15 for further discussion about the definition of $T_*$.

*Step 2.* Let us show, then, that there exists a solution in $\mathbb{R}^3 \times (0, T_*)$. Let $T_j$ be a strictly increasing sequence of times, with $\lim T_j = T_*$. Fix $j$, then since $T_j < T_*$ there is a subsequence $\{n_{j,k}\}$, $n_{j,k} \to \infty$ as $k \to \infty$, such that

$$\sup_k M(f_{\text{in}}, T, 10^{-n_{j,k}}) < \infty.$$

The above combined with Proposition 4.13 implies there is a constant $c = c(f_{\text{in}}, T_j)$ such that for all $k \in \mathbb{N}$ we have

$$A[f_{n_{j,k}}](v,t) \geq \frac{c(f_{\text{in}}, T_j)}{1+|v|^3}\mathbb{I}, \quad \forall (v,t) \in \mathbb{R}^3 \times (0, T_j).$$

The interior Hölder estimate (Theorem 2.2) then says that for any cylinder $Q \Subset \mathbb{R}^3 \times (0,T)$ we have

$$[f_{n_{j,k}}]_{C^{\alpha,\alpha/2}(Q)} \leq C(Q, T_j), \quad \forall k.$$

From here, the same argument as in Lemma 4.7 shows that $A[f_{n_{j,k}}]$ and $\nabla a[f_{n_{j,k}}]$ are $C^{\alpha,\alpha/2}$ uniformly in $k$ in compact subsets of $\mathbb{R}^3 \times (0, T_j)$. Accordingly, the uniform regularity of these coefficients together with the Schauder estimates (Theorem 2.4) guarantee that for every cylinder $Q \Subset \mathbb{R}^3 \times (0, T_j)$ we have a constant $C(Q, T_j)$ independent of $k$ such that

$$[f_{n_{j,k}}]_{C^{2+\alpha,1+\alpha/2}(Q)} \leq C(Q, T_j).$$

Then, the Arzelà–Ascoli theorem and a Cantor diagonalization argument yield local uniform convergence of $f_n$ to a function $f$ in $\mathbb{R}^3 \times (0,T)$ which will be differentiable in time and second-order differentiable in space. In particular, $\tilde{f}_j$ is a spherically symmetric, monotone solution to

$$\partial_t \tilde{f}_j = Q(\tilde{f}_j, \tilde{f}_j) \quad \text{in } \mathbb{R}^3 \times (0, T_j), \qquad \tilde{f}_j(\cdot, 0) = f_{\text{in}},$$

with $f_{\text{in}}$ as in (1-8). We can take this argument one step further and apply the Arzelà–Ascoli theorem one more time to the sequence $\{\tilde{f}_j\}_j$ and conclude that along a subsequence they (along with their derivatives) converge uniformly in compact subsets of $\mathbb{R}^3 \times (0, T_*)$ to a function

$$f : \mathbb{R}^3 \times (0, T_*) \to \mathbb{R}$$

which is again a solution. In summary, we have constructed a function $f : \mathbb{R}^3 \times (0, T_*)$ which is differentiable in time and second-order differentiable in space, such that

$$\partial_t f = Q(f, f)$$

and

$$\int_{\mathbb{R}^3} f(v,t)\phi(v)\,dv - \int_{\mathbb{R}^3} f_{\text{in}}(v)\phi(v)\,dv = -\int_0^t \int (B[f]\nabla f - f\nabla a[f], \nabla\phi)\,dv\,dt, \quad \forall \phi \in C_c^\infty(\mathbb{R}^3),\ t \in (0, T_*). \tag{4-23}$$

Moreover, the function $f$ has the property that for every $T < T_*$, there is a sequence $n_k \to \infty$ such that the functions $f_{n_k}$ defined in Step 1 converge to $f$ locally uniformly in $\mathbb{R}^3 \times [0, T]$.

*Step 3.* It remains to show that if $T_* < \infty$, then the solution built in Step 2 blows up in $L^\infty$ as time approaches $T_*$. We argue by contradiction, similarly to the proof of Theorem 4.12, but with slight modifications accounting for the fact that we do not know whether the functions $f_n$ have a unique limit as $n \to \infty$ (see Remark 4.15 for further discussion). Suppose $C > 0$ is a constant such that

$$\lim_{T \to T_*} \|f\|_{L^\infty(\mathbb{R}^3 \times [0,T])} < C.$$

Let $\varepsilon > 0$ be a small number (to be determined). According to Step 2, there is a sequence $n_k \to \infty$ such that $f_{n_k} \to f$ locally uniformly in $\mathbb{R}^3 \times [0, T_* - \varepsilon/2]$. In particular, there must be some $k_0 > 0$ such that

$$\|f_{n_k}\|_{L^\infty(B_1 \times [0, T_* - \varepsilon])} < 2C, \quad \forall k > k_0.$$

As in the proof of Theorem 4.14, choosing $\varepsilon$ such that $2\varepsilon(2C) < \frac{1}{2}$, the differential inequality argument guarantees that

$$\|f_{n_k}\|_{L^\infty(\mathbb{R}^3 \times [0, T_* + \varepsilon])} \leq 4C, \quad \forall k > k_0.$$

This shows there is a positive $\varepsilon > 0$ such that

$$\liminf_n M(f_{\text{in}}, T, 10^{-n}) < \infty, \quad \forall T < T_* + \varepsilon.$$

This is impossible, since $T_*$ is the infimum of $\{T \mid \liminf_n M(f_{\text{in}}, T, 10^{-n}) = \infty\}$. This contradiction shows that

$$\lim_{T \to T_*} \|f\|_{L^\infty(\mathbb{R}^3 \times [0,T])} = \infty,$$

and the theorem is proved at least for $f_{\text{in}}$, for which (4-1) holds.

*Step 4.* In order to remove (4-1), given $f_{\text{in}}$ for which only (1-8) holds, let $f_{\text{in}}^{(n)}$ be a sequence of functions such that (4-1) holds for each $f_{\text{in}}^{(n)}$ (with a constant $c$ that may depend on $n$) and such that

$$\lim_n \|f_{\text{in}} - f_{\text{in}}^n\|_{L^\infty} = \lim_n \|f_{\text{in}} - f_{\text{in}}^n\|_{L^1} = 0.$$

Let $f^{(n)}$ be a corresponding sequence of solutions as constructed in Steps 1–4 above. Then each $f^{(n)}$ is defined up to some time $T_{*,n}$. The times $T_{*,n}$ are bounded uniformly away from 0 since $f_{\text{in}} \in L^\infty$. The functions $f^{(n)}$ enjoy uniform local a priori estimates, therefore the same compactness argument from Step 2 allows us to pick a subsequence $n_k \to \infty$ and a time $T_*$ such that the functions $f^{(n_k)}$ and their derivatives have a local uniform limit as $k \to \infty$ to a function $f : \mathbb{R}^3 \times (0, T_*) \to \mathbb{R}$ which is a smooth solution to the nonlinear equation and which blows up in $L^\infty$ as time approaches $T_*$. Finally, fixing a test function $\phi$ and $t \in (0, T_*)$, we may apply (4-23) to each $f^{(n_k)}$ and conclude that $f$ satisfies the respective relation in the limit, proving the theorem. □

**Remark 4.15.** It is worth comparing the definition of $T_*^\delta$ in Theorem 4.12 with that of $T_*$ in Theorem 4.14. In the present situation, a priori it is unclear whether the sequence $f_n$ has a unique limit as $n \to \infty$. Hence, if we define

$$T_* := \sup\{T \mid \sup_n M(f_{\text{in}}, T, 10^{-n}) < \infty\},$$

the existence of a subsequence bounded for times strictly greater then $T^*$ does not contradict the definition of $T^*$. However, the contradiction holds if $T^*$ is defined via the lim inf as in (4-22). In the proof of the

former theorem, matters were simplified by the fact that $\{f_k^\delta\}_k$ was a Cauchy sequence (for $\delta$ fixed), meaning in particular that if it is shown that a subsequence of $f_k^\delta$ remains bounded in $[0, T]$, then the entire sequence remains bounded. This was key in proving the maximality of the interval of existence $(0, T_*^\delta)$.

## 5. Pointwise bounds and proof of Theorem 1.1

***Conditional pointwise bound.*** The first lemma of this section (Lemma 5.2) is the key argument for the proofs of Theorem 1.1 and Theorem 1.3. It consists of a barrier argument based on the observation that the function $U(v) = |v|^{-\gamma}$ is a supersolution for the elliptic operator $Q(f, \cdot)$ under certain assumptions on $f$ (this is where the radial symmetry and monotonicity is needed). It affords control of certain spatial $L^p$ norms of the solution, and from these higher regularity will follow by standard elliptic estimates (Lemma 5.5).

First, we prove an elementary proposition that will be of use in proving the key lemma.

**Proposition 5.1.** *If $h$ is a nonnegative, radially symmetric and decreasing function, then*

$$\frac{h(v)}{a[h](v)} \le 8 \sup_{r \le |v|} \left\{ r^2 \frac{\int_{B_r} h(w)\,dw}{\int_{B_r} a[h](w)\,dw} \right\} |v|^{-2}, \quad \forall v \in \mathbb{R}^3.$$

*Proof.* First of all, since $h$ is radially symmetric and decreasing,

$$\frac{1}{|B_{|v|}(0)|} \int_{B_{|v|}(0)} h(w)\,dw \ge h(v).$$

On the other hand, since $h \ge 0$ and (in particular) $a[h]$ is superharmonic,

$$a[h](v) \ge \frac{1}{|B_{2|v|}(v)|} \int_{B_{2|v|}(v)} a[h](w)\,dw = \frac{2^{-3}}{|B_{|v|}(0)|} \int_{B_{|v|}} a[h](w)\,dw, \quad \forall v \in \mathbb{R}^3.$$

Therefore,

$$\frac{h(v)}{a[h](v)} \le 8 \frac{\int_{B_{|v|}} h(w)\,dw}{\int_{B_{|v|}(0)} a[h](w)\,dw},$$

which implies that

$$\frac{h(v)}{a[h](v)} \le 8|v|^{-2} \sup_{r \le |v|} \left\{ r^2 \frac{\int_{B_r} h(w)\,dw}{\int_{B_r} a[h](w)\,dw} \right\}. \qquad \square$$

**Lemma 5.2.** *Suppose $f : \mathbb{R}^3 \times [0, T] \to \mathbb{R}_+$ is a classical solution of (2-1). Let $\gamma \in (0, 1)$ and suppose there exists some $R_0 > 0$ such that*

$$r^2 \frac{\int_{B_r} g(w, t)\,dw}{\int_{B_r} a[g](w, t)\,dw} \le \frac{1}{24}\gamma(1-\gamma), \quad \forall r \le R_0,\ t \le T. \tag{5-1}$$

*Then*

$$f(v, t) \le \max\left\{ \frac{3}{4\pi} R_0^{\gamma-3}, \left(\frac{3}{4\pi}\right)^{\gamma/3} \|f_{\text{in}}\|_{L^{3/\gamma}_{\text{weak}}} \right\} |v|^{-\gamma}, \quad \text{in } B_{R_0} \times [0, T].$$

*In particular, the conclusion of the lemma holds for some $R_0 > 0$ whenever there is a modulus of continuity $\omega(r)$ and some $R_1 > 0$ such that*

$$\sup_{r<|v|}\sup_{t\in[0,T]}\left\{r^2\frac{\int_{B_r} g(w,t)\,dw}{\int_{B_r} a[g](w,t)\,dw}\right\} \le \omega(|v|), \quad \forall 0<|v|\le R_1. \tag{5-2}$$

**Remark 5.3.** It is easy to see that for any radially decreasing function $h(v)$, the condition that $h$ belongs to $L^p_{\text{weak}}(\mathbb{R}^3)$ implies that $h$ lies below a power function of the form $1/|v|^{3/p}$, and vice versa. More precisely,

$$\|h(v)\|_{L^p_{\text{weak}}} \le C \Leftrightarrow h(v) \le C\Big(\frac{3}{4\pi}\Big)^{1/p}|v|^{-3/p}. \tag{5-3}$$

*Proof of Lemma 5.2.* Let $U_\gamma = |v|^{-\gamma}$. Then Proposition 3.4 says that

$$Q(g, U_\gamma) \le U_\gamma a[g]\Big(-\frac{1}{3}\gamma(1-\gamma)|v|^{-2} + \frac{g}{a[g]}\Big).$$

Applying Proposition 5.1 with $h = g(\cdot, t)$,

$$\frac{g}{a[g]}(v,t) \le 8|v|^{-2}\sup_{r\le|v|}\left\{r^2\frac{\int_{B_r} g(w,t)\,dw}{\int_{B_r} a[g](w,t)\,dw}\right\} \le \frac{1}{3}\gamma(1-\gamma)|v|^{-2},$$

where we used (5-1) to get the last inequality. It follows that

$$Q(g, U_\gamma) \le 0, \quad \text{in } B_{R_0}\times[0,T]. \tag{5-4}$$

In particular, if there is a modulus of continuity as in (5-2), then $Q(g, U_\gamma) \le 0$ in $B_{R_0}\times[0,T]$ provided $R_0$ is chosen so that $\omega(R_0) \le \frac{1}{24}$.

On the other hand, given that $f(v,t)$ is radially decreasing and lies in $L^1$ (see (5-3)),

$$f(v,t) \le \frac{3}{4\pi|v|^3}\|f\|_{L^1(\mathbb{R}^3)} = \frac{3}{4\pi|v|^3}, \quad \forall v\in\mathbb{R}^3,\ t\in[0,T], \tag{5-5}$$

where we used that $\|f(\cdot,t)\|_{L^1(\mathbb{R}^3)} = 1$ for all $t$. Finally, the function $\widetilde{U}_\gamma(v)$ defined by

$$\widetilde{U}_\gamma(v) := \max\Big\{\frac{3}{4\pi}R_0^{\gamma-3}, \Big(\frac{3}{4\pi}\Big)^{\gamma/3}\|f_{\text{in}}\|_{L^{3/\gamma}_{\text{weak}}}\Big\}|v|^{-\gamma}$$

is a supersolution for the equation solved by $f$ in $B_{r_0}\times[0,T]$. Moreover, clearly $\widetilde{U}_\gamma$ lies above $f_{\text{in}}$ in $B_{R_0}$, while by (5-5), $\widetilde{U}_\gamma$ lies above $f$ in $\partial B_{R_0}\times[0,T]$. Then the comparison principle implies that $f \le \widetilde{U}_\gamma$ in $B_{r_0}\times[0,T]$, and the lemma is proved. □

The next lemma deals specifically with solutions to the nonlinear equations (1-4) or (1-7). It controls from below the integral of a solution in some ball $B_R$. For the case of the Landau equation (1-4), the constant is independent of time (by conservation of mass and second moment), while for the Krieger–Strain equation (1-7) the bound decays exponentially in time.

**Lemma 5.4.** *For $f$ solving (1-4), there is a constant $R > 0$ such that*

$$\int_{B_R} f(v,t)\,dv \ge \frac{1}{2}, \quad t>0. \tag{5-6}$$

*For $f$ solving (1-7) and any radii $R > r > 0$, there are $\beta > 0$ and $C_0 > 0$ such that*

$$\int_{B_R \setminus B_r} f(v,t)\,dv \geq C_0 e^{-\beta t} \int_{B_{4R}\setminus B_{r/4}} f_{\text{in}}(v)\,dv, \quad t > 0. \tag{5-7}$$

*Proof.* If $f$ solves (1-4), then

$$\int_{B_R(0)^c} f(v,t)\,dv \leq R^{-2}\int_{B_R(0)^c} f(v,t)|v|^2\,dv \leq 3R^{-2}.$$

Thus

$$\int_{B_R(0)} f(v,t)\,dv = 1 - \int_{B_R(0)^c} f(v,t)\,dv \geq 1 - 3R^{-2}.$$

Estimate (5-6) follows by choosing $R$ large enough. The corresponding estimate (5-7) for $f$ solving (1-7) follows a similar argument used in [Krieger and Strain 2012], and the derivation of the estimate is done in detail in the Appendix. □

The next lemma says that any solution $f$ to (1-4) or (1-7) is a bounded function for all times, provided that $f$ satisfies (5-2).

**Lemma 5.5.** *Let $f : \mathbb{R}^3 \times [0,T] \to \mathbb{R}$ be a radially symmetric, radially decreasing solution to (1-4) (or (1-7)) with initial data as in (1-8) and such that for some $R_0 > 0$, we have*

$$r^2 \frac{\int_{B_r} f(w,t)\,dw}{\int_{B_r} a[f](w,t)\,dw} \leq \frac{1}{24}\gamma(1-\gamma), \quad \forall r \leq R_0,\ t \leq T.$$

*Or, assume that there is some modulus of continuity $\omega(r)$ such that*

$$\sup_{r<|v|}\sup_{t\in[0,T]} \left\{ r^2 \frac{\int_{B_r} f(w,t)\,dw}{\int_{B_r} a[f](w,t)\,dw} \right\} \leq \omega(|v|), \quad \forall 0 < r \leq R_0. \tag{5-8}$$

*Then*

$$\sup_{t\in[T/2,T]} \|f(\cdot,t)\|_{L^\infty(\mathbb{R}^3)} \leq C_0 \tag{5-9}$$

*for some constant $C_0$ depending only on $f_{\text{in}}$, $T$ and $R_0$.*

*Proof.* The assumptions of the lemma are simply the same as those of Lemma 5.2 with $g(v,t) = f(v,t)$, from which it follows, using also (5-3), that

$$\begin{aligned}\sup_{t\in[0,T]} \|f(\cdot,t)\|_{L^p_{\text{weak}}(B_{R_0})} &\leq \max\left\{ \frac{3}{4\pi} R_0^{-3(1-1/p)}, \left(\frac{3}{4\pi}\right)^{1/p} \|f_{\text{in}}\|_{L^p_{\text{weak}}} \right\} \big\| |v|^{-3/p} \big\|_{L^p_{\text{weak}}} \\ &=: C_0(f_{\text{in}}, R_0, p)\end{aligned}$$

for some $p > 6$. By interpolation and the Sobolev embedding, it follows that $\|f(\cdot,t)\|_{L^6(\mathbb{R}^3)}$ and $\|\nabla a[f(\cdot,t)]\|_{L^\infty(\mathbb{R}^3)}$ are bounded by some constant $C$ determined by $C_0(f_{\text{in}}, R_0, p)$. Then, applying (2-5) from Theorem 2.3 with $Q = B_{R_0} \times [0,T]$, we arrive at

$$\|f\|_{L^\infty(B_{R_0/2}\times[T/2,T])} \leq C\left\{ \|f\|_{L^2(Q)} + R_0^2 \|\nabla a[f]\|_{L^\infty(Q)} \right\} < \infty$$

for some $C = C(f_{\text{in}}, R_0, T)$, and the lemma is proved. □

*Proof of Theorem 1.2.* According to Theorem 4.12, for $f_{\text{in}} \in L^\infty$, there exists a time $T_0 > 0$ and a solution $f(v,t)$ to (1-4) defined in $\mathbb{R}^3 \times [0, T_0)$ and with initial values $f_{\text{in}}$.

The time $T_0$ is maximal, in the sense that $T_0 = \infty$ or else

$$\lim_{t \to T_0^-} \|f(\cdot, t)\|_{L^\infty(\mathbb{R}^3)} = \infty. \tag{5-10}$$

Moreover, since $f \in L^\infty$ in $\mathbb{R}^3 \times [0, t]$ for every $t < T_0$, interior regularity estimates (see Theorem 2.2 and Theorem 2.4) show that $f$ must be twice differentiable in $v$ and differentiable in $t$ as long as $t \in (0, T)$.

Finally, arguing by contradiction, let us assume that

$$\limsup_{r \to 0^+} \sup_{t \in (0,T_0)} \left\{ r^2 \frac{\int_{B_r} f(v,t)\, dv}{\int_{B_r} a[f](v,t)\, dv} \right\} < \frac{1}{96}.$$

In this case, there must be some $R_0 > 0$ such that

$$\sup_{t \in (0,T_0)} \left\{ r^2 \frac{\int_{B_r} f(v,t)\, dv}{\int_{B_r} a[f](v,t)\, dv} \right\} \leq \frac{1}{96}, \quad \forall r \leq R_0.$$

This means Lemma 5.5 can be applied with $T = T_0$, and it follows that

$$\sup_{t \in [T_0/2, T_0]} \|f(\cdot, t)\|_{L^\infty(\mathbb{R}^3)} < \infty,$$

which contradicts (5-10), and the theorem is proved. □

*Proof of Theorem 1.1.* As in the proof of Theorem 1.2, we have a solution $f(v,t)$ defined up to some maximal time $T_0$. In case $T_0 < \infty$, we know that $\|f(\cdot, t)\|_{L^\infty}$ goes to infinity as $t \to T_0^-$. As before, this $f(v,t)$ is twice differentiable in $v$ and differentiable in $t$ for $t \in (0, T)$.

Now assume the $L^{3/2}$ norm of $f(\cdot, t)$ does not concentrate at 0 as $t \to T^-$. That is, suppose there is a modulus of continuity $\omega(\cdot)$ such that

$$\sup_{t \in (0,T_0)} \|f(\cdot, t)\|_{L^{3/2}(B_r)} \leq \omega(r).$$

Then there is some $C > 0$ such that

$$r^2 \frac{\int_{B_r} f(v,t)\, dv}{\int_{B_r} a[f](v,t)\, dv} = \frac{\frac{4\pi}{3r} \int_{B_r} f(v,t)\, dv}{\frac{1}{|B_r|} \int_{B_r} a[f](v,t)\, dv} \leq C \frac{1}{r} \int_{B_r} f(v,t)\, dv, \quad \forall r > 0,\ t \in (0, T_0).$$

Then Hölder's inequality says that

$$r^2 \frac{\int_{B_r} f(v,t)\, dv}{\int_{B_r} a[f](v,t)\, dv} \leq C' \|f(\cdot, t)\|_{L^{3/2}(B_r)} \leq C' \omega(r).$$

It follows that if $R_0 > 0$ is chosen so that $C'\omega(R_0) < \frac{1}{96}$, then Lemma 5.5 can be applied to conclude again that

$$\sup_{t \in [T_0/2, T_0]} \|f(\cdot, t)\|_{L^\infty(\mathbb{R}^3)} < \infty,$$

which as before directly contradicts $\lim_{t \to T_0^-} \|f(\cdot, t)\|_{L^\infty} = \infty$, and the theorem is proved. □

To end this section, we present a computation indicating that for an arbitrary function $f$ the quotient appearing in the assumption of Theorem 1.2 is always smaller than or equal to 3.

**Proposition 5.6.** *Let $h \in L^1(\mathbb{R}^3)$ be a nonnegative function. Then*

$$r^2 \frac{\int_{B_r} h(v)\,dv}{\int_{B_r} a[h](v)\,dv} \le 3, \quad \forall r > 0.$$

**Remark 5.7.** It could be of use in understanding the blow-up or (non-blow-up) of (1-4) to characterize those $h$ for which the above quotient goes to 0 as $r$ approaches 0. In particular, it would be useful to understand this when $h$ is not necessarily in a regular enough $L^p$ space or Morrey space, namely when $h$ is such that

$$h \notin L^{3/2}_{\mathrm{loc}} \quad \text{or} \quad \sup_{r>0} \frac{1}{r} \int_{B_r} h\,dv = \infty.$$

*Proof of Proposition 5.6.* Let us write $a(v) = a[h](v)$ for the sake of brevity. Note that

$$\int_{B_r} a(v)\,dx = \int_{\mathbb{R}^3} a(v)\chi_B(v)\,dv = \frac{1}{4\pi} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} h(w)|v-w|^{-1}\chi_{B_r}(v)\,dw\,dv.$$

The goal is to compare the two integrals

$$\frac{1}{4\pi} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} h(w)|v-w|^{-1}\chi_{B_r}(v)\,dw\,dv \quad \text{and} \quad r^2 \int_{\mathbb{R}^3} h(v)\chi_{B_r}(v)\,dv.$$

Note that

$$\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} h(w)|v-w|^{-1}\chi_{B_r}(v)\,dw\,dv = \int_{\mathbb{R}^3} h(v)(\chi_{B_r} * \Phi)(v)\,dv, \quad \Phi(v) = (4\pi|v|)^{-1}.$$

It is not hard to compute $\Phi_B := \chi_{B_r} * \Phi$ directly. Indeed, it is the unique $C^{1,1}$ solution of

$$\Delta \Phi_{B_r} = -\chi_{B_r}, \quad \Phi_{B_r} \to 0 \text{ at } \infty,$$

which has the simple expression

$$\Phi_{B_r}(x) = \begin{cases} -\frac{1}{6}|v|^2 + \frac{1}{2}r^2 & \text{in } B_r, \\ \frac{1}{3}r^3|v|^{-1} & \text{in } B_r^c. \end{cases}$$

It follows that

$$\int_{B_r} a(v)\,dv = \int_{B_r} \left(\tfrac{1}{2}r^2 - \tfrac{1}{6}|v|^2\right)h(v)\,dv + \frac{r^3}{3}\int_{B_r^c} h(v)|v|^{-1}\,dv \ge \int_{B_r} \left(\tfrac{1}{2}r^2 - \tfrac{1}{6}|v|^2\right)h(v)\,dv.$$

This proves the stated bound, since the last inequality guarantees that

$$\int_{B_r} a(v)\,dv \ge \frac{r^2}{3}\int_{B_r} h(v)\,dv.$$

□

## 6. Mass comparison and proof of Theorem 1.3

In this section we apply the ideas from previous sections to construct global solutions (in the radial, monotone case) for (1-7), namely

$$\partial_t f = a[f]\Delta f + f^2.$$

In view of Lemma 5.5, the fact that $T_0 = \infty$ in Theorem 1.1 results from a bound of any $L^p(\mathbb{R}^3)$ norm of $f$ with $p > \frac{3}{2}$. For (1-7), the bound of any $L^p(\mathbb{R}^3)$ norm of $f$ with $p > \frac{3}{2}$ will be proven by a barrier argument done at the level of the *mass function* of $f(v,t)$, which is defined by

$$M_f(r,t) = \int_{B_r} f(v,t)\,dv, \quad (r,t) \in \mathbb{R}_+ \times (0, T_0).$$

Depending on which problem $f$ solves, the associated function $M_f(r,t)$ solves a one-dimensional parabolic equation with diffusivity given by $A^*[f]$ or $a[f]$.

**Proposition 6.1.** *Let $f$ be a solution of (1-4) or (1-7) in $\mathbb{R}^3 \times [0, T_0]$. Then $M(r,t)$ solves, respectively,*

$$\partial_t M_f = A^* \partial_{rr} M_f + \frac{2}{r}\left(\frac{M_f}{8\pi r} - A^*\right)\partial_r M_f \quad \text{in } \mathbb{R}_+ \times (0, T_0) \tag{6-1}$$

$$\partial_t M_f = a \partial_{rr} M_f + \frac{2}{r}\left(\frac{M_f}{8\pi r} - a\right)\partial_r M_f \qquad \text{in } \mathbb{R}_+ \times (0, T_0). \tag{6-2}$$

*Proof.* We briefly show how to obtain (6-2); for (6-1), the calculations are identical. Using the divergence theorem and the divergence expression in (1-7), we get

$$\partial_t M_f = \int_{\partial B_r} (a[f]\nabla f - f\nabla a[f], n)\,d\sigma = 4\pi r^2 (a[f]\partial_r f - f\,\partial_r a[f]).$$

Furthermore, straightforward differentiation yields the formulas

$$4\pi r^2 \partial_r f = r^2 \partial_r (r^{-2}\partial_r M_f), \quad \partial_r a[f] = -(4\pi r^2)^{-1} M_f.$$

Substituting these in the expression for $\partial_t M_f$ above, we get

$$\partial_t M_f = a[f] r^2 \partial_r \left(\frac{1}{r^2}\partial_r M_f\right) + \frac{1}{4\pi r^2} M_f \partial_r M_f.$$

Expansion and rearrangement of the terms result in

$$\partial_t M_f = a\left(-\frac{2}{r}\partial_r M_f + \partial_{rr} M_f\right) + \frac{M_f}{4\pi r^2}\partial_r M_f = a\partial_{rr} M_f + \frac{2}{r}\left(\frac{M_f}{8\pi r} - a\right)\partial_r M_f,$$

and the conclusion follows. □

Define the linear parabolic operator $L$ in $\mathbb{R}_+ \times (0, T)$ as

$$Lh := \partial_t h - a\partial_{rr} h - \frac{2}{r}\left(\frac{M_f}{8\pi r} - a[f]\right)\partial_r h.$$

The above proposition simply says that $LM_f = 0$ in $\mathbb{R}_+ \times (0, T)$. The next proposition identifies suitable supersolutions for $L$.

**Proposition 6.2.** *If $m \in [0,2]$ and $h(r,t) = r^m$, then $Lh \geq 0$ in $\mathbb{R}_+ \times (0,T)$.*

*Proof.* By direct computation we see that

$$Lh = -mr^{m-2}\left((m-1)a + 2\left(\frac{M_f}{8\pi r} - a[f]\right)\right).$$

On the other hand,

$$a[f](r) = \frac{1}{4\pi r}\int_{B_r} f\,dv + \int_{B_r^c} \frac{f}{4\pi|v|}\,dv \geq \frac{M_f}{4\pi r},$$

which guarantees that $\frac{1}{2}a[f](r) \geq \frac{M_f}{8\pi r}$. Thus,

$$Lh = mr^{m-2}\left((1-m)a[f] + 2\left(a[f] - \frac{M_f}{8\pi r}\right)\right) \geq mr^{m-2}(2-m)a[f] \geq 0,$$

the last inequality being true for $m \leq 2$. □

*Proof of Theorem 1.3.* Assume $f_{\text{in}} \in L^\infty$, in which case Theorem 4.12 yields a solution $f(v,t)$ that exists for some time $T_0 > 0$ (possibly infinite). As the bound for $f(v,t)$ will not rely on the $L^\infty$ norm of $f_{\text{in}}$ but an $L^p_{\text{weak}}$ norm of $f_{\text{in}}$, the existence of a solution for unbounded initial data in $L^p$ ($p > 6$) will follow by a standard density argument.

Since $p > 6$, there is some $\alpha > 0$ and some $C_0 > 0$ (depending only on $\|f\|_{L^p_{\text{weak}}}$) such that

$$M_{f_{\text{in}}}(r,0) = \int_{B_r} f_{\text{in}}\,dv \leq C_0 r^{1+\alpha}.$$

Moreover, since $f(\cdot,t)$ has total mass 1 for every $t > 0$, we also have

$$M_f(r,t) \leq 1, \quad \forall r > 0,\ t \in (0,T).$$

Proposition 6.2 says that $h = Cr^{1+\alpha}$ is a supersolution of the parabolic equation solved by $M_f$ in $\mathbb{R}_+ \times (0,T)$. Then, choosing $C := \max\{C_0, 1\}$, the comparison principle yields

$$M_f(r,t) \leq h(r) = Cr^{1+\alpha}, \quad \forall r \in (0,1),\ t \in (0,T). \tag{6-3}$$

Since $f(v,t)$ is radially symmetric and decreasing, bound (6-3) implies that $f(|v|,t) \leq \frac{3C}{4\pi}\frac{1}{|v|^{2-\alpha}}$ for $v \in B_1$ and $t \in (0,T)$; hence there is some $p' > \frac{3}{2}$ and some $C_{p'} > 0$ such that

$$\|f(\cdot,t)\|_{L^{p'}(B_1)} \leq C_{p'}, \quad \forall t \in (0,T).$$

Then Lemma 5.2 says that $f(v,t)$ is bounded in $\mathbb{R}^3 \times (0,T_0)$. By Lemma 5.5 and the characterization of $T_0$ in Theorem 4.12, it follows that $T_0 = +\infty$, so the solution is global in time. □

The method of the proof for Theorem 1.3 falls short in preventing finite time blow-up for (1-4). In any case, it at least gives another criterion for blow-up, the proof of which is essentially the same as that of Theorem 1.3.

**Corollary 6.3.** *Suppose that for all $t \in [0, T_0]$ there is some $r_0 > 0$ and $0 < \lambda < 8\pi$ such that*

$$M_f(r,t) \le \lambda r A^*(r,t), \quad \forall r < r_0.$$

*Then any solution to (1-4) is bounded for any $t > 0$.*

## Appendix

*Proof of Proposition 3.1.* The radial symmetry of any solution $f$ to (2-1) follows by the uniqueness property of (2-1) and by the fact that $Q(g, f)$ commutes with rotations, as shown below. We first rewrite the collision operator as

$$Q(g,f) = \operatorname{div}(A[g]\nabla f - f\nabla a[g]) = a[g]\Delta f - \operatorname{div}(\tilde{A}[g]\nabla f) + fg,$$

with

$$\tilde{A}[g]\nabla f := \int \frac{g(|v-y|)}{|y|^3}\langle \nabla f(v), y\rangle y\, dy.$$

Let $\mathbb{T}$ be a rotation operator. Since $g$ is radially symmetric, so is $a[g]$. Hence

$$a[g]\Delta(f\circ\mathbb{T}) = a[g\circ\mathbb{T}]\Delta(f\circ\mathbb{T}) = (a[g]\circ\mathbb{T})(\Delta f\circ\mathbb{T}) = (a[g]\Delta f)\circ\mathbb{T},$$

taking into account that the Laplacian operator commutes with rotations. Moreover,

$$\begin{aligned}
\operatorname{div}(\tilde{A}[g]\nabla f(\mathbb{T}v)) &= \operatorname{div}\left(\int \frac{g(|v-y|)}{|y|^3}\langle \nabla f(\mathbb{T}v), y\rangle y\, dy\right)\\
&= \operatorname{div}\left(\int \frac{g(|v-y|)}{|y|^3}\langle \mathbb{T}^*\nabla_z f(z)|_{z=\mathbb{T}v}, y\rangle y\, dy\right)\\
&= \operatorname{div}\left(\int \frac{g(|\mathbb{T}(v-y)|)}{|y|^3}\langle \nabla_z f(z)|_{z=\mathbb{T}v}, \mathbb{T}y\rangle \mathbb{T}^*\mathbb{T}y\, dy\right)\\
&= \operatorname{div}\Bigg(\mathbb{T}^*\underbrace{\int \frac{g(|\mathbb{T}v-y)|)}{|y|^3}\langle \nabla_z f(z)|_{z=\mathbb{T}v}, y\rangle y\, dy}_{=:V(\mathbb{T}v)}\Bigg)\\
&= \operatorname{Tr}(\mathbb{T}^*\operatorname{Jac}(V)|_{z=\mathbb{T}v}\mathbb{T}) + \underbrace{\nabla(\operatorname{Tr}(\mathbb{T}^*))}_{=0}\cdot V(\mathbb{T}v)\\
&= \operatorname{Tr}(\mathbb{T}\mathbb{T}^*\operatorname{Jac}(V)|_{z=\mathbb{T}v})\\
&= \operatorname{Tr}(\mathbb{I}\operatorname{Jac}(V)_{|_{z=\mathbb{T}v}})\\
&= \operatorname{div}\left(\int \frac{g(|z-y|)}{|y|^3}\langle \nabla_z f(z), y\rangle y\, dy\right)\circ\mathbb{T}.
\end{aligned}$$

Hence $Q(g, f(\mathbb{T}v)) = Q(g,f)\circ\mathbb{T}$. Now we rewrite the linear equation (2-1) in spherical coordinates:

$$\partial_t f = A^*\partial_{rr} f + \frac{a - A^*}{r}\partial_r f + fg, \tag{6-4}$$

with $A^*[g](v) := (A[g](v)\hat{v}, \hat{v})$, $\hat{v} := v/|v|$ and differentiate (6-4) with respect to $r$. The function $w := \partial_r f$ satisfies the inequality

$$\partial_t w \le A^* \partial_{rr} w + \frac{a - A^*}{r} \partial_r w + wg + \partial_r A^* \partial_r w + \partial_r \Big( \frac{a - A^*}{r} \Big) w.$$

If $w(\cdot, 0) \le 0$ it follows from the maximum principle that $w(\cdot, t) \le 0$ for all $t \ge 0$. In other words, the (negative) sign of $\partial_r f$ is preserved in time. □

*Proof of Proposition 3.2.* The identity (3-3) is classical and a proof can be found in [Lieb and Loss 2001, Section 9.7]. To prove (3-2), let $v \in \mathbb{R}^3$ be nonzero and $r := |v|$. Then

$$(A[g](v)\hat{v}, \hat{v}) = \frac{1}{8\pi} \int_{\mathbb{R}^3} \frac{1}{|v - w|} g(w) \left( \left( \mathbb{I} - \frac{v - w}{|v - w|} \otimes \frac{v - w}{|v - w|} \right) \hat{v}, \hat{v} \right) dw.$$

Note that

$$\left( \left( \mathbb{I} - \frac{v - w}{|v - w|} \otimes \frac{v - w}{|v - w|} \right) \hat{v}, \hat{v} \right) = 1 - \cos(\hat{\theta}(w))^2,$$

where $\hat{\theta}$ denotes the angle between $w - v$ and $v$. Consider, for $0 \le t, r$, the function

$$I(r, t) := \int_{\partial B_t} \frac{1 - \cos(\hat{\theta})^2}{|v - w|} \, dw.$$

The function $I(r, t)$ encodes all the information about $A^*$. In particular, integration in spherical coordinates yields the expression

$$A^*[h](v) = \frac{1}{8\pi} \int_0^\infty f(t) I(|v|, t) \, dt.$$

As it turns out, $I(r, t)$ has rather different behavior according to whether $r < t$ or not. By averaging in the $v$ variable, it is not hard to see that

$$I(r, t) = \frac{t^2}{r^4} I(t, r), \quad \forall r < t.$$

Accordingly, we focus on $I(r, t)$ when $r > t$. To do so, denote by $\theta$ the angle between $w$ and $v$ and observe that

$$1 - \cos(\hat{\theta})^2 = \sin(\hat{\theta})^2 = \frac{t^2 - t^2 \cos(\theta)^2}{|v - w|^2} = \frac{t^2 - w_1^2}{|v - w|^2},$$

where $w_1 = (w, \hat{v})$. Thus,

$$\begin{aligned} I(r, t) &= \int_{\partial B_t} \frac{t^2 - w_1^2}{|v - w|^3} \, dw = \int_{\partial B_t} \frac{t^2 - w_1^2}{(t^2 - w_1^2 + (r - w_1)^2)^{3/2}} \, dw \\ &= \int_{\partial B_t} \frac{t^2 - w_1^2}{(t^2 - 2r w_1 + r^2)^{3/2}} \, dw = \int_{\partial B_1} \frac{t^2 (1 - z_1^2)}{t^3 \big(1 - 2\big(\frac{r}{t}\big) z_1 + \big(\frac{r}{t}\big)^2\big)^{3/2}} t^2 \, dz \\ &= \int_{\partial B_1} \frac{1 - z_1^2}{\big(1 - 2\big(\frac{r}{t}\big) z_1 + \big(\frac{r}{t}\big)^2\big)^{3/2}} t \, dz. \end{aligned}$$

This surface integral can be written entirely as an integral in terms of the variable $z_1 \in (-1, 1)$:

$$I(r,t) = 2\pi t \int_{-1}^{1} \frac{1-z_1^2}{\left(1-2\left(\frac{r}{t}\right)z_1+\left(\frac{r}{t}\right)^2\right)^{3/2}}\,dz_1.$$

For brevity, set for now $s = r/t$. Then

$$\begin{aligned}\int_{-1}^{1} \frac{1-z_1^2}{(1-2sz_1+s^2)^{3/2}}\,dz_1 &= \frac{-2s^4+2s^3+2s-2}{3s^3\sqrt{s^2-2s+1}} - \frac{-2s^4-2s^3-2s-2}{3s^3\sqrt{s^2+2s+1}}\\ &= \frac{-2s^4+2s^3+2s-2}{3s^3(s-1)} - \frac{-2s^4-2s^3-2s-2}{3s^3(s+1)}\\ &= \frac{-2s^4+2s^3+2s-2}{3s^3(s-1)} + \frac{2s^4+2s^3+2s+2}{3s^3(s+1)}.\end{aligned}$$

Furthermore,

$$\begin{aligned}\frac{-2s^4+2s^3+2s-2}{3s^3(s-1)} + \frac{2s^4+2s^3+2s+2}{3s^3(s+1)} &= \frac{2}{3s^3}\left(\frac{-s^4+s^3+s-1}{s-1} + \frac{s^4+s^3+s+1}{s+1}\right)\\ &= \frac{2}{3s^3}\frac{(-s^4+s^3+s-1)(s+1)+(s^4+s^3+s+1)(s-1)}{s^2-1}\\ &= \frac{2}{3s^3}\frac{2s^2-2}{s^2-1} = \frac{4}{3s^3}.\end{aligned}$$

Then, since $s = r/t$, we conclude that

$$\begin{aligned}I(r,t) &= 8\pi\frac{t^4}{3r^3}, \quad \text{for } t < r,\\ I(r,t) &= 8\pi\frac{1}{3t}, \quad \text{for } t > r.\end{aligned}$$

Going back to $A^*[h]$, the above leads to

$$\begin{aligned}A^*[h](v) &= \int_0^r h(t)I(r,t)\,dt + \int_r^\infty h(t)I(r,t)\,dt\\ &= \frac{1}{3r^3}\int_0^r h(t)t^4\,dt + \frac{1}{3}\int_r^\infty h(t)t\,dt.\end{aligned}$$

□

*Proof of Lemma 5.4.* This argument is inspired by the one in [Krieger and Strain 2012, Section 2.6]. For $\beta$, $R$, $r$ (with $0 < r < R$, $0 < \beta$), consider the function

$$\Phi(v,t) := e^{-\beta t}(|v|-R)^2(|v|-r)^2.$$

Since $\Phi$ is a $C^{1,1}$ function with compact support, we have

$$\frac{d}{dt}\int_{\mathbb{R}^3} f(v,t)\Phi(v)\,dv = -\int_{\mathbb{R}^3}(a\nabla f - f\nabla a, \nabla\Phi)\,dv = \int_{\mathbb{R}^3} f\,\mathrm{div}(a\nabla\Phi)\,dv + \int_{\mathbb{R}^3} f(\nabla a, \nabla\Phi)\,dv.$$

Hence,

$$\begin{aligned}\operatorname{div}(a\nabla\Phi)+(\nabla a,\nabla\Phi)&=a\Delta\Phi+2(\nabla a,\nabla\Phi)\\&=a\Phi''+\frac{2}{|v|}(a+|v|a')\Phi'=a\Phi''+\frac{2}{|v|}\Phi'\int_{|v|}^{+\infty}sf(s,t)\,ds.\end{aligned}$$

We have

$$\begin{gathered}\Phi'(s)=2(R-s)(s-r)(-(s-r)+R-s)=2(R-s)(s-r)(R+r-2s),\\ \Phi''(s)=2(R-s)(r+R-2s)-2(s-r)(r+R-2s)-4(R-s)(s-r),\\ \Phi'(r)=\Phi'(R)=0,\qquad \Phi''(r)=\Phi''(R)=2(R-r)^2,\\ |\Phi''|,\ |\Phi'|\le C_{\delta,r,R}\Phi,\qquad |v|\in((1+\delta)r,(1-\delta)R).\end{gathered}$$

Hence in a small neighborhood of $|v|=R$ and $|v|=r$ one can show that $\dfrac{d}{dt}\displaystyle\int_{\mathbb{R}^3}f(v,t)\Phi(v)\,dv\ge 0$; more precisely,

$$\operatorname{div}(a\nabla\Phi)+(\nabla a,\nabla\Phi)\ge 0\quad\text{in } B_R\setminus B_{(1-\delta)R}\cup B_{(1+\delta)r}\setminus B_r.$$

Since $a[g](v)\le\dfrac{\|g\|_{L^1(\mathbb{R}^3)}}{|v|}$, it follows that

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{R}^3}f(v,t)\Phi(v)\,dv&\ge -C_{\delta,r,R}\frac{\|g\|_{L^1(\mathbb{R}^3}}{r}\int_{B_{(1-\delta)R}\setminus B_{(1+\delta)r}}f(v,t)\Phi(v)\,dv\\&\ge-\frac{\|g\|_{L^1(\mathbb{R}^3)}}{r}C_{\delta,r,R}\int_{\mathbb{R}^3}f(v,t)\Phi(v)\,dv.\end{aligned}$$

This above differential inequality implies

$$\int_{\mathbb{R}^3}f(v,t)\Phi(v)\,dv\ge e^{-\beta T}\int_{\mathbb{R}^3}f_{\text{in}}\Phi(v)\,dv,\quad\forall t<T,$$

where $\beta=C_{r,R,\alpha}\|g\|_{L^1}$. Finally, since

$$\Phi(v)\le\tfrac14(R-r)^2\quad\text{in } B_R\setminus B_r,\qquad\Phi(v)\ge\tfrac14R^2r^2,$$

we conclude that

$$\int_{B_R\setminus B_r}f(v,t)\,dv\ge\frac{R^2r^2}{(R-r)^4}e^{-\beta T}\int_{B_{R/2}\setminus B_{2r}}f_{\text{in}}(v)\Phi(v)\,dv,\quad\forall t<T.\qquad\square$$

**Acknowledgements**

Gualdani is supported by NSF DMS-1310746 and DMS-1412748. Guillen is supported by NSF-DMS 1201413. Both authors would like to thank MSRI for the hospitality during the program Free Boundary Problems, Theory and Applications in the spring of 2011, where this work was started. We also would like to thank Luis Silvestre and Cedric Villani for many fruitful communications. Gualdani would like to thank NCTS Mathematics Division Taipei for their kind hospitality. Last but not least, the authors would

like to express their gratitude to the anonymous referees for their thoughtful and thorough reading of the original manuscript, as well as for their many useful suggestions.

## References

[Alexandre et al. 2015] R. Alexandre, J. Liao, and C. Lin, "Some a priori estimates for the homogeneous Landau equation with soft potentials", *Kinet. Relat. Models* **8**:4 (2015), 617–650. MR 3375485 Zbl 1320.35104

[Arsen'ev and Peskov 1977] A. A. Arsen'ev and N. V. Peskov, "The existence of a generalized solution of Landau's equation", *Ž. Vyčisl. Mat. i Mat. Fiz.* **17**:4 (1977), 1063–1068, 1096. In Russian; translated in *U.S.S.R. Computational Math. and Math. Phys.* **17**:4 (1977), 241–246. MR 0470442 Zbl 0383.35064

[Desvillettes and Villani 2000a] L. Desvillettes and C. Villani, "On the spatially homogeneous Landau equation for hard potentials, I: Existence, uniqueness and smoothness", *Comm. Partial Differential Equations* **25**:1-2 (2000), 179–259. MR 1737547 Zbl 0946.35109

[Desvillettes and Villani 2000b] L. Desvillettes and C. Villani, "On the spatially homogeneous Landau equation for hard potentials, II: $H$-theorem and applications", *Comm. Partial Differential Equations* **25**:1-2 (2000), 261–298. MR 1737548 Zbl 0951.35130

[Fournier 2010] N. Fournier, "Uniqueness of bounded solutions for the homogeneous Landau equation with a Coulomb potential", *Comm. Math. Phys.* **299**:3 (2010), 765–782. MR 2718931 Zbl 1198.35274

[Fournier and Guérin 2009] N. Fournier and H. Guérin, "Well-posedness of the spatially homogeneous Landau equation for soft potentials", *J. Funct. Anal.* **256**:8 (2009), 2542–2560. MR 2502525 Zbl 1165.35467

[Giga and Kohn 1985] Y. Giga and R. V. Kohn, "Asymptotically self-similar blow-up of semilinear heat equations", *Comm. Pure Appl. Math.* **38**:3 (1985), 297–319. MR 784476 Zbl 0585.35051

[Gressman et al. 2012] P. T. Gressman, J. Krieger, and R. M. Strain, "A non-local inequality and global existence", *Adv. Math.* **230**:2 (2012), 642–648. MR 2914961 Zbl 1248.35005

[Gualdani and Guillen $\geq$ 2016] M. Gualdani and N. Guillen, "On Ap weights and regularization effects for homogeneous Landau equations", in preparation.

[Guo 2002] Y. Guo, "The Landau equation in a periodic box", *Comm. Math. Phys.* **231**:3 (2002), 391–434. MR 1946444 Zbl 1042.76053

[Krieger and Strain 2012] J. Krieger and R. M. Strain, "Global solutions to a non-local diffusion equation with quadratic non-linearity", *Comm. Partial Differential Equations* **37**:4 (2012), 647–689. MR 2901061 Zbl 1247.35087

[Ladyženskaja et al. 1968] O. A. Ladyženskaja, V. A. Solonnikov, and N. N. Ural'ceva, *Linear and quasilinear equations of parabolic type*, Translations of Mathematical Monographs **23**, American Mathematical Society, Providence, RI, 1968. MR 0241822 Zbl 0164.12302

[Lieb and Loss 2001] E. H. Lieb and M. Loss, *Analysis*, 2nd ed., Graduate Studies in Mathematics **14**, American Mathematical Society, Providence, RI, 2001. MR 1817225 Zbl 0966.26002

[Lieberman 1996] G. M. Lieberman, *Second order parabolic differential equations*, World Sci. Pub., River Edge, NJ, 1996. MR 1465184 Zbl 0884.35001

[Schwab and Silvestre 2016] R. W. Schwab and L. Silvestre, "Regularity for parabolic integro-differential equations with very irregular kernels", *Anal. PDE* **9**:3 (2016), 727–772. MR 3518535 Zbl 06600511

[Silvestre 2016] L. Silvestre, "A new regularization mechanism for the Boltzmann equation without cut-off", *Comm. Math. Phys.* **348**:1 (2016), 69–100. MR 3551261 Zbl 06642348

[Villani 1998] C. Villani, "On a new class of weak solutions to the spatially homogeneous Boltzmann and Landau equations", *Arch. Rational Mech. Anal.* **143**:3 (1998), 273–307. MR 1650006 Zbl 0912.45011

[Villani 2002] C. Villani, "A review of mathematical topics in collisional kinetic theory", pp. 71–305 in *Handbook of mathematical fluid dynamics*, vol. I, edited by S. Friedlander and D. Serre, North-Holland, Amsterdam, 2002. MR 1942465 Zbl 1170.82369

[Wu 2014] K.-C. Wu, "Global in time estimates for the spatially homogeneous Landau equation with soft potentials", *J. Funct. Anal.* **266**:5 (2014), 3134–3155. MR 3158719 Zbl 1296.35112

Received 4 May 2015. Revised 15 Jun 2016. Accepted 28 Aug 2016.

MARIA PIA GUALDANI: gualdani@gwu.edu
*Department of Mathematics, George Washington University, 2115 G Street NW, Monroe Hall 272, Washington, DC 20052, United States*

NESTOR GUILLEN: nguillen@math.umass.edu
*Department of Mathematics and Statistics, University of Massachussetts Amherst, Amherst, MA 01003-9305, United States*

ANALYSIS AND PDE
Vol. 9, No. 8, 2016
dx.doi.org/10.2140/apde.2016.9.1811

# FORWARD SELF-SIMILAR SOLUTIONS OF THE NAVIER–STOKES EQUATIONS IN THE HALF SPACE

MIKHAIL KOROBKOV AND TAI-PENG TSAI

For the incompressible Navier–Stokes equations in the 3D half space, we show the existence of forward self-similar solutions for arbitrarily large self-similar initial data.

## 1. Introduction

Let $\mathbb{R}^3_+ = \{x = (x_1, x_2, x_3) : x_3 > 0\}$ be a half space with boundary $\partial\mathbb{R}^3_+ = \{x = (x_1, x_2, 0)\}$. Consider the 3D incompressible Navier–Stokes equations for velocity $u : \mathbb{R}^3_+ \times [0, \infty) \to \mathbb{R}^3$ and pressure $p : \mathbb{R}^3_+ \times [0, \infty) \to \mathbb{R}$,

$$\partial_t u - \Delta u + (u \cdot \nabla)u + \nabla p = 0, \quad \operatorname{div} u = 0, \tag{1-1}$$

in $\mathbb{R}^3_+ \times [0, \infty)$, coupled with the boundary condition

$$u|_{\partial\mathbb{R}^3_+} = 0, \tag{1-2}$$

and the initial condition

$$u|_{t=0} = a, \quad \operatorname{div} a = 0, \quad a|_{\partial\mathbb{R}^3_+} = 0. \tag{1-3}$$

The system (1-1) enjoys a scaling property: if $u(x, t)$ is a solution, then so is

$$u^{(\lambda)}(x, t) := \lambda u(\lambda x, \lambda^2 t) \tag{1-4}$$

for any $\lambda > 0$. We say that $u(x, t)$ is *self-similar* (SS) if $u = u^{(\lambda)}$ for every $\lambda > 0$. In that case,

$$u(x, t) = \frac{1}{\sqrt{2t}} U\left(\frac{x}{\sqrt{2t}}\right), \tag{1-5}$$

where $U(x) = u\big(x, \frac{1}{2}\big)$. It is called *discretely self-similar* (DSS) if $u = u^{(\lambda)}$ for one particular $\lambda > 1$. To get self-similar solutions $u(x, t)$ we usually assume the initial data $a(x)$ is also self-similar, i.e.,

$$a(x) = \frac{a(\hat{x})}{|x|}, \quad \hat{x} = \frac{x}{|x|}. \tag{1-6}$$

In view of the above, it is natural to look for solutions satisfying

$$|u(x, t)| \le \frac{C(C_*)}{|x|} \quad \text{or} \quad \|u(\cdot, t)\|_{L^{3,\infty}} \le C(C_*), \tag{1-7}$$

*MSC2010:* 35Q30, 76D05.
*Keywords:* Forward self-similar solutions, Navier-Stokes equations, half space.

where $C_*$ is some norm of the initial data $a$. For $1 \le q, r \le \infty$, we denote the Lorentz spaces by $L^{q,r}$. In such classes, with sufficiently small $C_*$, the unique existence of mild solutions — solutions of the integral equation version of (1-1)–(1-3) via a contraction mapping argument — has been obtained by Giga and Miyakawa [1989] and refined by Kato [1992], Cannone, Meyer and Planchon [Cannone et al. 1994; Cannone and Planchon 1996], and Barraza [1996]. It is also obtained in the broader class $\mathrm{BMO}^{-1}$ in [Koch and Tataru 2001]. In the context of the half space (and smooth exterior domains), it follows from [Yamazaki 2000]. As a consequence, if $a(x)$ is SS or DSS with small norm $C_*$ and $u(x,t)$ is a corresponding solution satisfying (1-7) with small $C(C_*)$, the uniqueness property ensures that $u(x,t)$ is also SS or DSS, because $u^{(\lambda)}$ is another solution with the same bound and same initial data $a^{(\lambda)} = a$. For large $C_*$, mild solutions still make sense but there is no existence theory since perturbative methods like the contraction mapping no longer work.

Alternatively, one may try to extend the concept of weak solutions (which requires $u_0 \in L^2(\mathbb{R}^3)$) to more general initial data. One such theory is local-Leray solutions in $L^2_{\mathrm{uloc}}$, constructed by Lemarié-Rieusset [2002]. However, there is no uniqueness theorem for them and hence the existence of large SS or DSS solutions was unknown. Recently, Jia and Šverák [2014] constructed SS solutions for every SS $u_0$ which is locally Hölder continuous. Their main tool is a local Hölder estimate for local-Leray solutions near $t = 0$, assuming minimal control of the initial data in the large. This estimate enables them to prove a priori estimates of SS solutions, and then to show their existence by the Leray–Schauder degree theorem. This result is extended by Tsai [2014] to the existence of discretely self-similar solutions.

When the domain is the half space $\mathbb{R}^3_+$, however, there is so far no analogous theory of local-Leray solutions. Hence the method of [Jia and Šverák 2014; Tsai 2014] is not applicable.

In this note, our goal is to construct SS solutions in the half space for arbitrary large data. By $BC_w$ we denote bounded and weak-* continuous functions. Our main theorem is the following.

**Theorem 1.1.** *Let $\Omega = \mathbb{R}^3_+$ and let $A$ be the Stokes operator in $\Omega$* (see (5-5)–(5-7)). *For any self-similar vector field $a \in C^1_{\mathrm{loc}}(\overline{\Omega}\backslash\{0\})$ satisfying* $\operatorname{div} a = 0$, $a|_{\partial\Omega} = 0$, *there is a smooth self-similar mild solution $u \in BC_w([0,\infty); L^{3,\infty}_\sigma(\Omega))$ of (1-1) with $u(0) = a$ and*

$$\|u(t) - e^{-tA}a\|_{L^2(\Omega)} = Ct^{1/4}, \quad \|\nabla(u(t) - e^{-tA}a)\|_{L^2(\Omega)} = Ct^{-1/4}, \quad \forall t > 0. \tag{1-8}$$

*Comments on Theorem 1.1:*

(1) There is no restriction on the size of $a$.

(2) It is concerned only with existence. There is no assertion on uniqueness.

(3) Our approach also gives a second construction of large self-similar solutions in the whole space $\mathbb{R}^3$, but for initial data more restrictive ($C^1$) than those of [Jia and Šverák 2014]. In fact, it would show the existence of self-similar solutions in the cones

$$K_\alpha = \{0 \le \phi \le \alpha\}, \quad \text{for } 0 < \alpha \le \pi,$$

(in spherical coordinates), if one could verify Assumption 3.1 for $e^{-A/2}a$. We are able to verify it only for $\alpha = \frac{\pi}{2}$ and $\alpha = \pi$.

(4) We have the uniform bound (1-7) for $u_0(t) = e^{-tA}a$ and we show $|u_0(x,t)| \lesssim \big(\sqrt{t}+|x|\big)^{-1}$ in Section 6. We expect $u_0(t) \notin L^q(\Omega)$ for any $q \le 3$, and $\|u_0(t)\|_{L^q} \to \infty$ as $t \to 0_+$ for $q > 3$. The difference $v = u - u_0$ is more localized: by interpolating (1-8), $\|v(t)\|_{L^q} \to 0$ as $t \to 0_+$ for all $q \in [2,3)$. Although $\|v(t)\|_{L^3(\Omega)} = C$ for $t > 0$, $v(t)$ weakly converges to 0 in $L^3$ as $t \to 0_+$, as easily shown by approximating the test function by $L^2 \cap L^{3/2}$ functions. Both $u_0(t)$ and $v(t)$ belong to $L^\infty(\mathbb{R}_+; L^{3,\infty}(\mathbb{R}^3_+))$.

We now outline our proof. Unlike previous approaches based on the evolution equations, we directly prove the existence of the profile $U$ in (1-5). It is based on the a priori estimates for $U$ using the classical Leray–Schauder fixed point theorem and the Leray reductio ad absurdum argument (which has been fruitfully applied in recent papers of Korobkov, Pileckas and Russo [Korobkov et al. 2013; 2014a; 2014b; 2015a; 2015b] on the boundary value problem of stationary Navier–Stokes equations). Specifically, the profile $U(x)$ satisfies the Leray equations

$$-\Delta U - U - x\cdot\nabla U + (U\cdot\nabla)U + \nabla P = 0, \quad \operatorname{div} U = 0 \tag{1-9}$$

in $\mathbb{R}^3_+$ with zero boundary condition and, in a suitable sense,

$$U(x) \to U_0(x) := (e^{-A/2}a)(x) \quad \text{as } |x| \to \infty. \tag{1-10}$$

System (1-9) was proposed by Leray [1934], with the opposite sign for $U + x\cdot\nabla U$, for the study of singular *backward* self-similar solutions of (1-1) in $\mathbb{R}^3$ of the form $u(x,t) = U\big(x/\sqrt{-2t}\big)/\sqrt{-2t}$. Their triviality was first established in [Nečas et al. 1996] if $U \in L^3(\mathbb{R}^3)$, in particular if $U \in H^1(\mathbb{R}^3)$ as assumed in [Leray 1934], and then extended in [Tsai 1998] to $U \in L^q(\mathbb{R}^3)$, $3 \le q \le \infty$. In the forward case and in the whole space setting, we have

$$|U_0(x)| \sim |x|^{-1}, \quad V(x) := U(x) - U_0(x), \quad |V(x)| \lesssim |x|^{-2} \quad \text{for } |x| > 1; \tag{1-11}$$

see [Jia and Šverák 2014; Tsai 2014]. In the half space setting, it is not clear if one can show a pointwise decay bound for $V$. We show, however, that $V(x)$ is a priori bounded in $H^1_0(\mathbb{R}^3_+)$, and use this a priori bound to construct a solution. Due to lack of compactness of $H^1_0$ at spatial infinity, we use the *invading method* introduced by Leray [1933]: we approximate $\Omega = \mathbb{R}^3_+$ by $\Omega_k = \Omega \cap B_k$, $k = 1, 2, 3, \dots$, where $B_k$ is an increasing sequence of concentric balls, construct solutions $V_k$ in $\Omega_k$ of the difference equation (3-3) with zero boundary condition, and extract a subsequence converging to a desired solution $V$ in $\mathbb{R}^3_+$.

Our proof is structured as follows. We first recall some properties for Euler flows in Section 2, and then use it to show that the $V_k$ are uniformly bounded in $H^1_0(\Omega_k)$ in Section 3. In Section 4, we construct $V_k$ using the a priori bound and a linear version of the Leray–Schauder theorem, and extract a weak limit $V$ using the uniform bound. The arguments in Sections 2–4 are valid as long as one can show that $U_0 = e^{-A_\Omega/2}a$, $A_\Omega$ being the Stokes operator in $\Omega$, satisfies certain decay properties to be specified in Assumption 3.1. In Section 5 we show that, for $\Omega = \mathbb{R}^3_+$ and those initial data $a$ considered in Theorem 1.1, $U_0$ indeed satisfies Assumption 3.1. We finally verify that $u(x,t)$ defined by (1-5) satisfies the assertions of Theorem 1.1 in Section 6.

Because our existence proof does not use the evolution equation, we do not need the nonlinear version of the Leray–Schauder theorem as in [Jia and Šverák 2014; Tsai 2014]. As a side benefit, we do not need to check the small-large uniqueness (cf. [Tsai 2014, Lemma 4.1]).

## 2. Some properties of solutions to the Euler system

For $q \geq 1$, denote by $D^{1,q}(\Omega)$ the set of functions $f \in W^{1,q}_{\mathrm{loc}}(\Omega)$ such that $\|f\|_{D^{1,q}}(\Omega) = \|\nabla f\|_{L^q(\Omega)} < \infty$. Recall, that by the Sobolev embedding theorem, if $q < n$ then for any $f \in D^{1,q}(\mathbb{R}^n)$ there exists a constant $c \in \mathbb{R}$ such that $f - c \in L^p(\mathbb{R}^n)$ with $p = nq/(n-q)$. In particular,

$$f \in D^{1,2}(\mathbb{R}^3) \Rightarrow f - c \in L^6(\mathbb{R}^3), \qquad f \in D^{1,3/2}(\mathbb{R}^3) \Rightarrow f - c \in L^3(\mathbb{R}^3). \tag{2-1}$$

Further, denote by $D^{1,2}_0(\Omega)$ the closure of the set of all smooth functions having compact supports in $\Omega$ with respect to the norm $\|\cdot\|_{D^{1,2}(\Omega)}$, and $H(\Omega) = \{\boldsymbol{v} \in D^{1,2}_0(\Omega) : \operatorname{div} \boldsymbol{v} = 0\}$. In particular,

$$H(\Omega) \hookrightarrow L^6(\Omega). \tag{2-2}$$

(Recall that by the Sobolev inequality, $\|f\|_{L^6(\mathbb{R}^3)} \leq C\|\nabla f\|_{L^2(\mathbb{R}^3)}$ holds for every function $f \in C^\infty_c(\mathbb{R}^3)$ having compact support in $\mathbb{R}^3$; see [Adams and Fournier 2003, Theorem 4.31].)

Assume that the following conditions are fulfilled:

**(E)** Let $\Omega$ be a domain in $\mathbb{R}^3$ with (possibly unbounded) connected Lipschitz boundary $\Gamma = \partial\Omega$, and the functions $\boldsymbol{v} \in H(\Omega)$ and $p \in D^{1,3/2}(\Omega) \cap L^3(\Omega)$ satisfy the Euler system

$$\begin{cases} (\boldsymbol{v} \cdot \nabla)\boldsymbol{v} + \nabla p = 0 & \text{in } \Omega, \\ \operatorname{div} \boldsymbol{v} = 0 & \text{in } \Omega, \\ \boldsymbol{v} = 0 & \text{on } \partial\Omega. \end{cases} \tag{2-3}$$

The next statement was proved in [Kapitanskiĭ and Piletskas 1983, Lemma 4] and in [Amick 1984, Theorem 2.2]; see also [Amirat et al. 1999, Lemma 4].

**Theorem 2.1.** *Let the conditions (E) be fulfilled. Then*

$$\exists\, \hat{p}_0 \in \mathbb{R} : \quad p(x) \equiv \hat{p}_0 \quad \textit{for } \mathfrak{H}^2\textit{-almost all } x \in \partial\Omega. \tag{2-4}$$

Here and henceforth we denote by $\mathfrak{H}^m$ the $m$-dimensional Hausdorff measure $\mathfrak{H}^m(F) = \lim_{t\to 0+} \mathfrak{H}^m_t(F)$, where $\mathfrak{H}^m_t(F) = \inf\{\sum_{i=1}^\infty (\operatorname{diam} F_i)^m : \operatorname{diam} F_i \leq t,\, F \subset \bigcup_{i=1}^\infty F_i\}$.

## 3. A priori bound for Leray equations

Recall that the profile $U(x)$ in (1-5) satisfies Leray equations (1-9) with zero boundary condition and $U(x) \to U_0(x)$ at spatial infinity. Decompose

$$U = U_0 + V, \quad U_0 = e^{-A/2}a. \tag{3-1}$$

Because $a$ is self-similar, $u_0(\cdot, t) = e^{-tA}a$ is also self-similar, i.e., $u_0(x,t) = \lambda u_0(\lambda x, \lambda^2 t)$ for all $\lambda > 0$. Differentiating in $\lambda$ and evaluating at $\lambda = 1$ and $t = \frac{1}{2}$, we get

$$0 = U_0 + x \cdot \nabla U_0 + \partial_t u_0\big(x, \tfrac{1}{2}\big) = U_0 + x \cdot \nabla U_0 + \Delta U_0 - \nabla P_0 \tag{3-2}$$

for some scalar $P_0$. Thus, the difference $V(x)$ satisfies

$$-\Delta V - V - x \cdot \nabla V + \nabla P = F_0 + F_1(V), \quad \operatorname{div} V = 0 \tag{3-3}$$

for some scalar $P$, where

$$F_0 = -U_0 \cdot \nabla U_0, \tag{3-4}$$

$$F_1(V) = -(U_0 + V) \cdot \nabla V - V \cdot \nabla U_0, \tag{3-5}$$

and $V$ vanishes at the boundary and the spatial infinity.

For a Sobolev function $f \in W^{1,2}(\Omega)$, set

$$\|f\|_{H^1(\Omega)} := \left( \int_\Omega |\nabla f|^2 + \tfrac{1}{2}|f|^2 \right)^{1/2}. \tag{3-6}$$

Denote by $H^1_0(\Omega)$ the closure of the set of all smooth functions having compact supports in $\Omega$ with respect to the norm $\|\cdot\|_{H^1(\Omega)}$, and

$$H^1_{0,\sigma}(\Omega) = \{f \in H^1_0(\Omega) : \operatorname{div} f = 0\}.$$

Note that $H^1_0(\Omega) = \{f \in W^{1,2}(\Omega) : f|_{\partial\Omega} = 0,\ \|f\|_{H^1(\Omega)} < \infty\}$ for bounded Lipschitz domains.

We assume the following.

**Assumption 3.1** (boundary data at infinity). Let $\Omega$ be a domain in $\mathbb{R}^3$. The vector field $U_0 : \Omega \to \mathbb{R}^3$ satisfies $\operatorname{div} U_0 = 0$ and

$$\|U_0\|_{L^6(\Omega)} < \infty, \qquad \|\nabla U_0\|_{L^2(\Omega)} < \infty. \tag{3-7}$$

Note that from Assumption 3.1 and (3-4) it follows, in particular, that

$$\left| \int_\Omega F_0 \cdot \eta \right| \le C, \qquad \left| \int_\Omega (\eta \cdot \nabla) U_0 \cdot \eta \right| \le C \tag{3-8}$$

for any $\eta \in H^1_{0,\sigma}(\Omega)$ with $\|\eta\|_{H^1_{0,\sigma}(\Omega)} \le 1$ (by virtue of the evident imbedding $H^1_{0,\sigma}(\Omega) \hookrightarrow L^p$ for all $p \in [2, 6]$).

If it is valid in $\Omega$, it is also valid in any subdomain of $\Omega$ with the same constant $C$. We show in Section 5 that for $\Omega = \mathbb{R}^3_+$ and $a$ satisfying (5-1), $U_0 = e^{-A/2}a$ satisfies (5-3) and hence Assumption 3.1. This is also true if $\Omega = \mathbb{R}^3$ and $a$ is self-similar, divergence free, and locally Hölder continuous.

**Theorem 3.2** (a priori estimate for bounded domain). *Let $\Omega$ be a bounded domain in $\mathbb{R}^3$ with connected Lipschitz boundary $\partial\Omega$, and assume Assumption 3.1 for $U_0$. Then for any function $V \in H^1_0(\Omega)$ satisfying*

$$-\Delta V + \nabla P = \lambda(V + x \cdot \nabla V + F_0 + F_1(V)), \quad \operatorname{div} V = 0 \tag{3-9}$$

*for some* $\lambda \in [0, 1]$, *we have the a priori bound*

$$\|V\|^2_{H^1(\Omega)} = \int_\Omega \left(|\nabla V|^2 + \tfrac{1}{2}|V|^2\right) \le C(U_0, \Omega).$$

**Remark.** Note that $C(U_0, \Omega)$ is independent of $\lambda \in [0, 1]$.

*Proof.* Let the assumptions of the theorem be fulfilled. Suppose that its assertion is not true. Then there exists a sequence of numbers $\lambda_k \in [0, 1]$ and functions $V_k \in H^1_0(\Omega)$ such that

$$-\Delta V_k - \lambda_k V_k - \lambda_k x \cdot \nabla V_k + \nabla P_k = \lambda_k (F_0 + F_1(V_k)), \quad \operatorname{div} V_k = 0, \tag{3-10}$$

and moreover,

$$J_k^2 := \int_\Omega |\nabla V_k|^2 \to \infty. \tag{3-11}$$

Multiplying (3-10) by $V_k$ and integrating by parts in $\Omega$, we obtain the identity

$$J_k^2 + \frac{\lambda_k}{2} \int_\Omega |V_k|^2 = \lambda_k \int_\Omega (F_0 - V_k \cdot \nabla U_0) V_k. \tag{3-12}$$

Consider the normalized sequence of functions

$$\widehat{V}_k = \frac{1}{J_k} V_k, \quad \widehat{P}_k = \frac{1}{\lambda_k J_k^2} P_k. \tag{3-13}$$

Since

$$\int_\Omega |\nabla \widehat{V}_k|^2 \equiv 1,$$

we could extract a subsequence, still denoted by $\widehat{V}_k$, which converges weakly in $W^{1,2}(\Omega)$ to some function $V \in H^1_0(\Omega)$, and strongly in $L^3(\Omega)$. Also we could assume without loss of generality that $\lambda_k \to \lambda_0 \in [0, 1]$.

Multiplying the identity (3-12) by $1/J_k^2$ and taking a limit as $k \to \infty$, we have

$$1 + \frac{\lambda_0}{2} \int_\Omega |V|^2 = -\lambda_0 \int_\Omega (V \cdot \nabla U_0) V = \lambda_0 \int_\Omega (V \cdot \nabla V) U_0. \tag{3-14}$$

In particular, $\lambda_k$ is separated from zero for large $k$.

Multiplying (3-10) by $1/(\lambda_k J_k^2)$, we see that the pairs $(\widehat{V}_k, \widehat{P}_k)$ satisfy the equation

$$\widehat{V}_k \cdot \nabla \widehat{V}_k + \nabla \widehat{P}_k = \frac{1}{J_k}\Big(\frac{1}{\lambda_k} \Delta \widehat{V}_k + \widehat{V}_k + x \cdot \nabla \widehat{V}_k + \frac{1}{J_k} F_0 - U_0 \cdot \nabla \widehat{V}_k - \widehat{V}_k \cdot \nabla U_0\Big). \tag{3-15}$$

Take an arbitrary function $\eta \in C^\infty_{c,\sigma}(\Omega)$. Multiplying (3-15) by $\eta$, integrating by parts and taking a limit, we obtain finally

$$\int_\Omega (V \cdot \nabla V) \cdot \eta = 0. \tag{3-16}$$

Since $\eta \in C^\infty_{c,\sigma}(\Omega)$ is arbitrary, we see that $V$ is a weak solution to the Euler equation

$$\begin{cases} (V \cdot \nabla) V + \nabla P = 0 & \text{in } \Omega, \\ \operatorname{div} V = 0 & \text{in } \Omega, \\ V = 0 & \text{on } \partial\Omega, \end{cases} \tag{3-17}$$

for some $P \in D^{1,3/2}(\Omega) \cap L^3(\Omega)$. By Theorem 2.1, there exists a constant $\hat{p}_0 \in \mathbb{R}$ such that $P(x) \equiv \hat{p}_0$ on $\partial\Omega$. Of course, we can assume without loss of generality that $\hat{p}_0 = 0$, i.e., $P(x) \equiv 0$ on $\partial\Omega$. Then by (3-14) and the first line of (3-17), we get

$$1 + \frac{\lambda_0}{2} \int_\Omega |V|^2 = -\lambda_0 \int_\Omega U_0 \cdot \nabla P = -\lambda_0 \int_\Omega \operatorname{div}(P \cdot U_0) = 0.$$

The obtained contradiction finishes the proof of the theorem. □

**Theorem 3.3** (a priori bound for invading method). *Let $\Omega = \mathbb{R}^3_+$, and assume Assumption 3.1 for $U_0$. Take a sequence of balls $B_k = B(0, R_k) \subset \mathbb{R}^3$ with $R_k \to \infty$, and consider half-balls $\Omega_k = \Omega \cap B_k$. Then for functions $V_k \in H^1_0(\Omega_k)$ satisfying*

$$-\Delta V_k - V_k - x \cdot \nabla V_k + \nabla P_k = F_0 + F_1(V_k), \quad \operatorname{div} V_k = 0, \tag{3-18}$$

*we have the a priori bound*

$$\int_{\Omega_k} \left(|\nabla V_k|^2 + \tfrac{1}{2}|V_k|^2\right) \le C(U_0),$$

*where the constant $C(U_0)$ is independent of $k$.*

*Proof.* Let the assumptions of the theorem be fulfilled. Suppose that its assertion is not true. Then there exists a sequence of domains $\Omega_k$ and a sequence of solutions $V_k \in H^1_0(\Omega_k)$ of (3-18) such that

$$J_k^2 := \|V_k\|^2_{H^1(\Omega_k)} = \int_{\Omega_k} \left(|\nabla V_k|^2 + \tfrac{1}{2}|V_k|^2\right) \to \infty. \tag{3-19}$$

Multiplying (3-18) by $V_k$ and integrating by parts in $\Omega_k$, we obtain the identity

$$J_k^2 = \int_{\Omega_k} (F_0 - V_k \cdot \nabla U_0) V_k. \tag{3-20}$$

Consider the normalized sequence of functions

$$\widehat{V}_k = \frac{1}{J_k} V_k, \quad \widehat{P}_k = \frac{1}{J_k^2} P_k. \tag{3-21}$$

Multiplying (3-18) by $1/J_k^2$, we see that the pairs $(\widehat{V}_k, \widehat{P}_k)$ satisfy the equation

$$\widehat{V}_k \cdot \nabla \widehat{V}_k + \nabla \widehat{P}_k = \frac{1}{J_k}(\Delta \widehat{V}_k + \widehat{V}_k + x \cdot \nabla \widehat{V}_k + F_0 - U_0 \cdot \nabla \widehat{V}_k - \widehat{V}_k \cdot \nabla U_0). \tag{3-22}$$

Since

$$\int_{\Omega_k} \left(|\nabla \widehat{V}_k|^2 + \tfrac{1}{2}|\widehat{V}_k|^2\right) \equiv 1,$$

we could extract a subsequence, still denoted by $\widehat{V}_k$, which converges weakly in $W^{1,2}(\Omega)$ to some function $V \in H^1_0(\Omega)$, and strongly in $L^2(E)$ for any $E \Subset \overline{\Omega}$.

Multiplying the identity (3-20) by $1/J_k^2$ and taking a limit as $k \to \infty$, we have

$$1 = \int_\Omega (-V \cdot \nabla U_0) V. \tag{3-23}$$

Take an arbitrary function $\eta \in C^\infty_{c,\sigma}(\Omega)$. Multiplying (3-22) by $\eta$, integrating by parts and taking a limit, we obtain finally

$$\int_\Omega (V \cdot \nabla V) \cdot \eta = 0. \tag{3-24}$$

Since $\eta \in C^\infty_{c,\sigma}(\Omega)$ is arbitrary, we see that $V$ is a weak solution to the Euler equation

$$\begin{cases} (V \cdot \nabla)V + \nabla P = 0 & \text{in } \Omega, \\ \operatorname{div} V = 0 & \text{in } \Omega, \\ V = 0 & \text{on } \partial\Omega, \end{cases} \tag{3-25}$$

with some $P \in D^{1,3/2}(\Omega) \cap L^3(\Omega)$. More precisely, since $V, \nabla V \in L^2(\Omega)$, we have $P \in D^{1,q}(\Omega)$ for every $q \in \left[1, \frac{3}{2}\right]$. Consequently, $P \in L^s(\Omega)$ for each $s \in \left[\frac{3}{2}, 3\right]$. In particular, $P \in L^3(\Omega)$ and $\nabla P \in L^{9/8}(\Omega)$. Furthermore,

$$\begin{aligned} \int_{S_R^+} |P|^{4/3} &= -R^2 \int_R^\infty \int_{S_1^+} \frac{d}{dr}\left(|P(r\omega)|^{4/3}\right) d\omega\, dr \\ &\lesssim \int_{|x|>R} |P|^{1/3} |\nabla P| \le \left(\int_{|x|>R} |P|^3\right)^{1/9} \left(\int_{|x|>R} |\nabla P|^{9/8}\right)^{8/9}, \end{aligned}$$

where $S_R^+ = \{x \in \Omega : |x| = R\}$ is the corresponding half-sphere. Hence, we conclude that

$$\int_{S_R^+} |P|^{4/3} \to 0 \quad \text{as } R \to \infty. \tag{3-26}$$

Analogously, from the assumption $U_0 \in L^6(\Omega)$, $\nabla U \in L^2(\Omega)$, it is very easy to deduce that

$$\int_{S_R^+} |U_0|^4 \to 0 \quad \text{as } R \to \infty. \tag{3-27}$$

On the other hand, by (3-23) and the first line of (3-25) we obtain

$$1 = \int_\Omega (V \cdot \nabla)V \cdot U_0 = -\int_\Omega \nabla P \cdot U_0 = -\lim_{R\to\infty} \int_{\Omega_R} \operatorname{div}(P \cdot U_0) = -\lim_{R\to\infty} \int_{S_R^+} P(U_0 \cdot \boldsymbol{n}) = 0, \tag{3-28}$$

where $\Omega_R = \Omega \cap B(0, R)$ and the last equality follows from (3-26)–(3-27). The obtained contradiction finishes the proof of the theorem. □

## 4. Existence for Leray equations

The proof of the existence theorem for the system of equations (3-3)–(3-5) in bounded domains $\Omega$ is based on the following fundamental fact.

**Theorem 4.1** (Leray–Schauder theorem). *Let $S : X \to X$ be a continuous and compact mapping of a Banach space $X$ into itself, such that the set*

$$\{x \in X : x = \lambda Sx \text{ for some } \lambda \in [0, 1]\}$$

*is bounded. Then $S$ has a fixed point $x_* = Sx_*$.*

Let $\Omega$ be a domain in $\mathbb{R}^3$ with connected Lipschitz boundary $\Gamma = \partial\Omega$, and set $X = H^1_{0,\sigma}(\Omega)$.

For functions $V_1, V_2 \in H^1_{0,\sigma}(\Omega)$, write $\langle V_1, V_2\rangle_H = \int_\Omega \nabla V_1 \cdot \nabla V_2$. Then the system (3-3)–(3-5) is equivalent to the following identities:

$$\langle V, \zeta\rangle_H = \int_\Omega G(V)\cdot\zeta, \quad \forall \zeta \in C^\infty_{c,\sigma}(\Omega), \tag{4-1}$$

where $G(V) = V + x\cdot\nabla V + F(V)$, $F(V) = F_0 + F_1(V)$,

$$F_0(x) = -U_0\cdot\nabla U_0, \tag{4-2}$$

$$F_1(V) = -(U_0+V)\cdot\nabla V - V\cdot\nabla U_0. \tag{4-3}$$

Since $H^1_{0,\sigma}(\Omega) \hookrightarrow L^6(\Omega)$, by the Riesz representation theorem, for any $f \in L^{6/5}(\Omega)$ there exists a unique mapping $T(f) \in H^1_{0,\sigma}(\Omega)$ such that

$$\langle T(f), \zeta\rangle_H = \int_\Omega f\cdot\zeta, \quad \forall \zeta \in C^\infty_{c,\sigma}(\Omega), \tag{4-4}$$

and moreover,

$$\|T(f)\|_H \le \|f\|_{X'},$$

where

$$\|f\|_{X'} = \sup_{\zeta\in C^\infty_{c,\sigma}(\Omega),\, \|\zeta\|_H\le 1} \int_\Omega f\cdot\zeta.$$

Then the system (3-3)–(3-5)$\sim$(4-1) is equivalent to the equality

$$V = T(G(V)). \tag{4-5}$$

**Theorem 4.2** (compactness). *If $\Omega$ is a bounded domain in $\mathbb{R}^3$ with connected Lipschitz boundary $\Gamma = \partial\Omega$, and Assumption 3.1 holds for $U_0$, then for $X = H^1_{0,\sigma}(\Omega)$ the operator $S : X \ni V \mapsto T(G(V)) \in X$ is continuous and compact.*

*Proof.* (i) For $V, \tilde V \in X$, setting $v = \tilde V - V$,

$$F(\tilde V) - F(V) = -(U_0 + V + v)\cdot\nabla v - v\cdot\nabla(U_0+V).$$

Thus we have

$$\begin{aligned}
&\|S(\tilde V) - S(V)\|_X \\
&\lesssim \|v\|_{L^2} + \|\nabla v\|_{L^2} + \|F(\tilde V) - F(V)\|_{L^{6/5}} \\
&\lesssim \|v\|_{L^2} + \|\nabla v\|_{L^2} + \|U_0\|_{L^3}\|\nabla v\|_{L^2} + \|V+v\|_{L^3}\|\nabla v\|_{L^2} + \|\nabla U_0\|_{L^2}\|v\|_{L^3} + \|v\|_{L^3}\|\nabla V\|_{L^2} \\
&\lesssim (1 + \|V\|_X + \|v\|_X)\|v\|_X.
\end{aligned} \tag{4-6}$$

(ii) By the Sobolev theorems, we have the compact embedding $X \hookrightarrow L^r(\Omega)$ for all $r \in [1, 6)$. Thus if a sequence $V_k \in X$ is bounded in $X$, i.e., $\|V_k\|_{L^2(\Omega)} + \|\nabla V_k\|_{L^2(\Omega)} \le C$, then we can extract a subsequence $V_{k_l}$ which converges to some $V \in X$ in $L^3(\Omega)$ norm: $\|V_{k_l} - V\|_{L^3(\Omega)} \to 0$ as $l \to \infty$. Then using the condition $V_{k_l} \equiv V \equiv 0$ on $\partial\Omega$ and integration by parts, it is easy to see that $\|F(V_{k_l}) - F(V)\|_{X'} \to 0$ and, consequently, $\|G(V_{k_l}) - G(V)\|_{X'} \to 0$ as $l \to \infty$. □

**Corollary 4.3** (existence in bounded domains). *Let $\Omega$ be a bounded domain in $\mathbb{R}^3$ with connected Lipschitz boundary $\partial\Omega$, and assume Assumption 3.1 for $U_0$. Then the system (3-3)–(3-5) has a solution $V \in H^1_{0,\sigma}(\Omega)$.*

*Proof.* This is a direct consequence of Theorems 4.1–4.2 and 3.2. □

**Theorem 4.4** (existence in unbounded domains). *Let $\Omega = \mathbb{R}^3_+$, and assume Assumption 3.1 for $U_0$. Then the system (3-3)–(3-5) has a solution $V \in H^1_{0,\sigma}(\Omega)$.*

*Proof.* Take balls $B_k = B(0,k)$ and consider the increasing sequence of domains $\Omega_k = \Omega \cap B_k$ from Theorem 3.3. By Corollary 4.3 there exists a sequence of solutions $V_k \in H^1_{0,\sigma}(\Omega_k)$ of the system (3-3)–(3-5) in $\Omega_k$. By Theorem 3.3, the norms $\|V_k\|_{H^1_{0,\sigma}(\Omega)}$ are uniformly bounded, thus we can extract a subsequence $V_{k_l}$ such that the weak convergence $V_{k_l} \rightharpoonup V$ in $W^{1,2}(\Omega')$ holds for any bounded subdomain $\Omega' \subset \Omega$. It is easy to check that the limit function $V$ is a solution of the system (3-3)–(3-5) in $\Omega$. □

## 5. Boundary data at infinity in the half space

In this section we restrict ourselves to the half space $\Omega = \mathbb{R}^3_+$ with boundary $\Sigma = \partial\mathbb{R}^3_+$ and study the decay property of $U_0 = e^{-A/2}a$. Our goal is to prove the following lemma, which ensures Assumption 3.1 under the conditions of Theorem 1.1.

Write $x^* = (x', -x_3)$ given $x = (x', x_3) \in \mathbb{R}^3$, and $\langle z\rangle = (1+|z|^2)^{1/2}$ for $z \in \mathbb{R}^m$.

**Lemma 5.1.** *Suppose $a$ is a vector field in $\Omega = \mathbb{R}^3_+$ satisfying*

$$\begin{gathered} a \in C^1_{\mathrm{loc}}(\overline{\Omega}\backslash\{0\};\mathbb{R}^3), \quad \operatorname{div} a = 0, \quad a|_{\partial\Omega} = 0, \\ a(x) = \lambda a(\lambda x), \quad \forall x \in \Omega,\ \forall \lambda > 0. \end{gathered} \tag{5-1}$$

*Let $U_0 = e^{-A/2}a$, where $A$ is the Stokes operator in $\Omega$. Then*

$$|\nabla^k U_0(x)| \le c_k [a]_1 (1+x_3)^{-\min(1,k)} (1+|x|)^{-1}, \quad \forall k \in \mathbb{Z}_+ = \{0,1,2,\dots\}, \tag{5-2}$$

*and, for any $0 < \delta \ll 1$,*

$$|\nabla U_0(x)| \le c_\delta [a]_1 x_3^{-\delta} \langle x\rangle^{2\delta-2}, \tag{5-3}$$

*where $[a]_m = \sup_{k\le m,\,|x|=1} |\nabla^k a(x)|$.*

*If we further assume $a \in C^m_{\mathrm{loc}}$, $m \ge 2$, and $\partial_3^k a|_\Sigma = 0$ for $k < m$, then $|\nabla^k U_0(x)| \le c_k [a]_m \langle x_3\rangle^{-k} \langle x\rangle^{-1}$ for $k \le m$.*

Estimates (5-2) and (5-3) imply, in particular,

$$U_0 \in L^4(\Omega) \cap L^\infty(\Omega), \quad \nabla U_0 \in L^2(\Omega), \tag{5-4}$$

and hence Assumption 3.1 for $U_0$ is satisfied.

***Green tensor for the nonstationary Stokes system in the half space.*** Consider the nonstationary Stokes system in the half space $\mathbb{R}^3_+$,

$$\partial_t v - \Delta v + \nabla p = 0, \quad \operatorname{div} v = 0, \quad \text{for } x \in \mathbb{R}^3_+,\ t > 0, \tag{5-5}$$

$$v|_{x_3=0} = 0, \quad v|_{t=0} = a. \tag{5-6}$$

In our notation,

$$v(t) = e^{-tA}a. \tag{5-7}$$

It is shown by Solonnikov [2003, §2] that, if $a = \breve{a}$ satisfies

$$\operatorname{div} \breve{a} = 0, \quad \breve{a}_3|_{x_3=0} = 0, \tag{5-8}$$

then

$$v_i(x,t) = \int_{\mathbb{R}^3_+} \breve{G}_{ij}(x,y,t)\breve{a}_j(y)\,dy \tag{5-9}$$

with

$$\begin{aligned} &\breve{G}_{ij}(x,y,t) = \delta_{ij}\Gamma(x-y,t) + G^*_{ij}(x,y,t), \\ &G^*_{ij}(x,y,t) = -\delta_{ij}\Gamma(x-y^*,t) - 4(1-\delta_{j3})\frac{\partial}{\partial x_j}\int_{\mathbb{R}^2\times[0,x_3]} \frac{\partial}{\partial x_i}E(x-z)\,\Gamma(z-y^*,t)\,dz, \end{aligned} \tag{5-10}$$

where $E(x) = 1/(4\pi|x|)$ and $\Gamma(x,t) = (4\pi t)^{-3/2}e^{-|x|^2/(4t)}$ are the fundamental solutions of the Laplace and heat equations in $\mathbb{R}^3$. (A sign difference occurs since $E(x) = -1/(4\pi|x|)$ in [Solonnikov 2003].) Moreover, $G^*_{ij}$ satisfies the pointwise bound

$$|\partial_t^s D_x^k D_y^\ell G^*_{ij}(x,y,t)| \lesssim t^{-s-\ell_3/2}\big(\sqrt{t}+x_3\big)^{-k_3}\big(\sqrt{t}+|x-y^*|\big)^{-3-|k'|-|\ell'|}e^{-cy_3^2/t} \tag{5-11}$$

for all $s \in \mathbb{N} = \{0,1,2,\dots\}$ and $k, \ell \in \mathbb{N}^3$ [Solonnikov 2003, (2.38)].

Note that $\breve{G}_{ij}$ is not the Green tensor in the strict sense since it requires (5-8). There is no known pointwise estimate for the Green tensor; cf. [Solonnikov 1964; Kang 2004].

We now estimate $U_0 = e^{-A/2}a$ for $a$ satisfying (5-1). By (5-9) and (5-10),

$$U_{0,i}(x) = \int_{\mathbb{R}^3_+} \Gamma\big(x-y,\tfrac{1}{2}\big)a_i(y)\,dy + \int_{\mathbb{R}^3_+} G^*_{ij}\big(x,y,\tfrac{1}{2}\big)a_j(y)\,dy =: U_{1,i}(x) + U_{2,i}(x). \tag{5-12}$$

By (5-11), for $k \in \mathbb{Z}_+$ and using only $|a(y)| \lesssim 1/|y'|$,

$$\begin{aligned} |\nabla^k U_2(x)| &\lesssim \int_{\mathbb{R}^3_+} (1+x_3)^{-k}(1+x_3+|x'-y'|)^{-3}e^{-cy_3^2}\frac{1}{|y'|}\,dy \\ &\lesssim (1+x_3)^{-k}\int_{\mathbb{R}^2}(1+x_3+|x'-y'|)^{-3}\frac{1}{|y'|}\,dy' \\ &= (1+x_3)^{-k-2}\int_{\mathbb{R}^2}(1+|\bar{x}-z'|)^{-3}\frac{1}{|z'|}\,dz' \\ &\lesssim (1+x_3)^{-k-2}(1+|\bar{x}|)^{-1} \\ &= (1+x_3)^{-k-1}(1+x_3+|x'|)^{-1}, \end{aligned} \tag{5-13}$$

where $\bar{x} = x'/(1+x_3)$. To estimate $U_1$, fix a cut-off function $\zeta(x) \in C_c^\infty(\mathbb{R}^3)$ with $\zeta(x) = 1$ for $|x| < 1$. We have

$$\nabla^k U_{1,i}(x) = \int_{\mathbb{R}^3_+} \Gamma\big(x-y,\tfrac{1}{2}\big)\nabla_y^k\big((1-\zeta(y))a_i(y)\big)\,dy + \int_{\mathbb{R}^3_+} \nabla_x^k\Gamma\big(x-y,\tfrac{1}{2}\big)(\zeta(y)a_i(y))\,dy, \tag{5-14}$$

using $a|_\Sigma = 0$. Hence, for $k \le 1$,

$$|\nabla^k U_1(x)| \lesssim \int_{\mathbb{R}^3} e^{-|x-y|^2/2} \langle y\rangle^{-1-k}\, dy + e^{-x^2/4} \lesssim \langle x\rangle^{-1-k}. \tag{5-15}$$

We can get the same estimate for $k \ge 2$ if we assume $\nabla^k a$ is defined and has the same decay. On the other hand, we can show $|\nabla_x^k U_1(x)| \lesssim \langle x\rangle^{-2}$ for $k \ge 2$ if we place the extra derivatives on $\Gamma$ in the first integral of (5-14).

Combining (5-13) and (5-15), we get (5-2) and the last statement of Lemma 5.1.

Write

$$\Omega_- = \{x \in \Omega : 1 + x_3 > |x'|\}, \quad \Omega_+ = \{x \in \Omega : 1 + x_3 \le |x'|\}. \tag{5-16}$$

By (5-13) and (5-15), we have shown (5-3) in $\Omega_-$ (with $\delta = 0$).

It remains to show (5-3) in $\Omega_+$.

***Estimates using boundary layer integrals.*** Set $\varepsilon_j = 1$ for $j < 3$ and $\varepsilon_3 = -1$. Thus $x_j^* = \varepsilon_j x_j$. Let $\bar a(x)$ be an extension of $a(x)$ to $x \in \mathbb{R}^3$ with

$$\bar a_j(x) = \varepsilon_j a_j(x^*), \quad \text{if } x_3 < 0.$$

Since $\operatorname{div} a = 0$ in $\mathbb{R}^3_+$ and $a|_\Sigma = 0$, it follows that $\operatorname{div} \bar a = 0$ in $\mathbb{R}^3$. Let $u(x,t)$ be the solution of the nonstationary Stokes system in $\mathbb{R}^3$ with initial data $\bar a$, given simply by

$$u_i(x,t) = \int_{\mathbb{R}^3} \Gamma(y,t)\bar a_i(x-y)\, dy.$$

It follows that $u_i(x,t) = \varepsilon_i u_i(x^*,t)$. Thus

$$\partial_3 u_i(x,t)|_\Sigma = 0, \quad \text{for } i < 3; \qquad u_3(x,t)|_\Sigma = 0. \tag{5-17}$$

We have $|\nabla^k a(y)| \lesssim |y|^{-1-k}$ for $k \le 1$. By the same estimates leading to (5-15) for $U_1$, we have

$$\left|\nabla_x^k u_i\left(x, \tfrac12\right)\right| \lesssim \langle x\rangle^{-1-\min(1,k)}, \quad \text{for } k \le 2. \tag{5-18}$$

Thus $u\left(x, \frac12\right)$ satisfies (5-3).

Using the self-similarity condition

$$u(x,t) = \lambda u(\lambda x, \lambda^2 t), \quad \forall \lambda > 0, \tag{5-19}$$

from (5-18) we get

$$|\nabla_x^m u_i(x,t)| \lesssim \begin{cases} \left(|x| + \sqrt{t}\right)^{-1-m}, & m \le 1, \\ t^{-1/2}\left(|x| + \sqrt{t}\right)^{-2}, & m = 2. \end{cases} \tag{5-20}$$

Now decompose

$$v = u - w.$$

Then $w$ satisfies the nonstationary Stokes system in $\mathbb{R}^3_+$ with zero force, zero initial data, and has boundary value

$$w_j(x,t)|_{x_3=0} = u_j(x',0,t), \quad \text{if } j < 3; \qquad w_3(x,t)|_{x_3=0} = 0. \tag{5-21}$$

Using (5-21), it is given by the boundary layer integral

$$w_i(x,t) = \sum_{j=1,2} \int_0^t \int_\Sigma K_{ij}(x - z', s)\, u_j(z', 0, t-s)\, dz'\, ds, \tag{5-22}$$

where, for $j < 3$,

$$K_{ij}(x,t) = -2\delta_{ij}\partial_3\Gamma - \frac{1}{\pi}\partial_j \mathcal{C}_i, \tag{5-23}$$

$$\mathcal{C}_i(x,t) = \int_{\Sigma\times[0,x_3]} \partial_3\Gamma(y,t)\frac{y_i - x_i}{|y-x|^3}\, dy \tag{5-24}$$

[Solonnikov 1964, pp. 40, 48]. (Note that the $K_{i3}$ $(j = 3)$ have extra terms.) They satisfy for $j < 3$

$$|\partial_t^m D_{x'}^\ell \partial_{x_3}^k \mathcal{C}_i(x,t)| \le c t^{-m-(1/2)}\big(x_3 + \sqrt{t}\big)^{-k}\big(|x| + \sqrt{t}\big)^{-2-\ell} \tag{5-25}$$

[Solonnikov 1964, pp. 41, 48].

We now show (5-3) for $w\big(x, \frac{1}{2}\big)$ in the region $\Omega_+ : 1 + x_3 \le |x'|$.

For $t = \frac{1}{2}$ and $i, k \in \{1, 2, 3\}$,

$$\begin{aligned}\partial_{x_k} w_i\big(x, \tfrac{1}{2}\big) &= -\sum_{j=1,2} \int_0^{1/2}\int_\Sigma \frac{1}{\pi}\partial_k \mathcal{C}_i(x - z', s)\partial_{z_j} u_j\big(z', 0, \tfrac{1}{2} - s\big)\, dz'\, ds \\ &\quad - \mathbf{1}_{i<3}\int_0^{1/2}\int_\Sigma 2\partial_k\partial_3\Gamma(x - z', s)\, u_i\big(z', 0, \tfrac{1}{2} - s\big)\, dz'\, ds \\ &= I_1 + I_2.\end{aligned} \tag{5-26}$$

Above, we have integrated by parts in tangential directions $x_j$ in $I_1$.

By (5-20) and (5-25),

$$|I_1| \lesssim \int_0^{1/2}\int_\Sigma s^{-1/2}\big(x_3 + \sqrt{s}\big)^{-1}\big(|x - z'| + \sqrt{s}\big)^{-2}\big(|z'| + \sqrt{\tfrac{1}{2} - s}\big)^{-2}\, dz'\, ds.$$

Fix $0 < \varepsilon \le \frac{1}{2}$. Splitting $\big(0, \frac{1}{2}\big)$ as $\big(0, \frac{1}{4}\big] \cup \big(\frac{1}{4}, \frac{1}{2}\big)$, and making the change of variable $s \to \frac{1}{2} - s$ in $\big(\frac{1}{4}, \frac{1}{2}\big)$, we get

$$\begin{aligned}|I_1| \lesssim \int_0^{1/4}\int_\Sigma x_3^{-2\varepsilon} s^{-1+\varepsilon}\big(|x' - z'| + x_3 + \sqrt{s}\big)^{-2}(|z'| + 1)^{-2}\, dz'\, ds \\ + \int_0^{1/4}\int_\Sigma (x_3 + 1)^{-1}(|x' - z'| + x_3 + 1)^{-2}\big(|z'| + \sqrt{s}\big)^{-2}\, dz'\, ds.\end{aligned}$$

Integrating first in time and using, for $0 < b < \infty$, $0 \le a < 1 < a + b$, and $0 < N < \infty$, that

$$\int_0^1 \frac{ds}{s^a(N+s)^b} \le \frac{C}{N^{a+b-1}(N+1)^{1-a}}, \tag{5-27}$$

$$\int_0^1 \frac{ds}{s^a(N+s)^{1-a}} \le C\min\left(\frac{1}{N^{1-a}}, \log\frac{2N+2}{N}\right), \tag{5-28}$$

where the constant $C$ is independent of $N$, we get

$$|I_1| \lesssim \int_\Sigma x_3^{-2\varepsilon}(|x'-z'|+x_3)^{-2+2\varepsilon}(|x'-z'|+x_3+1)^{-2\varepsilon}(|z'|+1)^{-2}\,dz' + \int_\Sigma (x_3+1)^{-1}(|x'-z'|+x_3+1)^{-2}\min\left(\frac{1}{|z'|^2}, \log\frac{2|z'|^2+2}{|z'|^2}\right)dz'.$$

Dividing the integration domain into $|z'| < \frac{1}{2}|x'|$, $\frac{1}{2}|x'| < |z'| < 2|x'|$, and $|z'| > 2|x'|$, we get

$$|I_1| \lesssim x_3^{-2\varepsilon}\langle x\rangle^{-2+\delta}, \quad \text{for } x \in \Omega_+ \tag{5-29}$$

for any $0 < \delta \ll 1$. Taking $\varepsilon = \frac{1}{2}\delta$ and $\varepsilon = \frac{1}{2}$, we get

$$(1+x_3)|I_1| \lesssim x_3^{-\delta}\langle x\rangle^{-2+2\delta}, \quad \text{for } x \in \Omega_+. \tag{5-30}$$

To estimate $I_2$ for $i < 3$ (note $I_2 = 0$ if $i = 3$), we separate two cases. If $k < 3$, integration by parts gives

$$I_2 = -\int_0^{1/2}\int_\Sigma 2\partial_3\Gamma(x-z', s)\,\partial_{z_k}u_i\big(z', 0, \tfrac{1}{2}-s\big)\,dz'\,ds.$$

Using $ue^{-u^2} \le C_\ell(1+u)^{-\ell}$ for $u > 0$ and any $\ell > 0$,

$$\partial_3\Gamma(x,s) = cs^{-2}\frac{x_3}{\sqrt{s}}e^{-x^2/4s} \le cs^{-2}\left(1+\frac{|x|}{\sqrt{s}}\right)^{-3} = cs^{-1/2}\big(|x|+\sqrt{s}\big)^{-3}. \tag{5-31}$$

Hence $I_2$ can be estimated in the same way as $I_1$, and (5-30) is valid if $I_1$ is replaced by $I_2$ and $k < 3$.

When $k = 3$, by $\partial_t\Gamma = \Delta\Gamma$ and integration by parts,

$$\begin{aligned} I_2 &= \int_0^{1/2}\int_\Sigma 2\Big(\sum_{j<3}\partial_j^2 - \partial_t\Big)\Gamma(x-z', s)\,u_i\big(z', 0, \tfrac{1}{2}-s\big)\,dz'ds \\ &= \sum_{j<3}\int_0^{1/2}\int_\Sigma 2\partial_j\Gamma(x-z', s)\,\partial_{z_j}u_i\big(z', 0, \tfrac{1}{2}-s\big)\,dz'\,ds \\ &\qquad + \int_0^{1/2}\int_\Sigma 2\Gamma(x-z', s)\,\partial_t u_i\big(z', 0, \tfrac{1}{2}-s\big)\,dz'\,ds \\ &\qquad - \lim_{\mu\to 0_+}\left(\int_\Sigma 2\Gamma\big(x-z', \tfrac{1}{2}-\mu\big)u_i(z', 0, \mu)\,dz - \int_\Sigma 2\Gamma(x-z', \mu)\,u_i\big(z', 0, \tfrac{1}{2}-\mu\big)\,dz\right) \\ &= I_3 + I_4 + \lim_{\mu\to 0_+}(I_{5,\mu} + I_{6,\mu}). \end{aligned}$$

Here $I_3$ can be estimated in the same way as $I_1$, and (5-30) is valid if $I_1$ is replaced by $I_3$. For $I_4$, since $\partial_t u_i = \Delta u_i$, by estimate (5-20) for $\nabla^2 u$,

$$|I_4| \lesssim \int_0^{1/2}\int_\Sigma s^{-3/2}\left(1+\frac{|x-z'|^2}{4s}\right)^{-3/2}\left(\frac{1}{2}-s\right)^{-1/2}\left(|z'|+\sqrt{\frac{1}{2}-s}\right)^{-2}dz'\,ds. \tag{5-32}$$

We have a similar estimate as $I_1$ with the following difference: we have to use the estimate (5-27) during the integration over each subinterval $s \in [0, \frac{1}{4}]$ and $s \in [\frac{1}{4}, \frac{1}{2}]$; for the second subinterval we apply (5-27) with $a = \frac{1}{2}$, $b = 1$, $N = |z'|^2$.

For the boundary terms, the integrand of $I_{5,\mu}$ is bounded by $e^{-|x-z'|^2/2}|z'|^{-1}$ and converges to 0 as $\mu \to 0_+$ for each $z' \in \Sigma$. Thus $\lim I_{5,\mu} = 0$ by the Lebesgue dominated convergence theorem. For $I_{6,\mu}$,

$$|I_{6,\mu}| \lesssim \mu^{-1/2} e^{-x_3^2/(4\mu)} \int_\Sigma \Gamma_{\mathbb{R}^2}(x' - z', \mu) \frac{1}{\langle z' \rangle} \, dz' \lesssim \mu^{-1/2} e^{-x_3^2/(4\mu)} \frac{1}{\langle x' \rangle}, \tag{5-33}$$

which converges to 0 as $\mu \to 0_+$ for any $x \in \Omega$.

We conclude that, for either $k < 3$ or $k = 3$, (5-30) is valid if $I_1$ is replaced by $I_2$ and hence, for any $0 < \delta \ll 1$,

$$(1 + x_3)\big|\partial_k w_i\big(x, \tfrac{1}{2}\big)\big| \lesssim x_3^{-\delta} \langle x \rangle^{-2+2\delta}, \quad \forall x \in \Omega_+, \ \forall i, k \le 3. \tag{5-34}$$

Combining (5-18) and (5-34), we have shown (5-3) in $\Omega_+$, concluding the proof of Lemma 5.1. □

## 6. Self-similar solutions in the half space

In this section we first complete the proof of Theorem 1.1, and then give a few comments.

*Proof of Theorem 1.1.* By Lemma 5.1, for those $a$ satisfying the assumptions of Theorem 1.1, $U_0 = e^{-A/2}a$ satisfies (5-2) and (5-3), and hence Assumption 3.1 is satisfied. By Theorem 4.4, there is a solution $V \in H^1_{0,\sigma}(\mathbb{R}^3_+)$ of the system (3-3)–(3-5).

Noting $U_0 \in C^\infty(\mathbb{R}^3_+)$ by (5-2), the system (3-3)–(3-5) is a perturbation of the stationary Navier–Stokes system with smooth coefficients. The regularity theory for the Navier–Stokes system implies that $V \in C^\infty_{\mathrm{loc}}\big(\overline{\mathbb{R}^3_+}\big)$. The vector field $U = U_0 + V$ is thus a smooth solution of the Leray equations (1-9) in $\mathbb{R}^3_+$.

The vector field $u(x,t)$ defined by (1-5), $u(x,t) = U\big(x/\sqrt{2t}\big)/\sqrt{2t}$, is thus smooth and self-similar. Moreover,

$$v(x,t) = u(x,t) - e^{-tA}a = \frac{1}{\sqrt{2t}} V\left(\frac{x}{\sqrt{2t}}\right)$$

satisfies

$$\|v(t)\|_{L^q(\mathbb{R}^3_+)} = \|V\|_{L^q(\mathbb{R}^3_+)} (2t)^{(3/2q)-(1/2)} \quad \text{and} \quad \|\nabla v(t)\|_{L^2(\mathbb{R}^3_+)} = \|\nabla V\|_{L^2(\mathbb{R}^3_+)} (2t)^{-1/4}.$$

This finishes the proof of Theorem 1.1. □

**Remark.** Let $u_0(x,t) = (e^{-tA}a)(x) = U_0\big(x/\sqrt{2t}\big)/\sqrt{2t}$. We have $u_0(\cdot, t) \to a$ as $t \to 0_+$ in $L^{3,\infty}(\mathbb{R}^3_+)$. Indeed, by (5-2), $|U_0(x)| \lesssim \langle x \rangle^{-1} \in L^{3,\infty} \cap L^q$, $q > 3$. We have $\|u_0(t)\|_{L^q(\mathbb{R}^3_+)} = \|U_0\|_{L^q(\mathbb{R}^3_+)} (2t)^{(3/2q)-(1/2)}$, which remains finite as $t \to 0_+$ only if $q = (3, \infty)$, and

$$|u_0(x,t)| \lesssim \frac{1}{\sqrt{t}} \cdot \frac{1}{1 + |x|/\sqrt{t}} = \frac{1}{\sqrt{t} + |x|}. \tag{6-1}$$

This is consistent with the whole space case $\Omega = \mathbb{R}^3$.

For the difference $V(x)$, we only have its $L^q(\mathbb{R}^3_+)$ bounds, and not pointwise bounds as (1-11) in [Jia and Šverák 2014; Tsai 2014].

## Acknowledgments

The authors are deeply indebted to K. Pileckas and R. Russo for valuable discussions. In particular, they pointed out the way to simplify the proofs for a stronger result.

This work was initiated when Korobkov visited the Center for Advanced Study in Theoretical Sciences (CASTS) of the National Taiwan University. The research of Korobkov was partially supported by the Russian Foundation for Basic Research (projects 14-01-00768-a and 15-01-08275-a) and by the Dynasty Foundation. The research of Tsai was partially supported by the Natural Sciences and Engineering Research Council of Canada grant 261356-13.

## References

[Adams and Fournier 2003] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, 2nd ed., Pure and Applied Mathematics **140**, Elsevier, Amsterdam, 2003. MR 2424078 Zbl 1098.46001

[Amick 1984] C. J. Amick, "Existence of solutions to the nonhomogeneous steady Navier–Stokes equations", *Indiana Univ. Math. J.* **33**:6 (1984), 817–830. MR 763943 Zbl 0563.35059

[Amirat et al. 1999] Y. Amirat, D. Bresch, J. Lemoine, and J. Simon, "Existence of semi-periodic solutions of steady Navier–Stokes equations in a half space with an exponential decay at infinity", *Rend. Sem. Mat. Univ. Padova* **102** (1999), 341–365. MR 1739546 Zbl 0940.35158

[Barraza 1996] O. A. Barraza, "Self-similar solutions in weak $L^p$-spaces of the Navier–Stokes equations", *Rev. Mat. Iberoamericana* **12**:2 (1996), 411–439. MR 1402672 Zbl 0860.35092

[Cannone and Planchon 1996] M. Cannone and F. Planchon, "Self-similar solutions for Navier–Stokes equations in $\mathbf{R}^3$", *Comm. Partial Differential Equations* **21**:1-2 (1996), 179–193. MR 1373769 Zbl 0842.35075

[Cannone et al. 1994] M. Cannone, Y. Meyer, and F. Planchon, "Solutions auto-similaires des équations de Navier–Stokes", pp. VIII-1–10 in *Séminaire: Équations aux dérivées partielles, 1993–1994*, École Polytech., Palaiseau, 1994. MR 1300903 Zbl 0882.35090

[Giga and Miyakawa 1989] Y. Giga and T. Miyakawa, "Navier–Stokes flow in $\mathbf{R}^3$ with measures as initial vorticity and Morrey spaces", *Comm. Partial Differential Equations* **14**:5 (1989), 577–618. MR 993821 Zbl 0681.35072

[Jia and Šverák 2014] H. Jia and V. Šverák, "Local-in-space estimates near initial time for weak solutions of the Navier–Stokes equations and forward self-similar solutions", *Invent. Math.* **196**:1 (2014), 233–265. MR 3179576 Zbl 1301.35089

[Kang 2004] K. Kang, "On boundary regularity of the Navier–Stokes equations", *Comm. Partial Differential Equations* **29**:7-8 (2004), 955–987. MR 2097573 Zbl 1091.76012

[Kapitanskiĭ and Piletskas 1983] L. V. Kapitanskiĭ and K. I. Piletskas, "Spaces of solenoidal vector fields and boundary value problems for the Navier–Stokes equations in domains with noncompact boundaries", *Trudy Mat. Inst. Steklov.* **159** (1983), 5–36. In Russian; translated in *Proc. Math. Inst. Steklov* **159** (1984), 3–34. MR 720205 Zbl 0528.76029

[Kato 1992] T. Kato, "Strong solutions of the Navier–Stokes equation in Morrey spaces", *Bol. Soc. Brasil. Mat.* (*N.S.*) **22**:2 (1992), 127–155. MR 1179482 Zbl 0781.35052

[Koch and Tataru 2001] H. Koch and D. Tataru, "Well-posedness for the Navier–Stokes equations", *Adv. Math.* **157**:1 (2001), 22–35. MR 1808843 Zbl 0972.35084

[Korobkov et al. 2013] M. V. Korobkov, K. Pileckas, and R. Russo, "On the flux problem in the theory of steady Navier–Stokes equations with nonhomogeneous boundary conditions", *Arch. Ration. Mech. Anal.* **207**:1 (2013), 185–213. MR 3004771 Zbl 1260.35115

[Korobkov et al. 2014a] M. Korobkov, K. Pileckas, and R. Russo, "The existence of a solution with finite Dirichlet integral for the steady Navier–Stokes equations in a plane exterior symmetric domain", *J. Math. Pures Appl.* (9) **101**:3 (2014), 257–274. MR 3168911 Zbl 1331.35263

[Korobkov et al. 2014b] M. Korobkov, K. Pileckas, and R. Russo, "The existence theorem for the steady Navier—Stokes problem in exterior axially symmetric 3D domains", preprint, 2014. Submitted to *Math. Ann.* arXiv 1403.6921

[Korobkov et al. 2015a] M. Korobkov, K. Pileckas, and R. Russo, "An existence theorem for steady Navier–Stokes equations in the axially symmetric case", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **14**:1 (2015), 233–262. MR 3379492 Zbl 1321.35137

[Korobkov et al. 2015b] M. V. Korobkov, K. Pileckas, and R. Russo, "Solution of Leray's problem for stationary Navier–Stokes equations in plane and axially symmetric spatial domains", *Ann. of Math.* (2) **181**:2 (2015), 769–807. MR 3275850 Zbl 1318.35065

[Lemarié-Rieusset 2002] P. G. Lemarié-Rieusset, *Recent developments in the Navier–Stokes problem*, Chapman & Hall/CRC Research Notes in Mathematics **431**, Chapman & Hall/CRC, Boca Raton, FL, 2002. MR 1938147 Zbl 1034.35093

[Leray 1933] J. Leray, "Étude de diverses équations intégrales non linéaires et de quelques problèmes que pose l'hydrodynamique", *J. Math. Pures Appl.* **12** (1933), 1–82. MR 3533002 Zbl 0006.16702

[Leray 1934] J. Leray, "Sur le mouvement d'un liquide visqueux emplissant l'espace", *Acta Math.* **63**:1 (1934), 193–248. MR 1555394 Zbl 59.0763.02

[Nečas et al. 1996] J. Nečas, M. Ružička, and V. Šverák, "On Leray's self-similar solutions of the Navier–Stokes equations", *Acta Math.* **176**:2 (1996), 283–294. MR 1397564 Zbl 0884.35115

[Solonnikov 1964] V. A. Solonnikov, "Estimates for solutions of a non-stationary linearized system of Navier–Stokes equations", *Trudy Mat. Inst. Steklov.* **70** (1964), 213–317. In Russian; translated in *Amer. Math. Transl.* (2) **75** (1968), 6–121. MR 0171094

[Solonnikov 2003] V. A. Solonnikov, "Estimates for solutions of the nonstationary Stokes problem in anisotropic Sobolev spaces and estimates for the resolvent of the Stokes operator", *Uspekhi Mat. Nauk* **58**:2(350) (2003), 123–156. In Russian; translated in *Russian Math. Surveys* **58** (2003), 331–365. MR 1992567 Zbl 1059.35101

[Tsai 1998] T.-P. Tsai, "On Leray's self-similar solutions of the Navier–Stokes equations satisfying local energy estimates", *Arch. Rational Mech. Anal.* **143**:1 (1998), 29–51. MR 1643650 Zbl 0916.35084

[Tsai 2014] T.-P. Tsai, "Forward discretely self-similar solutions of the Navier–Stokes equations", *Comm. Math. Phys.* **328**:1 (2014), 29–44. MR 3196979 Zbl 1293.35218

[Yamazaki 2000] M. Yamazaki, "The Navier–Stokes equations in the weak-$L^n$ space with time-dependent external force", *Math. Ann.* **317**:4 (2000), 635–675. MR 1777114 Zbl 0965.35118

Received 16 Aug 2015. Revised 20 Jun 2016. Accepted 13 Sep 2016.

MIKHAIL KOROBKOV: korob@math.nsc.ru
*Sobolev Institute of Mathematics, 4 Acad. Koptyug Avenue, 630090 Novosibirsk, Russia*

and

*Novosibirsk State University, 2 Pirogova Str., 630090 Novosibirsk, Russia*

TAI-PENG TSAI: ttsai@math.ubc.ca
*Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada*

and

*Center for Advanced Study in Theoretical Sciences, National Taiwan University, Taipei, Taiwan*

ANALYSIS AND PDE
Vol. 9, No. 8, 2016
dx.doi.org/10.2140/apde.2016.9.1829

# DECAY OF SOLUTIONS OF MAXWELL–KLEIN–GORDON EQUATIONS WITH ARBITRARY MAXWELL FIELD

SHIWU YANG

In the author's previous work, it has been shown that solutions of Maxwell–Klein–Gordon equations in $\mathbb{R}^{3+1}$ possess some form of global strong decay properties with data bounded in some weighted energy space. In this paper, we prove pointwise decay estimates for the solutions for the case when the initial data are merely small on the scalar field but can be arbitrarily large on the Maxwell field. This extends the previous result of Lindblad and Sterbenz, in which smallness was assumed both for the scalar field and the Maxwell field.

## 1. Introduction

In this paper, we study the pointwise decay of solutions to the Maxwell–Klein–Gordon equations on $\mathbb{R}^{3+1}$ with large Cauchy data. To define the equations, let $A = A_\mu\, dx^\mu$ be a 1-form. The covariant derivative associated to this 1-form is

$$D_\mu = \partial_\mu + \sqrt{-1}\, A_\mu,$$

which can be viewed as a $U(1)$ connection on the complex line bundle over $\mathbb{R}^{3+1}$ with the standard flat metric $m_{\mu\nu}$. Then the curvature 2-form $F$ associated to this connection is given by

$$F_{\mu\nu} = -\sqrt{-1}[D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu = (dA)_{\mu\nu}.$$

This is a closed 2-form, that is, $F$ satisfies the Bianchi identity

$$\partial_\gamma F_{\mu\nu} + \partial_\mu F_{\nu\gamma} + \partial_\nu F_{\gamma\mu} = 0. \tag{1}$$

The Maxwell–Klein–Gordon equations (MKG) is a system for the connection field $A$ and the complex scalar field $\phi$:

$$\begin{cases} \partial^\nu F_{\mu\nu} = \Im(\phi \cdot \overline{D_\mu \phi}) = J_\mu, \\ D^\mu D_\mu \phi = \Box_A \phi = 0. \end{cases} \tag{MKG}$$

These are Euler–Lagrange equations of the functional

$$L[A, \phi] = \int_{\mathbb{R}^{3+1}} \left(\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} + \tfrac{1}{2} D_\mu \phi \overline{D^\mu \phi}\right) dx\, dt.$$

A basic feature of this system is that it is gauge invariant under the gauge transformation

$$\phi \mapsto e^{i\chi}\phi, \quad A \mapsto A - d\chi.$$

*MSC2010:* 35Q61.
*Keywords:* Maxwell–Klein–Gordon, decay.

More precisely, if $(A, \phi)$ solves (MKG), then $(A - d\chi, e^{i\chi}\phi)$ is also a solution for any potential function $\chi$. Note that $U(1)$ is abelian. The Maxwell field $F$ is invariant under the above gauge transformation, and (MKG) is said to be an *abelian gauge theory*. For the more general theory when $U(1)$ is replaced by a compact Lie group, the corresponding equations are referred to as *Yang–Mills–Higgs equations*.

In this paper, we consider the Cauchy problem to (MKG). The initial data set $(E, H, \phi_0, \phi_1)$ consists of the initial electric field $E$ and the magnetic field $H$, together with initial data $(\phi_0, \phi_1)$ for the scalar field. In terms of the solution $(F, \phi)$, on the initial hypersurface, these are

$$F_{0i} = E_i, \quad {}^*F_{0i} = H_i, \quad \phi(0, x) = \phi_0, \quad D_t\phi(0, x) = \phi_1,$$

where ${}^*F$ is the Hodge dual of the 2-form $F$. In local coordinates $(t, x)$,

$$(H_1, H_2, H_3) = (F_{23}, F_{31}, F_{12}).$$

The data set is said to be *admissible* if it satisfies the compatibility condition

$$\operatorname{div}(E) = \Im(\phi_0 \cdot \overline{\phi_1}) = J_0|_{t=0}, \quad \operatorname{div}(H) = 0, \tag{2}$$

where the divergence is taken on the initial hypersurface $\mathbb{R}^3$. For solutions of (MKG), the energy

$$E[F, \phi](t) := \int_{\mathbb{R}^3} |E|^2 + |H|^2 + |D\phi|^2 \, dx$$

is conserved. Another important conserved quantity is the total charge

$$q_0 = \frac{1}{4\pi} \int_{\mathbb{R}^3} \Im(\phi \cdot \overline{D_t\phi}) \, dx = \frac{1}{4\pi} \int_{\mathbb{R}^3} \operatorname{div}(E) \, dx, \tag{3}$$

which can be defined at any fixed time $t$. The existence of nonzero charge plays a crucial role in the asymptotic behavior of solutions of (MKG). It makes the analysis more complicated and subtle. This is obvious from the above definition as the electric field $E_i = F_{0i}$ has a tail $q_0 r^{-3} x_i$ at any fixed time $t$.

The Cauchy problem to (MKG) has been studied extensively. One of the most remarkable results is due to Eardley and Moncrief [1982a; 1982b], in which it was shown that there is always a global solution to the general Yang–Mills–Higgs equations for sufficiently smooth initial data. This was later improved to data merely bounded in the energy space for MKG by Klainerman and Machedon [1994] and for the nonabelian case of Yang–Mills equations in, e.g., [Klainerman and Machedon 1995; Oh 2015; Selberg and Tesfahun 2010]. Since then there has been extensive literature on generalizations and extensions of this classical result, aiming at improving the regularity of the initial data in order to construct a global solution; see [Krieger et al. 2015; Keel et al. 2011; Krieger and Lührmann 2015; Machedon and Sterbenz 2004; Oh and Tataru 2016; Rodnianski and Tao 2004] and references therein. A common feature of all these works is to construct a local solution with rough data. Then the global well-posedness follows by establishing a priori bounds for some appropriate norms of the solution. For example, a local solution was constructed in [Eardley and Moncrief 1982a], while in [Eardley and Moncrief 1982b], they showed that the $L^\infty$ norm of the solution never blows up even though it may grow in time $t$. As a consequence, the solution can be extended to all time; however, the decay property of the solution is unknown. In view

of this, although the solution of (MKG) exists globally with rough initial data, very little is known about the decay properties.

Asymptotic behavior and decay estimates are well understood for linear fields (see, e.g., [Christodoulou and Klainerman 1990]) and nonlinear fields with sufficiently small initial data (see, e.g., [Choquet-Bruhat and Christodoulou 1981b; Shu 1991]). These results rely on the conformal symmetry of the system, either by conformally compactifying the Minkowski space or by using the conformal Killing vector field $(t^2+r^2)\partial_t + 2tr\partial_r$ as multiplier. Nevertheless the use of the conformal symmetry requires strong decay of the initial data, and thus in general does not allow the presence of nonzero charge except when the initial data are essentially compactly supported. For the case with nonzero charge, the first related work regarding the asymptotic properties was due to W. Shu [1992]. However, that work only considered the case when the solution is trivial outside a fixed forward light cone. Details for the general case were not carried out. A complete proof towards this program was later contributed by Lindblad and Sterbenz [2006]; also see the more recent work [Bieri et al. 2014].

The presence of nonzero charge has a long range effect on the asymptotic behavior of the solutions, at least in a neighborhood of the spatial infinity. This can be seen from the conservation law of the total charge as the electric field $E$ decays at most $r^{-2}$ as $r \to \infty$ at any fixed time. This weak decay rate makes the analysis more complicated even for small initial data. To deal with this difficulty, Lindblad–Sterbenz decomposed the Maxwell field into charged and chargeless components (see discussions in the end of this section) and made use of the fractional Morawetz estimates obtained by using the vector fields $u^p\partial_u + v^p\partial_v$ as multipliers. The latter work [Bieri et al. 2014] relied on the observation that the angular derivative of the Maxwell field has zero charge. The Maxwell field then can be estimated by using the Poincaré inequality.

The asymptotic behavior of solutions of MKG with general large data remains unknown until recently in [Yang 2015c] quantitative decay estimates were obtained for solutions with data bounded in some weighted energy space. Pointwise decay requires the energy estimates for the derivatives of the solution. However, commuting the equations with derivatives generates nonlinear terms. The aim of this paper is to identify a class of large data for MKG equations such that we can derive the pointwise decay of the solutions.

We define some necessary notations in order to state our main result. We use the standard polar local coordinate system $(t, r, \omega)$ of Minkowski space as well as the null coordinates $u = \frac{1}{2}(t-r)$, $v = \frac{1}{2}(t+r)$. Let $\overline{\nabla}$ denote the covariant derivative on $\mathbb{R}^3$ and $\Omega$ be the set of angular momentum vector fields $\Omega_{ij} = x_i\partial_j - x_j\partial_i$. Without loss of generality we only prove estimates in the future, i.e., $t \geq 0$. Next we introduce a null frame $\{L, \underline{L}, e_1, e_2\}$, where

$$L = \partial_v = \partial_t + \partial_r, \quad \underline{L} = \partial_u = \partial_t - \partial_r$$

and $\{e_1, e_2\}$ is an orthonormal basis of the sphere with constant radius $r$. We use $\not{D}$ to denote the covariant derivative associated to the connection field $A$ on the sphere with radius $r$. For any 2-form $F$, denote the null decomposition under the above null frame by

$$\alpha_i = F_{Le_i}, \quad \underline{\alpha}_i = F_{\underline{L}e_i}, \quad \rho = \tfrac{1}{2}F_{\underline{L}L}, \quad \sigma = F_{e_1e_2}, \quad i \in \{1,2\}. \tag{4}$$

We assume that the initial data set $(E, H, \phi_0, \phi_1)$ is admissible. Let $q_0$ be the charge defined in (3) which is uniquely determined by the initial data of the scalar field $(\phi_0, \phi_1)$. We assume that the data for the scalar field is small, but the data for the Maxwell field is arbitrary. However the data cannot be assigned freely. They satisfy the compatibility condition (2). To measure the size of the initial data for the scalar field and the Maxwell field, let $(E^{\mathrm{df}}, E^{\mathrm{cf}})$ be the Hodge decomposition of the electronic field $E$ with $E^{\mathrm{df}}$ the divergence-free part and $E^{\mathrm{cf}}$ the curl-free part. Then the compatibility condition (2) on $E$ is equivalent to

$$\operatorname{div} E^{\mathrm{cf}} = \Im(\phi_0 \cdot \overline{\phi_1}).$$

This implies that $E^{\mathrm{cf}}$ can be uniquely determined by $(\phi_0, \phi_1)$ (with a suitable decay assumption on $E$). Therefore, for the initial data set $(E, H, \phi_0, \phi_1)$ for (MKG) we can freely assign $\phi_0$, $\phi_1$ and $E^{\mathrm{df}}$, $H$ as long as $\operatorname{div} H = 0$, $\operatorname{div} E^{\mathrm{df}} = 0$. The total charge $q_0$ is a constant determined by $(\phi_0, \phi_1)$.

We now define the norms of the initial data. For some positive constant $0 < \gamma_0 < 1$, we define the second-order weighted Sobolev norm respectively for the initial data of the Maxwell field $(E, H)$ and the initial data of the scalar field $(\phi_0, \phi)$:

$$\mathcal{M} := \sum_{l\le 2} \int_{\mathbb{R}^3} (1+r)^{1+\gamma_0} \big( |\Omega^l E^{\mathrm{df}}|^2 + |\Omega^l H|^2 + |\overline{\nabla}^l E^{\mathrm{df}}|^2 + |\overline{\nabla}^l H|^2 \big)\, dx,$$

$$\mathcal{E} := \sum_{l\le 2} \int_{\mathbb{R}^3} (1+r)^{1+\gamma_0} \big( |\overline{\nabla}\Omega^l \phi_0|^2 + |\Omega^l \phi_1|^2 + |\overline{\nabla}^{l+1}\phi_0|^2 + |\overline{\nabla}^l \phi_1|^2 + |\phi_0|^2 \big)\, dx.$$

We remark here that the definition for $\mathcal{E}$ is not gauge invariant. The gauge invariant norm depends on the connection field $A$, which up to a gauge transformation can be determined by the initial data of the Maxwell field $(E^{\mathrm{df}}, H)$. However, in our setting $\mathcal{M}$ is arbitrarily large while $\mathcal{E}$ is assumed to be small depending on $\mathcal{M}$. To measure the smallness of the scalar field, we choose the above gauge dependent norm for the scalar field. We will show later (see Lemma 58 in Section 5) that the gauge invariant norm is in fact equivalent to the above Sobolev norm up to a constant depending only on $\mathcal{M}$.

We now can state our main theorem:

**Theorem 1.** *Consider the Cauchy problem to (MKG) with admissible initial data set $(E, H, \phi_0, \phi_1)$. There exists a positive constant $\epsilon_0$, depending on $\mathcal{M}$ and $\gamma_0$, such that for all $\mathcal{E} < \epsilon_0$, the solution $(F, \phi)$ of (MKG) satisfies the following decay estimates*:

$$\begin{aligned}
|D_{\underline{L}}(r\phi)|^2(u,v,\omega) &\le C\mathcal{E}(1+|u|)^{-1-\gamma_0}, && |r\underline{\alpha}|^2(u,v,\omega) \le C(1+|u|)^{-1-\gamma_0};\\
r^p(|D_L(r\phi)|^2 + |\not{D}(r\phi)|^2)(u,v,\omega) &\le C\mathcal{E}(1+|u|)^{p-1-\gamma_0}, && 0 \le p \le 1+\gamma_0;\\
r^p(|r\alpha|^2 + |r\sigma|^2)(u,v,\omega) &\le C(1+|u|)^{p-1-\gamma_0}, && 0 \le p \le 1+\gamma_0;\\
r^{p+2}|\rho - q_0 r^{-2}\chi_{\{t+R\le r\}}|^2(u,v,\omega) &\le C(1+|u|)^{p-1-\gamma_0}, && 0 \le p < 1;\\
r^p|\phi|^2(u,v,\omega) &\le C\mathcal{E}(1+|u|)^{p-2-\gamma_0}, && 1 \le p \le 2;\\
|D\phi|^2(t,x) + |\phi|^2(t,x) &\le C\mathcal{E}(1+t)^{-1-\gamma_0}, && |F|^2(t,x) \le C(1+t)^{-1-\gamma_0},\ \forall |x| \le R;
\end{aligned}$$

*for all* $(u, v, \omega) \in \mathbb{R}^{3+1} \cap \{|x| \geq R\}$ *and for some constant* $C$ *depending on* $\mathcal{M}$, $\gamma_0$, $p$. *Here* $q_0$ *is the total charge and* $\chi_{\{t+2\leq r\}}$ *is the characteristic function on the exterior region* $\{t+2 \leq r\}$.

We make several remarks.

**Remark 2.** The second-order derivatives of the initial data are the minimum regularity we need to derive the above pointwise decay of the solution. Similar decay estimates hold for the higher-order derivatives of the solution if higher-order weighted Sobolev norms of the initial data are known.

**Remark 3.** The restriction $0 < \gamma_0 < 1$ on $\gamma_0$ is merely for the sake of brevity. If $\gamma_0 \geq 1$, then the decay property of the solutions propagates in the exterior region $(t+2 \leq r)$. In other words, we have the same decay estimates as in the theorem for $\tau \leq 0$. However in the interior region where $\tau > 0$, the maximal decay rate is $\tau_+^{-2}$ (corresponding to $\gamma_0 = 1$), that is, the decay rate in the interior region for $\gamma_0 \geq 1$ in general cannot be better than that of $\gamma_0 = 1$.

Compared to the previous result of Lindblad and Sterbenz [2006], we have made the following improvements: first of all, we obtain pointwise decay estimates for solutions of (MKG) for a class of large initial data. We only require smallness on the scalar field. In particular our initial data for (MKG) can be arbitrarily large. Combining the method in [Yang 2015a], we can even make the data on the scalar field large in the energy space. Secondly, we have lower regularity on the initial data. In [Lindblad and Sterbenz 2006], it was assumed that the derivative of the initial data decays one order better, that is, $\overline{\nabla}^k(E^{\mathrm{df}}, H)$, $D^k(D\phi_0, \phi_1)$ belong to the weighted Sobolev space with weights $(1+r)^{1+\gamma_0+2|k|}$, while in this paper we only assume that the angular derivatives of the data obey this improved decay (see the definition of $\mathcal{M}$, $\mathcal{E}$). For the other derivatives, the weight is merely $(1+r)^{1+\gamma_0}$. This makes the analysis more delicate. Moreover, as the solution decays weaker initially, our decay rate is weaker than that in [Lindblad and Sterbenz 2006] (only decay rate in $u$, the decay in $r$ is the same). However if we assume the same decay of the initial data as in [Lindblad and Sterbenz 2006], then we are able to obtain the same decay for the solution.

We use a new approach developed in [Yang 2015c] to study the asymptotic behavior of solutions of (MKG). This new method was originally introduced by Dafermos and Rodnianski [2010] for the study of decay of linear waves on black hole spacetimes. This novel method starts by proving the energy flux decay of the solutions of linear equations through the forward light cone $\Sigma_\tau$ (see definitions in Section 2). The pointwise decay then follows by commuting the equation with $\partial_t$ and the angular momentum $\Omega$. In the abstract framework set by Dafermos and Rodnianski [2010], the energy flux decay relies on three kinds of basic ingredients and estimates: a uniform energy bound, an integrated local energy decay estimate and a hierarchy of $r$-weighted energy estimates in a neighborhood of the null infinity, which can be obtained by using the vector fields $\partial_t$, $f(r)\partial_r$, $r^p(\partial_t + \partial_r)$ as multipliers, respectively. Combining these three estimates, a pigeonhole argument then leads to the energy flux decay.

As the initial data for the scalar field is small, we can use the perturbation method to prove the pointwise decay of the solution. With a suitable bootstrap assumption on the nonlinearity $J[\phi] = \Im(\phi \cdot \overline{D\phi})$, we first can use the new method to prove energy decay estimates for the Maxwell field up to the second-order derivatives. Once we have these decay estimates for the Maxwell field, we then can show the energy

decay as well as pointwise decay for the scalar field, which can then be used to improve the bootstrap assumption. The smallness of the scalar field is used here to close the bootstrap assumptions.

The existence of nonzero charge has a long-range effect on the asymptotic behavior of the solution in the exterior region $\{t+2\le r\}$, which has been discussed in [Yang 2015c] when the charge is large. To deal with this difficulty, we define the chargeless 2-form

$$\overline{F} = F - \chi_{\{t+2\le r\}} q_0 r^{-2}\, dt\wedge dr.$$

We first carry out estimates for $\overline{F}$ on the exterior region $\{t+2\le r\}$, which in particular controls the energy flux through $\{t+2=r\}$ (the intersection of the interior region and the exterior region). We then can use the new method to obtain estimates for the Maxwell field $F$ in the interior region. The Maxwell equation commutes with the Lie derivatives of $F$ (see Lemma 4). It is not hard to obtain energy decay estimates for the derivatives of the Maxwell field under suitable bootstrap assumptions on the nonlinearity $J[\phi]$ by using the new approach.

The main difficulty lies in showing the energy decay estimates for the scalar field due to the fact that the covariant derivative $D$ does not commute with the covariant wave operator $\Box_A$. The interaction terms of the Maxwell field and the scalar field arise from the commutator. To control those interaction terms, previous results [Bieri et al. 2014; Lindblad and Sterbenz 2006] rely on the smallness of the Maxwell field, and those terms could be absorbed. The key observation allowing the Maxwell field to be large in this paper is that the robust new method makes use of the decay in $u$ (equivalent to $\tau$ up to a constant) and those terms could be controlled using Gronwall's inequality without any smallness assumption on the Maxwell field. Traditionally, Gronwall's inequality is used with respect to the foliation $t=$ constant. Therefore strong decay in $t$ is necessary. As the new method foliates the spacetime by using the null hypersurfaces $H_u$, it enables us to make use of the weaker decay in $u$ in order to apply Gronwall's inequality.

The paper is organized as follows. We define additional notations and derive the transport equations for the curvature components of the Maxwell field in Section 2. Since we only commute the equations with $\partial_t$ or the angular momentum $\Omega$, these transport equations will be used to recover the missing derivative in order to derive pointwise estimates for the Maxwell field. Section 3 is devoted to reviewing the energy estimates (an integrated local energy estimate and a hierarchy of $r$-weighted energy estimates) both for the scalar field verifying the linear covariant wave equation $\Box_A\phi=0$ and the linear Maxwell field. The idea to prove these estimates is very similar to that in the author's other preprint [Yang 2015c], in which decay properties of solutions of MKG are discussed with data merely bounded in some weighted energy space. There the energy estimates are carried out for the full solution $(A,\phi)$ of the nonlinear MKG equations, and one of the difficulties is to deal with the arbitrarily large charge $q_0$. This paper aims at the pointwise decay of the solutions with some special initial data. In particular, energy decay estimates are also necessary for the derivatives of the solutions. We thus need energy estimates for the linearized equations. To make this paper self-contained, we give detailed proof for these energy estimates in Section 3. In Section 4, we use the new method to obtain decay estimates for the linear Maxwell field and the linear scalar field. More specifically, in Section 4.1, we derive energy flux decay estimates for

the linear Maxwell fields under suitable assumptions on the inhomogeneous term $J_\mu = \nabla^\nu F_{\mu\nu}$. Then in Section 4.2, we obtain pointwise decay estimates by commuting the equation with vector fields in $\Gamma = \{\partial_t, \Omega\}$ merely twice. This lower regularity result relies on the elliptic estimates in the bounded region $\{r \le R\}$ and the transport equations for the curvature components when $r \ge R$. The most technical part of this paper lies in Section 4.3, in which energy decay estimates are obtained for the scalar field up to second-order derivatives. The main difficulty is that the covariant wave operator $\Box_A$ does not commute with the covariant derivative $D$. It heavily relies on the null structure of the commutators. Finally, in Section 5 we improve the bootstrap assumption and conclude our main theorem.

## 2. Preliminaries and notations

We define some additional notation used in the sequel. Recall the null frame $\{L, \underline{L}, e_1, e_2\}$ defined in the introduction. At any fixed point $(t, x)$, we may choose $e_1$, $e_2$ such that

$$[L, e_i] = -\frac{1}{r} e_i, \quad [\underline{L}, e_i] = \frac{1}{r} e_i, \quad [e_1, e_2] = 0, \quad i \in \{1, 2\}.$$

This helps to compute those geometric quantities which are independent of the choice of the local coordinates. We then can compute the covariant derivatives for the null frame at any fixed point:

$$\begin{gathered} \nabla_L L = 0, \quad \nabla_L \underline{L} = 0, \quad \nabla_L e_i = 0, \quad \nabla_{\underline{L}} \underline{L} = 0, \quad \nabla_{\underline{L}} e_i = 0, \\ \nabla_{e_i} L = r^{-1} e_i, \quad \nabla_{e_i} \underline{L} = -r^{-1} e_i, \quad \nabla_{e_1} e_2 = \nabla_{e_2} e_1 = 0, \quad \nabla_{e_i} e_i = -r^{-1} \partial_r. \end{gathered} \tag{5}$$

Here $\nabla$ is the covariant derivatives in Minkowski space and $\overline{\nabla}$ is the spatial component. We also use $\partial$ to abbreviate the partial derivatives $(\partial_t, \partial_1, \partial_2, \partial_3)$ in Minkowski space under the coordinates $(t, x)$ and $\not\nabla$ to denote the covariant derivative on the sphere with radius $r$.

Now we define the foliation of the spacetime $\{t \ge 0\}$. Let $H_u$ be the outgoing null hypersurface $\{t - r = 2u\}$ and $\underline{H}_v$ be the incoming null hypersurface $\{t + r = 2v\}$. Let $R > 1$ be a fixed constant. We now use this fixed constant $R$ to define the foliation. For all $\tau \in \mathbb{R}$, denote

$$\tau^* = \frac{\tau - R}{2}.$$

In the exterior region where $t + R \le r$, we use the foliation

$$\Sigma_\tau := H_{\tau^*} \cap \{t \ge 0\}, \quad \tau \le 0,$$

while in the interior region where $t + R \ge r$, the foliation is defined as

$$\Sigma_\tau := \{t = \tau, |x| \le R\} \cup (H_{\tau^*} \cap \{|x| \ge R\}).$$

Unless we specify it, in the following the outgoing null hypersurface $H_u$ stands for $H_u \cap \{t \ge 0\}$ in the exterior region and $H_u \cap \{|x| \ge R\}$ in the interior region. Note that the boundary of the region bounded by $\Sigma_{\tau_1}$ and $\Sigma_{\tau_2}$ is part of the future null infinity where the decay behavior of the solution is unknown. To make the energy estimates rigorous, we instead consider the finite truncated hypersurfaces

$$\Sigma_\tau^{v_0} := \Sigma_\tau \cap \{v \le v_0\}, \quad H_u^{v_0} := H_u \cap \{v \le v_0\}, \quad \underline{H}_v^{u_1, u_2} := \underline{H}_v \cap \{u_1 \le u \le u_2\}.$$

On the initial hypersurface $\{t = 0\}$, we denote the annulus with radii $r_1 < r_2$ by

$$B_{r_1}^{r_2} = \{r_1 \le |x| \le r_2\}, \quad B_r = B_r^{\infty}.$$

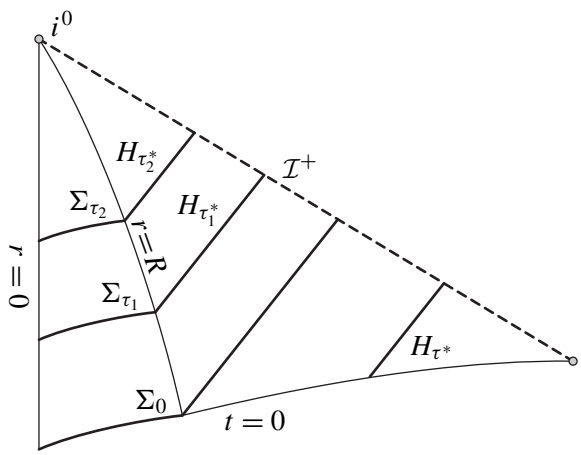
Next we define the domains. In the exterior region, for $\tau_2 \le \tau_1 \le 0$, define $\mathcal{D}_{\tau_1}^{\tau_2}$ to be the Cauchy development of the annulus $\{R - \tau_1 \le |x| \le R - \tau_2\}$, or more precisely

$$\mathcal{D}_{\tau_1}^{\tau_2} = \left\{(t, x) \mid \left||x| + \tau_1^* + \tau_2^*\right| + t \le \tau_1^* - \tau_2^*\right\}, \quad \tau_2 < \tau_1 \le 0,$$

while in the interior region, for any $\tau_2 > \tau_1 \ge 0$, we define $\mathcal{D}_{\tau_1}^{\tau_2}$ to be the region

$$\mathcal{D}_{\tau_1}^{\tau_2} = \{(t, x) \mid (t, x) \in \Sigma_\tau,\ 0 \le \tau_1 \le \tau \le \tau_2\}.$$

bounded by $\Sigma_{\tau_1}$ and $\Sigma_{\tau_2}$. We may also use $\overline{\mathcal{D}}_{\tau_1}^{\tau_2} = \mathcal{D}_{\tau_1}^{\tau_2} \cap \{|x| \ge R\}$ to denote the region outside the cylinder $\{r \le R\}$. Finally, we write $\mathcal{D}_\tau$ for the region $\mathcal{D}_\tau^{+\infty}$ if $\tau \ge 0$ or the region $\mathcal{D}_\tau^{-\infty}$ when $\tau < 0$. The following Penrose diagram may be of help for the various pieces of notation described above.




We use $E[\phi](\Sigma)$ to denote the energy flux of the complex scalar field $\phi$ and $E[F](\Sigma)$ for the energy flux of the 2-form $F$ through the hypersurface $\Sigma$ in Minkowski space. The derivative on the scalar field is with respect to the covariant derivative $D$. For our interested hypersurfaces, we can compute

$$E[\phi](\Sigma_\tau) = \int_{\{t=\tau, r\le R\}} |D\phi|^2\, dx + \int_{H_{\tau^*}} (|D_L\phi|^2 + |\not{D}\phi|^2) r^2\, dv\, d\omega,$$

$$E[F](\Sigma_\tau) = \int_{\{t=\tau, r\le R\}} \rho^2 + |\sigma|^2 + \frac{1}{2}(|\alpha|^2 + |\underline{\alpha}|^2)\, dx + \int_{H_{\tau^*}} (\rho^2 + |\sigma|^2 + |\alpha|^2) r^2\, dv\, d\omega,$$

$$E[\phi](\underline{H}_u) = \int_{\underline{H}_u} (|D_{\underline{L}}\phi|^2 + |\not{D}\phi|^2) r^2\, dv\, d\omega, \quad E[F](\underline{H}_u) = \int_{\underline{H}_u} (\rho^2 + |\sigma|^2 + |\underline{\alpha}|^2) r^2\, dv\, d\omega.$$

Here $\rho, \sigma, \alpha, \underline{\alpha}$ are the null components of the 2-form $F$ defined in line (4), and we recall that $\tau^* = \frac{1}{2}(\tau - R)$. Since we only consider estimates in the future when $t \ge 0$, the set $\{t = \tau, r \le R\}$ should be interpreted as the empty set when $\tau < 0$.

Next we define some useful weighted Sobolev norms either on domains or on surfaces. For any function $f$ (scalar or vector valued or tensors) we denote the spacetime integral on $\mathcal{D}$ in Minkowski space

$$I_q^p[f](\mathcal{D}) := \int_{\mathcal{D}} u_+^q r_+^p |f|^2, \quad r_+ = 1 + r, \quad u_+ = 1 + |u|$$

for any real numbers $p, q$. Here $\mathcal{D}$ can be the domain or hypersurface in the Minkowski space. For example, when $\mathcal{D}$ is $H_u$, then

$$I_q^p[f](H_u) := \int_{H_u} r_+^p u_+^q |f|^2 r^2 \, dv \, d\omega.$$

To define the norms of the derivatives of the solution, we need vector fields used as commutators which, in this paper, are the Killing vector field $\partial_t$ together with the angular momentum $\Omega$ with components $\Omega_{ij} = x_i \partial_j - x_j \partial_i$. We define the set

$$\Gamma = \{\partial_t, \Omega_{ij}\}.$$

For the scalar field, it is natural to take the covariant derivative $D_X = X^\mu D_\mu$ associated to the connection $A$ for any vector field $X = X^\mu \partial_\mu$. This covariant derivative has already been defined for the purpose of defining the equations in the beginning of the introduction. For the Maxwell field $F$, which is a 2-form, we define the Lie derivative

$$(\mathcal{L}_Z F)_{\mu\nu} = Z(F_{\mu\nu}) - F(\mathcal{L}_Z \nabla_\mu, \nabla_\nu) - F(\nabla_\mu, \mathcal{L}_Z \nabla_\nu), \quad (\mathcal{L}_Z J)_\mu = Z(J_\mu) - J(\mathcal{L}_Z \nabla_\mu)$$

for any 2-form $F$ and any 1-form $J$, respectively. Here $\mathcal{L}_Z X = [Z, X]$ for all vector fields $Z$, $X$.

If the vector field $Z$ is Killing, that is, $\nabla^\mu Z^\nu + \nabla^\nu Z^\mu = 0$ for all $\mu$, $\nu$, then we can show that

$$\begin{aligned} \nabla^\mu (\mathcal{L}_Z F)_{\mu\nu} &= Z(\nabla^\mu F_{\mu\nu}) + \nabla^\mu Z^\gamma \nabla_\gamma F_{\mu\nu} + \nabla_\mu Z^\gamma \nabla^\mu F_{\gamma\nu} + \nabla_\nu Z^\gamma \nabla^\mu F_{\mu\gamma} \\ &= Z(\nabla^\mu F_{\mu\nu}) + \nabla_\nu Z^\gamma \nabla^\mu F_{\mu\gamma} = (\mathcal{L}_Z \delta F)_\nu. \end{aligned}$$

Here we denote $\delta F_\nu = \nabla^\mu F_{\mu\nu}$ as the divergence of the 2-form $F$. We use $\mathcal{L}_Z^k$ or $D_Z^k$ to denote the $k$-th derivatives, that is,

$$\mathcal{L}_Z^k = \mathcal{L}_{Z^1} \mathcal{L}_{Z^2} \cdots \mathcal{L}_{Z^k}.$$

Similarly for $D_Z^k$. The vector fields $Z^j$ are any vector fields in the set $\Gamma = \{\partial_t, \Omega_{ij}\}$.

Based on these calculations, we have the following commutator lemma.

**Lemma 4.** *For any Killing vector field $Z$, we have*

$$\begin{aligned} [\Box_A, D_Z]\phi &= 2i Z^\nu F_{\mu\nu} D^\mu \phi + i \nabla^\mu (Z^\nu F_{\mu\nu}) \phi, \\ \nabla^\mu (\mathcal{L}_Z G)_{\mu\nu} &= (\mathcal{L}_Z \delta G)_\nu \end{aligned}$$

*for any complex scalar field $\phi$ and any 2-form $G$.*

For the energy estimates of the solutions of (MKG), the initial energies $\mathcal{M}$, $\mathcal{E}$ defined in the introduction cannot be used directly as $\mathcal{E}$ is not gauge invariant. Note that the vector fields used as commutators

are $\Gamma = \{\partial_t, \Omega\}$. For any 2-form $F$ satisfying the Bianchi identity and any scalar field $\phi$, for the given connection field $A$, we define the weighted $k$-th order initial energies

$$\begin{aligned} E_0^k[F] &:= \sum_{l\le k} \int_{\mathbb{R}^3} r_+^{1+\gamma_0} |\mathcal{L}_Z^l F|^2(0,x)\,dx, \\ E_0^k[\phi] &:= \sum_{l\le k, j\le 3} \int_{\mathbb{R}^3} r_+^{1+\gamma_0} |D_Z^l D_j \phi|^2(0,x)\,dx. \end{aligned} \tag{6}$$

Here $D_j$ denotes the spatial covariant derivative and $0 < \gamma_0 < 1$ is the constant in the main theorem. We remark here that $F$ may not be the full Maxwell field of the solution of (MKG). In application, it can be the chargeless part of the full solution which also satisfies the Bianchi identity. However, the connection field $A$ is associated to the full Maxwell field. In fact the full Maxwell field does not belong to this weighted Sobolev space due to the existence of nonzero charge.

We end this section by writing the Maxwell equation under the null frame $\{L, \underline{L}, e_1, e_2\}$. In other words, we derive the transport equations for the curvature components. Let $F_{\mu\nu}$ be the 2-form verifying the Bianchi identity. Let $J = \delta F$, that is, $J_\mu = \nabla^\nu F_{\mu\nu}$.

**Lemma 5.** *Under the null frame $\{L, \underline{L}, e_1, e_2\}$, the MKG equations are the following transport equations for the curvature components*:

$$\underline{L}(r^2\rho) - \not{\mathrm{div}}(r^2\underline{\alpha}) = r^2 J_{\underline{L}}, \quad L(r^2\rho) + \not{\mathrm{div}}(r^2\alpha) = r^2 J_L, \tag{7}$$

$$\nabla_L(r\underline{\alpha}_i) - r\not{\nabla}_{e_i}\rho - r\not{\nabla}_{e_j}F_{e_ie_j} = rJ_{e_i}, \quad i = 1,2, \tag{8}$$

$$\underline{L}(r^2\sigma) = r^2(e_2\underline{\alpha}_1 - e_1\underline{\alpha}_2), \quad L(r^2\sigma) = r^2(e_2\alpha_1 - e_1\alpha_2), \tag{9}$$

$$\nabla_{\underline{L}}(r\alpha_i) + r\not{\nabla}_{e_i}\rho - r\not{\nabla}_{e_j}F_{e_ie_j} = rJ_{e_i}, \quad i = 1,2. \tag{10}$$

*Here* $\not{\mathrm{div}}$ *is the divergence operator on the sphere with radius $r$.*

*Proof.* From the Maxwell equation, $J_L = (\delta F)(L)$. Use the formula

$$(\nabla_X F)(Y,Z) = XF(Y,Z) - F(\nabla_X Y, Z) - F(Y, \nabla_X Z)$$

for all vector fields $X$, $Y$, $Z$. By using (5), we then can compute

$$\begin{aligned} -(\delta F)(\underline{L}) &= -\tfrac{1}{2}(\nabla_L F)(\underline{L},\underline{L}) - \tfrac{1}{2}(\nabla_{\underline{L}} F)(L,\underline{L}) + (\nabla_{e_i}F)(e_i,\underline{L}) \\ &= \underline{L}\rho - e_i\underline{\alpha}_i - F(-2r^{-1}\partial_r, \underline{L}) - F(e_i, -r^{-1}e_i) \\ &= \underline{L}\rho - 2r^{-1}\rho - \not{\mathrm{div}}(\underline{\alpha}). \end{aligned}$$

Multiply both sides by $r^2$. We then get the first equation of (7). The second equation follows similarly.

For (8) and (10), we need to use the Bianchi identity (1) which is equivalent to

$$(\nabla_X F)(Y,Z) + (\nabla_Y F)(Z,X) + (\nabla_Z F)(X,Y) = 0$$

for all vector fields $X, Y, Z$. Let's only prove (8). We can show that

$$\begin{aligned}-(\delta F)(e_i) &= -\tfrac{1}{2}(\nabla_L F)(\underline{L}, e_i) - \tfrac{1}{2}(\nabla_{\underline{L}} F)(L, e_i) + (\nabla_{e_j} F)(e_j, e_i)\\ &= -\tfrac{1}{2} L\underline{\alpha}_i + \tfrac{1}{2}(\nabla_L F)(e_i, \underline{L}) + \tfrac{1}{2}(\nabla_{e_i} F)(\underline{L}, L) + e_j F_{e_j e_i} - F(-2r^{-1}\partial_r, e_i) - F(e_i, -r^{-1}\partial_r)\\ &= -L\underline{\alpha}_i + e_i\rho - \tfrac{1}{2}F(-r^{-1}e_i, L) - \tfrac{1}{2}F(\underline{L}, r^{-1}e_i) + e_j F_{e_j e_i} + r^{-1}F(\partial_r, e_i)\\ &= -L\underline{\alpha}_i + e_i\rho + e_j F_{e_j e_i} - r^{-1}\underline{\alpha}_i.\end{aligned}$$

This leads to (8).

The first transport equation (9) for $\sigma$ follows from the Bianchi identity:

$$\begin{aligned}0 &= (\nabla_{\underline{L}} F)(e_1, e_2) + (\nabla_{e_1} F)(e_2, \underline{L}) + (\nabla_{e_2} F)(\underline{L}, e_1)\\ &= \underline{L}\sigma - e_1\underline{\alpha}_2 - F(e_2, -r^{-1}e_1) + e_2\underline{\alpha}_1 - F(-r^{-1}e_2, e_1)\\ &= \underline{L}\sigma - e_1\underline{\alpha}_2 + e_2\underline{\alpha}_1 - 2r^{-1}\sigma.\end{aligned}$$

The dual one follows if we replace $\underline{L}$ with $L$. □

## 3. Energy method

In this section, we review the energy method for solutions of the covariant linear wave equations and Maxwell equations using the new method developed in [Yang 2015c]. This new method was originally introduced by Dafermos and Rodnianski [2010] for proving the decay of solutions of linear wave equations in Minkowski space. It has been successfully applied to MKG equations by the author in [Yang 2015c] to obtain the decay properties of the solutions for all data bounded in some weighted energy space. There the necessary new ingredients (see Propositions 7–10 in this section) were carried out for the full solution $(\phi, F)$. In this paper, the data for the scalar field are assumed to be small, and we also need to derive the decay estimates for the derivatives of the solutions in order to obtain the pointwise decay of the solutions. We thus need all the new ingredients both for the scalar field and the Maxwell field. The ideas to derive these new estimates are the same as those in [Yang 2015c]. For the readers' convenience, we repeat the proofs here.

**3.1.** ***Energy identity for the scalar field.*** Denote by $d\mathrm{vol}$ the volume form in the Minkowski space. In the local coordinate system $(t, x)$, we have $d\mathrm{vol} = dx\,dt$. Here we have chosen $t$ to be the time orientation. For any complex scalar field $\phi$, we define the associated energy momentum tensor

$$T[\phi]_{\mu\nu} = \Re(\overline{D_\mu\phi} D_\nu\phi) - \tfrac{1}{2} m_{\mu\nu}\overline{D^\gamma\phi} D_\gamma\phi.$$

Here $m_{\mu\nu}$ is the flat metric of Minkowski spacetime and the covariant derivative $D$ is defined with respect to the given connection field $A$. For any vector field $X$, we have the following identity

$$\nabla^\mu(T[\phi]_{\mu\nu}X^\nu) = \Re(\Box_A\phi X^\nu\overline{D_\nu\phi}) + X^\nu F_{\nu\gamma}J^\gamma[\phi] + T[\phi]^{\mu\nu}\pi^X_{\mu\nu},$$

where $\pi^X_{\mu\nu} = \frac{1}{2}\mathcal{L}_X m_{\mu\nu}$ is the deformation tensor of the vector field $X$ in Minkowski space, $\Box_A$ is the covariant wave operator associated to the connection $A$, $F = dA$ is the exterior derivative of the 1-form

$A$ which gives us a 2-form and $J^\gamma[\phi] = \Im(\phi \cdot \overline{D^\gamma \phi})$. For any function $\chi$, we have

$$\tfrac{1}{2}\nabla^\mu(\chi \nabla_\mu |\phi|^2 - \nabla_\mu \chi |\phi|^2) = \chi \overline{D_\mu \phi} D^\mu \phi - \tfrac{1}{2}\Box\chi \cdot |\phi|^2 + \chi \Re(\Box_A \phi \cdot \bar{\phi}).$$

We now define the vector field $\tilde{J}^X[\phi]$ with components

$$\tilde{J}^X_\mu[\phi] = T[\phi]_{\mu\nu} X^\nu - \tfrac{1}{2}\nabla_\mu \chi \cdot |\phi|^2 + \tfrac{1}{2}\chi \nabla_\mu |\phi|^2 + Y_\mu \tag{11}$$

for some vector field $Y$ which may also depend on the complex scalar field $\phi$. We then have the equality

$$\nabla^\mu \tilde{J}^X_\mu[\phi] = \Re(\Box_A \phi (\overline{D_X \phi} + \chi \bar{\phi})) + \operatorname{div}(Y) + X^\nu F_{\nu\mu} J^\mu[\phi] + T[\phi]^{\mu\nu} \pi^X_{\mu\nu} + \chi \overline{D_\mu \phi} D^\mu \phi - \tfrac{1}{2}\Box\chi \cdot |\phi|^2.$$

Here the operator $\Box$ is the wave operator in Minkowski space. Now for any region $\mathcal{D}$ in $\mathbb{R}^{3+1}$, using Stokes' formula, we derive the energy identity

$$\begin{aligned}\iint_{\mathcal{D}} \Re(\Box_A \phi (\overline{D_X \phi} + \chi \bar{\phi})) + \operatorname{div}(Y) + X^\nu F_{\nu\gamma} J^\gamma[\phi] + T[\phi]^{\mu\nu} \pi^X_{\mu\nu} + \chi \overline{D_\mu \phi} D^\mu \phi - \tfrac{1}{2}\Box\chi \cdot |\phi|^2 \, d\mathrm{vol} \\ = \iint_{\mathcal{D}} \nabla^\mu \tilde{J}^X_\mu[\phi] \, d\mathrm{vol} = \int_{\partial\mathcal{D}} i_{\tilde{J}^X[\phi]} \, d\mathrm{vol},\end{aligned} \tag{12}$$

where $\partial\mathcal{D}$ denotes the boundary of the domain $\mathcal{D}$ and $i_Z \, d\mathrm{vol}$ denotes the contraction of the volume form $d\mathrm{vol}$ with the vector field $Z$ which gives the surface measure of the boundary. For example, for any basis $\{e_1, e_2, \ldots, e_n\}$, we have

$$i_{e_1}(de_1 \wedge de_2 \wedge \cdots \wedge de_k) = de_2 \wedge de_3 \wedge \cdots \wedge de_k.$$

Throughout this paper, the domain $\mathcal{D}$ will be regular regions bounded by the $t$-constant slices, the outgoing null hypersurfaces $H_u$, the incoming null hypersurfaces $\underline{H}_v$ or the surface with constant $r$. We now compute $i_{\tilde{J}^X[\phi]} \, d\mathrm{vol}$ on each of these hypersurfaces.

On $t =$ constant slice, the surface measure is a function times $dx$. Recall the volume form

$$d\mathrm{vol} = dx\, dt = -dt\, dx.$$

Here note that $dx$ is a 3-form. We thus can show that

$$i_{\tilde{J}^X[\phi]} \, d\mathrm{vol} = -(\tilde{J}^X[\phi])^0 \, dx = -\big(\Re(\overline{D^t \phi} D_X \phi) - \tfrac{1}{2} X^0 \overline{D^\gamma \phi} D_\gamma \phi - \tfrac{1}{2}\partial^t \chi |\phi|^2 + \tfrac{1}{2}\chi \partial^t |\phi|^2 + Y^0\big) \, dx. \tag{13}$$

On the surface with constant $r$, the surface measure is $r^2 \, dt \, d\omega$. Therefore we have

$$i_{\tilde{J}^X[\phi]} \, d\mathrm{vol} = \big(\Re(\overline{D^r \phi} D_X \phi) - \tfrac{1}{2} X^r \overline{D^\gamma \phi} D_\gamma \phi - \tfrac{1}{2}\partial^r \chi |\phi|^2 + \tfrac{1}{2}\chi \partial^r |\phi|^2 + Y^r) r^2 \, dt \, d\omega. \tag{14}$$

On the outgoing null hypersurface $H_u$, we can write the volume form

$$d\mathrm{vol} = dx\, dt = r^2 \, dr\, dt\, d\omega = 2r^2 \, dv\, du\, d\omega = -2 \, du\, dv\, d\omega.$$

Here $d\omega$ is the standard surface measure on the unit sphere. Notice that $\underline{L} = \partial_u$. We can compute

$$i_{\tilde{J}^X[\phi]} \, d\mathrm{vol} = -2\big(\Re(\overline{D^{\underline{L}} \phi} D_X \phi) - \tfrac{1}{2} X^{\underline{L}} \overline{D^\gamma \phi} D_\gamma \phi - \tfrac{1}{2}\nabla^{\underline{L}} \chi |\phi|^2 + \tfrac{1}{2}\chi \nabla^{\underline{L}} |\phi|^2 + Y^{\underline{L}}\big) r^2 \, dv \, d\omega. \tag{15}$$

Similarly, on the $v$-constant incoming null hypersurfaces $\underline{H}_{\underline{u}}$, we have

$$i_{\tilde{J}^X[\phi]}\,d\mathrm{vol} = 2\big(\Re(\overline{D^L\phi}D_X\phi) - \tfrac{1}{2}X^L\overline{D^\gamma\phi}D_\gamma\phi - \tfrac{1}{2}\nabla^L\chi|\phi|^2 + \tfrac{1}{2}\chi\nabla^L|\phi|^2 + Y^L\big)r^2\,du\,d\omega. \tag{16}$$

We remark here that the above formula hold for any vector fields $X$, $Y$ and any function $\chi$.

**3.2.** ***Energy identities for the Maxwell field.*** Let $F$ be any 2-form satisfying the Bianchi identity (1). The associated energy momentum tensor is

$$T[F]_{\mu\nu} = F_{\mu\gamma}F_\nu^{\ \gamma} - \tfrac{1}{4}m_{\mu\nu}F_{\gamma\beta}F^{\gamma\beta}.$$

For any vector field $X$, we have the divergence formula

$$\nabla^\mu T[F]_{\mu\nu}X^\nu = \nabla^\mu F_{\mu\gamma}F_\nu^{\ \gamma}X^\nu + T[F]^{\mu\nu}\pi^X_{\mu\nu},$$

where as defined previously, $\pi^X_{\mu\nu} = \frac{1}{2}\mathcal{L}_X m_{\mu\nu}$ is the deformation tensor of the vector field $X$ in Minkowski space. Define the vector field $J^X[F]$ by

$$J^X[F]_\mu = T[F]_{\mu\nu}X^\nu.$$

Then for any domain $\mathcal{D}$ in $\mathbb{R}^{3+1}$, we have the following energy identity for the Maxwell field $F$:

$$\iint_{\mathcal{D}} \nabla^\mu F_{\mu\gamma}F_\nu^{\ \gamma}X^\nu + T[F]^{\mu\nu}\pi^X_{\mu\nu}\,d\mathrm{vol} = \iint_{\mathcal{D}} \nabla^\mu J^X_\mu[F]\,d\mathrm{vol} = \int_{\partial\mathcal{D}} i_{J^X[F]}\,d\mathrm{vol}. \tag{17}$$

For the terms on the boundary, similar to (13)–(16), we can compute

$$\begin{aligned}
\int_{\{t=\text{const.}\}} i_{J^X[F]}\,d\mathrm{vol} &= -\int_{\{t=\text{const.}\}} \Big(F^{0\mu}F_{\nu\mu}X^\nu - \frac{1}{4}X^0F_{\mu\nu}F^{\mu\nu}\Big)\,dx;\\
\int_{\{r=\text{const.}\}} i_{J^X[F]}\,d\mathrm{vol} &= \int_{\{r=\text{const.}\}} \Big(F^{r\mu}F_{\nu\mu}X^\nu - \frac{1}{4}X^rF_{\mu\nu}F^{\mu\nu}\Big)r^2\,dt\,d\omega;\\
\int_{H_u} i_{J^X[F]}\,d\mathrm{vol} &= -2\int_{H_u} \Big(F^{\underline{L}\mu}F_{\nu\mu}X^\nu - \frac{1}{4}X^{\underline{L}}F_{\mu\nu}F^{\mu\nu}\Big)r^2\,dv\,d\omega;\\
\int_{\underline{H}_v} i_{J^X[F]}\,d\mathrm{vol} &= 2\int_{\underline{H}_v} \Big(F^{L\mu}F_{\nu\mu}X^\nu - \frac{1}{4}X^LF_{\mu\nu}F^{\mu\nu}\Big)r^2\,du\,d\omega.
\end{aligned} \tag{18}$$

**3.3.** ***The integrated local energy estimates using the multiplier $f(r)\partial_r$.*** For the full solution $(\phi, F)$ of the Maxwell–Klein–Gordon equations, including the case with large charge, the integrated local energy estimates together with the $r$-weighted energy estimates in the next subsection have been studied in the author's work [Yang 2015c]. To obtain estimates for higher-order derivatives of the solutions, we need to commute the equations with derivatives, and hence nonlinear terms arise. Furthermore, in our setting, the data for the Maxwell field are large while the data for the complex scalar field are small. We thus need to obtain estimates separately for the Maxwell field and the scalar field.

We first consider the integrated local energy estimates for the scalar field. In the energy identity (12) for the scalar field, we choose the vector fields $X$, $Y$ as follows:

$$X = f(r)\partial_r, \quad Y = 0$$

for some function $f(r)$. We then can compute

$$T[\phi]^{\mu\nu}\pi^X_{\mu\nu} + \chi\,\overline{D_\mu\phi}D^\mu\phi - \tfrac{1}{2}\Box\chi|\phi|^2$$
$$= \big(r^{-1}f - \chi + \tfrac{1}{2}f'\big)|D_t\phi|^2 + \big(\chi + \tfrac{1}{2}f' - r^{-1}f\big)|D_r\phi|^2 + \big(\chi - \tfrac{1}{2}f'\big)|\not{D}\phi|^2 - \tfrac{1}{2}\Box\chi|\phi|^2.$$

The idea is to choose the functions $f$, $\chi$ so that the coefficients are positive. Let $\epsilon$ be a small positive constant, depending only on $\gamma_0$ (e.g., $\epsilon = 10^{-3}\gamma_0$). Construct the functions $f$ and $\chi$ so that

$$f(r) = 2\epsilon^{-1} - \frac{2\epsilon^{-1}}{(1+r)^\epsilon}, \quad \chi = r^{-1}f.$$

We can compute

$$\chi - r^{-1}f + \frac{1}{2}f' = r^{-1}f + \frac{1}{2}f' - \chi = \frac{1}{(1+r)^{1+\epsilon}}, \quad -\frac{1}{2}\Box\chi = \frac{1+\epsilon}{r(1+r)^{2+\epsilon}}.$$

When $r > 1$, we have the following improved estimate for $\chi - \frac{1}{2}f'$:

$$\chi - \frac{1}{2}f' \geq \frac{2\epsilon^{-1}}{r} - \frac{1+2\epsilon^{-1}}{r(1+r)^\epsilon} \geq \frac{1}{r}, \quad r \geq 1. \tag{19}$$

This improved estimate will be used to show the improved integrated local energy estimate for the covariant angular derivative of the scalar field $\phi$.

From the above calculation, we see that for this particular choice of vector field $X$ and the function $\chi$, the last three terms in the first line of (12) have positive signs. We treat the first two terms as nonlinear terms. To get an integrated local energy estimate for the scalar field $\phi$, it suffices to control the boundary terms arising from the Stokes' formula (12). This requires a version of Hardy's inequality. Before stating the lemma, we make a convention that the notation $A \lesssim_K B$ means that there exists a constant $C$, depending only on the constants $R$, $\gamma_0$, $\epsilon$ and the set $K$ such that $A \leq CB$. For the particular case when $K$ is empty, we omit the index $K$.

**Lemma 6.** *Assume $0 \leq \gamma < 1$ and the complex scalar field $\phi$ vanishes at null infinity, that is,*

$$\lim_{v\to\infty} \phi(v, u, \omega) = 0$$

*for all $u$, $\omega$. Then we have*

$$\int_{H_u} r^\gamma |\phi|^2\, dv\, d\omega \lesssim \int_\omega r^{1+\gamma}|\phi|^2(u, v(u), \omega)\, d\omega + \int_{H_u} r^\gamma |D_L(r\phi)|^2\, dv\, d\omega \tag{20}$$

*for all $u \in \mathbb{R}$. Here $v(u) = -u$ when $u \leq -\frac{1}{2}R$ that is in the exterior region and $v(u) = 2R + u$ when $u > -\frac{1}{2}R$ that is in the interior region. In particular, we have*

$$\int_{H_u} |\phi|^2\, dv\, d\omega \lesssim E[\phi](H_u), \quad \int_{\Sigma_\tau} |\phi|^2\, dv'\, d\omega \lesssim E[\phi](\Sigma_\tau). \tag{21}$$

*Here $v' = v$ when $r \geq R$ and $v' = r$ otherwise.*

*Proof.* It suffices to notice that the connection $D$ is compatible with the inner product $\langle\,,\rangle$ on the complex plane. Then the proof when $\gamma = 0$ goes the same as the case when the connection field $A$ is trivial; see, e.g., Lemma 2 of [Yang 2013] or Proposition 11.2 of [Dafermos and Rodnianski 2009]. Another quick way to reduce the proof of the lemma to the case with trivial connection field $A$ is to choose a particular gauge such that the scalar field $\phi$ is real. We can do this is due to the fact that all the norms in this paper are gauge invariant. For general $\gamma$, based on the above argument, the proof goes similar to the proof of the standard Hardy's inequality. Let $\psi = r\phi$. Note that $\gamma < 1$. We can show that

$$\begin{aligned}\int_{v_0}^{\infty}\int_{\omega} r^{\gamma-2}|\psi|^2\,dv\,d\omega &= \frac{1}{\gamma-1}\int_{v_0}^{\infty}\int_{\omega}|\psi|^2\,d\omega\,dr^{\gamma-1}\\ &= \frac{1}{\gamma-1}r^{\gamma-1}\int_{\omega}|\psi|^2\,d\omega\Big|_{v_0}^{\infty} + \frac{2}{1-\gamma}\int_{v_0}^{\infty}\int_{\omega} r^{\gamma-1}D_L\psi\cdot\psi\,dv\,d\omega\\ &\le \frac{1}{1-\gamma}\int_{\omega} r^{1+\gamma}|\phi|^2(u,v_0,\omega)\,d\omega + \frac{1}{2}\int_{v_0}^{\infty}\int_{\omega} r^{\gamma-2}|\psi|^2\,dv\,d\omega\\ &\quad + \frac{8}{(1-\gamma)^2}\int_{v_0}^{\infty}\int_{\omega} r^{\gamma}|D_L\psi|^2\,dv\,d\omega.\end{aligned}$$

The estimate (20) then follows by absorbing the second term and taking $v_0 = v(u)$. □

We then can derive the following integrated local energy estimate for the scalar field $\phi$.

**Proposition 7.** *Assume the complex scalar field $\phi$ vanishes at null infinity and the spatial infinity initially. Then in the interior region $\{r \le R + t\}$, we have the energy estimates*

$$\begin{aligned}&I_0^{-1-\epsilon}[\tilde{D}\phi](\mathcal{D}_{\tau_1}^{\tau_2}) + E[\phi](\Sigma_{\tau_2}) + E[\phi](\underline{H}_v^{\tau_1^*,\tau_2^*}) + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}}\frac{|\not{D}\phi|^2}{1+r}\,dx\,dt\\ &\qquad\lesssim E[\phi](\Sigma_{\tau_1}) + I_0^{1+\epsilon}[\Box_A\phi](\mathcal{D}_{\tau_1}^{\tau_2}) + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}}|F_{Lv}J^v[\phi]| + |F_{\underline{L}v}J^v[\phi]|\,dx\,dt \qquad (22)\end{aligned}$$

*for all $0 \le \tau_1 < \tau_2$ and $v \ge \frac{1}{2}(\tau_2 + R)$, where we let $\tilde{D}\phi = (D\phi, r_+^{-1}\phi)$ and $F = dA$. Similarly, in the exterior region $\{r > t + R\}$, we have*

$$\begin{aligned}&I_0^{-1-\epsilon}[\tilde{D}\phi](\mathcal{D}_{\tau_1}^{\tau_2}) + E[\phi]\big(H_{\tau_1^*}^{-\tau_2^*}\big) + E[\phi]\big(\underline{H}_{-\tau_2^*}^{\tau_2^*,\tau_1^*}\big)\\ &\qquad\lesssim E[\phi](B_{R-\tau_1}) + I_0^{1+\epsilon}[\Box_A\phi](\mathcal{D}_{\tau_1}^{\tau_2}) + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}}|F_{Lv}J^v[\phi]| + |F_{\underline{L}v}J^v[\phi]|\,dx\,dt \qquad (23)\end{aligned}$$

*for all $\tau_2 \le \tau_1 \le 0$. Here see the notations in Section 2 and $J^\mu[\phi] = \Im(\phi\cdot\overline{D^\mu\phi})$, $\tau^* = \frac{1}{2}(\tau - R)$.*

*Proof.* For all $v_0 \ge \frac{1}{2}(\tau_2 + R)$, take the region $\mathcal{D}$ to be $\mathcal{D}_{\tau_1}^{\tau_2} \cap \{v \le v_0\}$, which is bounded by the surfaces

$$\Sigma_{\tau_1}, \quad \Sigma_{\tau_2}, \quad \underline{H}_{v_0}^{\tau_1^*,\tau_2^*}$$

and the functions $f$, $\chi$ as above and the vector field $Y = 0$ in the energy identity (12). The boundary terms can be controlled by the energy flux according to Hardy's inequality of Lemma 6. For more details regarding this bound, we refer to, e.g., Proposition 1 in [Yang 2013]. Therefore the above calculations lead to the following integrated local energy estimate:

$$\iint_{\mathcal{D}_{\tau_1}^{\tau_2}\cap\{v\le v_0\}} \frac{|D\phi|^2}{(1+r)^{1+\epsilon}} + \frac{|\not{D}\phi|^2}{1+r} + \frac{|\phi|^2}{r(1+r)^{2+\epsilon}}\,dx\,dt$$
$$\lesssim E[\phi](\Sigma_{\tau_1}^{v_0}) + E[\phi](\Sigma_{\tau_2}^{v_0}) + E[\phi]\big(\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}\big) + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} |\Box_A\phi(\overline{D_X\phi} + \chi\bar{\phi})| + |F_{rv}J^v[\phi]|\,dx\,dt.$$

Next, we take the vector fields $X = \partial_t$, $Y = 0$ and the function $\chi = 0$ in the energy identity (12) for the scalar field. Consider the region $\mathcal{D}_{\tau_1}^{\tau_2}\cap\{v \le v_0\}$. We retrieve the classical energy estimate

$$E[\phi](\Sigma_{\tau_2}^{v_0}) + E[\phi]\big(\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}\big) = E[\phi](\Sigma_{\tau_1}^{v_0}) - 2\iint_{\mathcal{D}_{\tau_1}^{\tau_2}\cap\{v\le v_0\}} \Re(\Box_A\phi\overline{D_t\phi}) + F_{0v}J^v[\phi]\,dx\,dt.$$

Combined with the previous integrated local energy estimate and letting $v_0 \to \infty$, we derive that

$$I_0^{-1-\epsilon}[\tilde{D}\phi](\mathcal{D}_{\tau_1}^{\tau_2}) \lesssim E[\phi](\Sigma_{\tau_1}) + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} \big|\Box_A\phi\overline{\tilde{D}\phi}\big| + |F_{Lv}J^v[\phi]| + |F_{\underline{L}v}J^v[\phi]|\,dx\,dt.$$

We apply the Cauchy–Schwarz inequality to the integral of $\Box_A\phi\overline{\tilde{D}\phi}$:

$$2\big|\Box_A\phi\overline{\tilde{D}\phi}\big| \le \epsilon_1 r_+^{-1-\epsilon}|\tilde{D}\phi|^2 + \epsilon_1^{-1}r_+^{1+\epsilon}|\Box_A\phi|^2, \quad \forall\epsilon_1 > 0.$$

Choose $\epsilon_1$ to be sufficiently small depending only on $\epsilon$, $\gamma_0$, $R$ so that the integral of the first term can be absorbed. We thus can derive the integrated local energy estimate for the scalar field. Then in the above classical energy estimate, we can use the Cauchy–Schwarz inequality again to bound $\Re(\Box_A\phi\overline{D_t\phi})$ which gives control of the energy flux $E[\phi](H_{\tau_2^*})$. This energy estimate together with the previous integrated local energy estimate imply the energy estimate (22) of the proposition in the interior region. The improved estimate for the angular covariant derivative is due to the improve estimate (19).

The proof for the estimate (23) in the exterior region is similar. The only point we need to emphasize is that we use the fact that the $\phi$ goes to zero as $r \to \infty$ on the initial hypersurface. We thus can use the Hardy's inequality to control the integral of $|\phi|^2/(1+r)^2$. This is also the reason that we have $E[\phi](B_{R-\tau_1})$ instead of $E[\phi](B_{R-\tau_1}^{R-\tau_2})$ on the right-hand side of (23). □

In our setting, $F$ is the Maxwell field, which is no longer small. In particular this means that the integral of $|F_{Lv}J^v[\phi]|$ on the right-hand side of (22), (23) could not be absorbed. The key to controlling those terms is to use the $r$-weighted energy estimates in the next section.

Let $F$ be any 2-form satisfying the Bianchi identity (1). Let $J = \delta F$ or $J_\mu = \nabla^\nu F_{\nu\mu}$ be the divergence of $F$. This notation $J$ can be viewed as the inhomogeneous term of the linear Maxwell equation. In (MKG), this $J$ is identical to $J[\phi]$, which is quadratic in the scalar field $\phi$. Under the null frame $\{L, \underline{L}, e_1, e_2\}$, write $\not{J} = (J_{e_1}, J_{e_2})$. We derive an analogue of Proposition 7 for the Maxwell field $F$.

**Proposition 8.** *In the interior region* $\{r \le t + R\}$, *we have the integrated local energy estimates*

$$I_0^{-1-\epsilon}[F](\mathcal{D}_{\tau_1}^{\tau_2}) + \int_{\tau_1}^{\tau_2}\int_{\Sigma_\tau} \frac{\rho^2+|\sigma|^2}{1+r}\,dx\,d\tau + E[F]\big(\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}\big) + E[F](\Sigma_{\tau_2})$$
$$\lesssim E[F](\Sigma_{\tau_1}) + I_0^{1+\epsilon}[|J_L| + |\not{J}|](\mathcal{D}_{\tau_1}^{\tau_2}) + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} |J_{\underline{L}}||\rho|\,dx\,dt \quad (24)$$

*for all* $0 \le \tau_1 < \tau_2$ *and* $v_0 \ge \frac{1}{2}(\tau_2 + R)$. *Similarly, in the exterior region* $\{r \le R + t\}$, *for all* $\tau_2 < \tau_1 \le 0$ *we have*

$$I_0^{-1-\epsilon}[F](\mathcal{D}_{\tau_1}^{\tau_2}) + E[F]\big(\underline{H}_{-\tau_2^*}^{\tau_2^*,\tau_1^*}\big) + E[F]\big(H_{\tau_1^*}^{-\tau_2^*}\big)$$
$$\lesssim E[F](B_{R-\tau_1}^{R-\tau_2}) + I_0^{1+\epsilon}[|J_L| + |\not{J}|](\mathcal{D}_{\tau_1}^{\tau_2}) + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} |J_{\underline{L}}||\rho|\,dx\,dt. \quad (25)$$

*Proof.* The idea to prove this proposition is the same as that of the previous proposition for the scalar field. However, the calculations are slightly different for the Maxwell field $F$. In the energy identity (17) for the Maxwell field, we take the vector field

$$X = f(r)\partial_r = 2\epsilon^{-1}(1 - r_+^{-\epsilon})\partial_r.$$

Set $\omega_i = r^{-1}x_i$. We then can compute

$$T[F]^{\mu\nu}\pi^X_{\mu\nu} = T[F]^{ij}(f'\omega_i\omega_j + r^{-1}f\delta_{ij} - r^{-1}f\omega_i\omega_j)$$
$$= \tfrac{1}{4}(2r^{-1}f - f')F_{\mu\nu}F^{\mu\nu} + (f' - r^{-1}f)F_{r\nu}F^{r\nu} - r^{-1}fF_{0\nu}F^{0\nu},$$

where the Greek indices $\mu$, $\nu$ run from 0 to 3 and the Latin indices $i$, $j$ run from 1 to 3. Using the null decomposition of the 2-form under the null frame $\{L, \underline{L}, e_1, e_2\}$ defined in line (4), we can show that

$$F_{\mu\nu}F^{\mu\nu} = -2\rho^2 - 2\alpha\cdot\underline{\alpha} + 2|\sigma|^2,$$
$$F_{0\nu}F^{0\nu} = -\tfrac{1}{4}(4\rho^2 + 2\alpha\cdot\underline{\alpha} + |\alpha|^2 + |\underline{\alpha}|^2),$$
$$F_{r\nu}F^{r\nu} = -\tfrac{1}{4}(4\rho^2 + 2\alpha\cdot\underline{\alpha} - |\alpha|^2 - |\underline{\alpha}|^2).$$

Therefore we have

$$T[F]^{\mu\nu}\pi^X_{\mu\nu} = \big(r^{-1}f - \tfrac{1}{2}f'\big)(\rho^2 + |\sigma|^2) + \tfrac{1}{4}f'(|\alpha|^2 + |\underline{\alpha}|^2). \quad (26)$$

The calculations before line (19) imply that the coefficients $r^{-1}f - \frac{1}{2}f'$ and $f'$ have positive signs. To obtain the similar integrated local energy estimates for the Maxwell field $F$, we need to control the boundary terms arising from the Stokes' formula (17). Using the formula (18), we can compute that

$$2|i_{J^X[F]}\,d\mathrm{vol}| = f\big||\alpha|^2 - |\underline{\alpha}|^2\big|\,dx \le |F|^2\,dx = 2fi_{J^{\partial_t}[F]}\,d\mathrm{vol},$$
$$2|i_{J^X[F]}\,d\mathrm{vol}| = f\big|-\rho^2 + |\alpha|^2 - |\sigma|^2\big|r^2\,dv\,d\omega \le f(\rho^2 + |\alpha|^2 + |\sigma|^2)r^2\,dv\,d\omega = 2fi_{J^{\partial_t}[F]}\,d\mathrm{vol},$$
$$2|i_{J^X[F]}\,d\mathrm{vol}| = f\big|-\rho^2 + |\underline{\alpha}|^2 - |\sigma|^2\big|r^2\,du\,d\omega \le f(\rho^2 + |\underline{\alpha}|^2 + |\sigma|^2)r^2\,du\,d\omega = 2fi_{J^{\partial_t}[F]}\,d\mathrm{vol},$$

on the $t =$ constant slice, the outgoing null hypersurface and the incoming null hypersurface, respectively, for all positive functions $f$. This in particular implies that the boundary terms corresponding to the

vector field $f\partial_r$ can be bounded by the energy flux for all positive bounded functions $f$. Therefore, for the particular choice of vector field $X$, the energy identity (17) on the domain $\mathcal{D}^{\tau_2}_{\tau_1}\cap\{v\le v_0\}$ for all $0\le\tau_1<\tau_2$ and $v_0\ge\frac{1}{2}(\tau_2+R)$ leads to

$$\int_{\tau_1}^{\tau_2}\int_{\Sigma_\tau^{v_0}}\frac{|F|^2}{(1+r)^{1+\epsilon}}+\frac{\rho^2+|\sigma|^2}{1+r}\,dx\,d\tau\lesssim E[F](\Sigma^{v_0}_{\tau_1})+E[F](\Sigma^{v_0}_{\tau_2})+E[F]\big(\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}\big)+\int_{\tau_1}^{\tau_2}\int_{\Sigma_\tau}|J^\gamma||F_{L\gamma}-F_{\underline{L}\gamma})|\,dx\,d\tau.$$

Here notice that we have the improved estimate (19) for the coefficient of $\rho^2+|\sigma|^2$. If we take the vector field $X=\partial_t$ on the same domain, we then can derive the classical energy identity

$$\int_{\tau_1}^{\tau_2}\int_{\Sigma_\tau^{v_0}}J^\gamma(F_{L\gamma}+F_{\underline{L}\gamma})\,dx\,d\tau=E[F](\Sigma^{v_0}_{\tau_1})-E[F]\big(\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}\big)-E[F](\Sigma^{v_0}_{\tau_2}).$$

Let $v_0\to\infty$ and apply Cauchy–Schwarz to the inhomogeneous term $J^\mu(|F_{L\mu}|+|F_{\underline{L}\mu}|)$ for $\mu=\underline{L},e_1,e_2$:

$$|J^{\underline{L}}||F_{L\underline{L}}|+|J^{e_i}|(|F_{Le_i}|+|F_{\underline{L}e_i}|)\lesssim\epsilon_1^{-1}(|J_L|+|\not{J}|)r_+^{1+\epsilon}+\epsilon_1|F|^2r_+^{-1-\epsilon},\quad\epsilon_1>0.$$

The integral of the second term could be absorbed for sufficiently small $\epsilon_1$. For the component when $\mu=L$, we estimate

$$|J^L||F_{L\underline{L}}|\lesssim|J_{\underline{L}}||\rho|.$$

Then the above energy identity together with the integrated local energy estimates imply the integrated local energy estimate (24) in the interior region. The energy estimate (25) in the exterior region follows in a similar way. □

**3.4.** ***The r-weighted energy estimates using the multiplier $r^pL$.*** In this section, we establish the robust $r$-weighted energy estimates both for the scalar field and the Maxwell field. This estimate for solutions of linear wave equation in Minkowski space was first introduced by Dafermos and Rodnianski [2010]. We study the $r$-weighted energy estimate either in the exterior region $\{r\ge R+t\}$ for the domain $\mathcal{D}^{\tau_2}_{\tau_1}$ for $\tau_2\le\tau_1\le0$ or in the interior region for domain $\overline{\mathcal{D}}^{\tau_2}_{\tau_1}$ for $0\le\tau_1<\tau_2$ which is bounded by the outgoing null hypersurfaces $H_{\tau_1^*}$, $H_{\tau_2^*}$ and the cylinder $\{r=R\}$.

Through out this paper, we denote $\psi=r\phi$ as the $r$-weighted scalar field. We have the following $r$-weighted energy estimates for the complex scalar field.

**Proposition 9.** *Assume that the complex scalar field $\phi$ vanishes at null infinity. Then in the interior region, for all $0\le\tau_1<\tau_2$ and $v_0\ge\frac{1}{2}(\tau_2+R)$, we have the r-weighted energy estimate*

$$\begin{aligned}&\int_{H_{\tau_2^*}}r^p|D_L\psi|^2\,dv\,d\omega+\int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}}r^{p-1}(p|D_L\psi|^2+(2-p)|\not{D}\psi|^2)\,dv\,d\omega\,d\tau+\int_{\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}}r^p|\not{D}\psi|^2\,du\,d\omega\\&\lesssim\int_{H_{\tau_1^*}}r^p|D_L\psi|^2\,dv\,d\omega+I^{\max\{1+\epsilon,p\}}_{\min\{1+\epsilon,p\}}[\Box_A\phi](\mathcal{D}^{\tau_2}_{\tau_1})+E[\phi](\Sigma_{\tau_1})+I_0^{1+\epsilon}[\Box_A\phi](\mathcal{D}^{\tau_2}_{\tau_1})\\&\quad+\iint_{\mathcal{D}^{\tau_2}_{\tau_1}}|F_{\underline{L}\mu}J^\mu[\phi]|+|F_{L\mu}J^\mu[\phi]|\,dx\,dt+\iint_{\overline{\mathcal{D}}^{\tau_2}_{\tau_1}}r^p|F_{L\mu}J^\mu[\phi]|\,dx\,dt\end{aligned}\tag{27}$$

*for all* $0 \le p \le 2$. *Similarly, in the exterior region, for all* $\tau_2 < \tau_1 \le 0$, *we have*

$$\int_{H_{\tau_1^*}^{-\tau_2^*}} r^p |D_L\psi|^2\,dv\,d\omega + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} r^{p-1}(p|D_L\psi|^2 + (2-p)|\not{D}\psi|^2)\,dv\,d\omega\,du + \int_{\underline{H}_{-\tau_2^*}^{\tau_2^*,\tau_1^*}} r^p|\not{D}\psi|^2\,du\,d\omega$$
$$\lesssim \int_{B_{R-\tau_1}^{R-\tau_2}} r^p(|D_L\psi|^2 + |\not{D}\psi|^2)\,dr\,d\omega + I_{\min\{1+\epsilon,p\}}^{\max\{p,1+\epsilon\}}[\Box_A\phi](\mathcal{D}_{\tau_1}^{\tau_2}) + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} r^p|F_{L\mu}J^\mu[\phi]|\,dx\,dt \quad (28)$$

*for all* $0 \le p \le 2$. *Here* $\psi = r\phi$.

*Proof.* Apply the energy identity (12) to the region $\overline{\mathcal{D}}_{\tau_1}^{\tau_2} \cap \{v \le v_0\}$, which is bounded by $H_{\tau_1^*}$, $H_{\tau_2^*}$, $\{r = R\}$ and $\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}$ with the vector fields $X$, $Y$ and the function $\chi$ as follows:

$$X = r^p L, \quad Y = \tfrac{1}{2}pr^{p-2}|\phi|^2 L, \quad \chi = r^{p-1}.$$

Define $\psi = r\phi$ to be the weighted scalar field. We have the equalities

$$r^2|D_L\phi|^2 = |D_L\psi|^2 - L(r|\phi|^2),$$
$$r^2|\not{D}\phi|^2 = |\not{D}\psi|^2,$$
$$r^2|D_{\underline{L}}\phi|^2 = |D_{\underline{L}}\psi|^2 + \underline{L}(r|\phi|^2).$$

We then can compute

$$\begin{aligned}\operatorname{div}(Y) + T[\phi]^{\mu\nu}\pi^X_{\mu\nu} + \chi\,\overline{D^\mu\phi}D_\mu\phi - \tfrac{1}{2}\Box\chi|\phi|^2 &= \tfrac{1}{2}pr^{-2}L(r^p|\phi|^2) + \tfrac{1}{2}r^{p-1}\big(p|D_L\phi|^2 + (2-p)|\not{D}\phi|^2\big) - \tfrac{1}{2}p(p-1)r^{p-3}|\phi|^2\\ &= \tfrac{1}{2}r^{p-3}\big(p|D_L\psi|^2 + (2-p)|\not{D}\psi|^2\big).\end{aligned}$$

We next compute the boundary terms using the formula (18). We have

$$\int_{H_{\tau^*}^{v_0}} i_{\tilde{J}^X[\phi]}\,d\mathrm{vol} = \int_{H_{\tau^*}^{v_0}} r^p|D_L\psi|^2 - \tfrac{1}{2}L(r^{p+1}\phi)\,dv\,d\omega,$$
$$\int_{\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}} i_{\tilde{J}^X[\phi]}\,d\mathrm{vol} = -\int_{\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}} r^p|\not{D}\psi|^2 + \tfrac{1}{2}\underline{L}(r^{p+1}|\phi|^2)\,du\,d\omega,$$
$$\int_{\{r=R\}\cap\{\tau_1\le t\le\tau_2\}} i_{\tilde{J}^X[\phi]}\,d\mathrm{vol} = \int_{\tau_1}^{\tau_2}\int_\omega \tfrac{1}{2}r^p(|D_L\psi|^2 - |\not{D}\psi|^2) - \tfrac{1}{2}\partial_t(r^{p+1}|\phi|^2)\,d\omega\,dt.$$

Now notice that there is a cancellation for the boundary terms:

$$-\int_{H_{\tau_1^*}^{v_0}} L(r^{p+1}|\phi|^2)\,dv\,d\omega - \int_{\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}} \underline{L}(r^{p+1}|\phi|^2)\,du\,d\omega + \int_{H_{\tau_2^*}^{v_0}} L(r^{p+1}|\phi|^2)\,dv\,d\omega + \int_{\tau_1}^{\tau_2}\int_\omega \partial_t(r^{p+1}|\phi|^2)\,d\omega\,dt = 0.$$

Therefore in the interior region for the domain $\overline{\mathcal{D}}_{\tau_1}^{\tau_2} \cap \{v \le v_0\}$, the above calculations lead to the $r$-weighted energy identity

$$\int_{H_{\tau_2^*}^{v_0}} r^p |D_L\psi|^2 \, dv \, d\omega + \int_{\tau_1}^{\tau_2} \int_{H_{\tau^*}^{v_0}} r^{p-1}(p|D_L\psi|^2 + (2-p)|\not{D}\psi|^2) \, dv \, d\omega \, d\tau + \int_{\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}} r^p |\not{D}\psi|^2 \, du \, d\omega$$
$$= \int_{H_{\tau_1^*}^{v_0}} r^p |D_L\psi|^2 \, dv \, d\omega - \frac{1}{2} \int_{\tau_1}^{\tau_2} \int_{\omega} r^p (|D_L\psi|^2 - |\not{D}\psi|^2) \, d\omega \, dt$$
$$- \int_{\tau_1}^{\tau_2} \int_{H_{\tau^*}^{v_0}} r^{p-1} \Re(\Box_A \phi \overline{D_L \psi}) + r^p F_{L\mu} J^\mu[\phi] \, dx \, dt. \quad (29)$$

Similarly, in the exterior region $\{r \le R + t\}$ for the domain $\mathcal{D}_{\tau_1}^{\tau_2}$ for all $\tau_2 < \tau_1 \le 0$, we have

$$\int_{H_{\tau_1^*}^{-\tau_2^*}} r^p |D_L\psi|^2 \, dv \, d\omega + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} r^{p-1}(p|D_L\psi|^2 + (2-p)|\not{D}\psi|^2) \, dv \, d\omega \, du + \int_{\underline{H}_{-\tau_2^*}^{\tau_2^*,\tau_1^*}} r^p |\not{D}\psi|^2 \, du \, d\omega$$
$$= \frac{1}{2} \int_{B_{R-\tau_1}^{R-\tau_2}} r^p (|D_L\psi|^2 + |\not{D}\psi|^2) \, dr \, d\omega - \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} r^{p-1} \Re(\Box_A \phi \overline{D_L \psi}) + r^p F_{L\mu} J^\mu[\phi] \, dx \, dt. \quad (30)$$

For the inhomogeneous term, when $p \ge 1 + \epsilon$, we apply the Cauchy–Schwarz inequality directly:

$$2r^{p+1} |\Box_A \phi \cdot \overline{D_L \psi}| \lesssim r^p u_+^{-1-\epsilon} |D_L\psi|^2 + r^{p+2} u_+^{1+\epsilon} |\Box_A \phi|^2.$$

The integral of the first term in the above inequality can be controlled using Gronwall's inequality both in (29) and (30). In particular this shows that estimate (28) follows from (30).

When $p < 1 + \epsilon$, we note that

$$2p - 1 - \epsilon < p \cdot \frac{p}{1+\epsilon} + (p-1)\Big(1 - \frac{p}{1+\epsilon}\Big).$$

Then we can estimate the inhomogeneous term as follows:

$$\begin{aligned} 2r^{p+1} |\Box_A \phi \cdot \overline{D_L \psi}| &\le \epsilon_1 r^{2p-1-\epsilon} u_+^{-p} |D_L\psi|^2 + \epsilon_1^{-1} r^{1+\epsilon+2} u_+^p |\Box_A \phi|^2 \\ &\le \epsilon_1 (r^p u_+^{-1-\epsilon})^{p/(1+\epsilon)} (r^{p-1})^{1-p/(1+\epsilon)} |D_L\psi|^2 + \epsilon_1^{-1} r^{1+\epsilon+2} u_+^p |\Box_A \phi|^2 \\ &\le \epsilon_1 r^p u_+^{-1-\epsilon} |D_L\psi|^2 + \epsilon_1 r^{p-1} |D_L\psi|^2 + \epsilon_1^{-1} r^{1+\epsilon+2} u_+^p |\Box_A \phi|^2 \end{aligned}$$

for all $\epsilon_1 > 0$. The integral of the first term can be controlled using Gronwall's inequality. The integral of the second term can be absorbed for sufficiently small $\epsilon_1$. Then estimate (28) follows.

For the $r$-weighted energy estimate (27) in the interior region, we need to control the boundary term on $\{r = R\}$. It suffices to estimate it for $p = 0$ in (29) by making use of the energy estimate (22). From Hardy's inequality in Lemma 6, we note that

$$\int_{H_\tau} |D_L\psi|^2 \, d\omega \, dv \lesssim E[\phi](\Sigma_\tau).$$

By using the integrated local energy estimate (22), we therefore can show that

$$\begin{aligned}
&\left|\int_{\tau_1}^{\tau_2}\int_{\omega} r^p(|D_L\psi|^2-|\not{D}\psi|^2)\,d\omega\,dt\right| \\
&\qquad\lesssim R^p\int_{H^{v_0}_{\tau_2^*}}|D_L\psi|^2\,dv\,d\omega+\int_{\tau_1}^{\tau_2}\int_{H^{v_0}_{\tau^*}} r^{-1}|\not{D}\psi|^2\,dv\,d\omega\,d\tau+\int_{\underline{H}^{\tau_1^*,\tau_2^*}_{v_0}}|\not{D}\psi|^2\,du\,d\omega \\
&\qquad\qquad+\int_{H^{v_0}_{\tau_1^*}}|D_L\psi|^2\,dv\,d\omega+\int_{\tau_1}^{\tau_2}\int_{H^{v_0}_{\tau^*}} r^{-1}|\Re(\Box_A\phi\overline{D_L\psi})|+|F_{L\mu}J^\mu[\phi]|\,dx\,dt \\
&\qquad\lesssim E[\phi](\Sigma_{\tau_1})+I_0^{1+\epsilon}[\Box_A\phi](\mathcal{D}^{\tau_2}_{\tau_1})+\iint_{\mathcal{D}^{\tau_2}_{\tau_1}}|F_{\underline{L}\mu}J^\mu[\phi]|+|F_{L\mu}J^\mu[\phi]|\,dx\,dt.
\end{aligned}$$

The inhomogeneous term can be bounded using the Cauchy–Schwarz inequality together with the integrated local energy estimates. Once we have the bound for the boundary terms on $\{r=R\}$, the $r$-weighted energy estimate (27) follows from the identity (29) and Gronwall's inequality. □

Next we establish the $r$-weighted energy estimate for the Maxwell field.

**Proposition 10.** *Let $F$ be any 2-form satisfying the Bianchi identity (1). Then in the interior region, for all $0\le\tau_1<\tau_2$ and $v_0\ge\frac{1}{2}(\tau_2+R)$, we have the $r$-weighted energy estimate*

$$\begin{aligned}
&\int_{H_{\tau_2^*}} r^{p+2}|\alpha|^2\,dv\,d\omega \\
&\quad+\int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}} r^{p+1}(p|\alpha|^2+(2-p)(\rho^2+|\sigma|^2))\,dv\,d\omega\,d\tau+\int_{\underline{H}^{\tau_1^*,\tau_2^*}_{v_0}} r^{p+2}(\rho^2+|\sigma|^2)\,du\,d\omega \\
&\qquad\lesssim\int_{H_{\tau_1^*}} r^{p+2}|\alpha|^2\,dv\,d\omega+I^{\max\{p,1+\epsilon\}}_{\min\{1+\epsilon,p\}}[\not{J}](\mathcal{D}^{\tau_2}_{\tau_1})+(2-p)^{-1}I_0^{p+1}[J_L](\mathcal{D}^{\tau_2}_{\tau_1}) \\
&\qquad\qquad+E[F](\Sigma_{\tau_1})+I_0^{1+\epsilon}[|J_L|+|\not{J}|](\mathcal{D}^{\tau_2}_{\tau_1})+\iint_{\mathcal{D}^{\tau_2}_{\tau_1}}|J_{\underline{L}}||\rho|\,dx\,dt \qquad (31)
\end{aligned}$$

*for all $0\le p\le 2$. Similarly in the exterior region, for all $\tau_2<\tau_1\le 0$ and $0\le p\le 2$, we have*

$$\begin{aligned}
&\int_{H^{-\tau_2^*}_{\tau_1^*}} r^p|\alpha|^2r^2\,dv\,d\omega \\
&\quad+\iint_{\mathcal{D}^{\tau_2}_{\tau_1}} r^{p+1}(p|\alpha|^2+(2-p)(\rho^2+|\sigma|^2))\,dv\,d\omega\,du+\int_{\underline{H}^{\tau_2^*,\tau_1^*}_{-\tau_2^*}} r^p(\rho^2+|\sigma|^2)r^2\,du\,d\omega \\
&\qquad\lesssim\int_{B^{R-\tau_2}_{R-\tau_1}} r^p|F|^2\,dx+I^{\max\{p,1+\epsilon\}}_{\min\{p,1+\epsilon\}}[\not{J}](\mathcal{D}^{\tau_2}_{\tau_1})+(2-p)^{-1}I_0^{p+1}[J_L](\mathcal{D}^{\tau_2}_{\tau_1}). \qquad (32)
\end{aligned}$$

*Proof.* Take the vector field

$$X=r^pL=f\,\partial_t+f\,\partial_r$$

in the energy identity (17) for the Maxwell field. Using the computations before (26), we have

$$\begin{aligned} T[F]^{\mu\nu}\pi^X_{\mu\nu} &= T[F]^{\mu\nu}\pi^{f\partial_r}_{\mu\nu} + T[F]^{\mu\nu}\pi^{f\partial_t} \\ &= \big(r^{-1}f - \tfrac{1}{2}f'\big)(\rho^2+|\sigma|^2) + \tfrac{1}{4}f'(|\alpha|^2+|\underline{\alpha}|^2) + \tfrac{1}{4}f'(|\alpha|^2-|\underline{\alpha}|^2) \\ &= \tfrac{1}{2}r^{p-1}((2-p)(\rho^2+|\sigma|^2)+p|\alpha|^2). \end{aligned}$$

For the boundary terms corresponding to the vector field $X = r^p L$, we have

$$\begin{aligned} &i_{J^X[F]}\,d\mathrm{vol} = \tfrac{1}{2}r^p(|\alpha|^2+\rho^2+|\sigma|^2)\,dx, \quad && i_{J^X[F]}\,d\mathrm{vol} = \tfrac{1}{2}r^p(|\alpha|^2-\rho^2-|\sigma|^2)r^2\,dt\,d\omega, \\ &i_{J^X[F]}\,d\mathrm{vol} = r^p|\alpha|^2r^2\,dv\,d\omega, && i_{J^X[F]}\,d\mathrm{vol} = -r^p(\rho^2+|\sigma|^2)r^2\,du\,d\omega \end{aligned}$$

on $\{t=\tau\}$, $\{r=R\}$, $H_u$ and $\underline{H}_v$, respectively. Therefore, for all $0 \le \tau_1 < \tau_2$ and $v_0 \ge \frac{1}{2}(\tau_2+R)$, if we take the region $\mathcal{D}$ bounded by $H_{\tau_1^*}$, $H_{\tau_2^*}$, $\{r=R\}$, $\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}$, we get the $r$-weighted energy identity

$$\begin{aligned} &\int_{H_{\tau_2^*}^{v_0}} r^p|\alpha|^2r^2\,dv\,d\omega \\ &\quad + \int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}^{v_0}} r^{p-1}(p|\alpha|^2+(2-p)(\rho^2+|\sigma|^2))r^2\,dv\,d\omega\,d\tau + \int_{\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}} r^p(\rho^2+|\sigma|^2)r^2\,du\,d\omega \\ &\quad = \int_{H_{\tau_1^*}^{v_0}} r^p|\alpha|^2r^2\,dv\,d\omega - \frac{1}{2}\int_{\tau_1}^{\tau_2}\int_{\omega} r^p(|\alpha|^2-\rho^2-|\sigma|^2)r^2\,d\omega\,dt - \int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}^{v_0}} r^pJ_\nu F_L^\nu\,dx\,dt. \end{aligned} \tag{33}$$

Similarly, in the exterior region $\{r \ge R+t\}$, consider the region $\mathcal{D}_{\tau_1}^{\tau_2}$ for $\tau_2 < \tau_1 \le 0$. We have the following identity:

$$\begin{aligned} &\int_{H_{\tau_1^*}^{-\tau_2^*}} r^p|\alpha|^2r^2\,dv\,d\omega \\ &\quad + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} r^{p-1}(p|\alpha|^2+(2-p)(\rho^2+|\sigma|^2))\,dv\,d\omega\,du + \int_{\underline{H}_{-\tau_2^*}^{\tau_2^*,\tau_1^*}} r^p(\rho^2+|\sigma|^2)r^2\,du\,d\omega \\ &\quad = \frac{1}{2}\int_{B_{R-\tau_1}^{R-\tau_2}} r^p(|\alpha|^2+\rho^2+|\sigma|^2)\,dx - \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} r^pJ_\nu F_L^\nu\,dx\,dt. \end{aligned} \tag{34}$$

To obtain (32), we first note that under the null frame $\{L, \underline{L}, e_1, e_2\}$,

$$F_L^L = -\tfrac{1}{2}F_{L\underline{L}} = \rho, \quad F_L^{\underline{L}} = 0, \quad F_L^{e_j} = \alpha_j, \quad j=1,2.$$

We can use the same method to treat the term $r^p|J_{e_j}F_L^{e_j}|$ as that for $\Box_A\phi\cdot\overline{D_L\psi}$ in Proposition 9 (simply replace $\Box_A\phi$ with $J_{e_j}$ and $\overline{D_L}\psi$ with $r\alpha_j$). For the term involving $\rho$, we estimate

$$r^{p+2}|J_L\cdot\rho| \le \frac{1}{2}(2-p)r^{p+1}|\rho|^2 + \frac{2}{2-p}r^{p+3}|J_L|^2.$$

The integral of the first term could be absorbed. Then the $r$-weighted energy estimate (32) follows from the above $r$-weighted energy identity (34).

We can treat the inhomogeneous term the same way for the $r$-weighted energy estimate in the interior region from the $r$-weighted energy identity (33). Like the case for the scalar field, the boundary term on $\{r = R\}$ can be bounded by taking $p = 0$ in (33) and then by making use of the integrated local energy estimate (24):

$$\begin{aligned}\left|\int_{\tau_1}^{\tau_2}\int_{\omega} r^p(|\alpha|^2-\rho^2-|\sigma|^2)r^2\,d\omega\,dt\right| &\lesssim \int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}^{v_0}}(\rho^2+|\sigma|^2)r\,dv\,d\omega\,d\tau+\int_{\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}}(\rho^2+|\sigma|^2)r^2\,du\,d\omega\\ &\quad+\int_{H_{\tau_2^*}^{v_0}}|\alpha|^2r^2\,dv\,d\omega+\int_{H_{\tau_1^*}^{v_0}}|\alpha|^2r^2\,dv\,d\omega+\int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}^{v_0}}|J_\nu F_L^\nu|\,dx\,dt\\ &\lesssim E[F](\Sigma_{\tau_1})+I_0^{1+\epsilon}[|J_L|+|\not{J}|](\mathcal{D}_{\tau_1}^{\tau_2})+\iint_{\mathcal{D}_{\tau_1}^{\tau_2}}|J_{\underline{L}}||\rho|\,dx\,dt.\end{aligned}$$

This combined with Gronwall's inequality implies the $r$-weighted energy estimate for the Maxwell field in the interior region. □

## 4. Decay estimates for the linear solutions

In this section we derive energy flux decay for both the linear Maxwell field and the linear complex scalar field under appropriate assumptions. We use a bootstrap argument to construct global solutions of the nonlinear (MKG). The first step is to study the decay properties of the linear solutions. Recall that $F = dA$ with $A$ the connection used to define the covariant derivative $D$. Our strategy is that we make assumptions on $J_\mu = \nabla^\nu F_{\mu\nu}$ to obtain estimates for the linear solution $F$. We then use these estimates to derive estimates for the solutions of the linear covariant wave equation $\Box_A\phi = 0$. As in (MKG) the nonlinearity $J[\phi]$ is quadratic in $\phi$, so by making use of the smallness of the scalar field we then can improve the bootstrap assumption on $J$. The difficulties are that the Maxwell field $F$ is no longer small and that there exists nonzero charge.

Assume that the Maxwell field $F = dA$ has charge $q_0$ and splits into the charge part and chargeless part

$$F = \chi_{\{r>t+R\}}q_0r^{-2}\,dt\wedge dr+\overline{F}.$$

Let $J = \delta F$ be the divergence of $F$ and $\not{J} = (J_{e_1}, J_{e_2})$ be the angular component. Let

$$\begin{aligned}m_k &= \sum_{l\le k} I_{1+\epsilon}^{1+\gamma_0}[\mathcal{L}_Z^l\not{J}](\{r\ge R\})+I_0^{2+\gamma_0}[\mathcal{L}_Z^lJ_L](\{r\ge R\})+I_{1+\gamma_0}^{1+\epsilon}[|\mathcal{L}_Z^l\not{J}|+|\mathcal{L}_Z^lJ_L|](\{t\ge 0\})\\ &\quad+I_{1+\gamma_0+2\epsilon}^{1-\epsilon}[\mathcal{L}_Z^lJ_{\underline{L}}](\{t\ge 0\})+I_{1+\gamma_0}^{0}[\overline{\nabla}\mathcal{L}_Z^{l-1}J](\{r\le 2R\})+|q_0|\sup_{\tau\le 0}\tau_+^{1+\gamma_0}\iint_{\mathcal{D}_\tau^{-\infty}}|J_{\underline{L}}|r^{-2}\,dx\,dt,\\ M_k &= m_k+E_0^k[\overline{F}]+1+|q_0|,\end{aligned}\tag{35}$$

where we recall from (6) in Section 2 that $E_0^k[\bar{F}]$ denotes the weighted Sobolev norm of the Maxwell field $\bar{F}$ with weights $r_+^{1+\gamma_0}$ on the initial hypersurface $t=0$. The integral of $|J_{\underline{L}}|r^{-2}$ is used to control the interaction of the nonzero charge with the nonlinearity $J$ in the exterior region.

To derive the energy decay for the Maxwell field, we assume that $M_k$ is finite. This can be fulfilled as follows: the charge $q_0$ is a constant depending on the initial data of the scalar field. $E_0^k[\bar{F}]$ denotes the size of the initial data for the chargeless part of the Maxwell field. Recall that the nonlinearity $J$ is quadratic in the scalar field $\phi$. By using the bootstrap assumption, it is small.

**4.1.** ***Energy decay for the Maxwell field.*** We derive energy flux decay for the Maxwell field $F$ under the assumption that $M_k$ is finite.

**Proposition 11.** *In the interior region for all $0 \le \tau_1 < \tau_2$ and $v_0 \ge \frac{1}{2}(\tau_2 + R)$, we have the following energy flux decay for the Maxwell field*:

$$I_0^{-1-\epsilon}[F](\mathcal{D}_{\tau_1}^{\tau_2}) + \int_{\tau_1}^{\tau_2}\int_{\Sigma_\tau} \frac{\rho^2+|\sigma|^2}{1+r}\,dx\,d\tau + E[F]\big(\underline{H}_{v_0}^{\tau_1^*,\tau_2^*}\big) + E[F](\Sigma_{\tau_1}) \lesssim (\tau_1)_+^{-1-\gamma_0} M_0. \tag{36}$$

*In the exterior region $\{r \le R+t\}$ for all $\tau_2 < \tau_1 \le 0$ and $0 \le p \le 1+\gamma_0$, we have*

$$I_0^{-1-\epsilon}[\bar{F}](\mathcal{D}_{\tau_1}^{\tau_2}) + E[\bar{F}]\big(\underline{H}_{-\tau_2^*}^{\tau_2^*,\tau_1^*}\big) + E[\bar{F}](H_{\tau_1^*}) + (\tau_1)_+^{-p}\int_{H_{\tau_1^*}} r^{p+2}|\alpha|^2\,dv\,d\omega \lesssim (\tau_1)_+^{-1-\gamma_0} M_0. \tag{37}$$

*Here and throughout the paper, $\tau_+ = 1+|\tau|$ for all real numbers $\tau$.*

*Proof.* Let's first consider the estimates in the exterior region. By the definition of $M_0$, we derive that

$$\int_{B_{R-\tau_1}^{R-\tau_2}} r^p|\bar{F}|^2\,dx + I_{1+\epsilon}^p[\not{J}](\mathcal{D}_{\tau_1}^{\tau_2}) + I_0^{p+1}[J_L](\mathcal{D}_{\tau_1}^{\tau_2}) \lesssim (\tau_1)_+^{p-1-\gamma_0} M_0, \quad 0 \le p \le 1+\gamma_0.$$

Here note that in the exterior region, $r \ge \frac{1}{2}u_+$. Then the $r$-weighted energy estimate (32) implies that

$$\int_{H_{\tau_1^*}^{-\tau_2^*}} r^{p+2}|\alpha|^2\,dv\,d\omega + \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} r^{p+1}(|\alpha|^2+\bar{\rho}^2+|\sigma|^2)\,dv\,d\omega\,du \lesssim (\tau_1)_+^{p-1-\gamma_0} M_0.$$

This estimate can be used to bound the integral of $|J_{\underline{L}}||\rho|$ on the right-hand side of (25). Recall that $\rho = q_0 r^{-2} + \bar{\rho}$ when $r \ge R+t$. We then can show that

$$\iint_{\mathcal{D}_{\tau_1}^{\tau_2}} |J_{\underline{L}}||\rho|\,dx\,dt \lesssim \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} (|q_0||J_{\underline{L}}|r^{-2} + |\bar{\rho}|^2 r^{\epsilon-1}u_+^{-\epsilon} + |J_{\underline{L}}|^2 r^{1-\epsilon}u_+^{\epsilon})\,dx\,dt \lesssim M_0(\tau_1)_+^{-1-\gamma_0}.$$

The decay estimate (37) then follows from the energy estimate (25) as

$$E[\bar{F}](B_{R-\tau_1}^{R-\tau_2}) + I_0^{1+\epsilon}[|\not{J}| + |J_L|](\mathcal{D}_{\tau_1}^{\tau_2}) \lesssim (\tau_1)_+^{-1-\gamma_0} M_0.$$

For the decay estimates in the interior region, we use the pigeonhole argument in [Dafermos and Rodnianski 2010]. First, by interpolation, we derive from the definition of $M_0$ that

$$I_{\min\{p,1+\epsilon\}}^{\max\{p,1+\epsilon\}}[\not{J}](\mathcal{D}_{\tau_1}^{\tau_2}) + I_0^{p+1}[J_L](\mathcal{D}_{\tau_1}^{\tau_2}) \lesssim (\tau_1)_+^{p-1-\gamma_0} M_0$$

for all $\epsilon \le p \le 1+\gamma_0$. To bound $|J_{\underline{L}}||\rho|$, we use the Cauchy–Schwarz inequality:

$$\iint_{\mathcal{D}^{\tau_2}_{\tau_1}} |J_{\underline{L}}||\rho|\,dx\,dt \lesssim \iint_{\mathcal{D}^{\tau_2}_{\tau_1}} (\epsilon_1|\rho|^2 r_+^{\epsilon-1} + \epsilon_1^{-1} r_+^{1-\epsilon}|J_{\underline{L}}|^2)\,dx\,dt, \quad \forall \epsilon_1 > 0.$$

Here note that in the interior region, $\rho = \bar{\rho}$. For $\epsilon \le p \le 1+\gamma_0$ and sufficiently small $\epsilon_1$ the first term could be absorbed from the $r$-weighted energy estimates (31) and the second term is bounded above by $M_0(\tau_1)_+^{-1-\gamma_0}$ by the definition of $M_0$.

To apply the pigeonhole argument, we need to control the weighted energy flux through the initial hypersurface $\Sigma_0$ of the interior region. Note that $H_{-R/2} = H_{0^*}$. The bound for the weighted energy flux through $H_{0^*}$ follows from the decay estimate (37) in the exterior region:

$$E[F](H_{0^*}) + \int_{H_{0^*}} r^{3+\gamma_0}|\alpha|^2\,dv\,d\omega \lesssim M_0.$$

Here we note that on the boundary $H_{0^*}$ the charge part has bounded energy. Hence take $p = 1+\gamma_0$, $\tau_1 = 0$ in the $r$-weighted energy estimate (31). We derive that

$$\int_{H^*_{\tau_2}} r^{3+\gamma_0}|\alpha|^2\,dv\,d\omega + \int_0^{\tau_2}\int_{H_{\tau^*}} r^{\gamma_0+2}(|\alpha|^2+|\sigma|^2+\rho^2)\,dv\,d\omega \lesssim M_0, \quad \forall \tau_2 \ge 0.$$

We conclude that there exists a dyadic sequence $\{\tau_n\}$, $n \ge 3$ such that

$$\int_{H^*_{\tau_n}} r^{\gamma_0+2}|\alpha|^2\,dv\,d\omega \lesssim (\tau_n)_+^{-1} M_0, \quad \lambda^{-1}\tau_n \le \tau_{n+1} \le \lambda\tau_n$$

for some constant $\lambda$ depending only on $\gamma_0, \epsilon, R$. Interpolation implies that

$$\int_{H^*_{\tau_n}} r^{1+2}|\alpha|^2\,dv\,d\omega \lesssim (\tau_n)_+^{-\gamma_0} M_0.$$

To bound $|J_{\underline{L}}||\rho|$ on the right-hand side of the energy estimate (24), we interpolate $|\rho|$ between the integrated local energy estimate and the above $r$-weighted energy estimate:

$$\begin{aligned}\iint_{\mathcal{D}^{\tau_2}_{\tau_1}} |J_{\underline{L}}||\rho|\,dx\,dt &\lesssim \iint_{\mathcal{D}^{\tau_2}_{\tau_1}} \left(\epsilon_1|\rho|^2(r_+^{-\epsilon-1} + r_+^{\gamma_0}\tau_+^{-1-\gamma_0}) + \epsilon_1^{-1}\tau_+^{2\epsilon} r_+^{1-\epsilon}|J_{\underline{L}}|^2\right) dx\,dt \\ &\lesssim \epsilon_1 I_0^{-1-\epsilon}[F](\mathcal{D}^{\tau_2}_{\tau_1}) + \epsilon_1^{-1} M_0(\tau_1)_+^{-1-\gamma_0}, \quad \forall 1 > \epsilon_1 > 0.\end{aligned}$$

Here we have used the bound

$$r_+^{\epsilon-1}\tau_+^{-2\epsilon} \le r_+^{-1-\epsilon} + \tau_+^{-1-\gamma_0} r_+^{\gamma_0}.$$

Take $\epsilon_1$ to be sufficiently small. From the energy estimate (24), we then obtain

$$I_0^{-1-\epsilon}[F](\mathcal{D}^{\tau_2}_{\tau_1}) + E[F](\Sigma_{\tau_2}) \lesssim E[F](\Sigma_{\tau_1}) + (\tau_1)_+^{-1-\gamma_0} M_0$$

for all $0 \le \tau_1 < \tau_2$ and $0 < \epsilon_1 < 1$. In particular, we have

$$\int_{\tau_n}^{\tau_2}\int_{\{r\le R\}\cap\{t=\tau\}} |F|^2\,dx\,d\tau \lesssim E[F](\Sigma_{\tau_1}) + (\tau_1)_+^{-1-\gamma_0} M_0.$$

Then combine this integrated local energy estimate with the $r$-weighted energy estimate (31) with $p = 1$. For all $\tau_n \le \tau_2$, we derive that

$$\begin{aligned}\int_{\tau_n}^{\tau_2} E[F](\Sigma_\tau)\,d\tau &\lesssim \int_{\tau_n}^{\tau_2}\int_{\{r\le R\}\cap\{t=\tau\}} |F|^2\,dx\,d\tau + \int_{\tau_n}^{\tau_2}\int_{H_{\tau^*}} (|\alpha|^2+|\sigma|^2+\rho^2) r^2\,dv\,d\omega\\ &\lesssim \int_{H_{\tau_n^*}} r^{1+2}|\alpha|^2\,dv\,d\omega + E[F](\Sigma_{\tau_n}) + (\tau_n)_+^{-\gamma_0} M_0\\ &\lesssim E[F](\Sigma_{\tau_n}) + (\tau_n)_+^{-\gamma_0} M_0.\end{aligned}$$

On the other hand, for all $\tau < \tau_2$, we have

$$E[F](\Sigma_{\tau_2}) \le E[F](\tau) + (\tau)_+^{-1-\gamma_0} M_0 \lesssim E[F](\Sigma_0) + M_0 \lesssim M_0.$$

Then from the previous estimate, we can show that

$$(\tau_2 - \tau_n) E[F](\Sigma_{\tau_2}) \lesssim E[F](\Sigma_{\tau_n}) + (\tau_n)_+^{-\gamma_0} M_0 \lesssim M_0.$$

The above estimate holds for all $\tau_2 \ge \tau_n$. In particular, we obtain the coarse bound

$$E[F](\Sigma_\tau) \lesssim \tau_+^{-1} M_0, \quad \forall \tau \ge 0.$$

Based on this coarse bound, we can take $\tau_2 = \tau_{n+1}$ in the previous estimate. We then can show that

$$(\tau_{n+1} - \tau_n) E[F](\Sigma_{\tau_{n+1}}) \lesssim (\tau_n)^{-\gamma_0} M_0.$$

As $\{\tau_n\}$ is dyadic, we conclude that

$$E[F](\Sigma_{\tau_n}) \lesssim (\tau_n)^{-1-\gamma_0} M_0, \quad \forall n \ge 3.$$

Then using the energy estimate, we can show that for $\tau \in [\tau_n, \tau_{n+1}]$ we have

$$E[F](\Sigma_\tau) \lesssim E[F](\tau_n) + (\tau_n)_+^{-1-\gamma_0} M_0 \lesssim (\tau_n)_+^{-1-\gamma_0} M_0 \lesssim \tau_+^{-1-\gamma_0} M_0.$$

Having this energy flux decay, the integrated local energy decay (36) follows from the integrated local energy estimate (24). □

Since the Lie derivative $\mathcal{L}_Z$ commutes with the linear Maxwell equation from the commutator by Lemma 4, as a corollary of the above energy decay proposition, we also have the energy decay estimates for the higher-order derivatives of the Maxwell field.

**Corollary 12.** *We have the following energy flux decay for the $k$-th derivative of the Maxwell field*:

$$E[\mathcal{L}_Z^k \overline{F}](\Sigma_\tau) \lesssim (\tau)_+^{-1-\gamma_0} M_k, \quad \forall \tau \in \mathbb{R}. \tag{38}$$

This decay estimate then leads to the integrated local energy and $r$-weighted energy estimates for the Maxwell field.

**Remark 13.** By using the finite speed of propagation, the estimates in the above proposition and corollary in the exterior region depend only on the data and $J$ in the exterior region $\{t + R \le r\}$ instead of the whole spacetime. Therefore the quantity $M_k$ can be replaced by the corresponding one defined in the exterior region. However, the estimates in the interior region rely on the data in the whole space.

**4.2.** ***Pointwise bounds for the Maxwell field.*** The energy decay estimates derived in the previous section are sufficient to obtain pointwise bounds for the Maxwell field $F$ after commuting the equation with vector fields in $\Gamma = \{\partial_t, \Omega\}$ sufficiently many times; e.g., in [Yang 2015b], four derivatives were used to show the pointwise bound for the solution. The aim of this section is to derive the pointwise bound for the Maxwell field $F$ merely assuming $M_2$ is finite, that is, we commute the equation with $\Gamma$ only twice. The difficulty is that we are not able to use Klainerman–Sobolev embedding to derive the decay of the solution directly as in [Lindblad and Sterbenz 2006]. Our idea is that in the inner region $\{r \le R\}$ we rely on elliptic estimates. In the outer region $\{r \ge R\}$, we analyze the solutions under the null coordinates $(u, v, \omega)$. The angular momentum $\Omega$ can be viewed as the derivative on $\omega$. The pointwise bound then follows by using a trace theorem on the null hypersurfaces and a Sobolev embedding on the sphere. Since we do not commute the equation with $L$ nor $\underline{L}$, those necessary energy estimates heavily rely on the null equations given in Lemma 5.

Let's first consider the pointwise bound for the Maxwell field in the inner region $\{r \le R\}$. To derive the pointwise bound, we use the vector fields $\partial_t$ and the angular momentum $\Omega$ as commutators. Note that the angular momentum vanishes at $r = 0$. In particular we are not able to get the robust estimates for the solution in the bounded region $\{r \le R\}$ merely from the angular momentum. We thus rely on the Killing vector field $\partial_t$ and elliptic estimates. The following proposition gives the estimates for the Maxwell field $F$ on the bounded region $\{r \le R\}$.

**Proposition 14.** *For all* $0 \le \tau$ *and* $0 \le \tau_1 < \tau_2$, *we have*

$$\int_{\tau_1}^{\tau_2} \sup_{|x| \le R} |F|^2(\tau, x)\, d\tau \lesssim \int_{\tau_1}^{\tau_2} \int_{r \le R} |\nabla^2 F|^2\, dx\, dt \lesssim M_2 (\tau_1)_+^{-1-\gamma_0}, \tag{39}$$

$$|F|^2(\tau, x) \lesssim M_2 \tau_+^{-1-\gamma_0}, \quad \forall |x| \le R. \tag{40}$$

**Remark 15.** Estimate (40) gives the pointwise bound for $F$ in the inner region $\{r \le R\}$ but it is weaker than the integral version (39) in the sense of decay rate. It is this integrated decay estimate that allows us to control the nonlinearities in the inner region. In other words, it is not necessary to show the improved decay of the solution in the inner region by using our approach; see, e.g., [Luk 2010]. However this does not mean that our method is not able to obtain the improved decay in the inner region. The improved decay can be derived by commuting the equation with the vector field $L$. For details about this, we refer to [Schlue 2013].

*Proof of Proposition 14.* We use elliptic estimates to prove this proposition. At fixed time $t$, let $E$ and $H$ be the electric and magnetic parts of the Maxwell field $F$. Let $B_r$ be the ball with radius $r$, that is, $B_r = \{t \mid |x| \le r\}$. The Maxwell equation can be written as

$$\operatorname{div}(E) = J_0, \quad \partial_t H + \operatorname{curl}(E) = 0,$$
$$\operatorname{div}(H) = 0, \quad \partial_t E - \operatorname{curl}(H) = \bar{J},$$

where $\bar{J} = (J_1, J_2, J_3)$ is the spatial part of $J$. Therefore, using elliptic theory we derive that

$$\sum_{k\le 1} \|\partial_t^k F\|^2_{H^1_x(B_{3R/2})} \le \sum_{k\le 1} \|\partial_t^k H\|^2_{H^1_x(B_{3R/2})} + \|\partial_t^k E\|^2_{H^1_x(B_{3R/2})} \lesssim \sum_{k\le 1} \|\partial_t^k J\|^2_{L^2_x(B_{2R})} + \|\partial_t^{k+1} F\|^2_{L^2_x(B_{2R})}.$$

Make use of the above estimates with $k = 1$. Differentiate the linear Maxwell equation with the spatial covariant derivative $\overline{\nabla}$. Using elliptic estimates again, we then obtain

$$\|\nabla F\|^2_{H^1_x(B_R)} \lesssim \|\nabla J\|^2_{L^2_x(B_{2R})} + \|\partial_t^2 F\|^2_{L^2_x(B_{2R})}.$$

Here we omitted the lower-order terms. Integrate the above inequality from time $\tau_1$ to $\tau_2$. We derive

$$\begin{aligned}\int_{\tau_1}^{\tau_2}\int_{r\le R} |\nabla^2 F|^2\,dx\,dt &\lesssim \int_{\tau_1}^{\tau_2}\int_{r\le 2R} |\partial_t^2 F|^2 + |\nabla J|^2\,dx\,dt \\ &\lesssim I_0^{-1-\epsilon}[\partial_t^2 F](\mathcal{D}^{\tau_2}_{\tau_1^+}) + I_0^{-1-\epsilon}[\partial_t J](\mathcal{D}^{\tau_2}_{\tau_1^+}) + I_0^0[\overline{\nabla} J](\mathcal{D}^{\tau_2}_{\tau_1}\cap\{r\le 2R\}) \\ &\lesssim M_2(\tau_1)_+^{-1-\gamma_0}.\end{aligned}$$

Here $\tau_1^+ = \max\{\tau_1 - R, 0\}$. The estimate (39) then follows using Sobolev embedding.

For the pointwise bound (40), first we note that

$$\int_{r\le 2R} |\overline{\nabla} J|^2\,dx \lesssim \sum_{k\le 1}\int_{\tau}^{\tau+1} |\overline{\nabla}\mathcal{L}_Z^k J|^2\,dx\,dt \lesssim M_2\tau_+^{-1-\gamma_0}.$$

Consider the energy estimate on the region $\mathcal{D}_1$ bounded by $\Sigma_{\tau^+}$, $\tau^+ = \max\{\tau - R, 0\}$ and $t = \tau$, $\tau \ge 0$. From the energy estimate (24), we conclude that

$$\int_{r\le 2R} |\mathcal{L}_Z^2 F|^2\,dx = E[\mathcal{L}_Z^2 F](r\le 2R) \lesssim E[\mathcal{L}_Z^2 F](\Sigma_{\tau^+}) + I_0^{1+\epsilon}[\mathcal{L}_Z^2 J](\mathcal{D}_1) \lesssim M_2\tau_+^{-1-\gamma_0}.$$

Thus the pointwise bound (40) holds. □

To show the decay of the solution via the energy flux through the null hypersurface, we rely on the following trace theorem.

**Lemma 16.** *Let $f(r,\omega)$ be a smooth function defined on $[a,b]\times\mathbb{S}^2$. Then*

$$\left(\int_\omega |f|^4(r_0,\omega)\,d\omega\right)^{1/2} \le C\int_a^b\int_\omega |f|^2 + |\partial_r f|^2 + |\partial_\omega f|^2\,d\omega\,dr, \quad \forall r_0\in[a,b] \tag{41}$$

*for some constant $C$ independent of $r_0$.*

*Proof.* The condition implies that $f \in H^1_{r,\omega}$. By using the trace theorem, we have

$$\|f(r_0,\cdot)\|_{H^{1/2}_\omega} \le C\|f\|_{H^1_{r,\omega}}, \quad \forall r_0\in[a,b].$$

The lemma then follows using Sobolev embedding on the sphere. □

Using this lemma, we are now able to show the pointwise bound for the Maxwell field when $\{r \geq R\}$.

**Proposition 17.** *Let $\overline{\mathcal{D}}_{\tau_1} = \mathcal{D}_{\tau_1} \cap \{r \geq R\}$. Then we have*

$$\|r\mathcal{L}_Z^k \underline{\alpha}\|^2_{L^2_u L^\infty_v L^2_\omega(\overline{\mathcal{D}}_{\tau_1})} \lesssim M_2(\tau_1)_+^{-1-\gamma_0+2\epsilon}, \quad k = 0, 1, \tag{42}$$

$$|r\underline{\alpha}|^2(\tau, v, \omega) \lesssim M_2 \tau_+^{-1-\gamma_0}, \tag{43}$$

$$r^p(|r\alpha|^2 + |r\sigma|^2)(\tau, v, \omega) \lesssim M_2 \tau_+^{p-1-\gamma_0}, \quad 0 \leq p \leq 1 + \gamma_0, \tag{44}$$

$$r^p |r\bar{\rho}|^2(\tau, v, \omega) \lesssim M_2 \tau_+^{p-1-\gamma_0}, \quad 0 \leq p \leq 1 - \epsilon, \tag{45}$$

$$\|r\mathcal{L}_Z^k \sigma\|^2_{L^2_v L^\infty_u L^2_\omega(\overline{\mathcal{D}}_\tau)} \lesssim M_2 \tau_+^{-1-\gamma_0+\epsilon}, \quad k \leq 1. \tag{46}$$

*Here recall that $Z$ is a vector field in the set $\Gamma = \{\partial_t, \Omega_{ij}\}$.*

**Remark 18.** In terms of decay rate, the integral version (42) is stronger than the pointwise bound (43). We are not able to improve the $u$ decay of the Maxwell field due to the weak decay rate of the initial data. However the integral version improves one order of decay in $u$ (or $\tau$ as $u = \frac{1}{2}(\tau - R)$). This is the key point that allows us to construct the global solution with the weak decay rate of the initial data.

*Proof of Proposition 17.* For the integral estimate (42), we rely on the transport equation (8) for $\underline{\alpha}$. For the case in the exterior region, one can choose the initial hypersurface $\{t = 0\}$. In the interior region, for all $0 \leq \tau_1 < \tau_2$, we can choose the incoming null hypersurface $\underline{H}^{\tau_1^*, \tau_2^*}_{(\tau_2+R)/2}$. Let's only consider the case in the interior region. From (8) for $\underline{\alpha}$ under the null frame, for $k = 0$ or 1, we can show that

$$\begin{aligned}
\|r\mathcal{L}_Z^k \underline{\alpha}\|^2_{L^2_u L^\infty_v L^2_\omega(\overline{\mathcal{D}}^{\tau_2}_{\tau_1})} &\lesssim E[\mathcal{L}_Z^k \underline{\alpha}]\big(\underline{H}^{\tau_1^*, \tau_2^*}_{(\tau_2+R)/2}\big) + I_0^{-1-\epsilon}[\mathcal{L}_Z^k \underline{\alpha}](\overline{\mathcal{D}}^{\tau_2}_{\tau_1}) + \|r^{(1+\epsilon)/2} L\mathcal{L}_Z^k (r\underline{\alpha})\|^2_{L^2_u L^2_v L^2_\omega(\overline{\mathcal{D}}^{\tau_2}_{\tau_1})} \\
&\lesssim M_2(\tau_1)_+^{-1-\gamma_0} + \|r^{(1+\epsilon)/2}(|\mathcal{L}_Z^{k+1}\rho| + |\mathcal{L}_Z^{k+1}\sigma|)\|^2_{L^2_u L^2_v L^2_\omega(\overline{\mathcal{D}}^{\tau_2}_{\tau_1})} + I_0^{1+\epsilon}[\mathcal{L}_Z^k \not{J}](\overline{\mathcal{D}}^{\tau_2}_{\tau_1}) \\
&\lesssim M_2(\tau_1)_+^{-1-\gamma_0+2\epsilon}.
\end{aligned}$$

Here we use interpolation to bound $\rho$ and $\sigma$. Indeed, the integrated local energy estimate implies that

$$\iint_{\overline{\mathcal{D}}^{\tau_2}_{\tau_1}} r^{-\epsilon+1}(|\mathcal{L}_Z^{k+1}\rho|^2 + |\mathcal{L}_Z^{k+1}\sigma|^2)\, du\, dv\, d\omega \lesssim M_2(\tau_1)_+^{-1-\gamma_0}.$$

On the other hand, the $r$-weighted energy estimate shows that

$$\iint_{\overline{\mathcal{D}}^{\tau_2}_{\tau_1}} r^{2+\gamma_0}(|\mathcal{L}_Z^{k+1}\rho|^2 + |\mathcal{L}_Z^{k+1}\sigma|^2)\, du\, dv\, d\omega \lesssim M_2.$$

Interpolation then implies the estimate for $\rho$ and $\sigma$. Thus estimate (42) holds.

For the pointwise bound (43) for $\underline{\alpha}$, we rely on the energy flux on the incoming null hypersurface together with Lemma 16. Consider the point $(\tau, v, \omega)$. In the exterior region when $\tau < 0$, let $\underline{H}_\tau = \underline{H}_v^{\tau^*, -v}$ be the incoming null hypersurface extending to the initial hypersurface $\{t = 0\}$. In the interior region when $\tau \geq 0$, we instead let $\underline{H}_\tau$ be $\underline{H}_v^{\tau, 2v-R}$, which is the incoming null hypersurface truncated by $\{r = R\}$.

From the energy estimates (24) and (25), we conclude that

$$\int_{\underline{H}_\tau} |r\mathcal{L}_Z^k \underline{\alpha}|^2 \, du \, d\omega \lesssim E[\mathcal{L}_Z^k F](\underline{H}_\tau) \lesssim M_2 \tau_+^{-1-\gamma_0}, \quad \forall k \leq 2.$$

As $Z$ may be $\partial_t$ or the angular momentum $\Omega$, to apply Lemma 16, we need the energy flux of the tangential derivative $\underline{L}(\underline{\alpha})$. We make use of the structure equation (8), which implies that

$$\begin{aligned}
\int_{\underline{H}_\tau} |\underline{L}\mathcal{L}_Z^k(r\underline{\alpha})|^2 \, du \, d\omega &\lesssim \int_{\underline{H}_\tau} (|L\mathcal{L}_Z^k(r\underline{\alpha})|^2 + |\mathcal{L}_{\partial_t}\mathcal{L}_Z^k(r\underline{\alpha})|^2) \, du \, d\omega \\
&\lesssim \int_{\underline{H}_\tau} (|\mathcal{L}_Z^{k+1}\rho|^2 + |\mathcal{L}_Z^{k+1}\sigma|^2 + |\mathcal{L}_Z^k(r\not{J})|^2 + |\mathcal{L}_Z^{k+1}(r\underline{\alpha})|^2) \, du \, d\omega \\
&\lesssim E[\mathcal{L}_Z^{k+1}F](\underline{H}_\tau) + I_0^0[\mathcal{L}_Z^{k+1}\not{J}](\mathcal{D}_\tau) \\
&\lesssim M_2 \tau_+^{-1-\gamma_0}, \quad k \leq 1.
\end{aligned}$$

Here note that $\Omega = (re_1, re_2)$. Then by Lemma 16, for all $v$ and fixed $\tau$,

$$\left( \int_\omega |r\mathcal{L}_Z^k \underline{\alpha}|^4(\tau, v, \omega) \, d\omega \right)^{1/2} \lesssim M_2 \tau_+^{-1-\gamma_0}, \quad k \leq 1.$$

Estimate (43) then follows using Sobolev embedding on the sphere.

For the pointwise bound (44), (45) for $\alpha$, $\sigma$, $\bar{\rho}$, the proof for $\alpha$ is slightly different from that of $\sigma$ and $\bar{\rho}$. However, the idea is the same. Let's consider $\alpha$ first. Consider $H_{\tau^*}$, $\tau \in \mathbb{R}$. The $r$-weighted energy estimates (31), (32) imply that

$$\int_{H_{\tau^*}} r^p |r\mathcal{L}_Z^k \alpha|^2 \, dv \, d\omega \lesssim M_2 \tau_+^{p-1-\gamma_0}, \quad \forall 0 \leq p \leq 1 + \gamma_0, \quad k \leq 2.$$

To apply Lemma 16, we need the energy flux of the tangential derivative $L(r\alpha)$. Similar to the case of $\alpha$, we make use of (10) and the $\partial_t$ derivative:

$$\begin{aligned}
\int_{H_{\tau^*}} r^p |L(r\mathcal{L}_Z^k \alpha)|^2 \, dv \, d\omega &\lesssim \int_{H_{\tau^*}} r^p (|\underline{L}(r\mathcal{L}_Z^k \alpha)|^2 + |r\partial_t \mathcal{L}_Z^k \alpha|^2) \, dv \, d\omega \\
&\lesssim \int_{H_{\tau^*}} r^p (|\mathcal{L}_Z^{k+1}\rho|^2 + |\mathcal{L}_Z^{k+1}\sigma|^2 + |\mathcal{L}_Z^k(r\not{J})|^2 + |\mathcal{L}_Z^{k+1}(r\alpha)|^2) \, dv \, d\omega \\
&\lesssim M_2 \tau_+^{p-1-\gamma_0} + \int_{H_{\tau^*}} r^2 (|\mathcal{L}_Z^{k+1}\rho|^2 + |\mathcal{L}_Z^{k+1}\sigma|^2) + r^p |\mathcal{L}_Z^k(r\not{J})|^2 \, dv \, d\omega \\
&\lesssim M_2 \tau_+^{p-1-\gamma_0} + E[\mathcal{L}_Z^k F](H_{\tau^*}) + I_0^p[\mathcal{L}_Z^{k+1}\not{J}](\mathcal{D}_\tau) \\
&\lesssim M_2 \tau_+^{p-1-\gamma_0}
\end{aligned}$$

for $k \leq 1$. The estimate for $\alpha$ then follows from Lemma 16 together with Sobolev embedding on the unit sphere.

For $\bar{\rho}$, $\sigma$, we make use of the $r$-weighted energy estimates (31), (32) through the incoming null hypersurface $\underline{H}_\tau$ defined as above. First, we have

$$\int_{\underline{H}_\tau} r^{p-2}(|\mathcal{L}_Z^k(r^2\bar{\rho})|^2 + |\mathcal{L}_Z^k(r^2\sigma)|^2)\, du\, d\omega \lesssim M_2\tau_+^{p-1-\gamma_0}, \quad k \le 2.$$

To derive the tangential derivative $\underline{L}(r^2\bar{\rho})$, $\underline{L}(r^2\sigma)$, we use the equations (7) and (9). We can show that

$$\begin{aligned}\int_{\underline{H}_\tau} r^{p-2}(|\underline{L}(r^2\mathcal{L}_Z^k\bar{\rho})|^2\, du\, d\omega &\lesssim \int_{\underline{H}_\tau} r^{p-2}(|r\mathcal{L}_Z^{k+1}\underline{\alpha}|^2 + |r^2\mathcal{L}_Z^k J_{\underline{L}}|^2)\, du\, d\omega\\ &\lesssim E[\mathcal{L}_Z^{k+1}F](\underline{H}_\tau) + I_0^p[\mathcal{L}_Z^{k+1}J_{\underline{L}}](\mathcal{D}_\tau)\\ &\lesssim M_2\tau_+^{p-1-\gamma_0}, \quad k = 0, 1\end{aligned}$$

for all $0 \le p \le 1-\epsilon$. We cannot extend $p$ to the full range of $[0, 1+\gamma_0]$ due the weak assumption on $J_{\underline{L}}$. The equation (9) for $\sigma$ does not involve $J_{\underline{L}}$. We hence have the full range $0 \le p \le 1+\gamma_0$ for $\sigma$. Lemma 16 and Sobolev embedding on the sphere then lead to the pointwise bound for $\bar{\rho}$ and $\sigma$. We thus have shown estimates (44), (45).

Finally, for the integrated decay estimates (46), we proceed by integrating along the incoming null hypersurface. In the interior region case we integrate from $\{r = R\}$, while in the exterior region we integrate from the initial hypersurface $\{t = 0\}$. Let's only prove (46) for the interior region case. In particular, take $\overline{\mathcal{D}}_\tau$ to be $\overline{\mathcal{D}}_{\tau_1}^{\tau_2}$ for $0 \le \tau_1 < \tau_2$. First, using the decay estimate (39) for $F$ when $r \le R$, we can show that on the boundary $\{r = R\}$,

$$\int_{\tau_1}^{\tau_2}\int_\omega |\mathcal{L}_Z^k F|^2(\tau, R, \omega)\, d\omega\, d\tau \lesssim \int_{\tau_1}^{\tau_2}\int_{r\le R} |\nabla\mathcal{L}_Z^k F|^2\, dx\, d\tau \lesssim M_2(\tau_1)_+^{-1-\gamma_0}.$$

Then from the transport equations (7) and (9), we can show that

$$\begin{aligned}\|r\mathcal{L}_Z^k\sigma\|^2_{L_v^2L_u^\infty L_\omega^2(\overline{\mathcal{D}}_{\tau_1}^{\tau_2})} &\lesssim \int_{\tau_1}^{\tau_2} |\mathcal{L}_Z^k F|^2(\tau, R, \omega)\, d\omega\, d\tau + \iint_{\overline{\mathcal{D}}_{\tau_1}^{\tau_2}} (r|\mathcal{L}_Z^k\sigma|^2 + |\mathcal{L}_Z^k\sigma\cdot\underline{L}(r^2\mathcal{L}_Z^k\sigma)|)\, du\, dv\, d\omega\\ &\lesssim M_2(\tau_1)_+^{-1-\gamma_0} + \iint_{\overline{\mathcal{D}}_{\tau_1}^{\tau_2}} (r^{1+\epsilon}|\mathcal{L}_Z^k\sigma|^2 + r^{1-\epsilon}|\mathcal{L}_Z^{k+1}\underline{\alpha}|^2)\, du\, dv\, d\omega\\ &\lesssim M_2(\tau_1)_+^{-1-\gamma_0} + M_2(\tau_1)_+^{-1-\gamma_0+\epsilon} \lesssim M_2(\tau_1)_+^{-1-\gamma_0+\epsilon}.\end{aligned}$$

Here we have used the $r$-weighted energy estimates for $\sigma$ with $p = \epsilon$ and the integrated local energy estimates to bound $\underline{\alpha}$. This proves (46). □

**4.3.** ***Energy decay for the scalar field.*** In this section, we study the energy decay for the complex scalar field $\phi$ satisfying the linear covariant wave equation. When the connection field $A$ is trivial, the energy decay has been well studied using the new approach; see, e.g., [Yang 2013]. For a general connection field $A$, presumably not small, new difficulty arises as there are interaction terms between the curvature $dA$ and the scalar field. In the previous subsection, we derived the energy flux decay for the Maxwell

field $F = dA$ with appropriate bound on $J = \delta F$. The purpose of this section is to derive energy flux decay for the complex scalar field.

In addition to the assumption that $M_k$ is finite, for the general complex scalar field $\phi$, we assume the inhomogeneous term $\Box_A\phi$ and the initial data are bounded in the norm

$$\mathcal{E}_k[\phi] = E_0^k[\phi] + \sum_{l\le k} + I_{1+\epsilon}^{1+\gamma_0}[D_Z^l \Box_A\phi](\{t \ge 0\}) + I_{1+\gamma_0}^{1+\epsilon}[D_Z^l \Box_A\phi](\{t \ge 0\}). \tag{47}$$

Here in this section we will estimate the general complex scalar field $\phi$ in terms of the initial data and the inhomogeneous term $\Box_A\phi$. For solutions of (MKG), the complex scalar field $\phi$ verifies the linear covariant wave equation $\Box_A\phi = 0$. In particular, if $(\phi, A)$ solves (MKG), then $\mathcal{E}_k[\phi] = E_0^k[\phi]$, which denotes the weighted Sobolev norm of the initial data for the complex scalar field.

As the estimates in the interior region require information on the boundary $\Sigma_0$, which contains the boundary $H_{0*}$ of the exterior region, we need first to obtain the energy decay estimates in the exterior region. The main difficulty in the presence of a nontrivial connection field is to control the interaction term $(dA)_{\mu\nu}J^\nu[\phi]$ under mild assumptions on the curvature $dA$. In the integrated local energy estimate (23) for the scalar field, it is not possible to control or absorb those terms as there is no smallness assumption on $dA$. The idea is to make use of the null structure of $J^\nu[\phi]$ together with the $r$-weighted energy estimate (28). More precisely, we first control those terms in the $r$-weighted energy estimate via Gronwall's inequality. Then we estimate those terms in the integrated local energy estimates. Once we have control of those interaction terms, the decay of the energy flux follows from the standard argument of the new approach, similar to that of the energy decay for the Maxwell field in the previous section.

We first prove a lemma used to control the scalar field $\phi$ by using the $r$-weighted energy.

**Lemma 19.** *Assume $\phi$ vanishes at null infinity. In the exterior region on $H_u$, we have*

$$\int_\omega |r\phi|^2(u, v, \omega)\, d\omega \lesssim \int_\omega |r\phi|^2(u, -u, \omega)\, d\omega + \beta^{-1}u_+^{-\beta}\int_{-u}^{v}\int_\omega r^{1+\beta}|D_L(r\phi)|^2\, dv\, d\omega, \quad \forall \beta > 0. \tag{48}$$

*In the interior region on $\Sigma_\tau$, for $1 \le p \le 2$, we have*

$$\int_\omega r^p|\phi|^2\, d\omega \lesssim (E[\phi](\Sigma_\tau))^{\delta_p}(I_0^{1+\gamma_0}[r^{-1}D_L(r\phi)](H_{\tau*}))^{1-\delta_p}, \quad \delta_p = \frac{2+\gamma_0-p}{1+\gamma_0}. \tag{49}$$

*Moreover on $\Sigma_\tau$, $\tau \in \mathbb{R}$, we have*

$$r\int_\omega |\phi|^2\, d\omega \lesssim \epsilon_1^{-1}\int_\Sigma |\phi|^2\, d\tilde{v}\, d\omega + \epsilon_1 E[\phi](\Sigma_\tau) \tag{50}$$

*for all $0 < \epsilon_1 \le 1$. Here $(\tilde{v}, \omega) = (v, \omega)$ when $r \ge R$ or $(r, \omega)$ when $r < R$.*

*Proof.* Estimate (48) follows from the inequality

$$|r\phi|(u, v, \omega) \le |r\phi|(u, -u, \omega) + \int_{-u}^{v}|D_L(r\phi)|\, dv$$

followed by the Cauchy–Schwarz inequality.

In the interior region, the problem is that we cannot integrate from the initial hypersurface nor the boundary $H_{0^*}$ nor the null infinity as the behavior of $r\phi$ at null infinity is unknown (generically not zero). However, the scalar field $\phi$ vanishes at null infinity. We thus can bound $r|\phi|^2$ by the energy flux through $\Sigma_\tau$. More precisely, on $\Sigma$ we can show that

$$\begin{aligned} r\int_\omega |\phi|^2\,d\omega &\lesssim \int_{\Sigma_\tau} |\phi|^2\,d\tilde{v}\,d\omega + \int_{\Sigma_\tau} r|D_{\tilde{v}}\phi||\phi|\,d\tilde{v}\,d\omega \\ &\lesssim \epsilon_1 \int_{\Sigma_\tau} |D_{\tilde{v}}\phi|^2 r^2\,d\tilde{v}\,d\omega + (\epsilon_1^{-1}+1)\int_{\Sigma_\tau}|\phi|^2\,d\tilde{v}\,d\omega \\ &\lesssim \epsilon_1 E[\phi](\Sigma_\tau) + \epsilon_1^{-1}\int_{\Sigma_\tau}|\phi|^2\,d\tilde{v}\,d\omega. \end{aligned}$$

This gives estimate (50). In particular, for $\epsilon_1 = 1$, from Hardy's inequality (21) we conclude that estimate (49) holds for $p = 1$. To prove it for all $1 \le p \le 2$, it suffices to show the estimate with $p = 2$. Consider the sphere with radius $r = \frac{1}{2}(\tau^* + v)$ on $H_{\tau^*} \subset \Sigma_\tau$. Choose the sphere with radius $r_1 = \frac{1}{2}(\tau^* + v_1)$ such that

$$r_1^{1+\gamma_0} = E[\phi](\Sigma_\tau)^{-1}\int_{H_{\tau^*}} r^{1+\gamma_0}|D_L(r\phi)|^2\,dv\,d\omega.$$

If $r \le r_1$, then (49) with $p = 2$ follows from (49) with $p = 1$. Otherwise, we have $r_1 < r$. Then

$$\begin{aligned} \int_\omega |r\phi|^2(\tau^*, v, \omega)\,d\omega &\lesssim \int_\omega |r\phi|^2(\tau^*, v_1, \omega) + r_1^{-\gamma_0}\int_{H_{\tau^*}} r^{1+\gamma_0}|D_L(r\phi)|^2\,dv\,d\omega \\ &\lesssim r_1 E[\phi](\Sigma_\tau) + r_1^{-\gamma_0} I_0^{1+\gamma_0}[r^{-1}D_L\psi](H_{\tau^*}) \\ &\lesssim (E[\phi](\Sigma_\tau))^{\gamma_0/(1+\gamma_0)}(I_0^{1+\gamma_0}[r^{-1}D_L\psi](H_{\tau^*}))^{1/(1+\gamma_0)}. \end{aligned}$$

Here we recall the notation $I$ defined in Section 2. □

The following lemma is very simple but it turns out to be very useful.

**Lemma 20.** *Suppose $f(\tau)$ is smooth. Then for any $\beta \neq 0$, we have the identity*

$$\int_{\tau_1}^{\tau_2} s^\beta f(s)\,ds = \beta\int_{\tau_1}^{\tau_2}\tau^{\beta-1}\int_\tau^{\tau_2} f(s)\,ds\,d\tau + \tau_1^\beta\int_{\tau_1}^{\tau_2} f(s)\,ds.$$

**4.3.1.** *Energy decay in the exterior region.* In the exterior region, as $r \ge \frac{1}{3}u_+$, it suffices to consider the $r$-weighted energy estimate for the largest $p = 1 + \gamma_0$. First we can show the following proposition.

**Proposition 21.** *In the exterior region, for all $\tau_2 < \tau_1 \le 0$, we have*

$$\begin{aligned} \iint_{\mathcal{D}_{\tau_1}^{\tau_2}} r^{1+\gamma_0}|F_{L\mu}J^\mu[\phi]|\,dx\,dt \lesssim M_2 E_0^0[\phi] + M_2\int_u u_+^{-1-\epsilon}\int_v r^{1+\gamma_0}|D_L\psi|^2\,dv\,d\omega\,du \\ + |q_0|\iint_{\mathcal{D}_{\tau_1}^{\tau_2}} r^{\gamma_0}(|D_L(r\phi)|^2 + |\not{D}(r\phi)|^2)\,dv\,du\,d\omega. \end{aligned} \tag{51}$$

*Proof.* As $F = dA$ has different decay properties for different components, we estimate the integral according to the index $\mu$. Denote $\psi = r\phi$. Note that $r^2 J[\phi] = J[r\phi]$. For $\mu = \underline{L}$, we have

$$|F_{L\underline{L}} J^{\underline{L}}[\phi]| \lesssim r^{-2}|q_0|\,|D_L\psi||\psi| + |\bar{\rho}||D_L\psi||\psi|. \tag{52}$$

The first term on the right-hand side will be absorbed with the smallness assumption on the charge $q_0$ (as the data for the scalar field is small). Indeed, using Lemma 6 we can show that

$$\begin{aligned} 2\iint r^{\gamma_0-1}|D_L\psi||\psi|\,du\,dv\,d\omega &\le \iint r^{\gamma_0}|D_L\psi|^2\,dv\,du\,d\omega + \iint r^{\gamma_0}|\phi|^2\,dv\,du\,d\omega \\ &\lesssim \iint r^{\gamma_0}|D_L\psi|^2\,dv\,du\,d\omega + \int_u\int_\omega (r^{1+\gamma_0}|\phi|^2)(u,-u,\omega)\,d\omega\,du \\ &\lesssim \iint r^{\gamma_0}|D_L\psi|^2\,dv\,du\,d\omega + E_0^0[\phi]. \end{aligned}$$

For the second term on the right-hand side of (52), the idea is that we use the Cauchy–Schwarz inequality and make use of the $r$-weighted energy estimate. First, we can estimate that

$$2r^{1+\gamma_0}|\bar{\rho}||D_L\psi||\psi| \le r^{1+\gamma_0}|D_L\psi|^2 u_+^{-1-\epsilon} + u_+^{1+\epsilon} r^2|\bar{\rho}|^2 r^{1+\gamma_0}|\phi|^2.$$

The first term will be controlled through Gronwall's inequality. For the second term, we can first use Sobolev embedding on the unit sphere to bound $\bar{\rho}$ and then apply Lemma 19:

$$\begin{aligned} &\iint u_+^{1+\epsilon} r^2|\bar{\rho}|^2 r^{1+\gamma_0}|\phi|^2\,du\,dv\,d\omega \\ &\quad\lesssim \int_u u_+^{1+\epsilon}\int_v \sum_{j\le 2} r^2\int_\omega |\mathcal{L}_\Omega^j\bar{\rho}|^2\,d\omega\cdot\int_\omega r^{1+\gamma_0}|\phi|^2\,d\omega\,dv\,du \\ &\quad\lesssim \int_u u_+^{1+\epsilon-1} E[\mathcal{L}_Z^2\overline{F}](H_u)\left(u_+^{\gamma_0}\int_\omega |r\phi|^2(u,-u,\omega)\,d\omega + \int_v\int_\omega r^{1+\gamma_0}|D_L\psi|^2\,dv\,d\omega\right)du \\ &\quad\lesssim M_2\int_{|x|\ge R} r_+^{1+\gamma_0-\epsilon-2}|\phi|^2(0,x)\,dx + M_2\int_u u_+^{-1-\epsilon}\int_v r^{1+\gamma_0}|D_L\psi|^2\,dv\,d\omega\,du. \end{aligned}$$

The first term is bounded by the weighted Sobolev norm of the initial data. The second term can be controlled by using Gronwall's inequality. Thus estimate (51) holds for the case $\mu = L$.

For $\mu = e_1$ or $e_2$, first we can bound

$$r^{1+\gamma_0}|F_{Le_j}||J^{e_j}[\psi]| \le \epsilon_1 r^{\gamma_0}|\nabla\!\!\!/\psi|^2 + \epsilon_1^{-1} r^{3+\gamma_0}|\alpha|^2 r|\phi|^2, \quad \forall \epsilon_1 > 0.$$

We choose sufficiently small $\epsilon_1$ so that the integral of the first term can be absorbed. For the second term, we first use Sobolev embedding on the unit sphere to bound $\alpha$ and then Lemma 19 to control $\phi$:

$$\iint r^{3+\epsilon}|\alpha|^2 r^{1+\gamma_0-\epsilon}|\phi|^2\,du\,dv\,d\omega$$

$$\lesssim \int_u u_+^{\gamma_0-\epsilon-1}\int_v r^{3+\epsilon}\sum_{j\le 2}\int_\omega |\mathcal{L}_\Omega^j\alpha|^2\,d\omega\cdot\left(\int_\omega |r\phi|^2(u,-u,\omega)\,d\omega + u_+^{-\gamma_0}\int_v\int_\omega r^{1+\gamma_0}|D_L\psi|^2\,dv\,d\omega\right)du$$

$$\lesssim M_2\int_{|x|\ge R} r_+^{1+\gamma_0-\epsilon-2}|\phi|^2(0,x)\,dx + M_2\int_u u_+^{-1-\epsilon}\int_v r^{1+\gamma_0}|D_L\psi|^2\,dv\,d\omega\,du$$

$$\lesssim M_2 E_0^0[\phi] + M_2\int_u u_+^{-1-\epsilon}\int_v r^{1+\gamma_0}|D_L\psi|^2\,dv\,d\omega\,du.$$

As the data for the scalar field is small, the charge is also small. In particular, we can choose $\epsilon_1 = |q_0|$ (if $q_0 = 0$, let $\epsilon_1$ be small depending only on $\epsilon$, $\gamma_0$ and $R$). Therefore estimate (51) holds for the case when $\mu = e_1$ or $e_2$. This completes the proof. □

As a corollary, we show the $r$-weighted energy flux decay of the scalar field in the exterior region.

**Corollary 22.** *Assume that the charge $q_0$ is sufficiently small, depending only on $\epsilon$, $R$, $\gamma_0$. Then in the exterior region, we have the energy flux decay*

$$\int_{H_{\tau_1}^*} r^p|D_L\psi|^2\,dv\,d\omega + \iint_{\mathcal{D}_{\tau_1}} r^{p-1}(p|D_L\psi|^2 + |\not{D}\psi|^2)\,dv\,d\omega\,du + \int_{\underline{H}_{-\tau_2^*}^{\tau_2^*,\tau_1^*}} r^p|\not{D}\psi|^2\,du\,d\omega$$
$$\lesssim_{M_2} \mathcal{E}_0[\phi](\tau_1)_+^{p-1-\gamma_0},\quad \forall 0\le p\le 1+\gamma_0,\quad \forall \tau_2\le\tau_1\le 0,\quad \psi = r\phi. \tag{53}$$

*Proof.* It suffices to prove the corollary for $p = 1+\gamma_0$. For sufficiently small $q_0$ depending only on $\epsilon$, $\gamma_0$ and $R$, from the $r$-weighted energy estimate (28) and the estimate (51) for the error term, the integral of $r^{\gamma_0}(|D_L(r\phi)|^2 + |\not{D}(r\phi)|^2)$ can be absorbed. Then estimate (53) follows from Gronwall's inequality. □

Next we make use of the $r$-weighted energy decay to show the energy flux decay and the integrated energy decay for the scalar field in the exterior region. From the integrated energy estimate (23), it suffices to bound the interaction term of the gauge field and the scalar field.

**Proposition 23.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $\tau_2 < \tau_1 \le 0$, we have*

$$\iint_{\mathcal{D}_{\tau_1}^{\tau_2}} |F_{Lv}J^v[\phi]| + |F_{\underline{L}v}J^v[\phi]|\,dx\,dt$$
$$\lesssim \epsilon_1 I_0^{-1-\epsilon}[D\phi](\mathcal{D}_{\tau_1}^{\tau_2}) + C_{M_2,\epsilon_1}\left(\mathcal{E}_0[\phi](\tau_1)_+^{-1-\gamma_0} + (\tau_1)_+^{\epsilon}\int_{-\tau_1^*}^{-\tau_2^*} v^{-1-\epsilon}E[\phi]\big(\underline{H}_v^{-v,\tau_1^*}\big)\,dv\right) \tag{54}$$

*for all $\epsilon_1 > 0$ and some constant $C_{M_2,\epsilon_1}$ depending on $M_2$ and $\epsilon_1$.*

*Proof.* The integral of $(dA)_{Lv}J^v[\phi]$ has been controlled in the previous Proposition 21 as Corollary 22 implies that the right-hand side of (51) can be bounded by a constant depending on $M_2$, $\epsilon$, $\gamma_0$ and $R$.

Since in the exterior region $r \geq \frac{1}{3}u_+$, we easily obtain the desired bound:

$$\iint_{\mathcal{D}_{\tau_1}^{\tau_2}} |F_{L^\nu} J^\nu[\phi]|\, dx\, dt \lesssim_{M_2} (\tau_1)_+^{-1-\gamma_0} \mathcal{E}_0[\phi].$$

It remains to estimate the integral of $F_{\underline{L}^\nu} J^\nu[\phi]$. The $r$-weighted energy decay gives control for the "good" derivative of the scalar field. The problem is that we do not have any control for the "bad" derivative $D_{\underline{L}}\phi$. In addition, since the charge is nonzero, we are not able to absorb the charge part $q_0 r^{-2} J_{\underline{L}}[\phi]$ in the integrated local energy estimate (23) as there is a small $\epsilon$ loss of decay in $I_0^{-1-\epsilon}[\tilde{D}\phi]$ on the left-hand side. The idea to treat this term is to make use of the energy flux on the incoming null hypersurface $\underline{H}_{-u_2}^{u_2,u_1}$ and then apply Gronwall's inequality. Let's first consider the easier terms in the integral of $F_{\underline{L}^\nu} J^\nu[\phi]$. For $\nu = e_1$ or $e_2$, we have

$$|F_{\underline{L}^\nu} J^\nu[\phi]| \lesssim |\underline{\alpha}|\,|\not{D}\phi|\,|\phi|.$$

Note that from estimate (48) of Lemma 19 and Corollary 22, we obtain

$$\int_\omega |r\phi|^2(u, v, \omega)\, d\omega \lesssim_{M_2} u_+ \int_\omega (-2u)|\phi(0, -2u, \omega)|^2\, d\omega + \mathcal{E}_0[\phi] u_+^{-\gamma_0}.$$

Here we parametrize $\phi$ in $(t, r, \omega)$ coordinates. We then use Sobolev embedding on the initial hypersurface $\{t = 0\}$ to derive the decay of $\phi$:

$$\int_\omega |r\phi|^2(u, v, \omega)\, d\omega \lesssim_{M_2} \mathcal{E}_0[\phi] u_+^{-\gamma_0}. \tag{55}$$

From the $r$-weighted energy estimate (53), we have an estimate for the weighted angular derivative of the scalar field on the incoming null hypersurface:

$$\int_{\underline{H}_{-\tau_2^*}^{\tau_2^*,\tau_1^*}} r^{1+\gamma_0} |\not{D}(r\phi)|^2\, du\, d\omega \lesssim_{M_2} \mathcal{E}_0[\phi].$$

In the exterior region, note that $r \geq \frac{1}{2}v$. Therefore we can show that

$$\begin{aligned}
&\iint_{\mathcal{D}_{\tau_1}^{\tau_2}} |F_{\underline{L}e_j}|\,|J^{e_j}[\phi]|\, dx\, dt \\
&\quad\lesssim \int_{-\tau_1^*}^{-\tau_2^*} \int_{-v}^{\tau_1^*} \int_\omega r^2 |\underline{\alpha}|\,|\not{D}\phi|\,|\phi|\, d\omega\, du\, dv \\
&\quad\lesssim \int_{-\tau_1^*}^{-\tau_2^*} \int_{-v}^{\tau_1^*} r^{-\frac{1}{2}(3+\gamma_0)} \left( r^2 \sum_{j\leq 2} \int_\omega |\mathcal{L}_\Omega^j \underline{\alpha}|^2\, d\omega \right)^{\frac{1}{2}} \left( r^{3+\gamma_0} \int_\omega |\not{D}\phi|^2\, d\omega \cdot \int_\omega |r\phi|^2\, d\omega \right)^{\frac{1}{2}} du\, dv \\
&\quad\lesssim_{M_2} \mathcal{E}_0[\phi]^{\frac{1}{2}} (\tau_1)_+^{-\frac{1}{2}\gamma_0} \int_{-\tau_1^*}^{-\tau_2^*} v^{-\frac{1}{2}(3+\gamma_0)} \left( E[\mathcal{L}_Z^2 dA]\left(\underline{H}_v^{-v,\tau_1^*}\right)\right)^{\frac{1}{2}} \mathcal{E}_0[\phi]^{\frac{1}{2}}\, dv \\
&\quad\lesssim_{M_2} \mathcal{E}_0[\phi] (\tau_1)_+^{-\frac{1}{2}\gamma_0 - \frac{1}{2}(1+\gamma_0) - \frac{1}{2}} \lesssim_{M_2} \mathcal{E}_0[\phi] (\tau_1)_+^{-1-\gamma_0}.
\end{aligned}$$

When $\nu = L$, first we have

$$|F_{\underline{L}L}||J^L[\phi]| \lesssim |q_0| r^{-2} |D_{\underline{L}}\phi||\phi| + |\bar{\rho}||D_{\underline{L}}\phi||\phi|.$$

The second term is easy to bound. We may use the Cauchy–Schwarz inequality. Indeed,

$$2|\bar{\rho}||D_{\underline{L}}\phi||\phi| \le \epsilon_1 |D_{\underline{L}}\phi|^2 r^{-1-\epsilon} + \epsilon_1^{-1}|\bar{\rho}|^2|\phi|^2 r^{1+\epsilon}, \quad \forall \epsilon_1 > 0.$$

For sufficiently small $\epsilon_1$, the integral of the first term on the right-hand side can be absorbed from the integrated energy estimate (23). For the second term, we make use of estimate (55) to show that

$$\begin{aligned}
\iint_{\mathcal{D}_{\tau_1}^{\tau_2}} |\bar{\rho}|^2 r^{3+\epsilon} |\phi|^2 \, dv \, du \, d\omega &\lesssim \int_{\tau_1^*}^{\tau_2^*} \int_{-u}^{-\tau_2^*} \sum_{j \le 2} \int_\omega r^2 |\mathcal{L}_\Omega^j \bar{\rho}|^2 \, d\omega \cdot r^{1+\epsilon} \int_\omega |\phi|^2 \, d\omega \, dv \, du \\
&\lesssim_{M_2} \int_{\tau_1^*}^{\tau_2^*} E^2[\bar{F}]\big(H_u^{-\tau_2^*}\big) u_+^{-1+\epsilon-\gamma_0} \mathcal{E}_0[\phi] \, du \\
&\lesssim_{M_2} \mathcal{E}_0[\phi] \int_{u_1^*}^{u_2^*} u_+^{-2-\gamma_0} \, du \lesssim_{M_2} \mathcal{E}_0[\phi] (\tau_1)_+^{-1-\gamma_0}.
\end{aligned}$$

Finally, we need to bound the charge part, namely the integral of $|q_0| r^{-2} |D_{\underline{L}}\phi||\phi|$. As we have explained previously, this term cannot be absorbed even though the charge $q_0$ is small due to the loss of decay in the integrated local energy $I_0^{-1-\epsilon}[\tilde{D}\phi](\mathcal{D}_{\tau_1}^{\tau_2})$ in (23). The idea is to make use of the energy flux in the incoming null hypersurface $\underline{H}_{-\tau_2^*}^{\tau_2^*, \tau_1^*}$ and then apply Gronwall's inequality. From estimate (55) and noting that $r \ge \frac{1}{2} v$ in the exterior region, we can show that

$$\begin{aligned}
\iint_{\mathcal{D}_{\tau_1}^{\tau_2}} r^{-2} |D_{\underline{L}}\phi||\phi| \, dx \, dt &\lesssim \int_{-\tau_1^*}^{-\tau_2^*} \int_{-v}^{\tau_1} \int_\omega |D_{\underline{L}}\phi||\phi| \, d\omega \, du \, dv \\
&\lesssim \int_{-\tau_1^*}^{-\tau_2^*} \int_{-v}^{\tau_1} r^{-2} \bigg( r^2 \int_\omega |D_{\underline{L}}\phi|^2 \, d\omega \cdot \int_\omega |r\phi|^2 d\omega \bigg)^{\frac{1}{2}} du \, dv \\
&\lesssim_{M_2} \mathcal{E}_0[\phi]^{\frac{1}{2}} \int_{-\tau_1^*}^{-\tau_2^*} v^{-\frac{1}{2}(3+\gamma_0-\epsilon)} \int_{-v}^{\tau_1^*} r^{-\frac{1}{2}(1-\gamma_0+\epsilon)} \bigg( r^2 \int_\omega |D_{\underline{L}}\phi|^2 \, d\omega \bigg)^{\frac{1}{2}} u_+^{-\frac{1}{2}\gamma_0} \, du \, dv \\
&\lesssim_{M_2} \mathcal{E}_0[\phi]^{\frac{1}{2}} \int_{-\tau_1^*}^{-\tau_2^*} v^{-\frac{1}{2}(3+\gamma_0-\epsilon)} \big( E[\phi]\big(\underline{H}_v^{-v,\tau_1^*}\big) \big)^{\frac{1}{2}} (\tau_1)_+^{-\frac{1}{2}\epsilon} \, dv \\
&\lesssim_{M_2} \mathcal{E}_0[\phi] (\tau_1)_+^{-1-\gamma_0} + (\tau_1)_+^{\epsilon} \int_{-\tau_1^*}^{-\tau_2^*} v^{-1-\epsilon} E[\phi]\big(\underline{H}_v^{-v,\tau_1^*}\big) \, dv.
\end{aligned}$$

Combining all the previous estimates, we then have shown (54). □

As a corollary we then can show the energy flux decay as well as the integrated local energy decay of the scalar field in the exterior region.

**Corollary 24.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $\tau_2 < \tau_1 \le 0$, we have*

$$I_0^{-1-\epsilon}[\tilde{D}\phi](\mathcal{D}_{\tau_1}^{\tau_2}) + E[\phi]\big(H_{\tau_1^*}^{-\tau_2^*}\big) + E[\phi]\big(\underline{H}_{-\tau_2^*}^{\tau_2^*,\tau_1^*}\big) \lesssim_{M_2} (\tau_1)_+^{-1-\gamma_0}\mathcal{E}_0[\phi]. \tag{56}$$

*Proof.* First choose $\epsilon_1$ in the estimate (54) to be sufficiently small, depending only on $\epsilon$, $\gamma_0$ and $R$, so that after combining estimate (54) and the integrated energy estimate (23), the term $\epsilon_1 I_0^{-1-\epsilon}[D\phi](\mathcal{D}_{\tau_1}^{\tau_2})$ on the right-hand side of (54) can be absorbed by $I_0^{-1-\epsilon}[\tilde{D}\phi](\mathcal{D}_{\tau_1}^{\tau_2})$ on the left-hand side of (23). Then notice that we have the uniform bound

$$(\tau_1)_+^{\epsilon}\int_{-\tau_1^*}^{-\tau_2^*} v^{-1-\epsilon}\,dv \lesssim 1, \quad \forall \tau_2 < \tau_1 \le 0.$$

Using Gronwall's inequality (fix $\tau_1 \le 0$ and take $\tau_2 \le \tau_1$ as variable), we then obtain (56). □

**4.3.2.** *Energy decay in the interior region.* Once we have the energy flux and the $r$-weighted energy decay estimates for the scalar field in the exterior region, we in particular have the energy flux bound for the scalar field on the boundary $H_{-R/2}$. This is necessary to consider the energy flux decay in the interior region. Compared to the case in the exterior region, the charge is not a problem as the charge only effects the decay property of the Maxwell field in the exterior region. However, new difficulties arise in the interior region case. First of all there is no lower bound for $r/\tau_+$. That means we may need estimates for general $p$ for the $r$-weighted energy estimates instead of simply the largest $p$. Secondly, as we have explained before, we are not able to absorb the interaction term between the gauge field $A$ and the scalar field due to the fact that $dA$ is no longer small in our setting. Thus we need to rely on the $r$-weighted energy estimates and make use of the null structure of $J[\phi]$. In the exterior region, the idea is first to derive the $r$-weighted energy decay and then to obtain the integrated local energy and energy flux decay. In the interior region, we see from the $r$-weighted energy estimates (27) that the term $|F_{\underline{L}\mu}J^{\mu}[\phi]|$ also appears on the right-hand side. This suggests that we have to consider the $r$-weighted energy estimate and the integrated local energy estimates simultaneously.

We first estimate the interaction terms of $dA$ and $J[\phi]$ in the $r$-weighted energy estimate (27).

**Proposition 25.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then in the interior region, for all $0 \le \tau_1 < \tau_2$ and $1 \le p \le 1+\gamma_0$, we have*

$$\begin{aligned}\iint_{\overline{\mathcal{D}}_{\tau_1}^{\tau_2}} r^p |F_{L\mu}J^{\mu}[\phi]|^2\,dx\,dt \lesssim \epsilon_1 \iint_{\overline{\mathcal{D}}_{\tau_1}^{\tau_2}} r^{p-1}|\not{D}(r\phi)|^2\,dv\,d\omega\,d\tau + I_{-1-\epsilon}^{p}[r^{-1}D_L(r\phi)](\overline{\mathcal{D}}_{\tau_1}^{\tau_2})\\ + M_2\epsilon_1^{-1}\left(\delta_p\int_{\tau_1}^{\tau_2} E[\phi](\Sigma_\tau)\tau_+^{\delta_p^{-1}-1-\epsilon}\,d\tau + (1-\delta_p)I_{-1-\epsilon}^{1+\gamma_0}[r^{-1}D_L(r\phi)](\overline{\mathcal{D}}_{\tau_1}^{\tau_2})\right)\end{aligned} \tag{57}$$

*for all $\epsilon_1 > 0$. Here $\delta_p = (2+\gamma_0-p)/(1+\gamma_0)$ is given in Lemma 19 in line (49).*

*Proof.* Denote $\psi = r\phi$ and $F = dA$. First we have

$$2r^p|F_{L\mu}J^{\mu}[\phi]|r^2 \le r^p|D_L\psi|^2\tau_+^{-1-\epsilon} + r^p|\rho|^2|\psi|^2\tau_+^{1+\epsilon} + \epsilon_1 r^{p-1}|\not{D}\psi|^2 + \epsilon_1^{-1}r^{p+3}|\alpha|^2|\phi|^2$$

for all $\epsilon_1 > 0$. The first term can be absorbed using Gronwall's inequality. The third term will be absorbed for sufficiently small $\epsilon_1$ depending only on $\epsilon$, $\gamma_0$ and $R$. For the second term, we use the energy flux of $\rho$

on $H_{\tau^*}$ to bound $\rho$, and estimate (49) of Lemma 19 to bound $\phi$. For the last term, we use the $r$-weighted energy estimate to bound $\alpha$. Then similarly to the proof of Proposition 21 we can show that

$$\begin{aligned}\int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}} \tau_+^{1+\epsilon} r^p |\rho|^2|\psi|^2 + r^{p+3}|\alpha|^2|\phi|^2\, dv\, d\omega\, d\tau \\ &\lesssim \int_{\tau_1}^{\tau_2}\int_{2R+\tau^*}^{\infty}\sum_{j\le 2}\int_\omega \tau_+^{1+\epsilon} r^2|\mathcal{L}_\Omega^j\rho|^2 + r^3|\mathcal{L}_\Omega^j\alpha|^2\, d\omega \cdot \int_\omega r^p|\phi|^2\, d\omega\, dv\, du \\ &\lesssim M_2\int_{\tau_1}^{\tau_2} \tau_+^{-\epsilon}(E[\phi](\Sigma_\tau))^\delta (I_0^{1+\gamma_0}[r^{-1}D_L\psi](H_{\tau^*}))^{1-\delta}\, d\tau \\ &\lesssim M_2\left(\delta\int_{\tau_1}^{\tau_2} E[\phi](\Sigma_\tau)\tau_+^{\delta^{-1}-1-\epsilon}\, d\tau + (1-\delta) I_{-1-\epsilon}^{1+\gamma_0}[r^{-1}D_L\psi](\overline{\mathcal{D}}_{\tau_1}^{\tau_2})\right).\end{aligned}$$

The proposition then follows. □

Next we estimate the interaction terms in the energy estimate (22). We show the following:

**Proposition 26.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then we have*

$$\begin{aligned}\iint_{\mathcal{D}_{\tau_1}^{\tau_2}} |F_{Lv}J^v[\phi]| + |F_{\underline{L}v}J^v[\phi]|\, dx\, dt \\ &\lesssim \epsilon_1 I_0^{-1-\epsilon}[D\phi](\mathcal{D}_{\tau_1}^{\tau_2}) + \epsilon_1^{-1}\int_{\tau_1}^{\tau_2} g(\tau)E[\phi](\Sigma_\tau)\, d\tau + I_{-2-\gamma_0}^{1+\gamma_0}[r^{-1}D_L(r\phi)](\overline{\mathcal{D}}_{\tau_1}^{\tau_2}) \quad (58)\end{aligned}$$

*for all $0<\epsilon_1<1$, where*

$$g(\tau) := \sum_{j\le 2} I_{1+2\epsilon}^{-1-\epsilon}[\mathcal{L}_\Omega^j F](\Sigma_\tau) + \sum_{j\le 2}\int_{H_{\tau^*}} r^{2+\epsilon}(|\mathcal{L}_\Omega^j\alpha|^2 + |\mathcal{L}_\Omega^j\rho|^2)\, dv\, d\omega + \sup_{|x|\le R}|F|^2(\tau,x).$$

*Proof.* For the integral on $\{r\ge R\}$, we use Sobolev embedding on the unit sphere to bound the curvature, and the proof is quite similar to that of the previous proposition. On the finite region $\{r\le R\}$, we make use of the $L_t^2L_x^\infty$ norm of the curvature given in Proposition 14. For the case when $r\ge R$, first we have

$$\begin{aligned}|F_{Lv}J^v[\phi]| + |F_{\underline{L}v}J^v[\phi]| &\lesssim (|\rho|+|\alpha|)|D\phi||\phi| + |\underline{\alpha}||\not{D}\phi||\phi| \\ &\lesssim \epsilon_1 r_+^{-1-\epsilon}|D\phi|^2 + \epsilon_1^{-1}(|\rho|^2+|\alpha|^2)r_+^{1+\epsilon}|\phi|^2 + |\underline{\alpha}||\not{D}\phi||\phi|.\end{aligned}$$

The first term can be absorbed in the energy estimate (22) for sufficiently small $\epsilon_1$. For the second term, we can use estimate (49) to bound $\phi$ by the energy flux through $H_{\tau^*}$ and the $r$-weighted energy to control the curvature terms. The last term is the most difficult one to control. The reason is that we do not have powerful estimates for $\underline{\alpha}$. The estimates we have are the integrated local energy estimate and the energy flux decay through the incoming null hypersurface. Unlike the case in the exterior region, where we can make use of the energy flux through the incoming null hypersurface for $\underline{\alpha}$, that method fails in the interior region. The main reason is that the energy flux $E[F]\big(\underline{H}_v^{\tau_1^*,\tau_2^*}\big)$ decays in $\tau_1$ instead of $v$. A possible way to solve this issue is to assume a pointwise bound for $\underline{\alpha}$. However the problem is that the pointwise decay for $\underline{\alpha}$ is too weak (due to the assumption on the initial data, as explained in the introduction) to be useful. We thus can only rely on the integrated local energy estimate for $\underline{\alpha}$. As there is an $r^\epsilon$ decay

loss in the integrated local energy estimate for $\underline{\alpha}$, we are not able to bound $\phi$ simply by using the energy flux through $H_{\Sigma_{\tau^*}}$. Instead, we need to make use of the $r$-weighted energy estimate. This means that we cannot obtain a uniform energy bound from the energy estimate (22). We need to combine it with the $r$-weighted energy estimate.

For the integral of $|\underline{\alpha}||\not{D}\phi||\phi|$, from estimate (49) with $p = 1+\epsilon$, we can show that

$$\begin{aligned}
&\int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}} |\underline{\alpha}||\not{D}\phi||\phi| r^2\, d\omega\, dv\, d\tau \\
&\quad\lesssim \int_{\tau_1}^{\tau_2}\int_{2R+\tau^*}^{\infty}\Big(\sum_{j\le 2}\int_\omega r^{1-\epsilon}|\mathcal{L}_\Omega^j\underline{\alpha}|^2\, d\omega\Big)^{\frac12}\Big(\int_\omega r^2|\not{D}\phi|^2\, d\omega\cdot\int_\omega r^{1+\epsilon}|\phi|^2\, d\omega\Big)^{\frac12} dv\, d\tau \\
&\quad\lesssim \sum_{j\le 2}\int_{\tau_1}^{\tau_2}(I_0^{-1-\epsilon}[\mathcal{L}_\Omega^j\underline{\alpha}](\Sigma_\tau)E[\phi](\Sigma_\tau))^{\frac12}(E[\phi](\Sigma_\tau))^{\frac12\delta}(I_0^{1+\gamma_0}[r^{-1}D_L(r\phi)](H_{\tau^*}))^{\frac12-\frac12\delta}\, d\tau \\
&\quad\lesssim \sum_{j\le 2}\int_{\tau_1}^{\tau_2} I_{1+2\epsilon}^{-1-\epsilon}[\mathcal{L}_\Omega^j\underline{\alpha}](\Sigma_\tau)E[\phi](\Sigma_\tau)\, d\tau + \int_{\tau_1}^{\tau_2}\tau_+^{-1-\epsilon}E[\phi](\Sigma_\tau)\, d\tau + I_{-2-\gamma_0}^{1+\gamma_0}[r^{-1}D_L(r\phi)](\overline{\mathcal{D}}_{\tau_1}^{\tau_2}).
\end{aligned}$$

Here $\delta = (1+\gamma_0-\epsilon)/(1+\gamma_0)$, and in the last step we have used Jensen's inequality as well as the relation

$$\frac12+\epsilon-\frac12\delta(1+\epsilon)-(2+\gamma_0)\Big(\frac12-\frac12\delta\Big) = \frac12\frac{\epsilon}{1+\gamma_0} > 0.$$

In the above estimate the first two terms will be estimated using Gronwall's inequality. We keep the last term involving the $r$-weighted energy estimates. For the integral of $(|\rho|^2+|\alpha|^2)r_+^{1+\epsilon}|\phi|^2$, we use estimate (49) to bound $\phi$. We have

$$\begin{aligned}
&\int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}}(|\rho|^2+|\alpha|^2)r_+^{1+\epsilon}|\phi|^2 r^2\, d\omega\, dv\, d\tau \\
&\qquad\lesssim \int_{\tau_1}^{\tau_2}\int_{2R+\tau^*}^{\infty}\sum_{j\le 2}\int_\omega r^{2+\epsilon}(|\mathcal{L}_\Omega^j\alpha|^2+|\mathcal{L}_\Omega^j\rho|^2)\, d\omega\cdot\int_\omega r|\phi|^2\, d\omega\, dv\, d\tau \\
&\qquad\lesssim \sum_{j\le 2}\int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}} r^{2+\epsilon}(|\mathcal{L}_\Omega^j\alpha|^2+|\mathcal{L}_\Omega^j\rho|^2)\, dv\, d\omega\cdot E[\phi](\Sigma_\tau)\, d\tau.
\end{aligned}$$

This term will be controlled in the energy estimate (22) using Gronwall's inequality.

For the integral on the region $\{r\le R\}$, we can show that

$$\begin{aligned}
\int_{\tau_1}^{\tau_2}\int_{r\le R}|F_{L\nu}J^\nu[\phi]|+|F_{\underline{L}\nu}J^\nu[\phi]|\, dx\, d\tau &\lesssim \epsilon_1\int_{\tau_1}^{\tau_2}\int_{r\le R}|D\phi|^2\, dx\, d\tau+\epsilon_1^{-1}\int_{\tau_1}^{\tau_2}\int_{r\le R}|F|^2|\phi|^2\, dx\, d\tau \\
&\lesssim \epsilon_1\int_{\tau_1}^{\tau_2}\int_{r\le R}\frac{|D\phi|^2}{r_+^{1+\epsilon}}\, dx\, d\tau+\epsilon_1^{-1}\int_{\tau_1}^{\tau_2}\sup_{|x|\le R}|F|^2\cdot E[\phi](\Sigma_\tau)\, d\tau
\end{aligned}$$

for all $\epsilon_1>0$. The first term will be absorbed for small $\epsilon_1$. The second term can be controlled using Gronwall's inequality. Combining all these estimates above, we thus have shown estimate (58). □

As a corollary, the energy estimate (22) leads to the following:

**Corollary 27.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then in the interior region, we have the estimate*

$$I_0^{-1-\epsilon}[\tilde{D}\phi](\mathcal{D}^{\tau_2}_{\tau_1}) + E[\phi](\Sigma_{\tau_2}) + \iint_{\mathcal{D}^{\tau_2}_{\tau_1}} |F_{Lv}J^v[\phi]| + |F_{\underline{L}v}J^v[\phi]|\,dx\,dt$$
$$\lesssim_{M_2} E[\phi](\Sigma_{\tau_1}) + (\tau_1)_+^{-1-\gamma_0}\mathcal{E}_0[\phi] + I^{1+\gamma_0}_{-2-\gamma_0}[r^{-1}D_L(r\phi)](\overline{\mathcal{D}}^{\tau_2}_{\tau_1}). \quad (59)$$

*Proof.* First choose $\epsilon_1$ sufficiently small in the estimate (58) so that combining the energy estimate (22) with (58), the integrated local energy term $I_0^{-1-\epsilon}[D\phi](\mathcal{D}^{\tau_2}_{\tau_1})$ could be absorbed. By our notation, the smallness of $\epsilon_1$ depends only on $\epsilon$, $\gamma_0$ and $R$. Then for the second term on the right-hand side of (58), to apply Gronwall's inequality, we show that $g(\tau)$ (defined after line (58)) is integrable. From the integrated local energy estimates (36) and the $r$-weighted energy estimates (31) for the Maxwell field, we conclude from the previous section that

$$I_0^{-1-\epsilon}[\mathcal{L}_Z^k F](\mathcal{D}^{\tau_2}_{\tau_1}) \lesssim M_k(\tau_1)_+^{-1-\gamma_0},$$
$$\int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}} r^{2+\epsilon}(|\mathcal{L}_Z^k\alpha|^2 + |\mathcal{L}_Z^k\rho|^2)\,dv\,d\omega\,d\tau \lesssim M_k(\tau_1)_+^{-\gamma_0+\epsilon}.$$

Therefore, using Lemma 20 and Proposition 14, we can show that

$$\int_{\tau_1}^{\tau_2} g(\tau)\,d\tau \lesssim M_2(\tau_1)_+^{-\gamma_0+\epsilon} + \sum_{j\le 2} I^{-1-\epsilon}_{1+2\epsilon}[\mathcal{L}_\Omega^j F](\mathcal{D}^{\tau_2}_{\tau_1})$$
$$\lesssim M_2(\tau_1)_+^{-\gamma_0+\epsilon} + \sum_{j\le 2}\int_{\tau_1}^{\tau_2} \tau_+^{2\epsilon} I_0^{-1-\epsilon}[\mathcal{L}_\Omega^j F](\mathcal{D}^{\tau_2}_{\tau})\,d\tau + (\tau_1)_+^{1+2\epsilon} I_0^{-1-\epsilon}[\mathcal{L}_\Omega^j F](\mathcal{D}^{\tau_2}_{\tau_1})$$
$$\lesssim M_2(\tau_1)_+^{-\gamma_0+\epsilon} + M_2\int_{\tau_1}^{\tau_2}\tau_+^{-1-\gamma_0+2\epsilon}\,d\tau + M_2(\tau_1)_+^{-\gamma_0+2\epsilon}$$
$$\lesssim M_2(\tau_1)_+^{-\gamma_0+2\epsilon}.$$

By using this uniform bound, the second term on the right-hand side of (58) can be absorbed using Gronwall's inequality. The corollary then follows. □

We now can use Proposition 25 and the above corollary to obtain the necessary $r$-weighted energy estimates. To derive energy decay estimates, we at least need the $r$-weighted energy estimates with $p = 1$ and $p = 1+\gamma_0$ (some $p$ bigger than one, the decay rate depending on this largest $p$). In any case, we first choose $\epsilon_1$ in estimate (57) sufficiently small, so that combining it with the $r$-weighted energy estimate (27), the first term on the right-hand side of (57) can be absorbed (note that $\gamma_0 < 1$). The second term on the right-hand side of (57) can be controlled using Gronwall's inequality. Let's first combine the $r$-weighted energy estimate (27) for $p = 1$ with the integrated local energy estimate (59) to derive the bound for the integral of the energy flux.

**Proposition 28.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then in the interior region, for all $0 \le \tau_1 < \tau_2$, we have*

$$\int_{\tau_1}^{\tau_2} E[\phi](\Sigma_\tau)\, d\tau \lesssim_{M_2} \int_{H_{\tau_1^*}} r|D_L\psi|^2\, dv\, d\omega + E[\phi](\Sigma_{\tau_1}) + (\tau_1)_+^{-\gamma_0}\mathcal{E}_0[\phi] + I_{-2-\gamma_0}^{1+\gamma_0}[r^{-1}D_L(r\phi)](\overline{\mathcal{D}}_{\tau_1}^{\tau_2}). \quad (60)$$

*Proof.* In $(t, r, \omega)$ coordinates, using Sobolev embedding, we have

$$\int_\omega |\phi|^2(\tau, R, \omega)\, d\omega \lesssim \int_{r\le R} |\phi|^2 + |D\phi|^2\, dx.$$

Then we can show that

$$\begin{aligned}\int_{\tau_1}^{\tau_2} E[\phi](\Sigma_\tau)\, d\tau &\lesssim \int_{\tau_1}^{\tau_2}\int_{r\le R} |D\phi|^2\, dx\, d\tau + \int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}} |D_L(r\phi)|^2 + |\not{D}(r\phi)|^2\, dv\, d\omega\, d\tau \\ &\quad + \int_{\tau_1}^{\tau_2}\int_\omega |\phi|^2(\tau, R, \omega)\, d\omega \\ &\lesssim I_0^{-1-\epsilon}[\tilde{D}\phi](\mathcal{D}_{\tau_1}^{\tau_2}) + \int_{\tau_1}^{\tau_2}\int_{H_{\tau^*}} |D_L(r\phi)|^2 + |\not{D}(r\phi)|^2\, dv\, d\omega\, d\tau.\end{aligned}$$

Therefore, take $p = 1$ in the $r$-weighted energy estimate (27). From the above argument, we obtain the following bound for the integral of the energy flux:

$$\begin{aligned}\int_{\tau_1}^{\tau_2} E[\phi](\Sigma_\tau) d\tau &\lesssim \int_{H_{\tau_1^*}} r|D_L\psi|^2\, dv\, d\omega + M_2\int_{\tau_1}^{\tau_2} E[\phi](\Sigma_\tau)\tau_+^{-\epsilon}\, d\tau \\ &\quad + C_{M_2}\big(E[\phi](\Sigma_{\tau_1}) + (\tau_1)_+^{-\gamma_0}\mathcal{E}_0[\phi] + I_{-2-\gamma_0}^{1+\gamma_0}[r^{-1}D_L(r\phi)](\overline{\mathcal{D}}_{\tau_1}^{\tau_2})\big)\end{aligned}$$

for some constant $C_{M_2}$ depending on $M_2$. For the second term, we further can bound

$$\tau_+^{-\epsilon} = (\epsilon_1^{-1/\epsilon}\tau_+^{-1-\epsilon})^{\epsilon/(1+\epsilon)}\cdot(\epsilon_1)^{1/(1+\epsilon)} \le \frac{\epsilon}{1+\epsilon}\epsilon_1^{-1/\epsilon}\tau_+^{-1-\epsilon} + \frac{\epsilon_1}{1+\epsilon}, \quad \forall \epsilon_1 > 0.$$

Choose $\epsilon_1$ sufficiently small, so that the second term can be absorbed. Then the first term can be bounded using Corollary 27. Therefore, the previous estimate amounts to estimate (60). □

We have the following corollary.

**Corollary 29.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then we have*

$$\begin{aligned}\int_{\tau_1}^{\tau_2} \tau_+^{\gamma_0-\epsilon} E[\phi](\Sigma_\tau)\, d\tau &\lesssim_{M_2} \int_{H_{\tau_1^*}} r^{1+\gamma_0}|D_L\psi|^2\, dv\, d\omega + (\tau_1)_+^{1+\gamma_0-\epsilon}E[\phi](\Sigma_{\tau_1}) \\ &\quad + \mathcal{E}_0[\phi] + I_{-2-\epsilon}^{1+\gamma_0}[r^{-1}D_L(r\phi)](\overline{\mathcal{D}}_{\tau_1}^{\tau_2}). \quad (61)\end{aligned}$$

*Proof.* Using estimate (49) of Lemma 19, we have the bound

$$\int_{H_{\tau^*}} |D_L(r\phi)|^2\, dv\, d\omega \le \int_{H_{\tau^*}} |D_L\phi|^2 r^2\, dv\, d\omega + \lim_{r\to\infty}\int_\omega r|\phi|^2\, d\omega \lesssim E[\phi](\Sigma_\tau).$$

For all $\epsilon_1 > 0$, we have the inequality

$$\tau_+^{\gamma_0-1-\epsilon} r = (\epsilon_1^{-\gamma_0} r^{1+\gamma_0} \tau_+^{-1-\epsilon})^{1/(1+\gamma_0)} (\epsilon_1 \tau_+^{\gamma_0-\epsilon})^{\gamma_0/(1+\gamma_0)} \leq \frac{\epsilon_1^{-\gamma_0} r^{1+\gamma_0} \tau_+^{-1-\epsilon}}{1+\gamma_0} + \frac{\gamma_0 \epsilon_1 \tau_+^{\gamma_0-\epsilon}}{1+\gamma_0}.$$

In particular, the above inequality holds for $r = 1$. Moreover, we also have

$$(\tau_1)_+^{\gamma_0-\epsilon} r = (r^{1+\gamma_0})^{1/(1+\gamma_0)} \big((\tau_1)_+^{1+\gamma_0-\epsilon(1+\gamma_0)/\gamma_0}\big)^{\gamma_0/(1+\gamma_0)} \leq r^{1+\gamma_0} + (\tau_1)_+^{1+\gamma_0-\epsilon}.$$

Denote $\psi = r\phi$. From estimate (60), we can show that

$$\begin{aligned}\int_{\tau_1}^{\tau_2} \tau_+^{\gamma_0-1-\epsilon} \bigg( \int_{H_{\tau^*}} r |D_L \psi|^2 \, dv \, d\omega + E[\phi](\Sigma_\tau) + \tau_+^{-\gamma_0} \mathcal{E}_0[\phi] + I_{-2-\gamma_0}^{1+\gamma_0}[r^{-1} D_L(r\phi)](\overline{\mathcal{D}}_\tau^{\tau_2}) \bigg) d\tau \\ \lesssim \epsilon_1^{-\gamma_0} \int_{\tau_1}^{\tau_2} \tau_+^{-1-\epsilon} \int_{H_{\tau^*}} r^{1+\gamma_0} |D_L \psi|^2 \, dv \, d\omega \, d\tau + \epsilon_1 \int_{\tau_1}^{\tau_2} \tau_+^{\gamma_0-\epsilon} E[\phi](\Sigma_\tau) \, d\tau \\ + \epsilon_1^{-\gamma_0} \int_{\tau_1}^{\tau_2} \tau_+^{-1-\epsilon} E[\phi](\Sigma_\tau) \, d\tau + \mathcal{E}_0[\phi] + I_{-2-\epsilon}^{1+\gamma_0}[r^{-1} D_L(r\phi)](\overline{\mathcal{D}}_{\tau_1}^{\tau_2}).\end{aligned}$$

On the right-hand side of the above estimate, the first term can be grouped with the last term. The second term will be absorbed for small $\epsilon_1$. The third term can be bounded using estimate (59). Therefore, using Lemma 20 and Proposition 28, we can show that

$$\begin{aligned}\int_{\tau_1}^{\tau_2} \tau_+^{\gamma_0-\epsilon} E[\phi](\Sigma_\tau) \, d\tau \lesssim_{M_2} \epsilon_1 \int_{\tau_1}^{\tau_2} \tau_+^{\gamma_0-\epsilon} E[\phi](\Sigma_\tau) \, d\tau + \epsilon_1^{-\gamma_0} \int_{H_{\tau_1^*}} r^{1+\gamma_0} |D_L \psi|^2 \, dv \, d\omega + \epsilon_1^{-\gamma_0} \mathcal{E}_0[\phi] \\ + \epsilon_1^{-\gamma_0} (\tau_1)_+^{1+\gamma_0-\epsilon} E[\phi](\Sigma_{\tau_1}) + \epsilon_1^{-\gamma_0} I_{-2-\epsilon}^{1+\gamma_0}[r^{-1} D_L(r\phi)](\overline{\mathcal{D}}_{\tau_1}^{\tau_2}).\end{aligned}$$

Let $\epsilon_1$ be sufficiently small, depending on $M_2$, $\epsilon$, $\gamma_0$ and $R$. We obtain estimate (61). □

Estimate (61) can now be used to derive the $r$-weighted energy estimate with $p = 1 + \gamma_0$.

**Proposition 30.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then we have*

$$\begin{aligned}\int_{H_{\tau_2^*}} r^{1+\gamma_0} |D_L \psi|^2 \, dv \, d\omega + \int_{\tau_1}^{\tau_2} \int_{H_{\tau^*}} r^{\gamma_0} (|D_L \psi|^2 + |\not{D} \psi|^2) \, dv \, d\omega \, d\tau \\ \lesssim_{M_2} \int_{H_{\tau_1^*}} r^{1+\gamma_0} |D_L \psi|^2 \, dv \, d\omega + \mathcal{E}_0[\phi] + (\tau_1)_+^{1+\gamma_0-\epsilon} E[\phi](\Sigma_{\tau_1}), \quad (62)\end{aligned}$$

*where $\psi = r\phi$.*

*Proof.* By taking $\epsilon_1$ in estimate (57) to be sufficiently small and combining it with the $r$-weighted energy estimate (27) for $p = 1 + \gamma_0$, from Corollary 27 we obtain

$$\begin{aligned}\int_{H_{\tau_2^*}} r^{1+\gamma_0} |D_L \psi|^2 \, dv \, d\omega + \int_{\tau_1}^{\tau_2} \int_{H_{\tau^*}} r^{\gamma_0} (|D_L \psi|^2 + |\not{D} \psi|^2) \, dv \, d\omega \, d\tau \\ \lesssim \int_{H_{\tau_1^*}} r^{1+\gamma_0} |D_L \psi|^2 \, dv \, d\omega + \mathcal{E}_0[\phi] + M_2 \bigg( \int_{\tau_1}^{\tau_2} E[\phi](\Sigma_\tau) \tau_+^{\gamma_0-\epsilon} \, d\tau + I_{-1-\epsilon}^{1+\gamma_0}[r^{-1} D_L \psi](\overline{\mathcal{D}}_{\tau_1}^{\tau_2}) \bigg) \\ + C_{M_2} \big( E[\Phi](\Sigma_{\tau_1}) + (\tau_1)_+^{-1-\gamma_0} \mathcal{E}_0[\phi] + I_{-2-\gamma_0}^{1+\gamma_0}[r^{-1} D_L \psi](\overline{\mathcal{D}}_{\tau_1}^{\tau_2}) \big)\end{aligned}$$

for some constant $C_{M_2}$ depending on $M_2$. Estimate (62) then follows from estimate (61) together with Gronwall's inequality. □

Take $\tau_1 = 0$ in (62). From the energy estimate (56) and the $r$-weighted energy estimate (53) in the exterior region, we conclude that the right-hand side of (62) is bounded. Since $\tau_2 > \tau_1$ is arbitrary there, we in particular have the $r$-weighted energy estimate for the scalar field in the interior region.

**Corollary 31.** *Let $\psi = r\phi$. Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $0 \le \tau_1 < \tau_2$, we have*

$$\int_{H^*_{\tau_2}} r^{1+\gamma_0} |D_L \psi|^2 \, dv \, d\omega + \int_{\tau_1}^{\tau_2} \int_{H_{\tau^*}} r^{\gamma_0} (|D_L \psi|^2 + |\not{D}\psi|^2) \, dv \, d\omega \, d\tau \lesssim_{M_2} \mathcal{E}_0[\phi]. \tag{63}$$

*Proof.* From the $r$-weighted energy estimate (53) in the exterior region with $p = 1 + \gamma_0$, $\tau_1 = 0$, we derive

$$\int_{H_{0^*}} r^{1+\gamma_0} |D_L \psi|^2 \, dv \, d\omega = \int_{H_{-R/2}} r^{1+\gamma_0} |D_L \psi|^2 \lesssim_{M_2} \mathcal{E}_0[\phi].$$

The energy estimate (56) in the exterior region implies that

$$E[\phi](\Sigma_0) = E[\phi](\{t = 0, r \le R\}) + E[\phi](H_{-R/2}) \lesssim \mathcal{E}_0[\phi].$$

Then estimate (63) follows from (62) by taking $\tau_1 = 0$. □

This uniform bound for the $r$-weighted energy estimate in the interior region is crucial for the energy flux decay. It in particular implies that the terms involving the $r$-weighted energy flux on the right-hand side of the energy estimate (59) and the integral of the energy flux estimate (60) have the right decay in order to show the energy flux decay.

**Proposition 32.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then in the interior region, we have the energy flux decay*

$$E[\phi](\Sigma_\tau) \lesssim_{M_2} \mathcal{E}_0[\phi] \tau_+^{-1-\gamma_0}, \quad \forall \tau \ge 0. \tag{64}$$

*Proof.* Estimate (63) implies that

$$I^{1+\gamma_0}_{-2-\gamma_0}[r^{-1} D_L \psi](\overline{\mathcal{D}}^{\tau_2}_{\tau_1}) \lesssim_{M_2} (\tau_1)_+^{-1-\gamma_0} \mathcal{E}_0[\phi], \quad \forall 0 \le \tau_1 < \tau_2.$$

Then using a pigeonhole argument like in the proof of Proposition 11 for the energy flux decay of the Maxwell field in the interior, the energy decay estimate (64) for the scalar field follows from the energy estimate (59), the integral of the energy flux estimate (60) and the $r$-weighted energy estimate (63). For a detailed proof for this, we refer to Proposition 2 of [Yang 2015b]. □

**4.3.3.** *Energy decay estimates for the first-order derivative of the scalar field.* In this section, we derive the energy flux decay estimates for the derivative of the scalar field. The difficulty is that the covariant wave operator $\Box_A$ does not commute with $D_Z$. Commutators are quadratic in the Maxwell field and the scalar field. In our setting, the Maxwell field is large. In particular, those terms cannot be absorbed. The idea is to exploit the null structure of the commutators and to use Gronwall's inequality adapted to our foliation $\Sigma_\tau$.

In the following, we always use $\psi$ to denote the weighted scalar field $r\phi$, that is, $\psi = r\phi$. The first-order derivative of $\phi$ is abbreviated $\phi_1$, and the second-order derivative $\phi_2$. More precisely, we denote $\phi_1 = D_Z\phi$, $\phi_2 = D_Z^2\phi$ with $Z$ any vector field in the set $\Gamma = \{\partial_t, \Omega_{ij} = x_i\partial_j - x_j\partial_i\}$. We use the same notation for the weighted scalar field $\psi$, e.g., $\psi_1 = rD_Z\phi$. For any function $f$, under the null coordinates $(u, v, \omega)$, we define

$$\|f\|^2_{L^2_v L^\infty_u L^2_\omega(\mathcal{D})} := \int_v \sup_u \int_\omega |f|^2 \, d\omega \, dv,$$

where $(u, v, \omega)$ are the null coordinates on the region $\mathcal{D}$. Similarly, we have the notation $\|f\|_{L^2_u L^\infty_v L^2_\omega(\mathcal{D})}$. We can also define $L^p_u L^q_v L^r_\omega$ norms for general $p, q, r$.

To apply Corollary 24 for the exterior region and Proposition 32 for the interior region, it suffices to control the commutator terms. However, we are not able to bound the commutator terms directly by using the zero's order energy estimates. One has to make use of the energy flux of the first-order derivative of the solution and then apply Gronwall's inequality. However, for the energy estimate for the first-order derivative of the solution, the key is to understand the commutator $[\Box_A, D_Z]$ with $Z = \partial_t$ or the angular momentum. The cases of $\partial_t$ and the angular momentum are quite different. The main reason is that the angular momentum contains weights in $r$ while $\partial_t$ does not. For the case when $Z = \partial_t$, it is easy to bound $[\Box_A, D_{\partial_t}]\phi$. The only place we need to be careful is the charge part. For the case of $Z = \Omega$, the problem is that the commutator $[\Box_A, D_\Omega]$ produces a term of the form $Z^\nu F_{\mu\nu} D^\mu\phi$ which cannot be written as a linear combination of $D_Z\phi$. The estimate for the commutator terms heavily rely on the null structure. We first show the following lemma for the commutator terms.

**Lemma 33.** *When $|x| \geq R$, we have*

$$|[\Box_A, D_Z]\phi| \lesssim |\alpha||D_{\underline{L}}\psi| + (|\underline{\alpha}| + r^{-1}|\rho|)|D_L\psi| + |F||\not{D}\phi| + (|J| + r|\not{J}| + |\sigma| + r^{-1}|\rho|)|\phi|. \quad (65)$$

*When $r \leq R$, we have*

$$|[\Box_A, D_Z]\phi| \lesssim |F||\overline{D}\phi| + |J||\phi|. \quad (66)$$

*Here $F = dA$ and $J = \delta F$.*

**Remark 34.** In this paper, all the quantities involving $Z$ should be interpreted as the sum of the quantity for all possible vector fields $Z$ in $\Gamma$ unless otherwise specified.

*Proof.* Let $\psi = r\phi$. First, from Lemma 4 we can write

$$[\Box_A, D_Z]\phi = 2ir^{-1}Z^\nu F_{\mu\nu}D^\mu\psi + i\nabla^\mu F_{\mu\nu}Z^\nu\phi + i\phi(-2Z^\nu F_{\mu\nu}r^{-1}\nabla^\mu r + \nabla^\mu Z^\nu F_{\mu\nu}). \quad (67)$$

We need to exploit the null structure of the above commutator terms. The first term is the main one. Since we will rely on the $r$-weighted energy estimates, it suggests writing the main term in terms of the weighted solution $r\phi$. The second term is easy, as $\nabla^\mu F_{\mu\nu}$ is a nonlinear term of $\phi$ by the Maxwell equation. Let's first estimate the third term. When $Z = \Omega$, note that $r^{-1}\Omega$ is a linear combination of $e_1$ and $e_2$. We then can show that

$$|r^{-1}Z^\nu F_{\mu\nu}D^\mu(r\phi)| \lesssim |\alpha||D_{\underline{L}}(r\phi)| + |\underline{\alpha}||D_L(r\phi)|.$$

This is the null structure we need: the "bad" component $\underline{\alpha}$ of the curvature does not interact with the "bad" component $D_{\underline{L}}(r\phi)$ of the scalar field. Similarly, when $Z = \partial_t$, the "bad" term $r^{-1}\underline{\alpha} D_{\underline{L}}(r\phi)$ does not appear. More precisely, we have

$$|r^{-1} Z^\nu F_{\mu\nu} D^\mu (r\phi)| \lesssim r^{-1}(|\alpha| + |\underline{\alpha}|)|\not{D}(r\phi)| + r^{-1}|\rho||D_r(r\phi)|.$$

For the second term on the right-hand side of (67), we note that $\nabla^\mu F_{\mu\nu}$ is a nonlinear term of $\phi$. We have

$$|\nabla^\mu F_{\mu\nu} Z^\nu \phi| \lesssim (|J| + r|\not{J}|)|\phi|.$$

For the third term on the right-hand side of (67), we show that

$$|i\phi(-2Z^\nu F_{\mu\nu} r^{-1}\nabla^\mu r + \nabla^\mu Z^\nu F_{\mu\nu})| \lesssim (|\sigma| + r^{-1}|\rho|)|\phi|.$$

The case when $Z = \partial_t$ is trivial. To check the above inequality for the case when $Z = \Omega$, it suffices to prove it for the component $\Omega_{jk} = x_j\partial_k - x_k\partial_j$. Then we can show that

$$\begin{aligned} -2\Omega^\nu F_{\mu\nu} r^{-1}\nabla^\mu r + \nabla^\mu \Omega^\nu F_{\mu\nu} &= 2F_{jk} - 2F(\partial_r, \Omega_{ij}) \\ &= 2F(\omega_j\partial_r + \partial_j - \omega_j\partial_r, \omega_k\partial_r + \partial_k - \omega_k\partial_r) - 2F(\partial_r, \Omega_{jk}) \\ &= 2F(\partial_j - \omega_j\partial_r, \partial_k - \omega_k\partial_r). \end{aligned}$$

Here recall that $\omega_j = r^{-1}x_j$. Since $\partial_j - \omega_j\partial_r$ is orthogonal to $L$ and $\underline{L}$ for all $j = 1, 2, 3$, we conclude that $\partial_j - \omega_j\partial_r$ is a linear combination of $e_1$ and $e_2$. The desired estimate then follows, as the norm of the vector fields $\partial_j - \omega_j\partial_r$ is less than 1. □

We begin a series of propositions in order to estimate the weighted spacetime norm of the commutators. The estimates in the bounded region $\{r \le R\}$ are easy to obtain as the weights are finite. We now concentrate on the region $\{r \ge R\}$. Let $\overline{\mathcal{D}}_\tau = \mathcal{D}_\tau \cap \{|x| \ge R\}$ and recall that $D_\tau = D_\tau^{+\infty}$ when $\tau \ge 0$, or $D_\tau = D_\tau^{-\infty}$ otherwise. We first consider $|\alpha||D_{\underline{L}}(r\phi)|$.

**Proposition 35.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $\epsilon_1 > 0$, we have*

$$\|D_{\underline{L}}(r\phi)\|_{L^2_u L^\infty_v L^2_\omega(\overline{\mathcal{D}}_\tau)} \lesssim_{M_2} \mathcal{E}_0[\phi]\epsilon_1^{-1}\tau_+^{-1-\gamma_0} + \epsilon_1 I_0^{1+\epsilon}[r^{-1}D_L D_{\underline{L}}(r\phi)](\overline{\mathcal{D}}_\tau). \tag{68}$$

*Proof.* The idea is to bound $\sup|D_{\underline{L}}(r\phi)|$ by the $L^2$ norm of $D_L D_{\underline{L}}(r\phi)$. In the exterior region when $\overline{\mathcal{D}}_\tau = \mathcal{D}_\tau^{-\infty}$, we can integrate from the initial hypersurface $\{t = 0\}$. In the interior region, choose the incoming null hypersurface $\underline{H}^{\tau_1^*,\tau_2^*}_{(\tau_2+R)/2}$ as the starting surface. Denote $\psi = r\phi$. We show estimate (68) for the interior region case, that is, when $0 \le \tau_1 < \tau_2$. On the outgoing null hypersurface $H_{\tau^*}$, for all $0 \le \tau_1 \le \tau \le \tau_2$, we have

$$\begin{aligned} &\sup_{v \ge (\tau+R)/2} \int_\omega |D_{\underline{L}}(r\phi)|^2(\tau^*, v, \omega)\, d\omega \\ &\qquad \lesssim \int_\omega |D_{\underline{L}}(r\phi)|^2\Big(\tau^*, \frac{\tau_2+R}{2}, \omega\Big)\, d\omega + \int_{H_{\tau^*}} |D_L D_{\underline{L}}(r\phi)| \cdot |D_{\underline{L}}(r\phi)|\, dv\, d\omega. \end{aligned}$$

Integrate the above estimate from $\tau_1$ to $\tau_2$ and apply the Cauchy–Schwarz inequality to the last term. From the integrated local energy estimate (59) and the energy decay estimate (64), we then derive

$$\int_{\tau_1}^{\tau_2} \sup_{v\geq (R+\tau)/2} \int_{\omega} |D_{\underline{L}}(r\phi)|^2 \, d\omega \, d\tau \lesssim_{M_2} \mathcal{E}_0[\phi]\epsilon_1^{-1}(\tau_1)_+^{-1-\gamma_0} + \epsilon_1 I_0^{1+\epsilon}[r^{-1}D_L D_{\underline{L}}\psi](\overline{\mathcal{D}}_{\tau_1}^{\tau_2})$$

for all $\epsilon_1 > 0$. The case in the exterior region follows in a similar way. □

We also need the analogous estimate for $D_L(r\phi)$.

**Proposition 36.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $\epsilon_1 > 0$ and $0 \leq p \leq 1+\gamma_0$, we have*

$$\|r^{p/2} D_L(r\phi)\|^2_{L^2_{\bar{v}}L^\infty_u L^2_\omega(\overline{\mathcal{D}}_\tau)} \lesssim_{M_2} \epsilon_1^{-1}\mathcal{E}_0[\phi](\tau)_+^{-1-\gamma_0} + \epsilon_1 I_{p_2}^{p_1}[r^{-1}D_{\underline{L}}D_L(r\phi)](\overline{\mathcal{D}}_\tau). \tag{69}$$

*Here $p_1 = \max\{1+\epsilon, p\}$ and $p_2 = \min\{1+\frac{1}{2}\epsilon, p\}$.*

*Proof.* Similar to the proof of the previous proposition, we choose the starting surface for $D_L(r\phi)$ to be $H_{\tau_1^*}$ in the interior region and the initial hypersurface $\{t=0\}$ in the exterior region. We only prove the proposition for the exterior region case. Denote $\psi = r\phi$. On $\underline{H}_v^{-v,\tau^*}$, $v \geq -\tau^*$, we can show that

$$r^p \int_\omega |D_L\psi|^2 \, d\omega \lesssim \int_\omega (r^p|D_L\psi|^2)(-v, v, \omega)\, d\omega + \int_{\underline{H}_v^{-v,\tau^*}} (r^{p-1}|D_L\psi|^2 + r^p|D_L\psi||D_{\underline{L}}D_L(r\phi)|)\, du\, d\omega.$$

The integral of the first term can be bounded by the assumption on the data. We control the second term by using the $r$-weighted energy estimate. We bound the last term as follows:

$$r^p|D_L\psi||D_{\underline{L}}D_L\psi| \lesssim \epsilon_1 r^{p_1}u_+^{p_2}|D_{\underline{L}}D_L\psi|^2 + \epsilon_1^{-1}r^{2p-p_1}u_+^{-p_2}|D_L\psi|^2, \quad \forall \epsilon_1 > 0.$$

When $2p \geq p_1$, we can use the $r$-weighted energy estimate (53) to bound the weighted integral of $|D_L\psi|$. Otherwise one can use interpolation and the integrated local energy decay estimate (56). For any case, from the energy decay estimates (53), (56), (63) and (64) for $\phi$, one can always show that

$$\iint_{\mathcal{D}_\tau} r^{2p-p_1}u_+^{-p_2}|D_L\psi|^2 \, du\, dv\, d\omega \lesssim_{M_2} \mathcal{E}_0[\phi]\tau_+^{p-1-\gamma_0}.$$

Another way to understand the above estimate is to use interpolation. It suffices to show the above estimate with $p=0$ and $p = 1+\gamma_0$. The former case follows by using the integrated local energy estimates for $\phi$, while the later situation relies on the $r$-weighted energy estimate. Estimate (69) for the exterior region case then follows. The interior region case holds in a similar way. □

As we only commute the equation with $\partial_t$ or the angular momentum $\Omega$, to estimate the weighted spacetime integral of $D_L D_{\underline{L}}(r\phi)$ in terms of $D_Z\phi$, we use the equation of $\phi$ under the null frame.

**Lemma 37.** *Under the null frame, we can write the covariant wave operator $\Box_A$ as*

$$r\Box_A\phi = r D^\mu D_\mu\phi = -D_L D_{\underline{L}}(r\phi) + \not{D}^2(r\phi) - i\rho\cdot r\phi = -D_{\underline{L}}D_L(r\phi) + \not{D}^2(r\phi) + i\rho\cdot r\phi \tag{70}$$

*for any complex scalar field $\phi$. Here $\not{D}^2 = \not{D}^{e_1}\not{D}_{e_1} + \not{D}^{e_2}\not{D}_{e_2}$ and $\rho = \frac{1}{2}(dA)_{\underline{L}L}$.*

*Proof.* The lemma follows by direct computation. □

This lemma leads to the following estimate for $D_L D_{\underline{L}}(r\phi)$ and $D_{\underline{L}} D_L(r\phi)$.

**Proposition 38.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $1+\epsilon \le p \le 1+\gamma_0$, we have*

$$I^p_{2+\gamma_0-p-2\epsilon}[r^{-1}(|D_L D_{\underline{L}}\psi| + |D_L D_{\underline{L}}\psi|)](\overline{\mathcal{D}}_\tau) \lesssim_{M_2} \mathcal{E}_0[\phi] + I^{-1-\epsilon}_{1+\gamma_0-\epsilon}[D\phi_1](\overline{\mathcal{D}}_\tau) + I^{\gamma_0}_0[\not{D}\psi_1](\overline{\mathcal{D}}_\tau). \quad (71)$$

*Here $\phi_1 = D_Z\phi$, $\psi_1 = D_Z(r\phi)$ and $\psi = r\phi$.*

*Proof.* Let's only consider the estimate for $D_L D_{\underline{L}}(r\phi)$ in the interior region. The proof easily implies the estimates for $D_{\underline{L}} D_L(r\phi)$. The case in the exterior region is easier since in that region $r \ge \frac{1}{3}u_+$. It hence suffices to show the estimate for $p = 1+\gamma_0$, which is similar to the proof for the interior region case. Take $\overline{\mathcal{D}}_\tau$ to be $\overline{\mathcal{D}}^{\tau_2}_{\tau_1}$ for $0 \le \tau_1 = \tau < \tau_2$. From the equation (70) for $\phi$ under the null frame, we derive

$$r^p |D_L D_{\underline{L}}(r\phi)|^2 \lesssim r^p |\Box_A \phi|^2 r^2 + r^p |r\phi\rho|^2 + r^p |r^{-1}\not{D} D_\Omega \psi|^2.$$

Here we note that $|\not{D}^2\psi|^2 \lesssim |r^{-1}\not{D} D_\Omega \psi|$. The integral of the first term on the right-hand side can be bounded by $\mathcal{E}_0[\phi]$. For the second term, we control $\phi$ by using Lemma 19. The last term is favorable as it is a form of $\not{D} D_Z \psi$. We absorb those terms with the help of the small constant $\epsilon_1$ from Propositions 35 and 36. According to our notation in this section, let $\psi_1 = D_\Omega \psi$. For all $1+\epsilon \le p \le 1+\gamma_0$, we have

$$\tau_+^{2+\gamma_0-p-2\epsilon} r^{p-2} \lesssim r^{\gamma_0} + \tau_+^{1+\gamma_0-\epsilon} r^{-1-\epsilon}, \quad r \ge R.$$

Since the energy flux for $\phi$ decays from Proposition 32, using Lemma 19 we conclude that

$$\int_\omega r^p |\phi|^2 \, d\omega \lesssim_{M_2} \mathcal{E}_0[\phi](\tau_1)_+^{p-2-\gamma_0}.$$

Therefore, for all $1+\epsilon \le p \le 1+\gamma_0$ we can show that

$$\begin{aligned}
&\iint_{\overline{\mathcal{D}}^{\tau_2}_{\tau_1}} \tau_+^{2+\gamma_0-p-2\epsilon} r^p |D_L D_{\underline{L}}(r\phi)|^2 \, dv \, du \, d\omega \\
&\qquad \lesssim I^p_{2+\gamma_0-p-2\epsilon}[\Box_A\phi](\overline{\mathcal{D}}^{\tau_2}_{\tau_1}) + I^{-1-\epsilon}_{1+\gamma_0-\epsilon}[D\phi_1](\overline{\mathcal{D}}^{\tau_2}_{\tau_1}) + I^{\gamma_0}_0[\not{D}\psi_1](\overline{\mathcal{D}}^{\tau_2}_{\tau_1}) \\
&\qquad\qquad + \int_{\tau_1}^{\tau_2} \tau_+^{2+\gamma_0-p-2\epsilon} \int_{\frac{1}{2}(\tau+R)}^{\infty} \int_\omega r^p |\phi|^2 \, d\omega \cdot \sum_{j\le 2} \int_\omega r^2 |\mathcal{L}^j_\Omega \bar{\rho}|^2 \, d\omega \, dv \, du \\
&\qquad \lesssim_{M_2} \mathcal{E}_0[\phi] + I^{-1-\epsilon}_{1+\gamma_0-\epsilon}[D\phi_1](\overline{\mathcal{D}}^{\tau_2}_{\tau_1}) + I^{\gamma_0}_0[\not{D}\psi_1](\overline{\mathcal{D}}^{\tau_2}_{\tau_1})
\end{aligned}$$

This finishes the proof. □

Next we estimate the weighted spacetime norm of $|\alpha||D_{\underline{L}}(r\phi)|$.

**Proposition 39.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Denote $\psi = r\phi$. For all $1+\epsilon \le p \le 1+\gamma_0$, $\epsilon_1 > 0$, we have*

$$\iint_{\overline{\mathcal{D}}_\tau} u_+^{2+\gamma_0+\epsilon-p} r^p |\alpha|^2 |D_{\underline{L}}(r\phi)|^2 \, dx \, dt \lesssim_{M_2} \mathcal{E}_0[\phi] \epsilon_1^{-1} \tau_+^{-\gamma_0+\epsilon} + \epsilon_1 I^{1+\epsilon}_{1+\epsilon}[r^{-1} D_L D_{\underline{L}}(r\phi)](\overline{\mathcal{D}}_\tau). \quad (72)$$

*Proof.* Make use of Proposition 35. For all $1+\epsilon \le p \le 1+\gamma_0$, we can show that

$$\begin{aligned}\iint_{\overline{\mathcal{D}}_\tau} r^p |\alpha|^2 |D_{\underline{L}}(r\phi)|^2 \, dx\, dt &\lesssim \|D_{\underline{L}}\psi\|^2_{L^2_u L^\infty_v L^2_\omega(\overline{\mathcal{D}}_\tau)} \cdot \left\| r^{(p/2)+1}\alpha \right\|^2_{L^\infty_u L^2_v L^\infty_\omega(\overline{\mathcal{D}}_\tau)} \\ &\lesssim \|D_{\underline{L}}\psi\|^2_{L^2_u L^\infty_v L^2_\omega(\overline{\mathcal{D}}_\tau)} \cdot \sum_{j\le 2} \left\| r^{(p/2)+1}\mathcal{L}^j_\Omega \alpha \right\|^2_{L^\infty_u L^2_v L^2_\omega(\overline{\mathcal{D}}_\tau)} \\ &\lesssim_{M_2} \mathcal{E}_0[\phi]\epsilon_1^{-1}\tau_+^{p-2-2\gamma_0} + \epsilon_1 \tau_+^{p-1-\gamma_0} I_0^{1+\epsilon}[r^{-1}D_L D_{\underline{L}}\psi](\overline{\mathcal{D}}_\tau)\end{aligned}$$

for all $\epsilon_1 > 0$. As the above estimate holds for all $\tau \in \mathbb{R}$, from Lemma 20, we conclude that

$$\iint_{\overline{\mathcal{D}}_\tau} u_+^{2+\gamma_0+\epsilon-p} r^p |\alpha|^2 |D_{\underline{L}}(r\phi)|^2 \, dx\, dt \lesssim_{M_2} \mathcal{E}_0[\phi]\epsilon_1^{-1}\tau_+^{-\gamma_0+\epsilon} + \epsilon_1 I^{1+\epsilon}_{1+\epsilon}[r^{-1}D_L D_{\underline{L}}\psi](\overline{\mathcal{D}}_\tau).$$

This finishes the proof for estimate (72). □

Next we estimate the weighted spacetime integral of $(|\underline{\alpha}| + r^{-1}|\rho|)|D_L(r\phi)|$. One possible way to bound this term, in particular $\underline{\alpha}$, is to make use of the energy flux through the incoming null hypersurface. It turns out that we lose a little bit of decay in $u$ and we are not able to close the bootstrap argument later. An alternative way is to use $\sup_v \int_\omega |\underline{\alpha}|^2 \, d\omega$, which has to exploit the equation for $F$. For $\tau \in \mathbb{R}$, denote

$$h(\tau) = \sum_{k\le 1} \|\mathcal{L}^k_\Omega(r\underline{\alpha})\|^2_{L^\infty_v L^2_\omega(H_{\tau^*})} + \sum_{k\le 2} \int_{\Sigma_\tau} \frac{|\mathcal{L}^k_Z \overline{F}|^2}{r_+^{1+\epsilon}} r^2 \, d\tilde{v}\, d\omega. \tag{73}$$

Here $(\tilde{v}, \omega)$ are coordinates of $\Sigma_\tau$, that is, $(\tilde{v}, \omega) = (r, \omega)$ when $r \le R$ and $(\tilde{v}, \omega) = (v, \omega)$ otherwise. We cannot show that $h(\tau)$ decays in $\tau$. However, we can show the following:

**Corollary 40.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then the function $h(\tau)$ is integrable in $\tau$:*

$$\int_{\tau_1}^{\tau_2} \tau_+^{1+\epsilon} h(\tau)\, d\tau \lesssim M_2 \tau_+^{-\gamma_0+3\epsilon}, \qquad \int_{\tilde{\tau}\le\tau} \tilde{\tau}_+^{1+\epsilon} h(\tilde{\tau})\, d\tilde{\tau} \lesssim M_2 \tau_+^{-\gamma_0+3\epsilon} \tag{74}$$

*for all $0 \le \tau_1 < \tau_2$ and $\tau \le 0$.*

*Proof.* Using Lemma 20, the corollary follows from estimate (42) and the integrated local energy estimates (24) and (25) for the Maxwell field $F$. □

We now can estimate the weighted spacetime integral of $|\underline{\alpha}||D_L\psi|$.

**Proposition 41.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $1+\epsilon \le p \le 1+\gamma_0$, $\epsilon_1 > 0$, we have*

$$\begin{aligned}&\iint_{\overline{\mathcal{D}}_\tau} u_+^{2+\gamma_0+\epsilon-p} |\underline{\alpha}|^2 |D_L(r\phi)|^2 r^p \, dx\, dt \\ &\qquad \lesssim_{M_2} \epsilon_1 \iint_{\overline{\mathcal{D}}_\tau} \tilde{\tau}_+^{2+\gamma_0+\epsilon-p} h(\tilde{\tau}) r^p |D_L(r D_Z \phi)|^2 \, dv\, d\omega\, d\tilde{\tau} + \epsilon_1^{-1}\mathcal{E}_0[\phi]\tau_+^{-\gamma_0+3\epsilon}.\end{aligned} \tag{75}$$

*Proof.* Let $\psi = r\phi$ and $\psi_1 = D_Z(r\phi)$. We first use Sobolev embedding on the unit sphere to bound

$$\big\| |r\underline{\alpha}||D_L\psi| \big\|_{L^2_\omega}^2 \lesssim \big( \|r\underline{\alpha}\|_{L^2_\omega}^2 + \|r\mathcal{L}_\Omega \underline{\alpha}\|_{L^2_\omega}^2 \big) \cdot \big( \epsilon_1^{-1} \|D_L\psi\|_{L^2_\omega}^2 + \epsilon_1 \|D_\Omega D_L \psi\|_{L^2_\omega}^2 \big), \quad \epsilon_1 > 0.$$

The proof for this estimate for all connections $A$ follows from the case when $A$ is trivial, as the norm is gauge invariant. We in particular can choose a gauge so that the function is real, then make use of estimate (42) of Proposition 17. We therefore can show that

$$\begin{aligned}
&\big\| r^{p/2} u_+^{(2+\gamma_0+\epsilon-p)/2} |r\underline{\alpha}||D_L\psi| \big\|^2_{L^2_u L^2_v L^2_\omega(\overline{\mathcal{D}}_\tau)} \\
&\quad\lesssim \bigg\| r^{p/2} \sum_{k\le 1} \|\mathcal{L}_\Omega^k (r\underline{\alpha})\|_{L^2_\omega} \cdot u_+^{(2+\gamma_0+\epsilon-p)/2} \big( \epsilon_1^{1/2} \|D_\Omega D_L\psi\|_{L^2_\omega} + \epsilon_1^{-1/2} \|D_L\psi\|_{L^2_\omega} \big) \bigg\|^2_{L^2_u L^2_v} \\
&\quad\lesssim \big\| \tilde{\tau}_+^{2+\gamma_0+\epsilon-p} h(\tilde{\tau}) \big( \epsilon_1 \|r^{p/2} D_L D_\Omega \psi\|^2_{L^2_v L^2_\omega} + \epsilon_1 \|r^{p/2} r\alpha\psi\|^2_{L^2_v L^2_\omega} + \epsilon_1^{-1} \|r^{p/2} D_L\psi\|^2_{L^2_v L^2_\omega} \big) \big\|_{L^1_u} \\
&\quad\lesssim \epsilon_1 \int_{\tilde{\tau}} \tilde{\tau}_+^{2+\gamma_0+\epsilon-p} h(\tilde{\tau}) \int_{H_{\tilde{\tau}*}} r^p |D_L\psi_1|^2 \, dv \, d\omega \, d\tilde{\tau} + \epsilon_1^{-1} \|\tilde{u}_+^{1+\epsilon} h(\tilde{u})\|_{L^1_u} \|u_+^{1+\gamma_0-p/2} r^{p/2} D_L\psi\|^2_{L^\infty_u L^2_v L^2_\omega} \\
&\qquad\qquad + \epsilon_1 \|\tilde{u}_+^{1+\epsilon-\gamma_0} h(\tilde{u})\|_{L^1_u} \|u_+^{1+\gamma_0-p/2} r^{p/2} r\alpha\|^2_{L^\infty_u L^2_v L^\infty_\omega} \|u_+^{\gamma_0/2}\psi\|^2_{L^\infty_u L^\infty_v L^2_\omega} \\
&\quad\lesssim_{M_2} \epsilon_1 \iint_{\overline{\mathcal{D}}_\tau} \tilde{\tau}_+^{2+\gamma_0+\epsilon-p} h(\tilde{\tau}) r^p |D_L\psi_1|^2 \, dv \, d\omega \, du + \epsilon_1^{-1} \mathcal{E}_0[\phi] \tau_+^{-\gamma_0+3\epsilon}.
\end{aligned}$$

Here we have used the $r$-weighted energy estimates (27) and (28) and estimate (49) to bound $\phi$. □

For $|r^{-1}\rho||D_L\psi|$, we have extra decay in $r$, which allows us to use Proposition 36.

**Proposition 42.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $\epsilon_1 > 0$, we have*

$$\iint_{\overline{\mathcal{D}}_\tau} u_+^{1+\gamma_0} |r^{-1}\rho|^2 |D_L(r\phi)|^2 r^{1+\gamma_0} \, dx \, dt \lesssim_{M_2} \epsilon_1 I_\epsilon^{1+\epsilon}[r^{-1} D_{\underline{L}} D_L(r\phi)](\mathcal{D}_{\tau_1}^{\tau_2}) + \epsilon_1^{-1} \mathcal{E}_0[\phi]. \tag{76}$$

*Proof.* The idea is that we bound $\rho$ by using the energy flux through the incoming null hypersurface and $D_L(r\phi)$ by using Proposition 36. In the exterior region, we need to specially consider the effect of the nonzero charge. Other than that, the proof is the same for the interior region case. We thus take $\overline{\mathcal{D}}_\tau$ to be $\mathcal{D}_\tau$ with $\tau \le 0$. Let $\psi = r\phi$. For all $1+\epsilon \le p \le 1+\gamma_0$, we can show that

$$\begin{aligned}
\iint_{\mathcal{D}_\tau} |r^{-1}\rho|^2 |D_L\psi|^2 r^p \, dx \, dt &\lesssim \iint_{\mathcal{D}_\tau} |\bar{\rho}|^2 |D_L\psi|^2 r^p \, du \, dv \, d\omega + \iint_{\mathcal{D}_\tau} |q_0|^2 |D_L\psi|^2 r^{p-4} \, du \, dv \, d\omega \\
&\lesssim_{M_2} \|D_L\psi\|^2_{L^2_v L^\infty_u L^2_\omega(\mathcal{D}_\tau)} \|r\bar{\rho}\|^2_{L^\infty_v L^2_u L^\infty_\omega(\mathcal{D}_\tau)} + \mathcal{E}_0[\phi] \tau_+^{-1-2\gamma_0} \\
&\lesssim_{M_2} \epsilon_1 \tau_+^{-1-\gamma_0} I_0^{1+\epsilon}[r^{-1} D_{\underline{L}} D_L\psi](\mathcal{D}_\tau) + \epsilon_1^{-1} \mathcal{E}_0[\phi] \tau_+^{-1-2\gamma_0}.
\end{aligned}$$

The above estimate also holds for the interior region case when $\overline{\mathcal{D}}_\tau = \mathcal{D}^{\tau_2}_{\tau_1}$ for all $0 \le \tau = \tau_1 < \tau_2$. From Lemma 20, we then can show, taking the interior region for example, that

$$\begin{aligned}\iint_{\mathcal{D}^{\tau_2}_{\tau_1}} \tau_+^{1+\gamma_0} |r^{-1}\rho|^2 |D_L\psi|^2 r^p \, dx\, dt &\\ \lesssim_{M_2} \epsilon_1 I_0^{1+\epsilon}[r^{-1} D_{\underline{L}} D_L \psi](\mathcal{D}^{\tau_2}_{\tau_1}) + \epsilon_1 \int_{\tau_1}^{\tau_2} \tau_+^{-1} I_0^{1+\epsilon}[r^{-1} D_{\underline{L}} D_L \psi](\mathcal{D}^{\tau_2}_{\tau}) \, d\tau + \epsilon_1^{-1}\mathcal{E}_0[\phi] &\\ \lesssim_{M_2} \epsilon_1 I_\epsilon^{1+\epsilon}[r^{-1} D_{\underline{L}} D_L \psi](\mathcal{D}^{\tau_2}_{\tau_1}) + \epsilon_1^{-1}\mathcal{E}_0[\phi].&\end{aligned}$$

Here we note that $\ln \tau_+ \lesssim \tau_+^\epsilon$. □

Next we estimate $r^{-1}|F||\not{D}(r\phi)|$.

**Proposition 43.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $\epsilon_1 > 0$, we have*

$$\iint_{\overline{\mathcal{D}}_\tau} u_+^{1+\gamma_0} |r^{-1}F|^2 |\not{D}(r\phi)|^2 r^{1+\gamma_0} \, dx\, dt \lesssim_{M_2} \epsilon_1^{-1}\mathcal{E}_0[\phi] + \epsilon_1 \int_{\tilde{\tau}} \tilde{\tau}_+^{1+\gamma_0} h(\tilde{\tau}) E[D_Z\phi](H_{\tilde{\tau}^*}) \, d\tilde{\tau}. \tag{77}$$

*Proof.* The idea is to use the energy flux through the outgoing null hypersurface to bound $\not{D}(r\phi) = D_\Omega\phi$ and the integrated local energy estimate to control $F$. We only show the estimate in the exterior region. Take $\overline{\mathcal{D}_\tau}$ to be $\mathcal{D}_\tau$ for any $\tau \le 0$. In the exterior region we have the relation $r \ge \frac{1}{3}u_+$. Therefore, from estimate (50) and the definition (73) of $h(\tau)$, we can show that

$$\begin{aligned}\iint_{\mathcal{D}_\tau} u_+^{1+\gamma_0} |F|^2 |\not{D}(r\phi)|^2 r^{1+\gamma_0} \, du\, dv\, d\omega &\\ \lesssim \int_u u_+^{1+\gamma_0} \int_v \sum_{k\le 2} \left( r^{1-\epsilon} \int_\omega \left( |\mathcal{L}_\Omega^k \overline{F}|^2 + |q_0 r^{-2}|^2 \right) d\omega \right) \cdot \int_\omega r |D_\Omega\phi|^2 \, d\omega\, dv\, du &\\ \lesssim |q_0|^2 \iint_{\mathcal{D}_\tau} \frac{|\not{D}\phi|^2}{r^{1+\epsilon}} \, dx\, dt + \int_{\tilde{\tau}} (\tilde{\tau})_+^{1+\gamma_0} h(\tilde{\tau}) \left( \epsilon_1^{-1} \int_{H_{\tilde{\tau}^*}} |\not{D}\phi|^2 r^2 \, dv\, d\omega + \epsilon_1 E[D_Z\phi](H_{\tilde{\tau}^*}) \right) d\tilde{\tau} &\\ \lesssim_{M_2} \mathcal{E}_0[\phi]\tau_+^{-1-\gamma_0} + \int_{\tilde{\tau}} (\tilde{\tau})_+^{1+\gamma_0} h(\tilde{\tau}) \epsilon_1^{-1} E[\phi](H_{\tilde{\tau}^*}) \, d\tilde{\tau} + \epsilon_1 \int_u h(2u+R) E[D_Z\phi](H_u) \, du &\\ \lesssim_{M_2} \epsilon_1^{-1}\mathcal{E}_0[\phi]\tau_+^{-1-\gamma_0+2\epsilon} + \epsilon_1 \int_{\tilde{\tau}} \tilde{\tau}_+^{1+\gamma_0} h(\tilde{\tau}) E[D_Z\phi](H_{\tilde{\tau}^*}) \, d\tilde{\tau}.&\end{aligned}$$

Here we assumed that $\gamma_0 < 1$ and $\epsilon$ is sufficiently small. For the case $\gamma_0 = 1$, the above estimate also holds but in a different form where we have to rely on the $r$-weighted energy estimate. For the sake of simplicity, we do not discuss this in detail when $\gamma_0 \ge 1$. □

Finally, we estimate the weighted spacetime norm of $(|J| + |r\not{J}| + |\sigma| + |r^{-1}\rho|)|\phi|$. We show:

**Proposition 44.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $1+\epsilon \le p \le 1+\gamma_0$, we have*

$$\iint_{\overline{\mathcal{D}}_\tau} (|J|^2 + |r\not{J}|^2 + |\sigma|^2 + |r^{-1}\rho|^2)|\phi|^2 r^p u_+^{2+\gamma_0+\epsilon-p} \, dx\, dt \lesssim_{M_2} \mathcal{E}_0[\phi]\tau_+^{-\gamma_0+\epsilon}. \tag{78}$$

*Proof.* Let's first consider $(|J|^2+|\sigma|^2+|r^{-1}\rho|^2)|\phi|^2$. The idea is that we bound $\phi$ by the energy flux. Note that the nonzero charge only affects the estimates in the exterior region where $r \geq \frac{1}{3}u_+$. From the embedding (49) and the energy decay estimates (56) and (64), we can show that

$$\begin{aligned}
&\iint_{\overline{\mathcal{D}}_\tau} (|J|^2+|\sigma|^2+|r^{-1}\rho|^2)|\phi|^2 r^{p+2} u_+^{2+\gamma_0+\epsilon-p}\, du\, dv\, d\omega \\
&\quad \lesssim \int_u u_+^{2+\gamma_0+\epsilon-p} \int_v \sum_{k\leq 2} r^{p+1} \int_\omega \left(|\mathcal{L}_\Omega^k J|^2+|\mathcal{L}_\Omega^k \sigma|^2+|r^{-1}\mathcal{L}_\Omega^k \bar{\rho}|^2+|q_0 r^{-3}|^2\right) d\omega \cdot \int_\omega r|\phi|^2\, d\omega\, dv\, du \\
&\quad \lesssim_{M_2} \mathcal{E}_0[\phi] \int_u u_+^{1+\epsilon-p} \int_v \sum_{k\leq 2} r^{p+1} \int_\omega \left(|\mathcal{L}_\Omega^k J|^2+|\mathcal{L}_\Omega^k \sigma|^2+|r^{-1}\mathcal{L}_\Omega^k \bar{\rho}|^2+|q_0 r^{-3}|^2\right) d\omega\, dv\, du \\
&\quad \lesssim_{M_2} \tau_+^{-\gamma_0+\epsilon}.
\end{aligned}$$

Here we used the $r$-weighted energy estimates (31) and (32) to bound the curvature components and the definition for $M_2$ to control $J$. For $|r\not{J}|^2|\phi|^2$, the only difference is that we need to put more $r$ weights on $\phi$. By using the embedding inequality (49) and the energy decay estimates (56) and (64), we conclude that

$$\int_\omega r^{1+p-\gamma_0}|\phi|^2\, d\omega \lesssim_{M_2} \tau_+^{p-1-2\gamma_0}.$$

Therefore, we have

$$\begin{aligned}
&\iint_{\overline{\mathcal{D}}_\tau} u_+^{2+\gamma_0+\epsilon-p}|r\not{J}|^2|\phi|^2 r^{p+2}\, du\, dv\, d\omega \\
&\qquad \lesssim \int_u u_+^{2+\gamma_0+\epsilon-p} \int_v \sum_{k\leq 2} r^{3+\gamma_0} \int_\omega |\mathcal{L}_\Omega^k \not{J}|^2\, d\omega \cdot \int_\omega r^{1+p-\gamma_0}|\phi|^2\, d\omega\, du\, dv \\
&\qquad \lesssim_{M_2} \mathcal{E}_0[\phi] \sum_{k\leq 2} \iint_{\overline{\mathcal{D}}_\tau} u_+^{1-\gamma_0+\epsilon} r^{3+\gamma_0} |\mathcal{L}_\Omega^k \not{J}|^2\, d\omega\, dv\, du \\
&\qquad \lesssim_{M_2} \mathcal{E}_0[\phi]\tau_+^{-\gamma_0}.
\end{aligned}$$

Estimate (78) follows. □

Now it remains to consider the spacetime norm on the bounded region $\{r \leq R\}$.

**Proposition 45.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then on the bounded region $\{r \leq R\}$, for all $0 \leq \tau_1 < \tau_2$ we have*

$$\int_{\tau_1}^{\tau_2} \tau_+^{1+\gamma_0} \int_{r\leq R} |[\Box_A, D_Z]\phi|^2\, dx\, dt \lesssim_{M_2} \mathcal{E}_0[\phi](\tau_1)_+^{-1-\gamma_0}. \tag{79}$$

*Proof.* First we conclude from the energy estimate (64) that the energy flux of the scalar field decays:

$$E[\phi](\Sigma_\tau) \lesssim_{M_2} \tau_+^{-1-\gamma_0}, \quad \forall \tau \geq 0.$$

From the commutator estimate (66), we have

$$|[\Box_A, D_Z]\phi| \lesssim |F||\tilde{D}\phi| + |J||\phi|.$$

For the first term, we make use of estimate (39):

$$\begin{aligned}\int_{\tau_1}^{\tau_2}\tau_+^{1+\gamma_0}\int_{r\le R}|F|^2|\tilde{D}\phi|\,dx\,dt &\lesssim \int_{\tau_1}^{\tau_2}\sup_{|x|\le R}|F|^2(\tau,x)E[\phi](\Sigma_\tau)\tau_+^{1+\gamma_0}\,d\tau\\ &\lesssim_{M_2}\mathcal{E}_0[\phi]\int_{\tau_1}^{\tau_2}\sup_{|x|\le R}|F|^2(\tau,x)\,d\tau\\ &\lesssim_{M_2}\mathcal{E}_0[\phi](\tau_1)_+^{-1-\gamma_0}.\end{aligned}$$

For $|J||\phi|$, we use Sobolev embedding on the ball $B_R$ with radius $R$ at fixed time $\tau$:

$$\begin{aligned}\int_{\tau_1}^{\tau_2}\tau_+^{1+\gamma_0}\int_{r\le R}|J|^2|\phi|^2\,dx\,dt &\lesssim \int_{\tau_1}^{\tau_2}\tau_+^{1+\gamma_0}\|J\|^2_{H^1_x(B_R)}\cdot\|\phi\|^2_{H^1_x(B_R)}\,d\tau\\ &\lesssim_{M_2}\int_{\tau_1}^{\tau_2}\mathcal{E}_0[\phi]\int_{r\le R}|\overline{\nabla}J|^2+|J|^2\,dx\,d\tau\\ &\lesssim_{M_2}\mathcal{E}_0[\phi](\tau_1)_+^{-1-\gamma_0}.\end{aligned}$$

Thus, estimate (79) holds. □

Now, from Lemma 33, combine estimates (72) and (75)-(79). We can bound the first-order commutator.

**Corollary 46.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all positive constants $\epsilon_1<1$, we have*

$$\begin{aligned}&I^{1+\gamma_0}_{1+\epsilon}[[\Box_A,D_Z]\phi](\{t\ge0\})+I^{1+\gamma_0}_{1+\epsilon}[[\Box_A,D_Z]\phi](\{t\ge0\})\\ &\quad\lesssim_{M_2}\epsilon_1 I^{-1-\epsilon}_{1+\gamma_0-\epsilon}[D\phi_1](\{t\ge0\})+\epsilon_1 I^{\gamma_0}_0[\not{D}\psi_1](\{t\ge0\}\cap\{r\ge R\})+\mathcal{E}_0[\phi]\epsilon_1^{-1}\\ &\qquad+\epsilon_1\int_{\mathbb{R}}\tau_+^{1+\gamma_0}h(\tau)E[D_Z\phi](\Sigma_\tau)\,d\tau+\epsilon_1\int_{\mathbb{R}}\int_{H_{\tau^*}}\tau_+^{2+\gamma_0+\epsilon-p}h(\tau)r^p|D_L\psi_1|^2\,dv\,d\omega\,d\tau.\end{aligned}\tag{80}$$

*Here $\phi_1=D_Z\phi$, $\psi_1=D_Z(r\phi)$ and $Z\in\Gamma\{\partial_t,\Omega_{ij}\}$.*

*Proof.* From Lemma 33, estimate (80) is a consequence of estimates (72), (75), (77), (76), (78) and (79). The term $I^{1+\epsilon}_{1+\epsilon}[r^{-1}D_{\underline{L}}D_L\psi](D)$ can further be controlled by using Proposition 38 with $p=1+\epsilon$. □

Now we are able to derive the energy decay estimates for the first-order derivative of the scalar field. Based on the result for the decay estimates for $\phi$ in the previous subsection, it suffices to bound $\mathcal{E}_0[D_Z\phi]$.

**Proposition 47.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then we have the bound*

$$\mathcal{E}_0[D_Z\phi]\lesssim_{M_2}\mathcal{E}_1[\phi].\tag{81}$$

*Proof.* First, by definition,

$$\mathcal{E}_0[D_Z\phi]\lesssim\mathcal{E}_1[\phi]+I^{1+\gamma_0}_{1+\epsilon}[[\Box_A,D_Z]\phi](\{t\ge0\})+I^{1+\epsilon}_{1+\gamma_0}[[\Box_A,D_Z]\phi](\{t\ge0\}).$$

Then from the previous estimate (80), the above inequality leads to

$$\mathcal{E}_0[D_Z\phi] \lesssim_{M_2} \epsilon_1 I^{-1-\epsilon}_{1+\gamma_0-\epsilon}[DD_Z\phi](\{t\geq 0\}) + \epsilon_1 I^{\gamma_0}_0[\not{D}\psi_1](\{t\geq 0\}\cap\{r\geq R\}) + \mathcal{E}_1[\phi]\epsilon_1^{-1} + \epsilon_1\int_{\mathbb{R}} \tau_+^{1+\gamma_0} h(\tau) E[D_Z\phi](\Sigma_\tau)\,d\tau + \epsilon_1\int_{\mathbb{R}}\int_{H_{\tau^*}} \tau_+^{2+\gamma_0+\epsilon-p} h(\tau) r^p |D_L\psi_1|^2\,dv\,d\omega\,d\tau$$

for all $0<\epsilon_1<1$. Here $\phi_1 = D_Z\phi$, $\psi_1 = D_Z(r\phi)$ and the implicit constant is independent of $\epsilon_1$.

Now from the integrated local energy estimates (56) and (59) combined with Lemma 20, we can show that

$$I^{-1-\epsilon}_{1+\gamma_0-\epsilon}[DD_Z\phi](\{t\geq 0\}) \lesssim_{M_2} \mathcal{E}_0[D_Z\phi].$$

By the energy decay estimates (23) and (64), we have the energy decay for $D_Z\phi$:

$$E[D_Z\phi](\Sigma_\tau) \lesssim_{M_2} \mathcal{E}_0[D_Z\phi]\tau_+^{-1-\gamma_0}, \quad \forall \tau\in\mathbb{R}.$$

Moreover, the $r$-weighted energy estimates (53) and (63) imply that

$$\tau_+^{1+\gamma_0-p}\int_{H_{\tau^*}} r^p |D_L(rD_Z\phi)|^2\,dv\,d\omega + \int_{\mathbb{R}}\int_{H_{\tau^*}} r^{\gamma_0}|\not{D}(rD_Z\phi)|^2\,dv\,d\omega\,d\tau \lesssim_{M_2} \mathcal{E}_0[D_Z\phi].$$

Recall the definition for $h(\tau)$ in line (73). By Corollary 40, we then can demonstrate that

$$\int_{\mathbb{R}} \tau_+^{1+\gamma_0} h(\tau) E[D_Z\phi](\Sigma_\tau)\,d\tau \lesssim_{M_2} \mathcal{E}_0[D_Z\phi]\int_{\mathbb{R}} h(\tau)\,d\tau \lesssim_{M_2} \mathcal{E}_0[D_Z\phi],$$

$$\int_{\mathbb{R}}\int_{H_{\tau^*}} \tau_+^{2+\gamma_0+\epsilon-p} h(\tau) r^p |D_L\psi_1|^2\,dv\,d\omega\,d\tau \lesssim_{M_2} \mathcal{E}_0[D_Z\phi]\int_{\mathbb{R}} \tau_+^{1+\epsilon} h(\tau)\,d\tau \lesssim_{M_2} \mathcal{E}_0[D_Z\phi].$$

We therefore derive that

$$\mathcal{E}_0[D_Z\phi] \lesssim_{M_2} \epsilon_1\mathcal{E}_0[D_Z\phi] + \epsilon_1^{-1}\mathcal{E}_1[\phi], \quad \forall 0<\epsilon_1<1.$$

Take $\epsilon_1$ to be sufficiently small, depending only on $M_2$, $\gamma_0$, $R$ and $\epsilon$. We then obtain estimate (81). □

The above argument implies all the desired energy decay estimates for the first-order derivative of the scalar field in terms of $\mathcal{E}_1[\phi]$. Moreover, estimate (80) can be improved as follows:

**Corollary 48.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all positive constants $\epsilon_1<1$, we have*

$$I^{1+\gamma_0}_{1+\epsilon}[[\Box_A, D_Z]\phi](\{t\geq 0\}) + I^{1+\gamma_0}_{1+\epsilon}[[\Box_A, D_Z]\phi](\{t\geq 0\}) \lesssim_{M_2} \epsilon_1\mathcal{E}_1[\phi] + \mathcal{E}_0[\phi]\epsilon_1^{-1}. \tag{82}$$

**4.3.4.** *Energy decay estimates for the second-order derivatives of the scalar field.* In this subsection, we establish the energy decay estimates for the second-order derivative of the scalar field. Note that the definition of $M_2$ records the size and regularity of the connection field $A$, which is independent of the scalar field. In particular, Proposition 47 and Corollary 48 apply to $\phi_1 = D_Z\phi$:

$$\mathcal{E}_0[D_Z\phi_1] \lesssim_{M_2} \mathcal{E}_1[\phi_1],$$

$$I^{1+\gamma_0}_{1+\epsilon}[[\Box_A, D_Z]\phi_1](\{t\geq 0\}) + I^{1+\gamma_0}_{1+\epsilon}[[\Box_A, D_Z]\phi_1](\{t\geq 0\}) \lesssim_{M_2} \epsilon_1\mathcal{E}_1[\phi_1] + \mathcal{E}_1[\phi]\epsilon_1^{-1}$$

for all $0 < \epsilon_1 < 1$. Here $\mathcal{E}_0[\phi_1] \lesssim_{M_2} \mathcal{E}_1[\phi]$ by Proposition 47. To derive the energy decay estimates for the second-order derivative of the solution, it suffices to bound $\mathcal{E}_1[\phi_1]$. As $\phi_1 = D_Z\phi$, by definition

$$\begin{aligned}
\mathcal{E}_1[\phi_1] &= \mathcal{E}_0[\phi_1] + E_0^1[\phi_1] + I_{1+\epsilon}^{1+\gamma_0}[D_Z\Box_A\phi_1](\{t \geq 0\}) + I_{1+\gamma_0}^{1+\epsilon}[D_Z\Box_A\phi_1](\{t \geq 0\}) \\
&\lesssim \mathcal{E}_2[\phi] + I_{1+\epsilon}^{1+\gamma_0}[D_Z[\Box_A, D_Z]\phi](\{t \geq 0\}) + I_{1+\gamma_0}^{1+\epsilon}[D_Z[\Box_A, D_Z]\phi](\{t \geq 0\}) \\
&\lesssim \mathcal{E}_2[\phi] + I_{1+\epsilon}^{1+\gamma_0}[[D_Z, [\Box_A, D_Z]]\phi](\{t \geq 0\}) + I_{1+\gamma_0}^{1+\epsilon}[[D_Z, [\Box_A, D_Z]]\phi](\{t \geq 0\}) \\
&\quad + I_{1+\epsilon}^{1+\gamma_0}[[\Box_A, D_Z]D_Z\phi](\{t \geq 0\}) + I_{1+\gamma_0}^{1+\epsilon}[[\Box_A, D_Z]D_Z\phi](\{t \geq 0\}) \\
&\lesssim_{M_2} \mathcal{E}_2[\phi] + I_{1+\epsilon}^{1+\gamma_0}[[D_Z, [\Box_A, D_Z]]\phi](\{t \geq 0\}) + I_{1+\gamma_0}^{1+\epsilon}[[D_Z, [\Box_A, D_Z]]\phi](\{t \geq 0\}) \\
&\quad + \epsilon_1\mathcal{E}_1[\phi_1] + \mathcal{E}_1[\phi]\epsilon_1^{-1}
\end{aligned}$$

for $0 < \epsilon_1 < 1$. Let $\epsilon_1$ be sufficiently small. We then conclude that

$$\mathcal{E}_1[\phi_1] \lesssim_{M_2} \mathcal{E}_2[\phi] + I_{1+\epsilon}^{1+\gamma_0}[[D_Z, [\Box_A, D_Z]]\phi](\{t \geq 0\}) + I_{1+\gamma_0}^{1+\epsilon}[[D_Z, [\Box_A, D_Z]]\phi](\{t \geq 0\}).$$

Therefore, bounding $\mathcal{E}_1[\phi_1]$ is reduced to controlling the second-order commutator $[D_Z, [\Box_A, D_Z]]\phi$.

First, we have the following analogue of Lemma 33.

**Lemma 49.** *For all $X, Y \in \Gamma$, when $r \geq R$, we have*

$$|[D_X, [\Box_A, D_Y]]\phi| \lesssim |[\Box_{\mathcal{L}_Z A}, D_Z]\phi| + \big(|F|^2 + |r\alpha||r\underline{\alpha}| + |r\sigma|^2 + |r\rho|(|\underline{\alpha}| + |\alpha|)\big)|\phi|. \tag{83}$$

*When $r \leq R$, we have*

$$|[D_X, [\Box_A, D_Y]]\phi| \lesssim |[\Box_{\mathcal{L}_Z A}, D_Z]\phi| + |[\Box_A, D_Z]\phi| + |F|^2|\phi|. \tag{84}$$

*Here we note that $\mathcal{L}_Z F = \mathcal{L}_Z\, dA = d\mathcal{L}_Z A$.*

*Proof.* First, from Lemma 4, we can write

$$[\Box_A, D_X]\phi = 2iX^\nu F_{\mu\nu}D^\mu\phi + i\nabla^\mu(F_{\mu\nu}X^\nu)\phi.$$

We need to compute the double commutator $[D_Y, [\Box_A, D_X]]\phi$ for $X, Y \in \Gamma$. We can compute that

$$\begin{aligned}
[D_Y, [\Box_A, D_X]]\phi &= \mathcal{L}_Y(2iX^\nu F_{\mu\nu}D^\mu + i\nabla^\mu(F_{\mu\nu}X^\nu))\phi \\
&= 2i(\mathcal{L}_Y F)(D\phi, X) + 2iF([D_Y, D\phi], X) + 2iF(D\phi, [Y, X]) + iY(\nabla^\mu(F_{\mu\nu}X^\nu))\phi.
\end{aligned}$$

Here

$$\begin{aligned}
[D_Y, D\phi] &= D^\mu\phi[Y, \nabla_\mu] + [D_Y, D^\mu]\phi\partial_\mu \\
&= iF_{Y\nabla_\mu}\phi\nabla^\mu - (\nabla^\mu Y^\nu + \nabla^\nu Y^\mu)D_\mu\phi\partial_\nu.
\end{aligned}$$

As $X, Y \in \Gamma$ for $\Gamma = \{\partial_t, \Omega_{ij}\}$, we conclude that $X, Y$ are Killing:

$$\nabla^\mu X^\nu + \nabla^\nu X^\mu = 0, \quad \nabla^\mu Y^\nu + \nabla^\nu Y^\mu = 0.$$

This implies that the following term can be simplified:

$$Y(\nabla^\mu(F_{\mu\nu}X^\nu)) = [Y,\nabla^\mu]F(\nabla_\mu,X)+\nabla^\mu(\mathcal{L}_YF)(\nabla_\mu,X)+\nabla^\mu F(\mathcal{L}_Y\nabla_\mu,X)+\nabla^\mu F(\nabla_\mu,\mathcal{L}_YX)$$
$$= \nabla^\mu(\mathcal{L}_YF)(\nabla_\mu,X)+\nabla^\mu F(\nabla_\mu,[Y,X]).$$

Therefore, we can write the double commutator as

$$[D_Y,[\Box_A,D_X]]\phi = 2i(\mathcal{L}_YF)(D\phi,X)+i\nabla^\mu(\mathcal{L}_YF)(\nabla_\mu,X)$$
$$+2iF(D\phi,[Y,X])+i\nabla^\mu F(\nabla_\mu,[Y,X])\phi-2F^\mu_XF_{Y\mu}\phi.$$

Note that $[X,Y]\in\mathrm{span}\{\Gamma\}$ for $X,Y\in\Gamma=\{\partial_t,\Omega_{ij}\}$. We thus can write

$$2iF(D\phi,[Y,X])+i\partial^\mu F(\partial_\mu,[Y,X])\phi=[\Box_A,D_{[Y,X]}]\phi,$$

which can be bounded using Lemma 33. The term

$$2i(\mathcal{L}_YF)(D\phi,X)+i\nabla^\mu(\mathcal{L}_YF)(\nabla_\mu,X)$$

has the same form with $[\Box_A,D_X]\phi$ if we replace $F$ with $\mathcal{L}_YF$. In particular, the bound follows from Lemma 33. Therefore, to show this lemma, it remains to control $F^\mu_XF_{Y\mu}\phi$ for $X,Y\in\Gamma$. This term has crucial null structure we need to exploit when $r\geq R$. The main difficulty is that the angular momentum $\Omega$ contains weights in $r$. If both $X,Y\in\Omega$, then

$$|F^\mu_XF_{Y\mu}|\lesssim|r\alpha||r\underline{\alpha}|+|r\sigma|^2.$$

If $X=Y=\partial_t$, then

$$|F^\mu_XF_{Y\mu}|\lesssim|F|^2.$$

If one and only one of $X,Y$ is $\partial_t$, then the null structure is as follows:

$$|F^\mu_XF_{Y\mu}|\lesssim r|F^\mu_LF_{e_i\mu}|+r|F^\mu_{\underline{L}}F_{e_i\mu}|$$
$$\lesssim r(|\rho|+|\sigma|)(|\underline{\alpha}|+|\alpha|).$$

We see that the "bad" term $r|\underline{\alpha}|^2$ does not appear on the right-hand side. Hence

$$|F^\mu_XF_{Y\mu}|\lesssim|F|^2+|r\alpha||r\underline{\alpha}|+|r\sigma|^2+|r\rho|^2,\quad\forall X,Y\in\Gamma.$$

Therefore, estimate (83) holds. On the bounded region $\{r\leq R\}$, null structure is not necessary and estimate (84) follows trivially. □

The above lemma shows that the double commutator $[D_Z,[\Box_A,D_Z]]\phi$ consists of the quadratic part $[\Box_{\mathcal{L}_Z^kA},D_Z]\phi$, which can be bounded similarly to $[\Box_A,D_Z]\phi$ as we can put one more derivative $D_Z$ on the scalar field $\phi$ when we do Sobolev embedding. It thus suffices to control those cubic terms in (83).

**Proposition 50.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then we have*

$$\sum_{k\le 1} I^{1+\epsilon}_{1+\gamma_0}[[\Box_{\mathcal{L}_Z^k A}, D_Z]\phi](\{t\ge 0\}) + I^{1+\gamma_0}_{1+\epsilon}[[\Box_{\mathcal{L}_Z^k A}, D_Z]\phi](\{t\ge 0\})$$

$$\lesssim_{M_2} \epsilon_1 I^{-1-\epsilon}_{1+\gamma_0-\epsilon}[D\phi_2](\{t\ge 0\}) + \epsilon_1 I^{\gamma_0}_0[\not{D}\psi_2](\{t\ge 0\}\cap\{r\ge R\}) + \mathcal{E}_1[\phi]\epsilon_1^{-1}$$

$$+\epsilon_1\int_{\mathbb{R}} \tau_+^{1+\gamma_0} h(\tau) E[\phi_2](\Sigma_\tau)\, d\tau + \epsilon_1 \int_{\mathbb{R}}\int_{H_{\tau^*}} \tau_+^{2+\gamma_0+\epsilon-p} h(\tau) r^p |D_L\psi_2|^2\, dv\, d\omega\, d\tau \quad (85)$$

*for all positive constants $\epsilon_1$. Here $\phi_2 = D_Z^2\phi$, $\psi_2 = D_Z^2(r\phi)$. The function $h(\tau)$ is defined in (73).*

*Proof.* From Corollary 46 and the decay estimates for the first-order derivative of the scalar field, it suffices to consider estimate (85) with $k = 1$. The difference between estimate (80) and estimate (85) is that $F$ is replaced with $\mathcal{L}_Z F$ in (85). However, we are allowed to put one more derivative on the scalar field ($\phi_1 = D_Z\phi$ is replaced with $D_Z^2\phi$). Note that for the proof of estimate (80), the higher-order derivative comes in when we use Sobolev embedding on the sphere to bound $\|F\cdot D\phi\|_{L^2_\omega}$:

$$\|F\cdot D\phi\|_{L^2_\omega} \lesssim \sum_{k\le 2} \|\mathcal{L}_Z^k F\|_{L^2_\omega}\cdot\|D\phi\|_{L^2_\omega} \quad \text{or} \quad \sum_{k\le 1}\|\mathcal{L}_Z^k F\|_{L^2_\omega}\cdot\|DD_Z^k\phi\|_{L^2_\omega}.$$

For estimate (85), the corresponding term $\mathcal{L}_Z F\cdot D\phi$ can be bounded as follows:

$$\|\mathcal{L}_Z F\cdot D\phi\|_{L^2_\omega} \lesssim \sum_{k\le 1}\|\mathcal{L}_Z^k\mathcal{L}_Z F\|_{L^2_\omega}\cdot\|DD_Z^k\phi\|_{L^2_\omega} \quad \text{or} \quad \|\mathcal{L}_Z F\|_{L^2_\omega}\cdot\sum_{k\le 2}\|DD_Z^k\phi\|_{L^2_\omega}.$$

This is how we can transfer one derivative on $F$ to the scalar field $\phi$. In particular, estimate (85) holds. □

From Lemma 49, to bound the double commutator, it suffices to control the cubic terms in (83) and (84). We rely on the pointwise bound for the Maxwell field summarized in Propositions 14 and 17.

**Proposition 51.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $1+\epsilon\le p\le 1+\gamma_0$, we have*

$$I^p_{2+\gamma_0+\epsilon-p}[(|F|^2+|r\alpha||r\underline{\alpha}|+|r\sigma|^2+|r\rho|(|\underline{\alpha}|+|\alpha|))|\phi|](\{t\ge 0,\ r\ge R\})$$

$$+ I^p_{2+\gamma_0+\epsilon-p}[|F|^2|\phi|](\{t\ge 0,\ r\le R\}) \lesssim_{M_2} \mathcal{E}_1[\phi]. \quad (86)$$

*Proof.* On the bounded region $\{r\le R\}$, the weights $r^p$ have an upper bound. The Maxwell field $F$ can be bounded by using the pointwise estimate (40). We then can estimate the scalar field by using the integrated local energy estimates. Indeed, for all $0\le\tau_1<\tau_2$, we can show that

$$\int_{\tau_1}^{\tau_2}\int_{r\le R}\tau_+^{1+\gamma_0}|F|^4|\phi|^2\, dx\, d\tau \lesssim \int_{\tau_1}^{\tau_2}\tau_+^{1+\gamma_0}\sup|F|^4|\phi|^2\, dx\, d\tau$$

$$\lesssim_{M_2} (\tau_1)_+^{-2-2\gamma_0}\mathcal{E}_1[\phi].$$

For the cubic terms on the region $\{r \geq R\}$, let's first consider $|r\alpha||r\underline{\alpha}||\phi|$. We use the $r$-weighted energy estimates (31) and (32) for the Maxwell field to control $\alpha$, and the integrated decay estimate (42) of Proposition 17 to bound $\underline{\alpha}$. The reason that we cannot use the pointwise bound (43) is the weak decay rate there. The scalar field $\phi$ can be bounded by using Lemma 19. Indeed, for $1+\epsilon \leq p \leq 1+\gamma_0$, we can show that

$$
\begin{aligned}
&I^p_{2+\gamma_0+\epsilon-p}[|r\alpha||r\underline{\alpha}||\phi|](\{t \geq 0\} \cap \{r \geq R\}) \\
&\qquad \lesssim \iint u_+^{2+\gamma_0+\epsilon-p} r^{p+2} |r\alpha|^2 |r\underline{\alpha}|^2 |\phi|^2 \, du \, dv \, d\omega \\
&\qquad \lesssim \sum_{k\leq 1} \int_u \int_v u_+^{2+\gamma_0+\epsilon-p} r^{1+\gamma_0} \int_\omega |r\mathcal{L}_Z^k \alpha|^2 \, d\omega \cdot \int_\omega |r\mathcal{L}_Z^k \underline{\alpha}|^2 \, d\omega \cdot \int_\omega r^{p+1-\gamma_0} |\mathcal{L}_Z^k \phi|^2 \, d\omega \, dv \, du \\
&\qquad \lesssim_{M_2} \mathcal{E}_1[\phi] \sum_{k\leq 1} \int_u \int_v u_+^{1+\epsilon-\gamma_0} r^{1+\gamma_0} \int_\omega |r\mathcal{L}_Z^k \alpha|^2 \, d\omega \cdot \int_\omega |r\mathcal{L}_Z^k \underline{\alpha}|^2 \, d\omega \, dv \, du \\
&\qquad \lesssim_{M_2} \mathcal{E}_1[\phi] \sum_{k\leq 1} \int_u u_+^{1+\epsilon-\gamma_0} \int_v r^{1+\gamma_0} \int_\omega |r\mathcal{L}_Z^k \alpha|^2 \, d\omega \, dv \cdot \sup_v \int_\omega |r\mathcal{L}_Z^k \underline{\alpha}|^2 \, d\omega \, du \\
&\qquad \lesssim_{M_2} \mathcal{E}_1[\phi] \int_{\mathbb{R}} \tau_+^{1+\epsilon-\gamma_0} h(\tau) \, d\tau \\
&\qquad \lesssim_{M_2} \mathcal{E}_1[\phi].
\end{aligned}
$$

Here recall the definition of $h(\tau)$ in (73), and the last step follows from Corollary 40.

For $|F|^2|\phi|$, we use the pointwise estimates (43) and (44) of Proposition 17 to bound the Maxwell field $F$. The scalar field $\phi$ can be bounded using Lemma 19 as above. In the exterior region where the Maxwell field contains the charge part $q_0 r^{-2}\, dt \wedge dr$, we have the relation $r_+ \geq \frac{1}{2} u_+$. We can show that

$$
\begin{aligned}
&I^p_{2+\gamma_0+\epsilon-p}[|F|^2 \cdot \phi](\{t \geq 0\} \cap \{r \geq R\}) \\
&\qquad \lesssim \iint u_+^{2+\gamma_0+\epsilon-p} r^{p+2} |\overline{F}|^4 |\phi|^2 \, du \, dv \, d\omega + |q_0|^2 \iint_{t+R\leq r} u_+^{2+\gamma_0+\epsilon-p} r^{p+2-8} |\phi|^2 \, du \, dv \, d\omega \\
&\qquad \lesssim_{M_2} \int_u \int_v u_+^{2+\gamma_0+\epsilon-p-2-2\gamma_0} r^{p+2-4-1} \int_\omega r|\phi|^2 \, d\omega \, dv \, du + \mathcal{E}_0[\phi] \\
&\qquad \lesssim_{M_2} \mathcal{E}_1[\phi] \int_u \int_v u_+^{\epsilon-p-1-2\gamma_0} r^{p-3} \, dv \, du \\
&\qquad \lesssim_{M_2} \mathcal{E}_1[\phi].
\end{aligned}
$$

For $|r\sigma|^2|\phi|$, for the same reason as in the case of $|r\alpha||r\underline{\alpha}||\phi|$, we are not allowed to use the pointwise bound (44) to control $\sigma$ due to the strong $r$ weights here. Instead, we use the $r$-weighted energy estimate for $\sigma$ on the incoming null hypersurface together with the integrated decay estimate (46). We can show that

$$\iint_{\overline{\mathcal{D}}_{\tau_1}} r^p |r\sigma|^4 |\phi|^2 \, dx\, dt \lesssim \sum_{k\le 1} \int_u \int_v r^{1+\gamma_0} \int_\omega |r\mathcal{L}_Z^k \sigma|^2 \, d\omega \cdot \int_\omega |r\mathcal{L}_Z^k \sigma|^2 \, d\omega \cdot \int_\omega r^{p+1-\gamma_0} |\mathcal{L}_Z^k \phi|^2 \, d\omega \, dv \, du$$

$$\lesssim_{M_2} \mathcal{E}_1[\phi](\tau_1)_+^{-1+p-2\gamma_0} \sum_{k\le 1} \int_v \int_u r^{1+\gamma_0} \int_\omega |r\mathcal{L}_Z^k \sigma|^2 \, d\omega \, du \cdot \sup_u \int_\omega |r\mathcal{L}_Z^k \sigma|^2 \, d\omega \, dv$$

$$\lesssim_{M_2} \mathcal{E}_1[\phi](\tau_1)_+^{-1+p-2\gamma_0} \sum_{k\le 1} \|r\mathcal{L}_Z^k \sigma\|^2_{L_v^2 L_u^\infty L_\omega^2(\overline{\mathcal{D}}_{\tau_1})}$$

$$\lesssim_{M_2} \mathcal{E}_1[\phi](\tau_1)_+^{-2+p+\epsilon-3\gamma_0}.$$

This holds for all $\tau_1 \in \mathbb{R}$. Since

$$2 + 3\gamma_0 - \epsilon - p > 2 + \gamma_0 + \epsilon - p, \quad 0 \le p \le 1 + \gamma_0,$$

from Lemma 20, we obtain

$$I^p_{2+\gamma_0+\epsilon-p}[|r\sigma|^2\phi](\{t \ge 0\} \cap \{r \ge R\}) \lesssim_{M_2} \mathcal{E}_1[\phi].$$

Finally, for $|r\rho|(|\underline{\alpha}| + |\alpha|)|\phi|$, we need to take into consideration the charge effect in the exterior region. Except for this charge, the proof for the interior region case is the same. Let's merely estimate this cubic term in the exterior region. In particular, take $\overline{\mathcal{D}}_{\tau_1}$ to be $\mathcal{D}_{\tau_1}$ for some $\tau_1 < 0$. By using the $r$-weighted energy estimate for $\bar{\rho}$ and the pointwise bound (43) and (44) for $F$, for $0 \le p \le 1 + \gamma_0$ we then can show that

$$\iint_{\mathcal{D}_{\tau_1}} r^{p+2} |r\rho|^2 (|\alpha|^2 + |\underline{\alpha}|^2) |\phi|^2 \, du\, dv\, d\omega$$

$$\lesssim \iint_{\mathcal{D}_{\tau_1}} |q_0| r^p (|\alpha|^2 + |\underline{\alpha}|^2) |\phi|^2 \, du\, dv\, d\omega + \iint_{\mathcal{D}_{\tau_1}} r^{p+2} |r\bar{\rho}|^2 (|\alpha|^2 + |\underline{\alpha}|^2) |\phi|^2 \, du\, dv\, d\omega$$

$$\lesssim_{M_2} \mathcal{E}_1[\phi](\tau_1)_+^{-1-2\gamma_0} + \sum_{k\le 1} \int_u \int_v r^{p-1} \int_\omega |r\mathcal{L}_Z^k \bar{\rho}|^2 \, d\omega \cdot \sup(|r\underline{\alpha}|^2 + |r\alpha|^2) \cdot \int_\omega r |\mathcal{L}_Z^k \phi|^2 \, d\omega \, dv \, du$$

$$\lesssim_{M_2} \mathcal{E}_1[\phi](\tau_1)_+^{-1-2\gamma_0} + \mathcal{E}_1[\phi](\tau_1)_+^{-2-2\gamma_0} \sum_{k\le 1} \iint_{\mathcal{D}_{\tau_1}} r^{p-1} |r\mathcal{L}_Z^k \bar{\rho}|^2 \, du\, dv\, d\omega$$

$$\lesssim_{M_2} \mathcal{E}_1[\phi](\tau_1)_+^{-1-2\gamma_0}.$$

Here the last term is bounded by using the $r$-weighted energy estimates for $\bar{\rho}$. As $\tau_1$ is arbitrary, from Lemma 20, we derive that

$$I^p_{2+\gamma_0+\epsilon-p}[r\rho \cdot (|\underline{\alpha}| + |\alpha|) \cdot \phi](\{t \ge 0\} \cap \{r \ge R\}) \lesssim_{M_2} \mathcal{E}_1[\phi], \quad 1 + \epsilon \le p \le 1 + \gamma_0.$$

To summarize, we have shown (86). □

Propositions 50 and 51 together with Lemma 49 lead to the desired estimates for the double commutator $[D_X, [\Box_A, D_Y]]$ for $X, Y \in \Gamma$. Then by the argument at the beginning of this section, we have control of

$\mathcal{E}_0[D_X D_Y \phi]$. By using the same argument as Proposition 47, we then can bound $\mathcal{E}_0[D_X D_Y \phi]$ by $\mathcal{E}_2[\phi]$. This then implies the decay of the second-order derivative of the scalar field.

**Proposition 52.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then for all $X, Y \in \Gamma$, we have the bound*

$$\mathcal{E}_0[D_X D_Y \phi] \lesssim_{M_2} \mathcal{E}_2[\phi]. \tag{87}$$

*Proof.* From the argument at the beginning of this section (before Lemma 49), we derive that

$$\mathcal{E}_0[D_X D_Y \phi] \lesssim_{M_2} \mathcal{E}_2[\phi] + I_{1+\epsilon}^{1+\gamma_0}[[D_X, [\Box_A, D_Y]]\phi](\{t \geq 0\}) + I_{1+\gamma_0}^{1+\epsilon}[[D_X, [\Box_A, D_Y]]\phi](\{t \geq 0\}).$$

Then by Lemma 49 and Proposition 50, for all $0 < \epsilon_1 < 1$ and $X, Y \in \Gamma$, we conclude that

$$\begin{aligned}\mathcal{E}_0[D_X D_Y \phi] \lesssim_{M_2} {} & \epsilon_1 I_{1+\gamma_0-\epsilon}^{-1-\epsilon}[D\phi_2](\{t \geq 0\}) + \epsilon_1 I_0^{\gamma_0}[\not{D}\psi_2](\{t \geq 0\} \cap \{r \geq R\}) \\ & + \mathcal{E}_1[\phi]\epsilon_1^{-1} + \epsilon_1 \int_{\mathbb{R}} \tau_+^{1+\gamma_0} g(\tau) E[\phi_2](\Sigma_\tau)\, d\tau + \epsilon_1 \int_{\mathbb{R}} \int_{H_{\tau^*}} \tau_+^{2+\gamma_0+\epsilon-p} g(\tau) r^p |D_L \psi_2|^2\, dv\, d\omega\, d\tau,\end{aligned}$$

where $\phi_2 = D_X D_Y \phi$ and $\psi_2 = D_X D_Y (r\phi)$. The proposition then follows by the same argument as Proposition 47. □

**4.4.** ***Pointwise bound for the scalar field.*** Once we have the bound (87), from Proposition 32 and Corollary 24, we obtain the energy flux decay estimates as well as the $r$-weighted energy estimates for the second-order derivatives of the scalar field. In other words, simply assuming $M_2$ is finite (see the definition of $M_2$ in (35)) and the charge $q_0$ is small, we then can derive the energy decay estimates for the second-order derivatives of the scalar field. For the MKG equations, $J = \delta F = J[\phi]$ is quadratic in $\phi$. To construct global solutions, we need to bound these nonlinear terms. In this section, we show the pointwise bound for the scalar field with the assumption that $M_2$ is finite.

We start with an analogue of Proposition 14 regarding the pointwise bound of the scalar field in the finite region $\{r \leq R\}$. Similarly to the pointwise bound of the Maxwell field, we use elliptic estimates. However as the connection field $A$ is general, we are not able to apply the elliptic estimates for the flat case directly. We therefore establish an elliptic lemma for the operator $\Delta_A = \sum_{i=1}^3 D_i D_i$ first. Let $B_{R_1}$ be the ball with radius $R_1$ in $\mathbb{R}^3$. Define

$$\|\phi\|_{H^k(B_{R_1})} = \sum_{1 \leq j_l \leq 3} \|D_{j_1} D_{j_2} \cdots D_{j_k} \phi\|_{L^2(B_{R_1})} + \|\phi\|_{H^{k-1}(B_{R_1})}, \quad k \geq 1.$$

Then we have the following lemma.

**Lemma 53.** *We have the elliptic estimates*

$$\|\phi\|_{H^2(B_{R_1})} \lesssim_{M_2, R_1, R_2} \|\Delta_A \phi\|_{L^2(B_{R_2})} + \big(1 + \|F\|_{L^\infty(B_{R_2})} + \|J\|_{H^1(B_{R_2})}\big)\|\phi\|_{H^1(B_{R_2})} \tag{88}$$

*for all $R_1 < R_2$. Here the constant $M_2$ is defined in line (35) and $J = \delta(dA)$ or $J_j = \partial^i (dA)_{ij}$.*

*Proof.* The proof is similar to the case when the connection field $A$ is trivial. For the case when the scalar field $\phi$ is compactly supported in some ball $B_{R_1}$, using integration by parts we can show that

$$\begin{aligned}\int_{B_{R_1}} D_i D_j\phi \cdot \overline{D_i D_j\phi}\, dx &= -\int_{B_{R_1}} D_i D_i D_j\phi \cdot \overline{D_j\phi}\, dx \\ &= -\int_{B_{R_1}} D_j D_i D_i\phi \cdot \overline{D_j\phi}\, dx - \int_{B_{R_1}} [D_i D_i, D_j]\phi \cdot \overline{D_j\phi}\, dx \\ &= \int_{B_{R_1}} |\Delta_A\phi|^2\, dx - \int_{B_{R_1}} \sqrt{-1}\,(2F_{ij} D_i\phi + \partial_i F_{ij}\phi) \cdot \overline{D_j\phi}\, dx.\end{aligned}$$

Estimate (88) then follows.

For a general complex function $\phi$, we can choose a real cut-off function $\chi$ which is supported on the ball $B_{R_2}$ and equal to 1 on the smaller ball $B_{R_1}$. By direct computation, we can show that

$$\begin{aligned}\|\Delta_A(\chi\phi)\|_{L^2(B_{R_2})} &= \|\chi\Delta_A\phi + 2\partial_i\chi \cdot D_i\phi + \Delta\chi \cdot \phi\|_{L^2(B_{R_2})} \\ &\lesssim \|\Delta_A\phi\|_{L^2(B_{R_2})} + \|\phi\|_{H^1(B_{R_2})}.\end{aligned}$$

The lemma then follows from the above argument for the compactly supported case. □

We assume $\Box_A\phi$ verifies the extra bound

$$\int_{\tau_1}^{\tau_2}\int_{r\le 2R} |D\Box_A\phi|^2 + |D_Z D\Box_A\phi|^2\, dx\, d\tau \le C\mathcal{E}_2[\phi](\tau_1)_+^{-1-\gamma_0}, \quad 0 \le \tau_1 < \tau_2 \tag{89}$$

for some constant $C$ depending only on $R$. For solutions of (MKG), one has $\Box_A\phi = 0$ and the above bound trivially holds. The above elliptic estimate adapted to the connection field $A$ implies the following pointwise bound for the scalar field $\phi$ on the compact region $\{r \le R\}$.

**Proposition 54.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds and the inhomogeneous term $\Box_A\phi$ verifies the bound (89). Then for all $0 \le \tau$ and $0 \le \tau_1 < \tau_2$, we have*

$$\int_{\tau_1}^{\tau_2} \sup_{|x|\le R} (|D\phi|^2 + |\phi|^2)(\tau, x)\, d\tau \lesssim \int_{\tau_1}^{\tau_2}\int_{r\le R} |D^2 D\phi|^2 + |\phi|^2\, dx\, dt \lesssim_{M_2} \mathcal{E}_2[\phi](\tau_1)_+^{-1-\gamma_0}, \tag{90}$$

$$|D\phi|^2(\tau, x) + |\phi|^2(\tau, x) \lesssim_{M_2} \mathcal{E}_2[\phi]\tau_+^{-1-\gamma_0}, \quad \forall |x| \le R. \tag{91}$$

*Proof.* At the fixed time $\tau \ge 0$, consider the elliptic equation for the scalar field $\phi_k = D_Z^k\phi$:

$$\Delta_A\phi_k = D_t D_t\phi_k + D_k\Box_A\phi + [\Box_A, D_Z^k]\phi.$$

Proposition 14 together with Proposition 17 indicate that the Maxwell field $F$ is bounded. The definition of $M_2$ shows that

$$\|J\|_{H^1(B_{2R})}^2 \lesssim \int_\tau^{\tau+1} |\overline{\nabla} J|^2 + |\partial_t\overline{\nabla} J|^2 + |J|^2 + |\partial_t J|^2\, dx\, dt \lesssim M_2.$$

Here $B_{R_1}$ denotes the ball with radius $R_1$ at time $\tau$. Then by the previous Lemma 53, we conclude that

$$\|\phi_k\|^2_{H^2(B_{3R/2})} \lesssim_{M_2} \|D_t D_t \phi_k\|^2_{L^2(B_{2R})} + \|D_Z^k \Box_A \phi\|^2_{L^2(B_{2R})} + \|[\Box_A, D_Z^k]\phi\|^2_{L^2(B_{2R})} + \|\phi_k\|^2_{H^1(B_{2R})}.$$

This gives the $H^2$ estimates for $D_t\phi$ and $\phi$. To obtain estimates for $D_j\phi$, commute the equation with $D_j$:

$$\Delta_A D_j \phi = D_j D_t D_t \phi + D_j \Box_A \phi + [\Delta_A, D_j]\phi = D_j D_t D_t \phi + D_j \Box_A \phi + \sqrt{-1}(2F_{ij} D_i \phi + \partial_i F_{ij} \phi).$$

Then using Lemma 53 again, we obtain

$$\begin{aligned}\|D_j\phi\|^2_{H^2(B_R)} &\lesssim_{M_2} \|D_j\phi\|^2_{H^1(B_{3R/2})} + \|\Delta_A D_j \phi\|^2_{L^2(B_{3R/2})} \\ &\lesssim_{M_2} \|\phi\|^2_{H^2(B_{3R/2})} + \|D_j D_t^2 \phi\|^2_{L^2(B_{3R/2})} + \|D_j \Box_A \phi\|^2_{L^2(B_{3R/2})}.\end{aligned}$$

Here we have used the facts $|F|^2 \lesssim M_2$ and $\|J\|^2_{H^1(B_{2R})} \lesssim M_2$. Then for the pointwise bound (91), we need to show the energy flux decay through $B_{2R}$ at time $\tau$. This can be fulfilled by considering the energy estimate obtained by using the vector field $\partial_t$ as multiplier on the region bounded by $\{t = \tau\}$ and $\Sigma_{\tau - R}$ (recall that $\Sigma_\tau = H_{\tau^*}$ for negative $\tau < 0$). Corollary 27 together with Propositions 32 and 52 then imply that

$$E[D_Z^k\phi](B_{2R}) \lesssim E[D_Z^k\phi](\Sigma_{\tau-R}) + (\tau - R)_+^{-1-\gamma_0} \mathcal{E}_0[D_Z^k\phi] \lesssim_{M_2} \mathcal{E}_k[\phi]\tau_+^{-1-\gamma_0}, \quad k \le 2.$$

For the flux of the inhomogeneous term $D\Box_A\phi$ and the commutator term $[D_Z, \Box_A]\phi$, we can make use of the integrated local energy estimates. More precisely, combine the above $H^2$ estimates for $\phi_k = D_Z^k\phi$, $k = 0, 1$, and $D_j\phi$. We can show that

$$\begin{aligned}&\|D_j\phi\|^2_{H^2(B_R)} + \sum_{k\le 1} \|\phi_k\|^2_{H^2(B_R)} \\ &\quad\lesssim_{M_2} \sum_{l\le 2} E[D_Z^l\phi](B_{2R}) + \|D\Box_A\phi\|^2_{L^2(B_{2R})} + \|[\Box_A, D_Z]\phi\|^2_{L^2(B_{2R})} \\ &\quad\lesssim_{M_2} \mathcal{E}_2[\phi]\tau_+^{-1-\gamma_0} + \int_\tau^{\tau+1}\int_{r\le 2R} |D\Box_A\phi|^2 + |D_t D\Box_A\phi|^2\, dx\, d\tau + I_0^0[D_Z[\Box_A, D_Z]\phi](D^\tau_{\tau-R}) \\ &\quad\lesssim_{M_2} \mathcal{E}_2[\phi]\tau_+^{-1-\gamma_0}.\end{aligned}$$

Here we have used the bound

$$I^{1+\epsilon}_{1+\gamma_0}[D_Z^k[\Box_A, D_Z]\phi](\{t \ge 0\}) \lesssim_{M_2} \mathcal{E}_{k+1}[\phi], \quad k = 0, 1,$$

which is a consequence of the proof in the previous section (see the argument in the beginning of Section 4.3.4). Then Sobolev embedding implies the pointwise bound (91) for $\phi$.

For the integrated decay estimate (90), we integrate the $H^2$ norm of $D_j\phi$ from time $\tau_1$ to $\tau_2$:

$$\begin{aligned}\int_{\tau_1}^{\tau_2} \|D\phi\|^2_{H^2(B_R)}\,d\tau &\lesssim_{M_2} \int_{\tau_1}^{\tau_2} \sum_{l\le 2} \|D_Z^l\phi\|^2_{L^2(B_{2R})} + \|D\Box_A\phi\|^2_{L^2(B_{2R})} + \|[\Box_A, D_Z]\phi\|^2_{L^2(B_{2R})}\,d\tau\\ &\lesssim_{M_2} \sum_{l\le 2} I_0^{-1-\epsilon}[D_Z^l\phi](D^{\tau_2}_{\tau_1-R}) + \int_{\tau_1}^{\tau_2}\int_{r\le 2R} |D\Box_A\phi|^2\,dx\,d\tau + I_0^0[[D_Z,\Box_A]\phi](D^{\tau_2}_{\tau_1-R})\\ &\lesssim_{M_2} \mathcal{E}_2[\phi](\tau_1)_+^{-1-\gamma_0}.\end{aligned}$$

Here we have used the integrated local energy estimates for the second-order derivative of the scalar field. Then Sobolev embedding implies the integrated decay estimate (90). □

**Remark 55.** For the Sobolev embedding adapted to the connection $A$, it suffices to establish the $L^p$ embedding in terms of the $H^1$ norm. As the norm is gauge invariant, we can choose a particular gauge so that the function is real. For a real function $f$ we have the trivial bound $\|D_A f\|_{L^2} \ge \|\partial f\|_{L^2}$. This explains the Sobolev embedding we have used in this paper adapted to the general connection field $A$.

Next we consider the pointwise bound for the scalar field outside the cylinder $\{r \le R\}$. The decay estimate for $\phi$ easily follows from Lemma 19, as we have energy decay estimates for second-order derivatives of $\phi$. However, this does not apply to the derivative of $\phi$ due to the limited regularity (only two derivatives). Like with the Maxwell field in Proposition 17, we rely on Lemma 16.

**Proposition 56.** *Assume that the charge $q_0$ is sufficiently small, so that Corollary 22 holds. Then we have the pointwise bound*

$$\|D_{\underline{L}}(rD_Z^k\phi)\|_{L^2_u L^\infty_v L^2_\omega(\overline{\mathcal{D}}_\tau)} \lesssim_{M_2} \mathcal{E}_{k+1}[\phi](\tau_1)_+^{-1-\gamma_0+3\epsilon}, \qquad k=0,1, \tag{92}$$

$$\|r^{p/2}D_L(rD_Z^k\phi)\|^2_{L^2_v L^\infty_u L^2_\omega(\overline{\mathcal{D}}_\tau)} \lesssim_{M_2} \mathcal{E}_{k+1}[\phi](\tau_1)_+^{p+4\epsilon-1-\gamma_0}, \quad 0\le p\le 1+\gamma_0-4\epsilon,\ k=0,1, \tag{93}$$

$$r^p(|D_L(r\phi)|^2 + |\not{D}(r\phi)|^2)(\tau,v,\omega) \lesssim_{M_2} \mathcal{E}_2[\phi]\tau_+^{p-1-\gamma_0}, \qquad 0\le p\le 1+\gamma_0, \tag{94}$$

$$|D_{\underline{L}}(r\phi)|^2(\tau,v,\omega) \lesssim_{M_2} \mathcal{E}_2[\phi]\tau_+^{-1-\gamma_0}, \tag{95}$$

$$r^p|\phi|^2(\tau,v,\omega) \lesssim_{M_2} \mathcal{E}_2[\phi]\tau_+^{p-2-\gamma_0}, \qquad 1\le p\le 2. \tag{96}$$

**Remark 57.** If we have one more derivative (assume $M_3$), then we have a better estimate for $\not{D}(r\phi)$, as we can write it as $D_Z\phi$.

*Proof.* Estimate (92) follows from (68) and (71) together with the $r$-weighted energy and integrated local energy estimates for the scalar field $D_Z^k\phi$, $k\le 2$. Estimate (93) is a consequence of (69) and (71).

For the pointwise bound for the scalar field, let $\phi_k = D_Z^k\phi$, $\psi_k = D_Z^k(r\phi)$, $k\le 2$. First, the $r$-weighted energy estimates (53) and (63) imply that

$$\int_{H_{\tau^*}} r^p|D_L\psi_k|^2\,dv\,d\omega \lesssim_{M_2} \mathcal{E}_k[\phi]\tau_+^{p-1-\gamma_0}, \quad k\le 2, \quad 0\le p\le 1+\gamma_0.$$

From the $r$-weighted energy estimate for $F$ and Lemma 19, we can bound the commutator:

$$\begin{aligned}\int_{H_{\tau^*}} r^p |[D_Z^2, D_L]\psi|^2 \, dv \, d\omega &\lesssim \int_{H_{\tau^*}} r^p (|F_{ZL} D_Z \psi|^2 + |\mathcal{L}_Z F_{ZL} \psi|^2) \, dv \, d\omega \\ &\lesssim \sum_{l \le 1} \int_{H_{\tau^*}} r^p (|\mathcal{L}_Z^l \rho \psi_{1-l}|^2 + |r \mathcal{L}_Z^l \alpha| |\psi_{l-1}|) \, dv \, d\omega \\ &\lesssim_{M_2} \mathcal{E}_2[\phi] |q_0| \tau_+^{p-3-\gamma_0} + \mathcal{E}_2[\phi] \tau_+^{p-1-2\gamma_0} \\ &\lesssim_{M_2} \mathcal{E}_2[\phi] \tau_+^{p-1-\gamma_0}.\end{aligned}$$

Here the charge part only appears when $\tau < 0$. The previous two estimates lead to

$$\int_{H_{\tau^*}} r^p |D_Z^k D_L D_Z^l \psi|^2 \, dv \, d\omega \lesssim_{M_2} \mathcal{E}_2[\phi] \tau_+^{p-1-\gamma_0}, \quad k + l \le 2, \quad 0 \le p \le 1 + \gamma_0.$$

To apply Lemma 16, we need the energy flux for $D_L D_L \psi$. From the null equation (70) for the scalar field, on the outgoing null hypersurface $H_{\tau^*}$, for $k = 0, 1$, we can show that

$$\begin{aligned}\int_{H_{\tau^*}} r^p |D_{\underline{L}} D_L \psi_k|^2 \, dv \, d\omega &\lesssim \int_{H_{\tau^*}} r^p (|\rho \cdot r \phi_k|^2 + |r^{-1} \not{D} D_\Omega \psi_k|^2 + |r \Box_A \phi_k|^2) \, dv \, d\omega \\ &\lesssim_{M_2} \mathcal{E}_{k+1}[\phi] \tau_+^{p-1-\gamma_0}.\end{aligned}$$

Here the first term $\rho \cdot r \psi_k$ has been bounded in the above commutator estimate for $[D_Z^2, D_L]\psi$. The second term $|r^{-1} \not{D} D_\Omega \psi_k|^2$ can be bounded by the energy flux of $\mathcal{L}_Z^2 F$ through $H_{\tau^*}$ as $p \le 2$. The bound for $\Box_A \phi_k$ follows from the argument in Section 4.3.4 where we have shown that $\mathcal{E}_1[\phi_k] \lesssim_{M_2} \mathcal{E}_2[\phi_{k-1}]$ for $k = 0, 1$. Now commute $D_L$ with $\psi_k = D_Z^k \psi$. First, we can show that

$$\begin{aligned}|D_L[D_L, D_Z]\psi| &\lesssim |L F_{LZ}| |\psi| + |F_{ZL}| |D_L \psi| \\ &\lesssim (|\underline{L}(r\alpha)| + |r \mathcal{L}_Z \alpha| + |L\rho|) |\psi| + (|\rho| + |r\alpha|) |D_L \psi|.\end{aligned}$$

On the right-hand side, the second term is easy to bound as we can control the Maxwell field $\rho, r\alpha$ by the $L^\infty$ norm shown in Proposition 17 and the scalar field $\psi$ by the $r$-weighted energy estimates. For the first term, we have to use the null structure equations of Lemma 5 to control $L(r\alpha)$, $L\rho$. Indeed, we can show that

$$\begin{aligned}\int_{H_{\tau^*}} r^p |D_L[D_L, D_Z]\psi|^2 \, dv \, d\omega &\lesssim_{M_2} \int_{H_{\tau^*}} r^p |D_Z^2 D_L \psi|^2 + r^p (|\underline{L}(r\alpha)|^2 + |r \mathcal{L}_Z \alpha|^2 + |L\rho|^2) \mathcal{E}_2[\phi] \, dv \, d\omega \\ &\lesssim_{M_2} \mathcal{E}_2[\phi] \left( \tau_+^{p-1-\gamma_0} + \int_{H_{\tau^*}} r^p (|\mathcal{L}_\Omega(\rho, \sigma, \alpha)|^2 + |rJ|^2 + |\rho|^2) \, dv \, d\omega \right) \\ &\lesssim_{M_2} \mathcal{E}_2[\phi] \tau_+^{p-1-\gamma_0}.\end{aligned}$$

Here we can bound $\rho, \alpha, \sigma$ by the energy flux as $p \le 2$. For the inhomogeneous term $J$ we can use one more derivative $\mathcal{L}_{\partial_t}$. In particular, we can show that

$$\begin{aligned}\sum_{k\le 1}\int_{H_{\tau^*}} r^p(|D_L D_Z^k D_L\psi|^2 + |D_\Omega D_Z^k D_L\psi|^2 + |D_Z^k D_L\psi|^2)\,dv\,d\omega \\ \lesssim \sum_{l\le 2}\int_{H_{\tau^*}} r^p(|D_Z^l D_L\psi|^2 + |D_L[D_Z, D_L]\psi|^2 + |D_{\underline{L}} D_L D_Z\psi|^2 + |D_{\partial_t} D_L D_Z\psi|^2)\,dv\,d\omega \\ \lesssim_{M_2} \mathcal{E}_{k+1}[\phi]\tau_+^{p-1-\gamma_0}.\end{aligned}$$

Then using Lemma 16 and Sobolev embedding, we derive the pointwise estimate for $D_L\psi$ (see Remark 55 for the Sobolev embedding adapted to the connection $A$). This proves the first part of (94).

For $D_{\underline{L}}\psi$ and $\not{D}(r\phi)$, we make use of the energy flux through the incoming null hypersurface $\underline{H}_\tau$, which is defined as $\underline{H}_v^{-v,\tau^*}$ when $\tau < 0$ or $\underline{H}_v^{\tau^*, v}$ when $\tau \ge 0$. From the energy estimates (53), (56), (63) and (64), we obtain the energy flux decay

$$\int_{\underline{H}_\tau} |D_{\underline{L}} D_Z^k\psi|^2 + |\not{D} D_Z^k\psi|^2 + \tau_+^{-p} r^p |D_\Omega D_Z^k\phi|^2 + r^2|D_{\underline{L}} D_Z^k\phi|^2\,du\,d\omega \lesssim_{M_2} \mathcal{E}_2[\phi]\tau_+^{-1-\gamma_0}$$

for $k \le 2$ and $0 \le p \le 1+\gamma_0$. As $\not{D}(r\phi) = D_\Omega\phi$, the above estimates together with Lemma 16 indicate that

$$r^p|D_\Omega\phi|^2 \lesssim_{M_2} \mathcal{E}_2[\phi]\tau_+^{p-1-\gamma_0}, \quad 0 \le p \le 1+\gamma_0.$$

Thus the second part of (94) holds.

For $D_{\underline{L}}\psi$, we need to pass the $D_{\underline{L}}$ derivative to $\psi$. We can compute the commutator:

$$|[D_Z^2, D_{\underline{L}}]\psi| \lesssim (|r\mathcal{L}_Z\underline{\alpha}| + |\mathcal{L}_Z\rho|)|\psi| + (|r\underline{\alpha}| + |\rho|)|D_Z\psi|.$$

We can bound $\psi$ using Lemma 19 and $\rho, \underline{\alpha}$ using the energy flux through $\underline{H}_\tau$. Then the previous energy estimate implies that

$$\int_{\underline{H}_\tau} |D_Z^k D_{\underline{L}} D_Z^l\psi|^2 + |r^{-1} D_Z^{k+1}\psi|^2\,du\,d\omega \lesssim_{M_2} \mathcal{E}_2[\phi]\tau_+^{-1-\gamma_0}, \quad k+l \le 2. \tag{97}$$

To apply Lemma 16, we also need an estimate for $D_{\underline{L}} D_{\underline{L}}\psi$. We use the null equation (70) to show that

$$\int_{\underline{H}_\tau} |D_L D_{\underline{L}}\psi_k|^2\,du\,d\omega \lesssim_{M_2} \mathcal{E}_2[\phi]\tau_+^{-1-\gamma_0}, \quad k \le 1.$$

The proof of this estimate is similar to that through the outgoing null hypersurface we have done above. To pass the $D_{\underline{L}}$ derivative to $\psi$, we commute $D_{\underline{L}}$ with $\psi_1 = D_Z\psi$:

$$|D_{\underline{L}}[D_{\underline{L}}, D_Z]\psi| \lesssim |D_{\underline{L}}\psi|(|r\underline{\alpha}| + |\rho|) + |\psi|(|\underline{L}\rho| + |L(r\underline{\alpha})| + |\partial_t(r\underline{\alpha})|).$$

Again we can bound $D_{\underline{L}}\psi$ using the energy flux and $r\underline{\alpha}$, $\rho$ by the $L^\infty$ norm. For the second term, $\psi$ can be bounded using Lemma 19, and the curvature components $\underline{L}\rho$, $L(r\underline{\alpha})$ are controlled by using the null

structure equations (7) and (8). More precisely, we can show that

$$\sum_{k\le 1}\int_{\underline{H}_\tau}|D_{\underline{L}}D_Z^k D_{\underline{L}}\psi|^2\,du\,d\omega\lesssim\int_{\underline{H}_\tau}|D_{\underline{L}}[D_Z,D_{\underline{L}}]\psi|^2+|D_L D_{\underline{L}}D_Z^k\psi|^2+|D_{\partial_t}D_{\underline{L}}D_Z^k\psi|^2\,du\,d\omega$$

$$\lesssim_{M_2}\mathcal{E}_2[\phi]\tau_+^{-1-\gamma_0}.$$

This estimate and (97) combined with Lemma 16 imply the pointwise bound (95) for $D_{\underline{L}}\psi$.

The pointwise bound (96) for $\phi$ follows from Lemma 19:

$$\int_\omega r^p|D_Z^k\phi|^2(\tau,v,\omega)\,d\omega\lesssim_{M_2}\mathcal{E}_k[\phi]\tau_+^{p-2-\gamma_0},\quad k\le 2,\quad 1\le p\le 2,$$

together with Sobolev embedding on the sphere. □

## 5. Bootstrap argument

We use a bootstrap argument to prove the main theorem. In the exterior region, we decompose the full Maxwell field $F$ into the chargeless part and the charge part:

$$F=\bar{F}+q_0\chi_{\{r\ge t+R\}}r^{-2}\,dt\wedge dr.$$

We make the bootstrap assumption

$$m_2\le 2\mathcal{E}\tag{98}$$

on the nonlinearity $J_\mu=\nabla^\nu F_{\nu\mu}=\Im(\phi\cdot\overline{D_\mu\phi})=J_\mu[\phi]$. Here recall the definition of $m_2$ in (35). Since the nonlinearity $J$ is quadratic in $\phi$, $m_2$ has size $\mathcal{E}^2$. By assuming that $\mathcal{E}$ is sufficiently small, we then can improve the above bootstrap assumption and hence conclude our main theorem. The smallness of $\mathcal{E}$ depends on $\mathcal{M}$. Without loss of generality, we assume $\mathcal{E}\le 1$ and $\mathcal{M}>1$.

In the definition (35) for $M_2$, the main contribution is $E_0^2[\bar{F}]$ with $\bar{F}$ the chargeless part of the Maxwell field on the initial hypersurface $\{t=0\}$. As the scalar field $\phi$ solves the linear equation $\Box_A\phi=0$, we derive from the definition (47) for $\mathcal{E}_2[\phi]$ that $\mathcal{E}_2[\phi]=E_0^2[\phi]$. The definition for $E_0^k[\bar{F}]$ and $E_0^k[\phi]$ has been given in (6). To proceed, we need to bound $E_0^2[\bar{F}]$ and $\mathcal{E}_2[\phi]$ in terms of $\mathcal{M}$ and $\mathcal{E}$, which is shown in the following lemma.

**Lemma 58.** *Assume that the initial data set $(E,H,\phi_0,\phi_1)$ satisfies the compatibility condition (2) and that the norms $\mathcal{M},\mathcal{E}$ defined before Theorem 1 are finite. Then we can bound $E_0^2[\bar{F}]$ and $E_0^2[\phi]$ as follows*:

$$E_0^2[\bar{F}]\lesssim\mathcal{M},\quad E_0^2[\phi]\lesssim_{\mathcal{M}}\mathcal{E}.$$

*Proof.* To define the norm $E_0^k[\phi]$, we need to know the connection field $A$ on the initial hypersurface $\{t=0\}$. As the norm $E_0^k[\phi]$ is gauge invariant, we may choose a particular gauge. Let $\bar{A}=(A_1,A_2,A_3)(0,x)$, $A_0=A_0(0,x)$. We want to choose a particular connection field $(A_0,\bar{A})$ on the initial hypersurface to define the gauge invariant norm $E_0^k[\phi]$.

It is convenient to choose the Coulomb gauge to make use of the divergence-free part $E^{\mathrm{df}}$ and the curl-free part $E^{\mathrm{cf}}$ of $E$. More precisely, on the initial hypersurface $\{t=0\}$, we choose $(A_0,\bar{A})$ so that

$\operatorname{div}(\overline{A}) = 0$. Then the compatibility condition (2) is equivalent to

$$\Delta A_0 = -\Im(\phi_0 \cdot \overline{\phi_1}) = -J_0(0), \quad \overline{\nabla} \times \overline{A} = H.$$

Define the weighed Sobolev space

$$W^p_{s,\delta} := \left\{ f \;\Big|\; \sum_{|\beta| \le s} \left\| (1+|x|)^{\delta+|\beta|} |\partial^\beta f| \right\|_{L^p} < \infty \right\}.$$

For the special case $p = 2$, let $H_{s,\delta} = W^2_{s,\delta}$. Denote $\tilde{\Gamma} = \{\Omega, \partial_j\}$, $\delta = \frac{1}{2}(1+\gamma_0)$. By the definition of $\mathcal{M}$,

$$\|\mathcal{L}^k_{\tilde{Z}} H\|_{H_{0,\delta}} \lesssim \mathcal{M}^{1/2}, \quad k \le 2, \quad \tilde{Z} \in \tilde{\Gamma}.$$

Then from Theorem 0 of [McOwen 1979] or Theorem 5.1 of [Choquet-Bruhat and Christodoulou 1981a], we conclude that

$$\|\mathcal{L}^k_{\tilde{Z}} \overline{A}\|_{H_{1,\delta-1}} \lesssim \mathcal{M}^{1/2}, \quad k \le 2, \quad \tilde{Z} \in \tilde{\Gamma}.$$

This is the desired estimate for the gauge field $\overline{A}$. With this connection field $\overline{A}$, we then can define the covariant derivative $\overline{D} = \overline{\nabla} + \sqrt{-1}\,\overline{A}$ in the spatial direction. Therefore,

$$\begin{aligned} \|D\phi(0,\cdot)\|_{H_{0,\delta}} = \|\overline{D}\phi_0\|_{H_{0,\delta}} + \|\phi_1\|_{H_{0,\delta}} &\lesssim \mathcal{E}^{1/2} + \|\overline{A}\|_{W^3_{0,\delta}} \|\phi_0\|_{W^6_{0,\delta}} \\ &\lesssim \mathcal{E}^{1/2}(1 + \mathcal{M}^{1/2}) \lesssim \mathcal{E}^{1/2}\mathcal{M}^{1/2}. \end{aligned}$$

By the same argument, and commuting the equations with $D_{\tilde{Z}}$, we obtain the same estimates for $D_{\tilde{Z}}\phi$:

$$\|DD^k_{\tilde{Z}}\phi(0,\cdot)\|_{H_{0,\delta}} \lesssim \mathcal{E}^{1/2}\mathcal{M}^{1/2}, \quad k \le 2.$$

To define the covariant derivative $D_0$, we need estimates for $A_0$. The difficulty is the nonzero charge. Take a cut-off function $\chi(x) = \chi(|x|)$ such that $\chi = 1$ when $|x| \ge R$ and vanishes for $|x| \le \frac{1}{2}R$. Denote the chargeless part of $A_0$ and $J_0$ as follows:

$$\bar{A}_0 = A_0 + \chi q_0 r^{-1}, \quad \bar{J}_0(0) := J_0 - \Delta(\chi q_0 r^{-1}).$$

By the definition of the charge $q_0$, we then have

$$\Delta \bar{A}_0 = -\bar{J}_0(0), \quad \int_{\mathbb{R}^3} \bar{J}_0(0)\, dx = 0.$$

Recall that $J_0(0) = \Im(\phi_0 \cdot \overline{\phi_1})$. Using Sobolev embedding, we can bound

$$\|\bar{J}_0(0)\|_{W^{3/2}_{0,2\delta}} \lesssim |q_0| + \|\phi_1\|_{W^2_{0,\delta}} \|\phi_0\|_{W^6_{0,\delta}} \lesssim |q_0| + \|\phi_1\|_{W^2_{0,\delta}} \|\phi_0\|_{W^2_{1,\delta}} \lesssim \mathcal{E}.$$

Then from Theorem 0 of [McOwen 1979] again, we conclude that

$$\|\bar{A}_0\|_{W^{3/2}_{2,2\delta-2}} \lesssim \mathcal{E}.$$

Here the condition that $\bar{A}_0$ is chargeless guarantees $\bar{A}_0$ to belong to the above weighted Sobolev space. Then using the Gagliardo–Nirenberg interpolation inequality, we derive that

$$\|\overline{\nabla}\bar{A}_0\|_{H_{0,2\delta-1/2}} \lesssim \|\overline{\nabla}\bar{A}_0\|^{1/2}_{W^{3/2}_{0,2\delta-1}} \cdot \|\overline{\nabla}\overline{\nabla}\bar{A}_0\|^{1/2}_{W^{3/2}_{0,2\delta}} \lesssim \mathcal{E}.$$

By definition, one has $E = \partial_t \bar{A} - \overline{\nabla} A_0$. By our gauge choice, $\partial_t \bar{A}$ is divergence-free and $\overline{\nabla} A_0$ is curl-free. In particular, we derive that $E^{\mathrm{df}} = \partial_t \bar{A}$ and $E^{\mathrm{cf}} = -\overline{\nabla} A_0$. Take the chargeless part. We obtain that $\bar{E}^{\mathrm{cf}} = \overline{\nabla}\bar{A}_0$ when $|x| \geq R$. Therefore, we can bound the weighted Sobolev norm of the chargeless part of the Maxwell field $\bar{F}$ on the initial hypersurface as follows:

$$\begin{aligned}\|\bar{F}\|_{H_{0,\delta}} &\leq \|F\chi_{\{|x|\leq R\}}\|_{H_{0,\delta}} + \|(\bar{E}, H)\chi_{\{|x|\geq R\}}\|_{H_{0,\delta}} \\ &\lesssim \|F\chi_{\{|x|\leq R\}}\|_{H_{0,\delta}} + \|(E^{\mathrm{df}}, H)\chi_{\{|x|\geq R\}}\|_{H_{0,\delta}} + \|\bar{E}^{\mathrm{cf}}\chi_{\{|x|\geq R\}}\|_{H_{0,\delta}} \\ &\lesssim \mathcal{M}^{1/2} + \|\overline{\nabla}\bar{A}_0\|_{H_{0,\delta}} \lesssim \mathcal{M}^{1/2}.\end{aligned}$$

Similarly, we have the same estimates for $\mathcal{L}^k_{\tilde{Z}}\bar{F}$, $k \leq 2$, that is,

$$\|\mathcal{L}^k_{\tilde{Z}}\bar{F}\|_{H_{0,\delta}} \lesssim \mathcal{M}^{1/2}, \quad k \leq 2.$$

To derive estimates for $D^k_Z\phi$ and $\mathcal{L}^k_Z\bar{F}$ on the initial hypersurface, we use the equations

$$\partial_t E - \overline{\nabla} \times H = \Im(\phi \cdot \tilde{D}\phi), \quad \partial_t H + \overline{\nabla} \times E = 0, \quad D_t\phi_1 = \bar{D}\bar{D}\phi$$

to replace the time derivatives with the spatial derivatives. The inhomogeneous term $\Im(\phi \cdot \bar{D}\phi)$ or the commutators $[D_t, \bar{D}]$ could be controlled using Sobolev embedding together with Hölder's inequality. The lemma then follows. □

The above lemma then leads to the following corollary:

**Corollary 59.** *Let $(\phi, A)$ be the solution of (MKG). Under the bootstrap assumption (98), we have*

$$M_2 \lesssim \mathcal{M}, \quad \mathcal{E}_2[\phi] \lesssim_{\mathcal{M}} \mathcal{E}.$$

*Proof.* The corollary follows from the definition of $M_2$ and $\mathcal{E}_2[\phi]$ in (35) and (47) together Lemma 58. □

From now on, we allow the implicit constant in $\lesssim$ to also depend on $\mathcal{M}$, that is, $B \lesssim K$ means that $B \leq CK$ for some constant $C$ depending on $\gamma_0$, $R$, $\epsilon$ and $\mathcal{M}$. The rest of this section is devoted to improving the bootstrap assumption.

To improve the bootstrap assumption, we need to estimate $m_2$ defined in (35). On the finite region $\{r \leq R\}$, the null structure of $J[\phi]$ is not necessary as the weights of $r$ are bounded above. When $r \geq R$, the null structure of $J[\phi]$ plays a crucial role. Note that $J_L$ and $\not{J} = (J_{e_1}, J_{e_2})$ are easy to control as they already contain "good" components $\not{D}\phi$ or $D_L(r\phi)$. The difficulty is to exploit the null structure of the component $J_{\underline{L}}$ which is not a standard null form as defined in [Klainerman 1984; 1986]. The null structure of the system is that $J_{\underline{L}}$ does not interact with the "bad" component $\underline{\alpha}$ of the Maxwell field.

For nonnegative integers $k$, write $\phi_k = D^k_Z\phi$, $\psi_k = D^k_Z(r\phi)$, $F_k = \mathcal{L}^k_Z F$ in this section. First we expand the second-order derivative of $J[\phi] = \Im(\phi \cdot \overline{D\phi})$.

**Lemma 60.** *Let $X$ be $L$, $\underline{L}$, $e_1$, $e_2$. Then we have*

$$|\mathcal{L}_Z^2 J| + |\overline{\nabla}\mathcal{L}_Z J| \lesssim |D\phi_1||D\phi| + |\phi_1||D^2\phi| + |\phi||D^2\phi_1| + |\nabla F||\phi|^2 + |F||D\phi||\phi|, \quad |x| \le R;$$

$$r^2|\mathcal{L}_Z^2 J_X| \lesssim \sum_{k\le 2} |\psi_k||D_X\psi_{2-k}| + \sum_{l_1+l_2+l_3\le 1} |\mathcal{L}_Z^{l_1} F_{ZX}||\psi_{l_2}||\psi_{l_3}|, \qquad |x| > R.$$

*Proof.* By the definition of the Lie derivative $\mathcal{L}_Z$, we can compute

$$\begin{aligned}\mathcal{L}_Z J_X = Z(J_X) - J_{\mathcal{L}_Z X} &= \Im(D_Z\phi\cdot\overline{D_X\phi} + \phi\cdot\overline{D_Z D_X\phi} - \phi\cdot\overline{D_{[Z,X]\phi}})\\ &= \Im(\phi_1\cdot\overline{D_X\phi} + \phi\cdot\overline{D_X\phi_1} + \phi\cdot\overline{([D_Z, D_X] - D_{[Z,X]})\phi})\\ &= \Im(\phi_l\cdot\overline{D_X\phi_{1-l}}) - F_{ZX}|\phi|^2.\end{aligned}$$

Here we note that $[D_Z, D_X] - D_{[Z,X]} = \sqrt{-1}F_{ZX}$ for any vector fields $Z$, $X$, and we omitted the summation sign for $l = 0, 1$. Take one more derivative $\overline{\nabla}$ (recall that $\overline{\nabla}$ is the covariant derivative in the spatial direction). The estimate on the region $\{r \le R\}$ then follows.

Similarly, the second-order derivative expands as follows:

$$\begin{aligned}\mathcal{L}_Y\mathcal{L}_Z J_X &= Y\mathcal{L}_Z J_X - \mathcal{L}_Z J_{[Y,X]}\\ &= Y\Im(\phi_l\cdot\overline{D_X}\phi_{1-l}) - Y(F_{ZX}|\phi|^2) - \Im(\phi_l\cdot\overline{D_{[Y,X]}}\phi_{1-l}) + F_{Z[Y,X]}|\phi|^2\\ &= \Im(\phi_k\cdot\overline{D_X}\phi_{2-k}) - (\mathcal{L}_Y F_{ZX} + F_{[Y,Z]X})|\phi|^2 + \Im(\sqrt{-1}\phi_l\cdot\overline{F_{YX}\phi_{1-l}}) - F_{ZX}Y|\phi|^2\end{aligned}$$

for any vector fields $X, Y, Z \in \Gamma$. Here we have omitted the summation sign for $k = 0, 1, 2$ and $l = 0, 1$. Note that

$$\Im(\phi\cdot\overline{D_X\phi}) = r^{-2}\Im(r\phi\cdot\overline{D_X(r\phi)}), \quad [Y, Z] = 0 \text{ or } \in \Gamma.$$

The estimate on the region $\{r \ge R\}$ then follows. Thus the proof of the lemma is finished. □

Next we use the above bound for $J[\phi]$ to improve the bootstrap assumption.

**Proposition 61.** *Assume that the charge $q_0$ is sufficiently small, depending only on $\epsilon$, $R$ and $\gamma_0$, so that Corollary 22 holds. Then we have*

$$m_2 \le C\mathcal{E}^2 \tag{99}$$

*for some constant $C$ depending on $\mathcal{M}$, $\epsilon$, $R$ and $\gamma_0$.*

*Proof.* Since $M_2 \lesssim \mathcal{M}$, all the estimates in the previous section hold. In particular, we have the energy flux and the $r$-weighted energy decay estimates for the scalar field and the chargeless part of the Maxwell field up to second-order derivatives. Moreover, the pointwise estimates in Propositions 14, 17, 54 and 56 hold.

Let's first consider the estimate of $|J_{\underline{L}}|r^{-2}$ in the exterior region. We have the simple bound that $|J_{\underline{L}}| \le |D_{\underline{L}}\phi||\phi|$. We can control $D_{\underline{L}}\phi$ by using the energy flux through the incoming null hypersurface

$\underline{H}_v$ and $\phi$ by the $L^\infty$ norm. In particular, for any $\tau < 0$ we can show that

$$\begin{aligned}\iint_{\mathcal{D}_\tau^{-\infty}} |J_{\underline{L}}| r^{-2}\,dx\,dt &\lesssim \int_{-\tau^*}^{\infty}\left(\int_{\underline{H}_v} |D_{\underline{L}}\phi|^2 r^2\,du\,d\omega\right)^{1/2}\cdot\left(\int_{\underline{H}_v} |\phi|^2 r^{-2}\,du\,d\omega\right)^{1/2} dv\\ &\lesssim \mathcal{E}\int_{-\tau^*}^{\infty} \tau_+^{-(1+\gamma_0)/2}\left(\int_{\underline{H}_v} r^{-4}\tau_+^{-\gamma_0}\,du\,d\omega\right)^{1/2} dv\\ &\lesssim \mathcal{E}\int_{-\tau^*}^{\infty} \tau_+^{-(1+2\gamma_0)/2} r^{-3/2}\,dv \lesssim \mathcal{E}\tau_+^{-1-\gamma_0}.\end{aligned}$$

We remark here that we cannot use the integrated local energy to bound the above term due to the exact total decay rate of $|J_{\underline{L}}| r^{-2}$. As $|q_0| \lesssim \mathcal{E}$, we therefore obtain

$$|q_0| \sup_{\tau\le 0} \tau_+^{1+\gamma_0} \iint_{\mathcal{D}_\tau^{-\infty}} |J_{\underline{L}}| r^{-2}\,dx\,dt \lesssim \mathcal{E}^2, \quad \forall \tau \le 0.$$

Next we consider the estimates on the compact region $\{r \le 2R\}$. As $|\phi_1| = |D_Z\phi| \lesssim |D\phi|$ when $|x| \le R$, we can bound $\phi_1$, $\phi$, $D\phi$ and $F$ by the $L^\infty$ norm obtained in (40) and (91). Then $D^2\phi_1$ and $\nabla F$ can be controlled using the integral decay estimates (39) and (90) on $\{r \le R\}$. To derive estimates for $D^2\phi_k$ or $\nabla F$ on the region $\{R \le r \le 2R\}$, we use (MKG). From Lemma 37 and Lemma 5, we can show that

$$\begin{aligned}|D^2\phi_1| + \mathcal{E}|\nabla F| &\lesssim |D\phi_2| + |D_L D_{\underline{L}}\psi_1| + |F||\phi_1| + \mathcal{E}\big(|\mathcal{L}_Z F| + |L(r^2\rho, r^2\sigma, r\underline{\alpha})| + |L(r\underline{\alpha})|\big)\\ &\lesssim |D\phi_2| + |\Box_A\phi_1| + |F||\phi_1| + \mathcal{E}(|\mathcal{L}_Z F| + |J|).\end{aligned}$$

Here we omitted the easier lower-order terms. On the region $\{R \le r \le 2R\}$, the set $\Gamma$ only misses one derivative, which could be recovered from the equation. From Lemma 60, we can show that

$$\begin{aligned}&I^0_{1+\gamma_0+2\epsilon}[|\mathcal{L}_Z^2 J| + |\overline{\nabla}\mathcal{L}_Z J|](\{r \le 2R\})\\ &\lesssim \mathcal{E}\int_0^\infty \tau_+^{2\epsilon}\int_{r\le 2R} |D^2\phi_1|^2 + \mathcal{E}|\nabla F|^2 + |D\phi|^2\,dx\,d\tau\\ &\lesssim \mathcal{E}^2 + \mathcal{E}\int_0^\infty \tau_+^{2\epsilon}\int_{R\le r\le 2R} |D\phi_2|^2 + |\Box_A\phi_1|^2 + |F|^2|\phi_1|^2 + \mathcal{E}(|\mathcal{L}_Z F|^2 + |J|^2)\,dx\,d\tau\\ &\lesssim \mathcal{E}^2 + \mathcal{E} I^0_{2\epsilon}[\Box_A\phi_1](\{r \ge R\}) + \mathcal{E}^2 I^0_{2\epsilon}[J](\{r \ge R\}) \lesssim \mathcal{E}^2.\end{aligned}$$

Here the implicit constant also depends on $\mathcal{M}$ and we only consider the highest-order terms. The second to last step follows as the integral from time $\tau_1$ to $\tau_2$ decays in $\tau_1$. Hence the spacetime integral is bounded, using Lemma 20. The bound for $\Box_A\phi_1$ follows from Proposition 47 and the spacetime norm for $J$ is controlled by the bootstrap assumption.

Next, we consider the case when $|x| \ge R$, where the null structure of $J$ plays a crucial role. For $|\mathcal{L}_Z^2 J_{\underline{L}}|$, Lemma 60 implies that

$$r^2|\mathcal{L}_Z^2 J_{\underline{L}}| \lesssim |\psi_k||D_{\underline{L}}\psi_{2-k}| + (|r\mathcal{L}_Z^{l_1}\underline{\alpha}| + |\mathcal{L}_Z^{l_1}\rho|)|\psi_{l_2}||\psi_{1-l_1-l_2}|.$$

Here the indices $k, 2-k, 1-l_1-l_2, l_1, l_2$ are nonnegative integers and we only consider the highest-order term as the lower-order terms are easier and could be bounded in a similar way. On the right-hand side of the above inequality, after using Sobolev embedding on the sphere, we can bound $|\psi|$ using Lemma 19 and $D_{\underline{L}}\psi$, $|\underline{\alpha}|$, $\rho$ using the integrated local energy estimates. Indeed we can show that

$$
\begin{aligned}
&I^{1-\epsilon}_{1+\gamma_0+2\epsilon}[\mathcal{L}_Z^2 J_{\underline{L}}](\{r\geq R\})\\
&=\int_\tau\int_{H_{\tau^*}} r_+^{-\epsilon-1}\tau_+^{1+\gamma_0+2\epsilon}|r^2\mathcal{L}_Z^2 J_{\underline{L}}|^2\,du\,d\omega\,d\tau\\
&\lesssim \int_\tau\int_v r_+^{-1-\epsilon}\tau_+^{1+\gamma_0+2\epsilon}\int_\omega|\psi_2|^2\,d\omega\cdot\int_\omega|D_{\underline{L}}\psi_2|^2\,d\omega+\int_\omega|r\mathcal{L}_Z^2\underline{\alpha}|^2+|\mathcal{L}_Z^2\bar{\rho}|^2\,d\omega\cdot\left(\int_\omega|\psi_2|^2\,d\omega\right)^2 dv\,d\tau\\
&\quad+\int_{\tau\leq 0}\int_v r^{-1-\epsilon}|q_0|^2r^{-4}\tau_+^{1+\gamma_0+2\epsilon}\left(\int_\omega|\psi_2|^2 d\omega\right)^{1/2}dv\,d\tau\\
&\lesssim \mathcal{E}\int_\tau\tau_+^{1+2\epsilon}\int_{H_{\tau^*}}\frac{|\tilde{D}\phi_2|^2}{r_+^{1+\epsilon}}\,dx\,d\tau+\mathcal{E}^2\int_\tau\tau_+^{1+2\epsilon}\int_{H_{\tau^*}}\frac{|\mathcal{L}_Z^2\bar{F}|^2}{r_+^{1+\epsilon}}\,dx\,d\tau+\mathcal{E}^2|q_0|^2\int_{\tau\leq 0}\int_v r_+^{-4+\epsilon+\gamma_0}\,dv\,d\tau\\
&\lesssim \mathcal{E}I^{-1-\epsilon}_{1+2\epsilon}[\tilde{D}\phi_2](\{t\geq 0\})+\mathcal{E}^2I^{-1-\epsilon}_{1+2\epsilon}[\mathcal{L}_Z^2\bar{F}](\{t\geq 0\})+|q_0|^2\mathcal{E}^2\lesssim\mathcal{E}^2.
\end{aligned}
$$

Here, after using Sobolev embedding on the sphere, we dropped the lower-order terms like $\psi_1$, $\psi$. In the above estimate, we have used the decay estimates $\int_\omega|\psi_k|^2\,d\omega\lesssim\mathcal{E}\tau_+^{-\gamma_0}$ by Lemma 19. The last step follows from the integrated local energy decay (see, e.g., estimate (64)) and Lemma 20. We also note that in the exterior region, $r_+\geq\frac{1}{2}\tau_+$.

For $J_L$, Lemma 60 indicates that

$$
r^2|\mathcal{L}_Z^2 J_L|\lesssim|\psi_k||D_L\psi_{2-k}|+(|r\mathcal{L}_Z^{l_1}\alpha|+|\mathcal{L}_Z^{l_1}\rho|)|\psi_{l_2}||\psi_{1-l_1-l_2}|.
$$

Similarly, after using Sobolev embedding, we control $\psi_k$ by using Lemma 19. Then for $D_L\psi_k$, $|\mathcal{L}_Z^k\alpha|$ we can apply the $r$-weighted energy estimates. For $\rho$, we split it into the charge part $q_0r^{-2}$ and the chargeless part which can be bounded by using the energy flux decay estimates. More precisely, for $\epsilon\leq p\leq 1+\gamma_0$, using the estimate $r^{-1}\int_\omega|\psi_k|^2\,d\omega\lesssim\mathcal{E}\tau_+^{-1-\gamma_0}$ we can show that

$$
\begin{aligned}
&I^{1+p}_{1+\gamma_0+\epsilon-p}[\mathcal{L}_Z^2 J_L](\{r\geq R\})\\
&\qquad=\int_\tau\int_{H_{\tau^*}} r_+^{p-1}\tau_+^{1+\gamma_0+\epsilon-p}|r^2\mathcal{L}_Z^2 J_L|^2\,dv\,d\omega\,d\tau\\
&\qquad\lesssim\mathcal{E}\int_\tau\int_{H_{\tau^*}} r_+^p\tau_+^{\epsilon-p}|D_L\psi_2|^2\,d\omega\,dv\,d\tau+\mathcal{E}^2\int_\tau\int_{H_{\tau^*}} r_+^p\tau_+^{\epsilon-p-\gamma_0}(|r\mathcal{L}_Z^2\alpha|^2+|\mathcal{L}_Z^2\bar{\rho}|^2)\,d\omega\,dv\,d\tau\\
&\qquad\quad+\mathcal{E}^2\int_{\tau\leq 0}\int_v r^{p-1}|q_0|^2r^{-4}\tau_+^{1-\gamma_0+\epsilon-p}\,dv\,d\tau\\
&\qquad\lesssim\mathcal{E}^2\int_\tau\tau_+^{\epsilon-p-1-\gamma_0+p}+\tau_+^{\epsilon-p-\gamma_0-1-\gamma_0+p}+\tau_+^{\epsilon-p-1-\gamma_0}\,d\tau+\mathcal{E}^2|q_0|^2\lesssim\mathcal{E}^2.
\end{aligned}
$$

Next, for $\not{J}$, Lemma 60 shows that

$$r^2|\mathcal{L}_Z^2\not{J}| \lesssim |\psi_k||\not{D}\psi_{2-k}| + (|r\mathcal{L}_Z^{l_1}\sigma| + |\mathcal{L}_Z^{l_1}\alpha| + |\mathcal{L}_Z^{l_1}\underline{\alpha}|)|\psi_{l_2}||\psi_{1-l_1-l_2}|.$$

Like the previous estimates for $J_{\underline{L}}$, $J_L$, for all $\epsilon \le p \le \gamma_0$ we can show that

$$\begin{aligned}
&I^{1+p}_{1+\gamma_0+\epsilon-p}[\mathcal{L}_Z^2\not{J}_L](\{r \ge R\}) \\
&\quad = \int_\tau \int_{H_{\tau^*}} r_+^{p-1}\tau_+^{1+\gamma_0+\epsilon-p}|r^2\mathcal{L}_Z^2\not{J}|^2\, dv\, d\omega\, d\tau \\
&\quad \lesssim \mathcal{E}\int_\tau \int_{H_{\tau^*}} r_+^p\tau_+^{\epsilon-p}|\not{D}\psi_2|^2\, d\omega\, dv\, d\tau + \mathcal{E}^2\int_\tau\int_{H_{\tau^*}} r_+^p\tau_+^{\epsilon-p-\gamma_0}(|r\mathcal{L}_Z^{l_1}\sigma|^2 + |\mathcal{L}_Z^{l_1}(\underline{\alpha},\alpha)|^2)\, d\omega\, dv\, d\tau \\
&\quad \lesssim \mathcal{E}\int_\tau \int_{H_{\tau^*}} r_+^{\gamma_0}(|\not{D}\psi_2|^2 + \mathcal{E}|r\mathcal{L}_Z^{l_1}(\sigma,\alpha)|^2)\, d\omega\, dv\, d\tau + \mathcal{E}^2\int_\tau\int_{H_{\tau^*}} r_+^{1-\epsilon}|\mathcal{L}_Z^{l_1}\underline{\alpha}|^2\, d\omega\, dv\, d\tau \\
&\quad \lesssim \mathcal{E}^2.
\end{aligned}$$

Here $l_1 \le 1$. The last term is bounded by using the integrated local energy estimates. This relies on the assumption that $\gamma_0 \le 1-\epsilon < 1$. For $\gamma_0 \ge 1$, we then can use the improved integrated local energy estimate for the angular derivatives of $\phi$ or $\sigma$, or we can move the $r$ weights to $\phi_k$.

Combining the above estimates, we have (99). □

By choosing $\mathcal{E}$ sufficiently small depending only on $\mathcal{M}$, $\epsilon$, $R$ and $\gamma_0$, we then can improve the bootstrap assumption (98). To prove Theorem 1, we can choose $R = 2$. Then for sufficiently small $\mathcal{E}$, we can bound $m_2$ and $M_2$. The pointwise estimates in the main Theorem 1 follow from Propositions 14, 17, 54 and 56.

## Acknowledgments

The author would like to thank Pin Yu for helpful discussions. He is also indebted to the anonymous referee for plenty of helpful suggestions and comments on the manuscript.

## References

[Bieri et al. 2014] L. Bieri, S. Miao, and S. Shahshahani, "Asymptotic properties of solutions of the Maxwell Klein Gordon equation with small data", preprint, 2014. arXiv

[Choquet-Bruhat and Christodoulou 1981a] Y. Choquet-Bruhat and D. Christodoulou, "Elliptic systems in $H_{s,\delta}$ spaces on manifolds which are Euclidean at infinity", *Acta Math.* **146**:1-2 (1981), 129–150. MR Zbl

[Choquet-Bruhat and Christodoulou 1981b] Y. Choquet-Bruhat and D. Christodoulou, "Existence of global solutions of the Yang–Mills, Higgs and spinor field equations in $3+1$ dimensions", *Ann. Sci. École Norm. Sup.* (4) **14**:4 (1981), 481–506. MR Zbl

[Christodoulou and Klainerman 1990] D. Christodoulou and S. Klainerman, "Asymptotic properties of linear field equations in Minkowski space", *Comm. Pure Appl. Math.* **43**:2 (1990), 137–199. MR Zbl

[Dafermos and Rodnianski 2009] M. Dafermos and I. Rodnianski, "The red-shift effect and radiation decay on black hole spacetimes", *Comm. Pure Appl. Math.* **62**:7 (2009), 859–919. MR Zbl

[Dafermos and Rodnianski 2010] M. Dafermos and I. Rodnianski, "A new physical-space approach to decay for the wave equation with applications to black hole spacetimes", pp. 421–432 in *XVIth International Congress on Mathematical Physics*, edited by P. Exner, World Sci. Publ., Hackensack, NJ, 2010. MR Zbl

[Eardley and Moncrief 1982a] D. M. Eardley and V. Moncrief, "The global existence of Yang–Mills–Higgs fields in 4-dimensional Minkowski space, I: Local existence and smoothness properties", *Comm. Math. Phys.* **83**:2 (1982), 171–191. MR Zbl

[Eardley and Moncrief 1982b] D. M. Eardley and V. Moncrief, "The global existence of Yang–Mills–Higgs fields in 4-dimensional Minkowski space, II: Completion of proof", *Comm. Math. Phys.* **83**:2 (1982), 193–212. MR Zbl

[Keel et al. 2011] M. Keel, T. Roy, and T. Tao, "Global well-posedness of the Maxwell–Klein–Gordon equation below the energy norm", *Discrete Contin. Dyn. Syst.* **30**:3 (2011), 573–621. MR Zbl

[Klainerman 1984] S. Klainerman, "Long time behaviour of solutions to nonlinear wave equations", pp. 1209–1215 in *Proceedings of the International Congress of Mathematicians, Vol. 2* (Warsaw, 1983), edited by Z. Ciesielski and C. Olech, PWN, Warsaw, 1984. MR Zbl

[Klainerman 1986] S. Klainerman, "The null condition and global existence to nonlinear wave equations", pp. 293–326 in *Nonlinear systems of partial differential equations in applied mathematics, Part 1* (Santa Fe, NM, 1984), edited by B. Nicolaenko et al., Lectures in Appl. Math. **23**, American Mathematical Society, Providence, RI, 1986. MR Zbl

[Klainerman and Machedon 1994] S. Klainerman and M. Machedon, "On the Maxwell–Klein–Gordon equation with finite energy", *Duke Math. J.* **74**:1 (1994), 19–44. MR Zbl

[Klainerman and Machedon 1995] S. Klainerman and M. Machedon, "Finite energy solutions of the Yang–Mills equations in $\mathbf{R}^{3+1}$", *Ann. of Math.* (2) **142**:1 (1995), 39–119. MR Zbl

[Krieger and Lührmann 2015] J. Krieger and J. Lührmann, "Concentration compactness for the critical Maxwell–Klein–Gordon equation", *Ann. PDE* **1**:1 (2015), Art. 5, 208. MR

[Krieger et al. 2015] 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**:6 (2015), 973–1040. MR Zbl

[Lindblad and Sterbenz 2006] H. Lindblad and J. Sterbenz, "Global stability for charged-scalar fields on Minkowski space", *IMRP Int. Math. Res. Pap.* (2006), Art. ID 52976, 109. MR Zbl

[Luk 2010] J. Luk, "Improved decay for solutions to the linear wave equation on a Schwarzschild black hole", *Ann. Henri Poincaré* **11**:5 (2010), 805–880. MR Zbl

[Machedon and Sterbenz 2004] M. Machedon and J. Sterbenz, "Almost optimal local well-posedness for the $(3+1)$-dimensional Maxwell–Klein–Gordon equations", *J. Amer. Math. Soc.* **17**:2 (2004), 297–359. MR Zbl

[McOwen 1979] R. C. McOwen, "The behavior of the Laplacian on weighted Sobolev spaces", *Comm. Pure Appl. Math.* **32**:6 (1979), 783–795. MR Zbl

[Oh 2015] S.-J. Oh, "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**:9 (2015), 1669–1732. MR Zbl

[Oh and Tataru 2016] S.-J. Oh and D. Tataru, "Global well-posedness and scattering of the (4+1)-dimensional Maxwell–Klein–Gordon equation", *Invent. Math.* **205**:3 (2016), 781–877. MR

[Rodnianski and Tao 2004] 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**:2 (2004), 377–426. MR Zbl

[Schlue 2013] V. Schlue, "Decay of linear waves on higher-dimensional Schwarzschild black holes", *Anal. PDE* **6**:3 (2013), 515–600. MR Zbl

[Selberg and Tesfahun 2010] S. Selberg and A. Tesfahun, "Finite-energy global well-posedness of the Maxwell–Klein–Gordon system in Lorenz gauge", *Comm. Partial Differential Equations* **35**:6 (2010), 1029–1057. MR Zbl

[Shu 1991] W.-T. Shu, "Asymptotic properties of the solutions of linear and nonlinear spin field equations in Minkowski space", *Comm. Math. Phys.* **140**:3 (1991), 449–480. MR Zbl

[Shu 1992] W.-T. Shu, "Global existence of Maxwell–Higgs fields", pp. 214–227 in *Nonlinear hyperbolic equations and field theory* (Lake Como, 1990), edited by M. K. V. Murthy and S. Spagnolo, Pitman Research Notes in Mathematics Series **253**, Longman Sci. Tech., Harlow, 1992. MR Zbl

[Yang 2013] S. Yang, "Global solutions of nonlinear wave equations in time dependent inhomogeneous media", *Arch. Ration. Mech. Anal.* **209**:2 (2013), 683–728. MR Zbl

[Yang 2015a] S. Yang, "Global solutions of nonlinear wave equations with large data", *Selecta Math.* (*N.S.*) **21**:4 (2015), 1405–1427. MR Zbl

[Yang 2015b] S. Yang, "Global stability of solutions to nonlinear wave equations", *Selecta Math.* (*N.S.*) **21**:3 (2015), 833–881. MR Zbl

[Yang 2015c] S. Yang, "On global behavior of solutions of the Maxwell-Klein-Gordon equations", preprint, 2015. arXiv

Received 1 Nov 2015. Revised 7 Mar 2016. Accepted 28 Aug 2016.

SHIWU YANG: yangsw315@126.com
*Beijing International Center for Mathematical Research, Peking University, Beijing 100086, China*

ANALYSIS AND PDE
Vol. 9, No. 8, 2016
dx.doi.org/10.2140/apde.2016.9.1903

# INVARIANT DISTRIBUTIONS AND THE GEODESIC RAY TRANSFORM

GABRIEL P. PATERNAIN AND HANMING ZHOU

We establish an equivalence principle between the solenoidal injectivity of the geodesic ray transform acting on symmetric $m$-tensors and the existence of invariant distributions or smooth first integrals with prescribed projection over the set of solenoidal $m$-tensors. We work with compact simple manifolds, but several of our results apply to nontrapping manifolds with strictly convex boundary.

## 1. Introduction

The present paper studies the geodesic ray transform of a compact simply connected Riemannian manifold with no conjugate points and strictly convex boundary. Our main objective is to establish an equivalence principle between injectivity of the ray transform acting on solenoidal symmetric $m$-tensors and the existence of solutions to the transport equation (associated with the geodesic vector field) with prescribed projection over the set of solenoidal $m$-tensors.

The Radon transform in the plane is the most fundamental example of the geodesic ray transform. It packs the integrals of a function $f$ in $\mathbb{R}^2$ over straight lines:

$$Rf(s,\omega) = \int_{-\infty}^{\infty} f(s\omega + t\omega^{\perp})\, dt, \quad s \in \mathbb{R},\ \omega \in S^1.$$

Here $\omega^{\perp}$ is the rotation of $\omega$ by 90 degrees counterclockwise. The properties of this transform are well studied [Helgason 1999] and constitute the theoretical underpinnings for many medical imaging methods such as CT and PET. Generalizations of the Radon transform are often needed. In seismic and ultrasound imaging one finds ray transforms where the measurements are given by integrals over more general families of curves, often modeled as the geodesics of a Riemannian metric. Moreover, integrals of tensor fields over geodesics are ubiquitous in rigidity questions in differential geometry and dynamics.

In this paper we will relate the injectivity properties of the geodesic ray transform with a well-studied subject in classical mechanics: the existence of special first integrals of motion along geodesics. Some Riemannian metrics admit distinguished first integrals; e.g., the geodesic flow of an ellipsoid in $\mathbb{R}^3$ admits a nontrivial first integral which is quadratic in momenta. As recently shown in [Kruglikov and Matveev 2016], a generic metric does not admit a nontrivial first integral that is polynomial in momenta, but here we will show a complementary statement going in the opposite direction: from the injectivity of the geodesic ray transform on tensors, we will show that it is possible to construct a smooth first integral with *any* prescribed polynomial part. In other words, given a polynomial $F$ of degree $m$ in momenta

*MSC2010:* primary 53C65; secondary 58J40.
*Keywords:* geodesic ray transform, first integral, tensor tomography, invariant distribution.

satisfying a natural restriction condition (related with the transport equation, see Section 7), we will show that we can find a smooth function $G$ whose dependence on momenta is of order $> m$ such that $F + G$ is a first integral of the geodesic flow. Generically $G$ is nonvanishing and *not* polynomial in momenta.

Let us now explain our results in more detail. The geodesic ray transform acts on functions defined on the unit sphere bundle of a compact oriented $n$-dimensional Riemannian manifold $(M, g)$ with boundary $\partial M$ ($n \geq 2$). Let $SM$ denote the unit sphere bundle on $M$; i.e.,

$$SM := \{(x, \xi) \in TM : \|\xi\|_g = 1\}.$$

We define the volume form on $SM$ by $d\Sigma^{2n-1}(x, \xi) = |dV^n(x) \wedge d\Omega_x(\xi)|$, where $dV^n$ is the volume form on $M$ and $d\Omega_x(\xi)$ is the volume form on the fiber $S_xM$. The boundary of $SM$ is given by $\partial SM := \{(x, \xi) \in SM : x \in \partial M\}$. On $\partial SM$ the natural volume form is $d\Sigma^{2n-2}(x, \xi) = |dV^{n-1}(x) \wedge d\Omega_x(\xi)|$, where $dV^{n-1}$ is the volume form on $\partial M$. We define two subsets of $\partial SM$,

$$\partial_\pm SM := \big\{(x, \xi) \in \partial SM : \pm\langle \xi, \nu(x)\rangle_g \leq 0\big\},$$

where $\nu(x)$ is the outward unit normal vector on $\partial M$ at $x$. It is easy to see that

$$\partial_+ SM \cap \partial_- SM = S(\partial M).$$

Given $(x, \xi) \in SM$, we denote by $\gamma_{x,\xi}$ the unique geodesic with $\gamma_{x,\xi}(0) = x$ and $\dot\gamma_{x,\xi}(0) = \xi$ and let $\tau(x, \xi)$ be the first time when the geodesic $\gamma_{x,\xi}$ exits $M$.

We say that $(M, g)$ is *nontrapping* if $\tau(x, \xi) < \infty$ for all $(x, \xi) \in SM$.

**Definition 1.1.** The *geodesic ray transform* of a function $f \in C^\infty(SM)$ is the function

$$If(x, \xi) = \int_0^{\tau(x,\xi)} f\big(\gamma_{x,\xi}(t), \dot\gamma_{x,\xi}(t)\big)\, dt, \quad (x, \xi) \in \partial_+ SM.$$

Note that if the manifold $(M, g)$ is nontrapping and has strictly convex boundary, then $I : C^\infty(SM) \to C^\infty(\partial_+ SM)$, and Santaló's formula (see Section 2) implies that $I$ is also a bounded map $L^2(SM) \to L^2_\mu(\partial_+ SM)$, where $d\mu(x, \xi) = |\langle \nu(x), \xi\rangle| d\Sigma^{2n-2}(x, \xi)$ and $L^2_\mu(\partial_+ SM)$ is the space of functions on $\partial_+ SM$ with inner product

$$(u, v)_{L^2_\mu(\partial_+ SM)} = \int_{\partial_+ SM} u\bar v\, d\mu.$$

Given $f \in C^\infty(SM)$, what properties of $f$ may be determined from the knowledge of $If$? Clearly a general function $f$ on $SM$ is not determined by its geodesic ray transform alone, since $f$ depends on more variables than $If$. In applications one often encounters the transform $I$ acting on special functions on $SM$ that arise from symmetric tensor fields, and we will now consider this case.

We denote by $C^\infty(S^m(T^*M))$ the space of smooth covariant symmetric tensor fields of rank $m$ on $M$ with $L^2$ inner product

$$(u, v) := \int_M u_{i_1 \cdots i_m} \overline{v^{i_1 \cdots i_m}}\, dV^n,$$

where $v^{i_1\cdots i_m} = g^{i_1 j_1}\cdots g^{i_m j_m} v_{j_1\cdots j_m}$. There is a natural map

$$\ell_m : C^\infty(S^m(T^*M)) \to C^\infty(SM)$$

given by $\ell_m(f)(x,\xi) := f_x(\xi,\dots,\xi)$. We can now define the geodesic ray transform acting on symmetric $m$-tensors simply by setting $I_m := I \circ \ell_m$. Let $d = \sigma\nabla$ be the symmetric inner differentiation, where $\nabla$ is the Levi-Civita connection associated with $g$, and $\sigma$ denotes symmetrization. It is easy to check that if $v = dp$ for some $p \in C^\infty(S^{m-1}(T^*M))$ with $p|_{\partial M} = 0$, then $I_m v = 0$. The tensor tomography problem asks the following question: are such tensors the only obstructions for $I_m$ to be injective? If this is the case, then we say $I$ is *solenoidal injective or s-injective* for short. The problem is wide open for compact nontrapping manifolds with strictly convex boundary (but see [Uhlmann and Vasy 2016; Stefanov et al. 2014]). There are more results if one assumes the stronger condition of being *simple*, i.e., $(M, g)$ is simply connected, has no conjugate points and strictly convex boundary. For simple surfaces, the tensor tomography problem has been completely solved [Paternain et al. 2013]. For simple manifolds of any dimension, solenoidal injectivity is known for $I_0$ and $I_1$ [Muhometov 1977; Anikonov and Romanov 1997]. For $m$-tensors, $m \geq 2$, the tensor tomography problem is still open, but some substantial partial results were established under additional assumptions; see, e.g., [Pestov and Sharafutdinov 1988; Sharafutdinov 1994; Stefanov and Uhlmann 2005; Paternain et al. 2015a; Stefanov et al. 2014].

Let us explain a bit further the term "solenoidal injective". Consider the Sobolev space $H^k(S^m(T^*M))$ naturally associated with the $L^2$ inner product defined above. By [Sharafutdinov 1994; Sharafutdinov et al. 2005], there is an orthogonal decomposition of $L^2$ symmetric tensors fields. Given $v \in H^k(S^m(T^*M))$, $k \geq 0$, there exist uniquely determined $v^s \in H^k(S^m(T^*M))$ and $p \in H^{k+1}(S^{m-1}(T^*M))$ such that

$$v = v^s + dp, \quad \delta v^s = 0, \quad p|_{\partial M} = 0,$$

where $\delta$ is the divergence. We call $v^s$ and $dp$ the *solenoidal* part and *potential* part of $v$ respectively. Moreover, we denote by $H^k(S^m_{\mathrm{sol}}(T^*M))$ and $C^\infty(S^m_{\mathrm{sol}}(T^*M))$ the subspaces of $H^k(S^m(T^*M))$ and $C^\infty(S^m(T^*M))$ respectively whose elements are solenoidal symmetric tensor fields. Solenoidal injectivity of $I_m$ simply means that $I_m$ is injective when restricted to $C^\infty(S^m_{\mathrm{sol}}(T^*M))$.

Let $I^*$ denote the adjoint of $I$ using the $L^2$ inner products defined above; that is,

$$(Iu, \varphi) = (u, I^*\varphi)$$

for $u \in L^2(SM)$, $\varphi \in L^2_\mu(\partial_+ SM)$. A simple application of Santaló's formula yields

$$I^*\varphi = \varphi^\sharp,$$

where $\varphi^\sharp(x,\xi) := \varphi\big(\gamma_{x,\xi}(-\tau(x,-\xi)), \dot\gamma_{x,\xi}(-\tau(x,-\xi))\big)$ (see Section 2 for details). Observe that by definition, $\varphi^\sharp$ is constant along orbits of the geodesic flow. If we are now interested in $I_m^*$, we note that

$$I_m^* = \ell_m^* \circ I^*$$

and hence we just need to compute $\ell_m^*$. This is easy (see Section 2) and one finds

$$L_m f := \ell_m^* f(x)_{i_1\cdots i_m} := g_{i_1 j_1}\cdots g_{i_m j_m} \int_{S_x M} f(x,\xi)\xi^{j_1}\cdots\xi^{j_m}\, d\Omega_x(\xi).$$

The fundamental microlocal property of the geodesic ray transform is that, for simple manifolds, $I_m^* I_m$ is a pseudodifferential operator of order $-1$ on a slightly larger open manifold engulfing $M$. Moreover, it has a suitable ellipticity property when acting on solenoidal tensors [Sharafutdinov et al. 2005]. This has been exploited to great effect to derive surjectivity of $I_m^*$ knowing injectivity of $I_m$ [Pestov and Uhlmann 2005; Dairbekov and Uhlmann 2010] for $m = 0, 1$. Since the range of $I_m^*$ is contained in the space of solenoidal tensors, by saying $I_m^*$ is *surjective* we mean that the range of $I_m^*$ equals the latter. Surjectivity of $I_m^*$ for tensors of order 0 and 1 has been the key for the recent success in the solution of several long standing questions in 2D [Salo and Uhlmann 2011; Pestov and Uhlmann 2005; Paternain et al. 2012; 2013; 2014; Guillarmou 2014]. However, very little is known about surjectivity for $m \geq 2$ and this largely motivates the present paper.

The surjectivity properties of the adjoint of the geodesic ray transform reveal themselves in the existence of solutions $f$ to the transport equation $Xf = 0$ with prescribed values for $L_m f$ in the space of solenoidal tensors. Here $X$ is the geodesic vector field acting on distributions by duality (recall that $X$ preserves the volume form $d\Sigma^{2n-1}$). A distribution $f$ on $SM$ is said to be *invariant* if it satisfies $Xf = 0$. As we already mentioned, in this paper we mainly study the relation among the injectivity of $I_m$, the surjectivity of its adjoint $I_m^*$ on solenoidal tensor fields and the existence of some invariant distributions or smooth first integrals associated with solenoidal tensor fields. On a compact nontrapping manifold with strictly convex boundary, the geodesic ray transform $I_m$ is extendable to a bounded operator

$$I_m : H^k(S^m(T^*M)) \to H^k(\partial_+ SM)$$

for all $k \geq 0$ [Sharafutdinov 1994, Theorem 4.2.1]. Moreover, it can be easily checked that

$$I_m(H_0^k(S^m(T^*M))) \subset H_0^k(\partial_+ SM)$$

and hence we can define $I_m^*$ by duality acting on negative Sobolev spaces to obtain a bounded operator:

$$I_m^* : H^{-k}(\partial_+ SM) \to H^{-k}(S^m(T^*M)).$$

In other words, for $\varphi \in H^{-k}(\partial_+ SM)$, we have $I_m^*\varphi$ is defined by $(I_m^*\varphi, u) = (\varphi, I_m u)$ for all $u \in H_0^k(S^m(T^*M))$. Let $C_\alpha^\infty(\partial_+ SM)$ denote the set of smooth functions $\varphi$ for which $\varphi^\sharp$ is also smooth.

Our main result is the following theorem:

**Theorem 1.2.** *Let $M$ be a compact simple Riemannian manifold. Then the following are equivalent*:

(1) *$I_m$ is s-injective on $C^\infty(S^m(T^*M))$.*

(2) *For every $u \in L^2(S^m_{\mathrm{sol}}(T^*M))$, there exists $\varphi \in H^{-1}(\partial_+ SM)$ such that $u = I_m^*\varphi$.*

(3) *For every $u \in L^2(S^m_{\mathrm{sol}}(T^*M))$, there exists $f \in H^{-1}(SM)$ satisfying $Xf = 0$ and $u = L_m f$.*

(4) *For every $u \in C^\infty(S^m_{\mathrm{sol}}(T^*M))$, there exists $\varphi \in C_\alpha^\infty(\partial_+ SM)$ such that $u = I_m^*\varphi$.*

(5) *For every $u \in C^\infty(S^m_{\mathrm{sol}}(T^*M))$, there exists $f \in C^\infty(SM)$ with $Xf = 0$ such that $L_m f = u$.*

We observe that by [Sharafutdinov et al. 2005, Theorem 1.1], s-injectivity of $I_m$ on $L^2(S^m(T^*M))$ is equivalent to s-injectivity of $I_m$ on $C^\infty(S^m(T^*M))$.

Let us return to the subject of special first integrals associated with the geodesic flow. By considering the vertical Laplacian $\Delta$ on each fiber $S_xM$ of $SM$, we have a natural $L^2$ decomposition $L^2(SM) = \bigoplus_{m\geq 0} H_m(SM)$ into vertical spherical harmonics. We set $\Omega_m := H_m(SM) \cap C^\infty(SM)$. Then a function $u$ belongs to $\Omega_m$ if and only if $\Delta u = m(m+n-2)u$, where $n = \dim M$. The maps

$$\ell_m : C^\infty(S^m(T^*M)) \to \bigoplus_{k=0}^{[m/2]} \Omega_{m-2k}$$

and

$$L_m : \bigoplus_{k=0}^{[m/2]} \Omega_{m-2k} \to C^\infty(S^m(T^*M))$$

are isomorphisms. These maps give natural identification between functions in $\Omega_m$ and *trace-free* symmetric $m$-tensors (for details on this, see [Guillemin and Kazhdan 1980b; Dairbekov and Sharafutdinov 2010; Paternain et al. 2015a]). If $(M, g)$ is a simple manifold with $I_m$ s-injective, Theorem 1.2(5) says that given any $u \in C^\infty(S^m_{\mathrm{sol}}(T^*M))$ there is a first integral of the geodesic flow $f$ such that $L_m f = u$. In other words, if we let $F = L_m^{-1}u \in \bigoplus_{k=0}^{[m/2]} \Omega_{m-2k}$ and $G = f - F$, we see that $F$ is polynomial of degree $m$ in velocities and it can be completed by adding $G$ to obtain a first integral. We also see that (taking the even or odd part of $f$ if necessary) $G \in \bigoplus_{k\geq 1} \Omega_{m+2k}$. These were the functions mentioned earlier in the introduction. If $G$ were to be zero, then there would be a first integral that is polynomial in velocities and generically these do not exist. We note that the paper [Paternain et al. 2015a] also constructs invariant distributions (they are not smooth in general) with prescribed $m$-th polynomial component using a different method (a Beurling transform), but it requires nonpositive curvature for it to work. As already mentioned, here we use instead the normal operator $I_m^* I_m$.

The results in [Pestov and Uhlmann 2005; Dairbekov and Uhlmann 2010] prove that (1) implies (4) or (5) in Theorem 1.2 for $m = 0, 1$, so the main contribution in the theorem is to cover the case $m \geq 2$ and also to provide additional invariant distributions associated with $L^2$ solenoidal tensors. The proof of Theorem 1.2 relies on a solenoidal extension of tensor fields. For $m = 0$ no extension is needed and for $m = 1$ the situation is considerably simpler and an extension result is already available in [Kato et al. 2000]. Paradoxically the need for a solenoidal extension does not arise in the more complicated setting of Anosov manifolds since there is no boundary. In this setting, an analogous result to Theorem 1.2 (in the $L^2$ setting) has been recently proved by C. Guillarmou [2014, Corollary 3.7] and these ideas gave rise to a full solution to the tensor tomography problem on an Anosov surface.

Since in 2D the tensor tomography problem has been fully solved [Paternain et al. 2013], we derive:

**Corollary 1.3.** *Let $(M, g)$ be a compact simple surface. For every $u \in C^\infty(S^m_{\mathrm{sol}}(T^*M))$, there exists $f \in C^\infty(SM)$ with $Xf = 0$ such that $L_m f = u$.*

We shall also give an alternative proof of the corollary using results from [Paternain et al. 2015b]. The alternative proof avoids the smooth solenoidal extension and sheds some light on the relationship between the transport equation and the solenoidal condition.

The rest of the paper is organized as follows. Section 2 contains some preliminaries. In Section 3 we establish the $L^2$ and $C^\infty$ compactly supported solenoidal extension of tensor fields. This necessitates at some point the use of the generic nonexistence of nontrivial Killing tensor fields recently proved in [Kruglikov and Matveev 2016]. Section 4 uses the well-established microlocal analysis to prove a surjectivity result for $I_m^* I_m$ following the strategy in [Dairbekov and Uhlmann 2010]. Section 5 establishes various boundedness properties on Sobolev spaces that allow us to extend the relevant operators to negative Sobolev spaces (i.e., distributions). Section 6 bundles up everything together and proves Theorem 1.2. Section 7 gives an alternative proof of Corollary 1.3 and clarifies the connection between solenoidal tensors and the transport equation.

## 2. Preliminaries

In this section we provide details about the regularity properties of the operators introduced in the previous section. First we describe the basic notation we will use frequently in the rest of the paper. Given a compact Riemannian manifold $\mathcal{M}$ with boundary, we define

$$C_c^\infty(\mathcal{M}^{\mathrm{int}}) := \{f \in C^\infty(\mathcal{M}) : \operatorname{supp} f \subset \mathcal{M}^{\mathrm{int}}\},$$
$$H_c^k(\mathcal{M}^{\mathrm{int}}) := \{f \in H^k(\mathcal{M}) : \operatorname{supp} f \subset \mathcal{M}^{\mathrm{int}}\} \quad \text{for } k \in \mathbb{Z}.$$

Then for any $s > 0$, $s \in \mathbb{Z}$, we say $H_0^s(\mathcal{M})$ is the completion of $C_c^\infty(\mathcal{M}^{\mathrm{int}})$ under the $H^s$ norm.

Now let $M$ be a compact manifold. Given $f \in C^\infty(SM)$ and $u \in C^\infty(S^m(T^*M))$, we have

$$\begin{aligned}(\ell_m u, f) &= \int_{SM} u_{j_1\cdots j_m}(x)\xi^{j_1}\cdots\xi^{j_m} f(x,\xi)\, d\Sigma^{2n-1} \\ &= \int_M u_{j_1\cdots j_m}(x) \int_{S_xM} f(x,\xi)\xi^{j_1}\cdots\xi^{j_m}\, d\Omega_x(\xi)\, dV^n(x).\end{aligned}$$

This means that

$$L_m = \ell_m^* : C^\infty(SM) \to C^\infty(S^m(T^*M))$$

is given by

$$L_m f(x)_{i_1\cdots i_m} = g_{i_1 j_1}\cdots g_{i_m j_m} \int_{S_xM} f(x,\xi)\xi^{j_1}\cdots\xi^{j_m}\, d\Omega_x(\xi).$$

Since the metric tensor $g$ is smooth, for the sake of simplicity, we identify $L_m f$ with its dual,

$$L_m f(x)^{j_1\cdots j_m} = \int_{S_xM} f(x,\xi)\xi^{j_1}\cdots\xi^{j_m}\, d\Omega_x(\xi).$$

On the other hand, it is easy to see that the map $\ell_m$ can be extend to the bounded operator

$$\ell_m : H^k(S^m(T^*M)) \to H^k(SM)$$

for any integer $k \geq 0$. In particular $\ell_m(H_0^k(S^m(T^*M))) \subset H_0^k(SM)$. Therefore we can define

$$L_m : H^{-k}(SM) \to H^{-k}(S^m(T^*M)) \tag{1}$$

in the sense of distributions and it is bounded.

Next, if $M$ is compact nontrapping with strictly convex boundary, we study the properties of $I$ and its adjoint $I^*$. Recall a useful integral identity called Santaló's formula.

**Lemma 2.1** [Sharafutdinov 1999, Lemma 3.3.2]. *Let $M$ be a compact nontrapping Riemannian manifold with strictly convex boundary. For every function $f \in C(SM)$, the equality*

$$\int_{SM} f(x,\xi)\, d\Sigma^{2n-1}(x,\xi) = \int_{\partial_+ SM} d\mu(x,\xi) \int_0^{\tau(x,\xi)} f\left(\gamma_{x,\xi}(t), \dot\gamma_{x,\xi}(t)\right) dt$$

*holds.*

Notice that the definition of compact dissipative Riemannian manifold (CDRM) in [Sharafutdinov 1999] is equivalent to compact nontrapping manifolds with strictly convex boundary.

Now let $\varphi \in C_\alpha^\infty(\partial_+ SM)$ and $f \in C^\infty(SM)$. By Santaló's formula,

$$\begin{aligned}(If,\varphi) &= \int_{\partial_+ SM} \varphi(x,\xi)\, d\mu \int_0^{\tau(x,\xi)} f\left(\gamma_{x,\xi}(t), \dot\gamma_{x,\xi}(t)\right) dt \\ &= \int_{\partial_+ SM} d\mu \int_0^{\tau(x,\xi)} \varphi^\sharp\left(\gamma_{x,\xi}(t), \dot\gamma_{x,\xi}(t)\right) f\left(\gamma_{x,\xi}(t), \dot\gamma_{x,\xi}(t)\right) dt \\ &= \int_{SM} \varphi^\sharp f\, d\Sigma^{2n-1}.\end{aligned}$$

Thus $I^*\varphi = \varphi^\sharp$ with

$$I^* : C_\alpha^\infty(\partial_+ SM) \to C^\infty(SM)$$

bounded. By the proof of [Sharafutdinov 1994, Theorem 4.2.1], one can extend $I$ to a bounded operator

$$I : H^k(SM) \to H^k(\partial_+ SM)$$

and $I(H_0^k(SM)) \subset H_0^k(\partial_+ SM)$ for any integer $k \geq 0$ (notice that $I(C_c^\infty((SM)^{\text{int}})) \subset C_c^\infty((\partial_+ SM)^{\text{int}})$). Thus we can define the bounded operator

$$I^* : H^{-k}(\partial_+ SM) \to H^{-k}(SM) \tag{2}$$

in the sense of distributions.

Given $u \in H_0^k(S^m(T^*M))$ and $\varphi \in H^{-k}(\partial_+ SM)$, we have $I_m^*\varphi$ is defined in the sense of distributions:

$$(I_m^*\varphi, u) := (I^*\varphi, \ell_m u) = (\varphi, I \circ \ell_m u) = (\varphi, I_m u).$$

**Lemma 2.2.** *Given a compact nontrapping Riemannian manifold $M$ with strictly convex boundary,*

$$I_m^* = L_m \circ I^* : H^{-k}(\partial_+ SM)) \to H^{-k}(S^m(T^*M))$$

*is a bounded operator.*

To conclude this section, we briefly discuss $X$, the generating vector field of the geodesic flow on the unit sphere bundle $SM$, acting on distributions. Since $X$ is a differential operator on $SM$, it is obvious that

$$X : H^{k+1}(SM) \to H^k(SM), \quad k \geq 0.$$

For $f \in H^{-k}(SM)$ and $h \in H_0^{k+1}(SM)$ (so $Xh \in H_0^k(SM)$), we define $Xf \in H^{-k-1}(SM)$ in the sense of distributions (notice that the volume form $d\Sigma^{2n-1}$ is invariant under the geodesic flow):

$$(Xf, h) := (f, -Xh).$$

## 3. Solenoidal extensions

In the paper [Kato et al. 2000], the authors proved the existence of compactly supported solenoidal extensions of solenoidal 1-forms to some larger manifold in both $L^2$ and smooth cases.

**Proposition 3.1.** *Let $\Omega$ be a bounded simply connected domain, with smooth boundary, contained in some Riemannian manifold $\mathcal{M}$. Let $U$ be an open neighborhood of $\Omega$ with $\partial U$ smooth. Then there exists a bounded map $\mathcal{E} : L^2_{\mathrm{sol}}(T^*\Omega) \to L^2_{U,\mathrm{sol}}(T^*\mathcal{M})$ such that $\mathcal{E}|_\Omega = \mathrm{Id}$. Moreover, $\mathcal{E}(C^\infty_{\mathrm{sol}}(T^*\Omega)) \subset C^\infty_{U,\mathrm{sol}}(T^*\mathcal{M})$.*

Here $L^2_{U,\mathrm{sol}}(T^*\mathcal{M})$ and $C^\infty_{U,\mathrm{sol}}(T^*\mathcal{M})$ denote the subspaces of $L^2_{\mathrm{sol}}(T^*\mathcal{M})$ and $C^\infty_{\mathrm{sol}}(T^*\mathcal{M})$ respectively consisting of elements supported in $U$.

Our goal is to extend this result to symmetric tensor fields of higher rank. However, for tensor fields of higher rank, new ideas are required and the argument is more involved.

***$L^2$ solenoidal extensions.*** We first prove the extension in the $L^2$ category by solving a suitable elliptic system.

**Proposition 3.2.** *Let $\Omega$ be a bounded simply connected domain, with smooth boundary, contained in some Riemannian manifold $(\mathcal{M}, g)$. Let $U$ be an open neighborhood of $\Omega$ with $\partial U$ smooth. Then given $m \geq 2$, $K \geq 2$ and $\epsilon > 0$, there exist a Riemannian metric $\tilde{g}$ and a bounded map $\mathcal{E} : L^2(S^m_{\mathrm{sol}}(T^*_g\Omega)) \to L^2(S^m_{U,\mathrm{sol}}(T^*_{\tilde{g}}\mathcal{M}))$ such that $\|\tilde{g} - g\|_{C^K} < \epsilon$, $\tilde{g}|_{\overline{\Omega}} = g$ and $\mathcal{E}|_\Omega = \mathrm{Id}$.*

*Proof.* Suppose $u \in L^2(S^m_{\mathrm{sol}}(T^*_g\Omega))$, i.e., $\delta u = 0$ in the sense of distributions. By the Green's formula for symmetric tensor fields (see [Sharafutdinov 1994]) one can define the boundary contraction of $u$ with the outward unit normal vector $\nu$ on $\partial\Omega$ in the sense of distributions; i.e., for $v \in H^1(S^{m-1}(T^*_g\Omega))$ we have

$$(u, dv)_\Omega = (j_\nu u, v)_{\partial\Omega}. \tag{3}$$

Since the trace operator $T : H^1(S^{m-1}(T^*_g\Omega)) \to H^{1/2}(S^{m-1}(\partial T^*_g\Omega))$, $Tv = v|_{\partial\Omega}$, is surjective, $j_\nu u \in H^{-1/2}(S^{m-1}(\partial T^*_g\Omega))$ is well-defined, and in local coordinates

$$(j_\nu u)_{i_1 i_2 \cdots i_{m-1}} = u_{i_1 i_2 \cdots i_{m-1} j}\, \nu^j.$$

By (3), for $v \in H^1(S^{m-1}(T^*_g\Omega))$ with $dv = 0$ (Killing tensor fields on $\Omega$), we have $(j_\nu u, v)_{\partial\Omega} = 0$.

It is known that generic (in the $C^K$-topology for $K \geq 2$) metrics admit only trivial integrals polynomial in momenta [Kruglikov and Matveev 2016]; i.e., for a generic metric $h$, the only Killing tensor fields are of the form $ch^k$, where $c \in \mathbb{R}$ and

$$h^k = \sigma(h \underbrace{\otimes \cdots \otimes}_{k} h)$$

is the symmetric tensor product of $k$ copies of $h$. Thus given any $\epsilon > 0$ and $K \geq 2$, there is a smooth metric $\tilde{g}$ with $\|\tilde{g} - g\|_{C^K} < \epsilon$ and $\tilde{g}|_{\overline{\Omega}} = g$ so that $(U \backslash \overline{\Omega}, \tilde{g})$ (thus $(U, \tilde{g})$) does not have nontrivial Killing tensor fields.

Define

$$f = \begin{cases} -j_\nu u & \text{on } \partial\Omega, \\ 0 & \text{on } \partial U. \end{cases}$$

Let $D := U \backslash \overline{\Omega}$ and consider the following boundary value problem for systems of second-order partial differential equations:

$$\begin{cases} \delta dw = 0 \quad \text{in } D, \\ j_\mu dw = f \in H^{-1/2}(S^{m-1}(\partial T^*_{\tilde{g}} D)), \\ w \in H^1(S^{m-1}(T^*_{\tilde{g}} D)). \end{cases} \tag{4}$$

Here $\mu$ is the outward unit normal vector on $\partial D$ for $D$; notice $\mu|_{\partial\Omega} = -\nu$. We claim that the system (4) is a regular elliptic system (also called coercive in some texts). Assume that the claim is true for the moment and let us continue the proof.

Next, we study the solutions of the homogeneous problem. Let $\delta dv = 0$ and $j_\mu dv|_{\partial D} = 0$ for some $v \in H^1(S^{m-1}(T^*_{\tilde{g}} D))$; by ellipticity, $v$ is smooth. Applying Green's formula, one has

$$\int_D \langle dv, dv \rangle \, dV^n(x) = -\int_D \langle \delta dv, v \rangle \, dV^n(x) + \int_{\partial D} \langle j_\mu dv, v \rangle \, dV^n(x) = 0,$$

i.e., $dv \equiv 0$. So the solution set of the homogeneous problem is

$$\mathcal{K} = \{ v \in C^\infty(S^{m-1}(T^*_{\tilde{g}} D)) : dv \equiv 0 \},$$

the set of Killing tensor fields of rank $m-1$ on $D$.

Now by [McLean 2000, Theorem 4.11], (4) is solvable in $H^1(S^{m-1}(T^*_{\tilde{g}} D))$ for the given boundary condition $f$ if and only if $(v, f)_{\partial D} = 0$ for all $v \in \mathcal{K}$. Note that $(D, \tilde{g})$ does not have nontrivial Killing tensor fields. If $m$ is even, the only Killing $(m-1)$-tensor field is $v = 0$; then $(v, f)_{\partial D} = (0, f)_{\partial D} = 0$. If $m$ is odd, the Killing $(m-1)$-tensor fields in $D$ are of the form $v = c\tilde{g}^{(m-1)/2}|_D$. Thus we can extend $v$ to $v = c\tilde{g}^{(m-1)/2}|_U$, which is also a Killing tensor field in $\Omega$. By the definition of $f$,

$$(v, f)_{\partial D} = -(v, j_\nu u)_{\partial\Omega} = -(v, \delta u)_\Omega - (dv, u)_\Omega = 0,$$

since $\delta u = 0$, $dv = 0$ in $\Omega$.

Thus the system (4) is solvable. Let $w \in H^1(S^{m-1}(T^*_{\tilde{g}} D))$ be a solution of (4) (the set of all solutions is $w + \mathcal{K}$) and define

$$\mathcal{E}u = \begin{cases} u & \text{in } \Omega, \\ dw & \text{in } D, \\ 0 & \text{in } \mathcal{M} \backslash \overline{U}. \end{cases}$$

It is easy to see that $\mathcal{E}u \in L^2(S^m(T^*_{\tilde{g}}\mathcal{M}))$ and $\operatorname{supp}\mathcal{E}u \subset \overline{U}$. In particular, for $v \in H^1(S^{m-1}(T^*_{\tilde{g}}\mathcal{M}))$,

$$\begin{aligned}(\delta\mathcal{E}u, v)_{\mathcal{M}} &= -(\mathcal{E}u, dv)_{\mathcal{M}} = -(dw, dv)_D - (u, dv)_\Omega \\ &= -(j_\mu dw, v)_{\partial D} - (j_\nu u, v)_{\partial\Omega} \\ &= -(-j_\nu u, v)_{\partial\Omega} - (j_\nu u, v)_{\partial\Omega} \\ &= 0.\end{aligned}$$

Thus $\mathcal{E}u$ is solenoidal in the sense of distributions, and $\mathcal{E}u \in L^2(S^m_{U,\mathrm{sol}}(T^*_{\tilde{g}}\mathcal{M}))$.

Moreover, by [McLean 2000, Theorem 4.11], we have the stability estimate

$$\|\mathcal{E}u\|^2_{L^2(\mathcal{M})} = \|u\|^2_{L^2(\Omega)} + \|dw\|^2_{L^2(D)} \le \|u\|^2_{L^2(\Omega)} + C\,\|j_\nu u\|^2_{H^{-1/2}(\partial\Omega)} \le C'\|u\|^2_{L^2(\Omega)},$$

i.e., $\mathcal{E}$ is bounded. □

The only thing left to prove is the claim about ellipticity.

**Lemma 3.3.** *The system (4) above is a regular elliptic system.*

*Proof.* It is well known that $\delta d$ is a self-adjoint elliptic operator; see, for example, [Sharafutdinov 1994]. We just need to show that the Neumann boundary value problem satisfies the Lopatinskii condition.

To check the Lopatinskii condition, we follow a similar procedure to that in the proof of [Sharafutdinov 1994, Theorem 3.3.2]. We choose local coordinates $(x^1, x^2, \dots, x^{n-1}, x^n = t \ge 0)$ in a neighborhood $W$ of $x_0 = (x', 0) \in \partial D$ in $D$ so that $\partial D \cap W = \{t = 0\}$ and $g_{ij}(x_0) = \delta_{ij}$. Define $d_0 = \sigma_p d$ and $\delta_0 = \sigma_p \delta$, the principal symbols of $d$ and $\delta$ respectively. Then we need to show that the boundary value problem for systems of ordinary differential equations

$$\begin{cases} \delta_0(x', 0, \xi', D_t)d_0(x', 0, \xi', D_t)w(t) = 0, \\ j_{-\frac{\partial}{\partial t}} d_0(x', 0, \xi', D_t)w(t)|_{t=0} = f_0 \end{cases}$$

has a unique solution in $\mathcal{N}_+$ for all $\xi' \in \mathbb{R}^{n-1}\setminus\{0\}$ and $f_0 \in S^{m-1}(\mathbb{R}^n)$, symmetric $(m-1)$-tensors on $\mathbb{R}^n$. Here $D_t = -i\,d/dt$, and for the sake of simplicity, we drop the space variables $(x', 0)$ from the symbols so

$$\begin{aligned}\mathcal{N}_+ := \big\{ w \in S^{m-1}(\mathbb{R}^n)|_{\{x'\}\times[0,\infty)} : \delta_0(\xi', D_t)d_0(\xi', D_t)w = 0 \text{ and} \\ w \text{ decays rapidly together with all derivatives as } t \to +\infty \big\}.\end{aligned}$$

Since the equation $\det\big(\delta_0(\xi', \zeta)d_0(\xi', \zeta)\big) = 0$ has real coefficients with no real root for $\xi' \neq 0$, it is not difficult to see that $\dim\mathcal{N}_+ = \dim S^{m-1}(\mathbb{R}^n)$. Thus it is sufficient to show that the homogeneous problem

$$\begin{cases} \delta_0(\xi', D_t)d_0(\xi', D_t)w(t) = 0, \\ j_{-\frac{\partial}{\partial t}} d_0(\xi', D_t)w(t)|_{t=0} = 0 \end{cases} \tag{5}$$

has only the zero solution in $\mathcal{N}_+$.

By a similar computation to that in the proof of [Sharafutdinov 1994, Theorem 3.3.2], we have the following Green's formula. Let $v(t) \in C^\infty\big([0,\infty) \to S^m(\mathbb{R}^n)\big)$ and $w(t) \in C^\infty\big([0,\infty) \to S^{m-1}(\mathbb{R}^n)\big)$

such that both of them decay rapidly together with all derivatives as $t \to +\infty$. If $j_-\frac{\partial}{\partial t} v(0) = 0$ (notice that different from [Sharafutdinov 1994], here we use the Neumann boundary condition at $t = 0$) then

$$\int_0^\infty \langle \delta_0(\xi', D_t)v, w\rangle \, dt = -\int_0^\infty \langle v, d_0(\xi', D_t)w\rangle \, dt. \tag{6}$$

Now if $w(t) \in \mathcal{N}_+$ is a solution to (5), let $v(t) = d_0(\xi', D_t)w(t)$. By (6) we obtain

$$d_0(\xi', D_t)w(t) = 0.$$

Notice that

$$(d_0(\xi)w)_{i_1\cdots i_m} = \frac{i}{m}\sum_{k=1}^{m} \xi_{i_k} w_{i_1\cdots \widehat{i_k}\cdots i_m},$$

where the $\widehat{\ }$ over $i_k$ means this index is omitted. Let $i_m = n$ and $\xi = (\xi', D_t)$. We obtain the system of first-order ordinary differential equations

$$(d_0(\xi', D_t)w)_{ni_1\cdots i_{m-1}} = \frac{i}{m}\left\{(\ell+1)D_t w_{i_1\cdots i_{m-1}} + \sum_{i_k \neq n} \xi_{i_k} w_{ni_1\cdots \widehat{i_k}\cdots i_{m-1}}\right\} = 0,$$

where $\ell = \ell(i_1, \ldots, i_{m-1})$ is the number of occurrences of the index $n$ in $(i_1, \ldots, i_{m-1})$. Since $\lim_{t\to+\infty} w(t) = 0$, by induction on $\ell$, the only solution to the above first-order homogeneous system is $w \equiv 0$, and this shows that (4) satisfies the Lopatinskii condition. □

***Smooth solenoidal extensions.*** In this subsection we achieve $C^\infty$ solenoidal extensions for tensors of arbitrary rank. Observe that the approach we use is quite different from the one of [Kato et al. 2000].

**Proposition 3.4.** *Let $\Omega$ be a bounded connected domain, with smooth boundary, contained in some Riemannian manifold $(\mathcal{M}, g)$. Let $U$ be an open neighborhood of $\Omega$ with $\partial U$ smooth. Then given $m \geq 2$, $K \geq 2$ and $\epsilon > 0$, there exist a Riemannian metric $\tilde{g}$ and a bounded map $\mathcal{E} : H^k(S^m_{\mathrm{sol}}(T^*_g\Omega)) \to L^2(S^m_{U,\mathrm{sol}}(T^*_{\tilde{g}}\mathcal{M}))$ for some integer $k \geq 2$ such that $\|\tilde{g} - g\|_{C^K} < \epsilon$, $\tilde{g}|_{\overline{\Omega}} = g$, $\mathcal{E}|_\Omega = \mathrm{Id}$ and*

$$\mathcal{E}(C^\infty(S^m_{\mathrm{sol}}(T^*_g\Omega))) \subset C^\infty(S^m_{U,\mathrm{sol}}(T^*_{\tilde{g}}\mathcal{M})).$$

To prove the proposition, we start with the following lemma on the existence of solenoidal extensions that might not be compactly supported.

**Lemma 3.5.** *Let $\Omega$ be a bounded connected domain, with smooth boundary, contained in some Riemannian manifold $(\mathcal{M}, g)$. There exists an open neighborhood $U$ of $\Omega$ such that every $u \in C^\infty(S^m_{\mathrm{sol}}(T^*\overline{\Omega}))$ can be extended to $\tilde{u} \in C^\infty(S^m_{\mathrm{sol}}(T^*U))$ with $\tilde{u}|_{\overline{\Omega}} = u$.*

*Proof.* Let $u \in C^\infty(S^m_{\mathrm{sol}}(T^*\overline{\Omega}))$, i.e., $\delta u = 0$, in local coordinates $u = u_{j_1\cdots j_m} dx^{j_1} \otimes \cdots \otimes dx^{j_m}$ and

$$(\delta u)_{i_1\cdots i_{m-1}} = g^{jk}\nabla_j u_{ki_1\cdots i_{m-1}} = 0, \tag{7}$$

where

$$\nabla_j u_{ki_1\cdots i_{m-1}} = \partial_j u_{ki_1\cdots i_{m-1}} - \Gamma^\ell_{jk} u_{\ell i_1\cdots i_{m-1}} - \sum_{s=1}^{m-1} \Gamma^\ell_{ji_s} u_{\ell ki_1\cdots \widehat{i_s}\cdots i_{m-1}}. \tag{8}$$

Pick $x_0 \in \partial\Omega$. We follow the idea of the proof in [Stefanov and Uhlmann 2005, Lemma 4.1] and choose semigeodesic coordinates $(x^1, \dots x^{n-1}, x^n) = (x', x^n)$ near $x_0$ with $\partial\Omega = \{x^n = 0\}$ and $\partial_n = \nu$ the unit outward (with respect to $\Omega$) vector normal to $\partial\Omega$; thus

$$g^{kn} = \delta^k_n, \quad \Gamma^n_{kn} = \Gamma^k_{nn} = 0 \quad \text{for all } k = 1, 2, \dots, n.$$

We extend the components $u_{j_1\cdots j_m}$, $j_s < n$ for all $1 \le s \le m$, smoothly to $U$ (note that $U\setminus\overline{\Omega}$ is determined by the semigeodesic neighborhood of $\partial\Omega$), and denote the extensions by $v_{j_1\cdots j_m}$. We will construct the other components in $\{x^n > 0\}$ by induction on the number of appearances of $n$ in $j_1 \dots j_m$. By equations (7) and (8), if $i_1, \dots, i_{m-1} < n$,

$$\begin{aligned}\partial_n v_{ni_1\cdots i_{m-1}} - \sum_{s=1}^{m-1}\sum_{\ell<n}\Gamma^\ell_{ni_s} v_{\ell ni_1\cdots\widehat{i_s}\cdots i_{m-1}} - \sum_{j,k<n} g^{jk}\left(\Gamma^n_{jk} v_{ni_1\cdots i_{m-1}} + \sum_{s=1}^{m-1}\Gamma^n_{ji_s} v_{nki_1\cdots\widehat{i_s}\cdots i_{m-1}}\right)\\ = -\sum_{j,k<n} g^{jk}\left(\partial_j v_{ki_1\cdots i_{m-1}} - \sum_{\ell<n}\Gamma^\ell_{jk} v_{\ell i_i\cdots i_{m-1}} - \sum_{s=1}^{m-1}\sum_{\ell<n}\Gamma^\ell_{ji_s} v_{\ell ki_1\cdots\widehat{i_s}\cdots i_{m-1}}\right). \quad (9)\end{aligned}$$

Notice that the right side of (9) is known, so it gives a system of first-order linear ODEs. Given the initial values $u_{ni_1\cdots i_{m-1}}(x', 0) = v_{ni_1\cdots i_{m-1}}(x', 0)$, there exists a unique solution to (9). Thus we obtain continuous $v_{ni_1\cdots i_{m-1}}$ with $i_1, \dots, i_{m-1} < n$ near $\partial M$. In particular, $v_{ni_1\cdots i_{m-1}}(x', x^n)$ depends smoothly on $x'$, the first $n-1$ variables.

By differentiating (9) repeatedly with respect to $x^n$, we get that $\partial^s_n v_{ni_1\cdots i_{m-1}}(x', x^n)$, $s \ge 0$, are continuous in $x^n \ge 0$ and smooth with respect to $x'$. Moreover, by (9) and the fact that $u$ is solenoidal we carry out an induction on $s$, so

$$\begin{aligned}&\partial^s_n v_{ni_1\cdots i_{m-1}}(x', 0)\\ &\quad = G^s_{i_1\cdots i_{m-1}}\big(\partial^\ell_n v_{nj_1\cdots j_{m-1}}, \partial^\ell_n v_{j_1\cdots j_m}, \partial^\ell_n\partial_k v_{j_1\cdots j_m}; \ell < s; j_1, \dots, j_{m-1}, j_m, k < n\big)(x', 0)\\ &\quad = G^s_{i_1\cdots i_{m-1}}\big(\partial^\ell_n u_{nj_1\cdots j_{m-1}}, \partial^\ell_n u_{j_1\cdots j_m}, \partial^\ell_n\partial_k u_{j_1\cdots j_m}; \ell < s; j_1, \dots, j_{m-1}, j_m, k < n\big)(x', 0)\\ &\quad = \partial^s_n u_{ni_1\cdots i_{m-1}}(x', 0)\end{aligned}$$

for all $s \ge 0$; i.e., $\partial^s_n v_{ni_1\cdots i_{m-1}}$ are consistent with $\partial^s_n u_{ni_1\cdots i_{m-1}}$ at $(x', 0)$.

Next by induction on the number of appearances of $n$ and repeatedly using equations (7) and (8), one can get unique

$$v_{ni_1\cdots i_{m-1}}, \quad v_{nni_1\cdots i_{m-2}}, \quad \dots, \quad v_{n\cdots ni_1}, \quad v_{n\cdots n},$$

which together with their normal derivatives with respect to $x^n$ of all orders, are continuous (smooth with respect to $x'$) and consistent with the corresponding $\partial^m_n u_{j_1\cdots j_m}$ at $(x', 0)$. Therefore we get a smooth solenoidal $m$-tensor

$$\tilde{u} = \begin{cases} u & \text{on } \overline{\Omega},\\ v & \text{on } U\setminus\overline{\Omega}. \end{cases}$$

□

*Proof of Proposition 3.4.* There exist two precompact open neighborhoods $V, U$ of $\Omega$ which satisfy

$$\Omega \subset \overline{\Omega} \subset V \subset \overline{V} \subset U \subset \overline{U} \subset \mathcal{M}.$$

Given $u \in C^\infty(S^m_{\mathrm{sol}}(T^*\Omega))$, by Lemma 3.5, we can extend $u$ to get $u_V \in C^\infty(S^m_{\mathrm{sol}}(T^*V))$ with $u_V|_{\overline{\Omega}} = u$. Then we extend $u_V$ to a smooth $m$-tensor $w$ on $\mathcal{M}$ with $\operatorname{supp} w \subset U$. Let $f = \delta w$ and $D = U \backslash \overline{\Omega}$ open, so $\operatorname{supp} f \subset U \backslash V \subset D$.

Similar to the perturbation-of-metrics argument in the proof of Proposition 3.2, given any $\epsilon > 0$ and $K \geq 2$, there is a smooth metric $\tilde{g}$ with $\|\tilde{g} - g\|_{C^K} < \epsilon$ and $\tilde{g}|_V = g$ so that $(D, \tilde{g})$ does not have nontrivial Killing tensor fields. Now if $m$ is even, the only Killing $(m-1)$-tensor field on $(D, \tilde{g})$ is $v = 0$. Then

$$(v, f)_D = (0, f)_D = 0.$$

If $m$ is odd, Killing $(m-1)$-tensor fields on $(D, \tilde{g})$ are of the form $v = c\tilde{g}^{(m-1)/2}|_D$. Thus we can extend $v$ to $v = c\tilde{g}^{(m-1)/2}|_U$, which is also a Killing tensor field in $\Omega$. By Green's formula,

$$(v, f)_D = (v, \delta w)_D = -(dv, w)_D + (v, j_\mu w)_{\partial D} = -(v, j_\nu u)_{\partial\Omega} = -(v, \delta u)_\Omega - (dv, u)_\Omega = 0.$$

since $\delta u = 0$ and $dv = 0$ in $\Omega$. Here $\mu = -\nu$ is the unit outward normal vector on $\partial D$ and

$$(j_\mu w)_{i_1 i_2 \cdots i_{m-1}} = w_{i_1 i_2 \cdots i_{m-1} j} \mu^j.$$

Now by [Delay 2012, Theorem 1.3], there exist $u_D \in C^\infty(S^m(T^*\mathcal{M}))$ with $\operatorname{supp} u_D \subset U \backslash \Omega$ such that $\delta u_D = -f$. It is not difficult to check that the symmetric differentiation $d$ satisfies the kernel restriction condition (KRC) and the asymptotic Poincaré inequality (API) of [Delay 2012]. We define $\mathcal{E}u = w + u_D$. Then $\delta \mathcal{E} u = \delta w + \delta u_D = f - f = 0$; i.e., $\mathcal{E}u \in C^\infty(S^m_{U,\mathrm{sol}}(T^*_{\tilde{g}}\mathcal{M}))$. Moreover, $\mathcal{E}u|_\Omega = u$.

The argument above gives a construction for compactly supported smooth solenoidal extensions. One can further check that the extension can be constructed in a stable way. In view of the ODEs (9), the solution is controlled by the initial value and the nonhomogeneous term on the right side under Sobolev norms; see, e.g., [Han 2011]. By induction on the number of appearances of $n$ and repeatedly differentiating (9), we have that

$$\|u_V\|_{H^1(V\backslash\overline{\Omega})} \leq C\left(\|j_\mu u\|_{H^{k_1}(\partial\Omega)} + \sum_{i_s < n} \|(u_V)_{i_1 \cdots i_m}\|_{H^{k_2}(V\backslash\overline{\Omega})}\right)$$

for some $k_1, k_2 \geq 1$. Note that in boundary normal coordinates $\mu = -\partial_n$, and we have full freedom to control the elements $(u_V)_{i_1 \cdots i_m}$, with $i_s < n$ for all $1 \leq s \leq m$, by $u|_\Omega$ due to the fact that $\delta$ is an underdetermined elliptic operator. Thus

$$\|u_V\|_{H^1(V\backslash\overline{\Omega})} \leq C\|u\|_{H^k(\Omega)}$$

for some integer $k \geq 2$. Then $\|w\|_{H^1(U)} \leq C\|u\|_{H^k(\Omega)}$ by extending $u_V$ to $w$ in a stable way.

Next we control the $L^2$ norm of $u_D$. Roughly speaking, $u_D$ is the symmetric differentiation of some smooth $(m-1)$-tensor $p$, multiplied by a smooth nonnegative weight which vanishes exponentially at the boundary of $D$; concretely $u_D = \psi^2\phi^2 dp$ with $\phi$ a boundary-defining function on $D$ and $\psi$ vanishes exponentially at the boundary $\partial D$. By [Delay 2012, Lemma 10.2], $\|p\|_{H^2_{\phi,\psi}(D)} \leq C\|\psi^{-2}\delta w\|_{L^2_\psi(D)}$,

where $H^2_{\phi,\psi}$ and $L^2_{\psi}$ are some weighted Sobolev spaces; see [Delay 2012] for more details. Then one can check that the following inequality with unweighted Sobolev norms holds:

$$\|u_D\|_{L^2(D)} \leq C\|w\|_{H^1(U)}.$$

Now we combine the estimates above to obtain

$$\|\mathcal{E}u\|_{L^2(\mathcal{M})} \leq C_1\big(\|w\|_{L^2(U)} + \|u_D\|_{L^2(D)}\big) \leq C_2\|w\|_{H^1(U)} \leq C\|u\|_{H^k(\Omega)}$$

for some $C > 0$ independent of $u$. Since $C^\infty(S^m_{\mathrm{sol}}(T^*\Omega))$ is dense in $H^k(S^m_{\mathrm{sol}}(T^*\Omega))$ under the $H^k$ norm, we can extend $\mathcal{E}$ to a bounded map from $H^k$ to $L^2$ with the same properties, which completes the proof. □

**Remark 3.6.** We expect that the $L^2$ norm of $\mathcal{E}u$ can be bounded by the $L^2$ norm of $u|_\Omega$ through sharper estimates, similar to the result under the $L^2$ setting in the previous subsection. However, the $H^k$ space is enough for carrying out the argument under the smooth setting in the next section; see Lemma 4.3.

## 4. Surjectivity of the normal operator $I_m^* I_m$

Since $M$ is simple we can consider an extension $\widetilde{M}$ of $M$ which is open ($\widetilde{M} = \widetilde{M}^{\mathrm{int}}$) and whose compact closure is also simple. It is well known that the normal operator $N = I_m^* I_m$ is a pseudodifferential operator of order $-1$ on $\widetilde{M}$; see, for example, [Sharafutdinov 1994; Stefanov and Uhlmann 2004; 2008; Sharafutdinov et al. 2005]. Below is a lemma that, roughly speaking, gives a right parametrix for $N$ on the space of solenoidal tensor fields. The proof is similar to [Sharafutdinov et al. 2005, Theorem 3.1].

**Lemma 4.1.** *Let $S$ be a parametrix for the operator $\delta d$. There exists a pseudodifferential operator $Q$ of order 1 on the bundle of symmetric $m$-tensor fields $S^m(T^*\widetilde{M})$ such that*

$$E = NQ + dS\delta + K, \tag{10}$$

*where $E$ is the identity operator and $K$ is a smoothing operator.*

*Proof.* Let $\lambda(\xi)$ be the principal symbol of the pseudodifferential operator $N$ and

$$S^m_\xi(T^*_x\widetilde{M}) = \{u \in S^m(T^*_x\widetilde{M}) : j_\xi u = 0\},$$

where $j_\xi = -i\sigma_p(\delta) : S^m(T^*_x\widetilde{M}) \to S^{m-1}(T^*_x\widetilde{M})$. By [Sharafutdinov 1994, Theorem 2.12.1],

$$\lambda(\xi) : S^m_\xi(T^*_x\widetilde{M}) \to S^m_\xi(T^*_x\widetilde{M})$$

is an isomorphism for $\xi \neq 0$. Thus there exists $p(\xi)$ such that $\lambda(\xi)p(\xi) = \mathrm{Id}$ on $S^m_\xi(T^*_x\widetilde{M})$. Namely, we can find some pseudodifferential operator $P$ of order 1 such that on $S^m_\xi(T^*_x\widetilde{M})$,

$$NP = E - B$$

for some operator $B$ of order $-1$. Now multiplying both sides by the "solenoidal projection" $E - dS\delta$, which is of order 0, one has

$$NP(E - dS\delta) = E - dS\delta - R \tag{11}$$

defined on $S^m(T^*\widetilde{M})$.

Then we multiply both sides of (11) by $\delta$ to get $\delta R = R'$ with $R'$ some smoothing operator. Let $C = \sum_{k=0}^{\infty} R^k$, which is a pseudodifferential operator of order 0 and a parametrix for $E - R$. Write (11) as

$$NP(E - dS\delta) + dS\delta = E - R,$$

and multiply both sides by $C$ to get

$$NP(E - dS\delta)C + dS\delta + dS\delta \sum_{k=1}^{\infty} R^k = (E - R)C = E + R'',$$

with $R''$ a smoothing operator. Since $\delta R$ is smoothing, $dS\delta \sum_{k=1}^{\infty} R^k$ is smoothing too. We arrive at the equation

$$NP(E - dS\delta)C + dS\delta + K = E,$$

where $K$ is a smoothing operator. Denote $P(E - dS\delta)C$ by $Q$ (note that one can make $Q$ properly supported). Then we get (10), which finishes the proof. □

Let $U$ be a small open neighborhood of $M$ in $\widetilde{M}$. Denote the restriction operator from $\widetilde{M}$ to $M$ by $r_M$. Then the following holds:

**Lemma 4.2.** *Suppose $M$ is a compact simple Riemannian manifold, and assume $I_m$ is s-injective on $C^\infty(S^m(T^*M))$. Then the operator*

$$r_M N : H_c^{-1}(S^m(T^*\widetilde{M})) \to L^2(S^m_{\mathrm{sol}}(T^*M))$$

*is surjective.*

Note that elements in $H_c^{-1}(S^m(T^*\widetilde{M}))$ are defined in the sense of distributions, which are compactly supported in $\widetilde{M}$.

*Proof.* We adopt the approach of [Dairbekov and Uhlmann 2010] for showing the surjectivity of $N$ on 1-forms. By Lemma 4.1,

$$NQu = u + Ku$$

for all $u \in L^2_c(S^m_{\mathrm{sol}}(T^*\widetilde{M}))$ with $K$ a smoothing operator on $\widetilde{M}$. Since the simplicity is stable under small $C^2$-perturbations of the metric $g$, by Proposition 3.2, we perturb the metric of $\widetilde{M} \setminus \overline{M}$ a little bit (still denoted by $g$) so that under the new metric $\widetilde{M}$ is still simple and there exists a bounded operator $\mathcal{E} : L^2(S^m_{\mathrm{sol}}(T^*M)) \to L^2(S^m_{U,\mathrm{sol}}(T^*\widetilde{M}))$ such that on $L^2(S^m_{\mathrm{sol}}(T^*M))$,

$$r_M NQ\mathcal{E} = E + r_M K\mathcal{E}.$$

Since $K$ is a smoothing operator, $r_M K\mathcal{E}$ is compact on $L^2(S^m_{\mathrm{sol}}(T^*M))$, which implies that $E + r_M K\mathcal{E}$ has closed range and finite codimension. Thus we have $r_M NQ\mathcal{E} : L^2(S^m_{\mathrm{sol}}(T^*M)) \to L^2(S^m_{\mathrm{sol}}(T^*M))$ has closed range and finite codimension. By the inclusion relation

$$r_M NQ\mathcal{E}(L^2(S^m_{\mathrm{sol}}(T^*M))) \subset r_M N(H_c^{-1}(S^m(T^*\widetilde{M}))) \subset L^2(S^m_{\mathrm{sol}}(T^*M)),$$

the intermediate space $r_M N(H_c^{-1}(S^m(T^*\widetilde{M})))$ is also closed in $L^2(S^m_{\mathrm{sol}}(T^*M))$. Thus it suffices to show that the adjoint $(r_M N)^*$ is injective, which will imply the surjectivity of $r_M N$.

For $L^2$ symmetric $m$-tensor fields, we have the decomposition

$$L^2(S^m(T^*M)) = L^2(S^m_{\mathrm{sol}}(T^*M)) \oplus L^2(S^m_P(T^*M)), \tag{12}$$

where $L^2(S^m_P(T^*M))$ is the potential part. Thus the dual operator of $r_M N$ is

$$(r_M N)^* : L^2(S^m_{\mathrm{sol}}(T^*M)) \to (H_c^{-1}(S^m(T^*\widetilde{M})))^*.$$

For $u \in L^2(S^m_{\mathrm{sol}}(T^*M))$ and $v \in H_c^{-1}(S^m(T^*\widetilde{M}))$, if we denote by $\mathcal{E}_0 u$ the extension of $u$ to $\widetilde{M}$ by zero (note that generally $\mathcal{E}_0 u$ is not solenoidal on $\widetilde{M}$), we have

$$((r_M N)^* u, v) = (u, r_M N v) = (\mathcal{E}_0 u, N v) = (N\mathcal{E}_0 u, v),$$

i.e., $(r_M N)^* = N\mathcal{E}_0$.

Therefore given $u \in L^2(S^m_{\mathrm{sol}}(T^*M))$, if $N\mathcal{E}_0 u = 0$, then

$$0 = (N\mathcal{E}_0 u, \mathcal{E}_0 u) = \|I_m \mathcal{E}_0 u\|^2_{L^2(\partial_+ S\widetilde{M})} \quad \Longrightarrow \quad I_m \mathcal{E}_0 u = 0.$$

Since $\mathcal{E}_0 u = 0$ outside $M$ and $\widetilde{M}$ is simple, this implies

$$I_m u = 0.$$

By [Sharafutdinov et al. 2005, Theorem 1.1], $u$ is smooth and $\delta u = 0$. The s-injectivity assumption implies $u = 0$. This completes the proof of the lemma. □

Next we prove the lemma in the smooth setting:

**Lemma 4.3.** *Suppose $M$ is a compact simple Riemannian manifold, and assume $I_m$ is s-injective on $C^\infty(S^m(T^*M))$. Then the operator*

$$r_M N : C_c^\infty(S^m(T^*\widetilde{M})) \to C^\infty(S^m_{\mathrm{sol}}(T^*M))$$

*is surjective.*

*Proof.* By Lemma 4.1,

$$NQu = u + Ku$$

for all $u \in C_c^\infty(S^m_{\mathrm{sol}}(T^*\widetilde{M}))$ with $K$ a smoothing operator on $\widetilde{M}$. Since the simplicity is stable under small $C^2$-perturbations of the metric $g$, by Proposition 3.4, we perturb the metric of $\widetilde{M}\setminus\overline{M}$ a little bit (still denoted by $g$) so that under the new metric $\widetilde{M}$ is still simple and there exists a bounded operator $\mathcal{E} : H^k(S^m_{\mathrm{sol}}(T^*M)) \to L^2(S^m_{U,\mathrm{sol}}(T^*\widetilde{M}))$ for some integer $k \geq 2$ with $\mathcal{E}(C^\infty(S^m_{\mathrm{sol}}(T^*M))) \subset C^\infty(S^m_{U,\mathrm{sol}}(T^*\widetilde{M}))$ such that on $H^k(S^m_{\mathrm{sol}}(T^*M))$,

$$r_M NQ\mathcal{E} = E + r_M K\mathcal{E}.$$

Now the argument of [Dairbekov and Uhlmann 2010, Lemma 2.2] can be applied to tensors of any order to finish the proof. □

**Remark 4.4.** One can actually prove Lemmas 4.2 and 4.3 just by applying Lemma 3.5. Given a smooth solenoidal tensor $u$ on $M$, by Lemma 3.5 we first extend it to a smooth solenoidal tensor $\tilde{u}$ on an arbitrarily small open neighborhood $U$; then we extend $\tilde{u}$ smoothly to $\widetilde{M}$ with compact support, denoted by $\mathcal{E}u$. Note that generally $\mathcal{E}u$ is not solenoidal. Since the Schwartz kernel of the parametrix $S$ of $\delta d$ is smooth away from the diagonal $\Delta_{\widetilde{M}\times\widetilde{M}}$, we can choose $S$ to make the support of its Schwartz kernel sufficiently close to $\Delta_{\widetilde{M}\times\widetilde{M}}$ so that $dS\,\delta\mathcal{E}u = 0$ in an open neighborhood of $M$. This implies that $r_M\, dS\,\delta\mathcal{E}u = 0$, i.e., $r_M NQ\mathcal{E}u = u + r_M K\mathcal{E}u$. It also works for $L^2$ solenoidal tensors.

On the other hand, the original proof of [Dairbekov and Uhlmann 2010, Lemma 2.2] uses the existence of compactly supported solenoidal extensions of solenoidal 1-forms one more time at the very end to show that the adjoint $(r_M N)^*$ is injective. However, one can also avoid this. Notice that given a 1-form $f$ in the kernel of $(r_M N)^*$, by [Dairbekov and Uhlmann 2010, equation (2.33)], $f = dp$ for some distribution $p$ on $\widetilde{M}$ with $\operatorname{sing\,supp} p \subset \partial M$ and $p|_{\partial\widetilde{M}} = 0$. Moreover, since $\operatorname{supp} f \subset M$, we have $dp = 0$ outside $M$. As $p$ is smooth outside $M$ and $p = 0$ on $\partial\widetilde{M}$, strict convexity of $\partial M$ implies $p \equiv 0$ in $\widetilde{M}\setminus M$. Now given a smooth solenoidal 1-form $u$ on $M$, by Lemma 3.5 let $\mathcal{E}u$ be the smooth compactly supported extension of $u$ to $\widetilde{M}$ which is solenoidal in a small open neighborhood ($\neq \widetilde{M}$) of $M$. Since the supports of $\delta\mathcal{E}u$ and $p$ are disjoint, we have

$$(f, \mathcal{E}u) = (dp, \mathcal{E}u) = (p, \delta\mathcal{E}u) = 0,$$

which implies that $f = 0$, i.e., $(r_M N)^*$ has trivial kernel. The argument works for tensors of arbitrary rank.

At this point, we see that one can prove the surjectivity of $r_M N$ just using Lemma 3.5, without the need of knowing the generic absence of nontrivial Killing tensors [Kruglikov and Matveev 2016]. However, a perturbation of the metric seems still necessary so far for the proof of the existence of *compactly supported* solenoidal extensions, and Propositions 3.2 and 3.4 may find their applications in other areas.

## 5. Analysis of the adjoint $I_m^*$

Before proving the main result, we need to extend the definition of the geodesic ray transform $I_m$ so that it acts on negative Sobolev spaces. To this end, we will study the regularity property of the adjoint of the geodesic ray transform, $I_m^*$.

As discussed in the Introduction, given $M$ a compact nontrapping manifold with strictly convex boundary, the operator $I_m^* : C_\alpha^\infty(\partial_+ SM) \to C^\infty(S^m(T^*M))$ is the product of two operators, i.e., $I_m^* = L_m \circ I^*$. We instead study the regularity properties of $I^*$ and $L_m$. We start with the latter.

**Lemma 5.1.** *Given a compact Riemannian manifold $M$ (with or without boundary), the operator*

$$L_m : H^k(SM) \to H^k(S^m(T^*M))$$

*is bounded for every integer $k \geq 0$.*

*Proof.* Our purpose is to show that there exists a constant $C > 0$ such that for any $w \in H^k(SM)$, the following holds:

$$\|L_m f\|_{H^k} \leq C \|f\|_{H^k}. \tag{13}$$

Since $M$ is compact, by a partition of unit, it suffices to show the above inequality in local charts. Let $U$ be a domain in $SM$ with local coordinate system $(z^1, \dots, z^{2n-1})$. We assume supp $f \subset U$. Let $V$ be a domain in $M$ with local coordinate system $(x^1, \dots, x^n)$, and $\psi$ be a smooth function with support in $V$. We will show

$$\|\psi L_m f\|_{H^k(S^m(T^*V))} \le C\|f\|_{H^k(U)}.$$

By the definition of the $H^k$ norm of tensors, we only need to show the above inequality is true for each component of the tensor.

We start with $f \in C^\infty(SM)$ with support in $U$; then $L_m f$ is also smooth. Let $J = (j_1 \cdots j_m)$ and $\xi^J := \xi^{j_1} \cdots \xi^{j_m}$. Then

$$\begin{aligned}
&D_x^\alpha\big[\psi(x) L_m f(x)^J\big] \\
&\quad = D_x^\alpha\left[\psi(x)\int_{S_xM} f(x,\xi)\xi^J \, d\Omega_x(\xi)\right] \\
&\quad = D_x^\alpha\left[\psi(x)\int_{S^{n-1}} f(x,\xi(x,\eta))\xi^J(x,\eta) P(x,\eta)\, d\Omega(\eta)\right] \\
&\quad = \sum_{\alpha_1+\alpha_2+\alpha_3=\alpha} D_x^{\alpha_1}\psi(x)\int_{S^{n-1}} D_x^{\alpha_2} f(x,\xi(x,\eta))\cdot D_x^{\alpha_3}\big[\xi^J(x,\eta)P(x,\eta)\big]\, d\Omega(\eta) \\
&\quad = \sum_{\alpha_1+\alpha_2+\alpha_3=\alpha} D_x^{\alpha_1}\psi(x)\int_{S_xM} D_x^{\alpha_2} f(x,\xi)\cdot D_x^{\alpha_3}\big[\xi^J P(x,\eta(x,\xi))\big]\cdot P'(x,\xi)\, d\Omega_x(\xi). \quad (14)
\end{aligned}$$

Here $P$ and $P'$ are corresponding Jacobians.

For $|\alpha| \le k$, according to (14),

$$\begin{aligned}
\|D_x^\alpha[\psi(x) L_m f(x)_J]\|^2_{L^2(V)} &\le \sum_{\beta\le\alpha} C_{\beta,\alpha}\int_V\int_{S_xM} |D_x^\beta f(x,\xi)|^2\, d\Omega_x(\xi)\, dx \\
&\le \sum_{|\gamma|\le|\alpha|} C_{\gamma,\alpha}\int_U |D_z^\gamma f(z)|^2\, dz \le C\|f\|^2_{H^k(U)}.
\end{aligned}$$

Thus the estimate (13) is proved when $w \in C^\infty(SM)$.

For $f \in H^k(SM)$, since $C^\infty(SM)$ is dense in $H^k(SM)$, by an approximation argument, it is easy to show that $L_m f \in H^k(S^m(T^*M))$ and the estimate (13) holds too. This proves the lemma. □

Now we turn to the analysis of the operator $I^*$, which basically is an invariant extension, along the geodesic flow, of functions on $\partial_+SM$ to functions on $SM$. It is well known that given $\varphi \in C^\infty(\partial_+SM)$, $\varphi^\sharp = I^*(\varphi)$ is not necessarily in $C^\infty(SM)$. The following subspace of $C^\infty(\partial_+SM)$ has already been considered in the Introduction:

$$C^\infty_\alpha(\partial_+SM) := \big\{\varphi \in C^\infty(\partial_+SM) : \varphi^\sharp \in C^\infty(SM)\big\}.$$

In particular, by [Pestov and Uhlmann 2005, Lemma 1.1], if $M$ is compact nontrapping with strictly convex boundary,

$$C^\infty_\alpha(\partial_+SM) = \big\{\varphi \in C^\infty(\partial_+SM) : A\varphi \in C^\infty(\partial SM)\big\}$$

where

$$A\varphi(x,\xi)=\begin{cases}\varphi(x,\xi), & (x,\xi)\in\partial_+SM,\\ \varphi(\gamma_{x,\xi}(-\tau(x,-\xi)),\dot\gamma_{x,\xi}(-\tau(x,-\xi))), & (x,\xi)\in\partial_-SM.\end{cases}$$

Since $A\varphi$ is smooth in both $(\partial_+SM)^{\text{int}}$ and $(\partial_-SM)^{\text{int}}$, the singularities can only come from $S(\partial M)$. We introduce the space $H^k_\alpha(\partial_+SM)$, $k\ge 0$, to be the completion of $C^\infty_\alpha(\partial_+SM)$ under the $H^k$ norm. Obviously $H^0_\alpha(\partial_+SM)=L^2(\partial_+SM)$. It is easy to show that $C^\infty_c((\partial_+SM)^{\text{int}})\subset C^\infty_\alpha(\partial_+SM)$ (this is from the fact that $\partial_+SM$ is compact and the boundary $\partial M$ is strictly convex), which implies that $H^k_0(\partial_+SM)\subset H^k_\alpha(\partial_+SM)$.

**Lemma 5.2.** *Given a compact nontrapping manifold $M$ with strictly convex boundary, the operator*

$$I^*: H^k_\alpha(\partial_+SM)\to H^k(SM)$$

*is bounded for any integer $k\ge 0$.*

*Proof.* The idea is similar to the proof of Lemma 5.1. First we consider the case $\varphi\in C^\infty_\alpha(\partial_+SM)$; thus $\varphi^\sharp\in C^\infty(SM)$. Let $U$ be a domain in $\partial_+SM$ with local coordinate systems $(y^1,\dots,y^{2n-2})$. We assume $\operatorname{supp}\varphi\subset U$. Let $V$ be a domain in $SM$ with local coordinate systems $(z^1,\dots,z^{2n-1})$, and $\psi$ be a smooth function with support in $V$. Since $M$ is compact, it suffices to show

$$\|\psi\varphi^\sharp\|_{H^k(V)}\le C\|\varphi\|_{H^k(U)}.$$

Since

$$D^\alpha_z[\psi(z)\varphi^\sharp(z)]=\sum_{\beta+\gamma=\alpha}D^\gamma_z\psi(z)\cdot D^\beta_z\varphi^\sharp(z),$$

we obtain that for $|\alpha|\le k$,

$$\left\|D^\alpha_z[\psi(z)\varphi^\sharp(z)]\right\|^2_{L^2(V)}\le\sum_{\beta\le\alpha}C_{\beta,\alpha}\int_V|D^\beta_z\varphi^\sharp(z)|^2\,dz.$$

Now let $D=\{(y,t): y\in\partial_+SM,\ 0\le t\le\tau(y)\}$ be a closed domain in $\partial_+SM\times\mathbb{R}$. Define the map $\Psi: D\to SM$ by $z=\Psi(y,t)=\big(\gamma_y(t),\dot\gamma_y(t)\big)$. By [Sharafutdinov 1994, Lemma 4.2.2],

$$\begin{aligned}\int_V|D^\beta_z\varphi^\sharp(z)|^2\,dz&\le\sum_{|\sigma|+s=|\beta|}C_{\beta,\sigma,s}\int_U\int_0^{\tau(y)}\left|D^\sigma_yD^s_t\varphi^\sharp(z(y,t))\right|^2\left|\big\langle\xi(y),\nu(x(y))\big\rangle\right|dt\,dy\\&=\sum_{|\sigma|=|\beta|}C_{\beta,\sigma}\int_U\int_0^{\tau(y)}|D^\sigma_y\varphi^\sharp(y,t)|^2\,dt\,d\mu(y)\quad(\text{since } D^s_tD^\sigma_y\varphi^\sharp=D^\sigma_yD^s_t\varphi^\sharp)\\&=\sum_{|\sigma|=|\beta|}C_{\beta,\sigma}\int_U\tau(y)|D^\sigma_y\varphi(y)|^2\,d\mu(y)\\&\le\sum_{|\sigma|=|\beta|}C'_{\beta,\sigma}\int_U|D^\sigma_y\varphi(y)|^2\,d\mu(y)\le C\|\varphi\|^2_{H^k(U)}.\end{aligned}$$

Therefore, $\|\varphi^\sharp\|_{H^k(SM)}\le C\|\varphi\|_{H^k(\partial_+SM)}$ for $\varphi\in C^\infty_\alpha(\partial_+SM)$.

If $\varphi \in H^k_\alpha(\partial_+ SM)$, since $C^\infty_\alpha(\partial_+ SM)$ is dense in $H^k_\alpha(\partial_+ SM)$, by an approximation argument, it is easy to show that $\varphi^\sharp \in H^k(SM)$ and the operator $I^*$ is bounded, which proves the lemma. □

Combining the two lemmas above, we obtain the desired regularity property of $I^*_m$.

**Proposition 5.3.** *Given a compact nontrapping Riemannian manifold $M$ with strictly convex boundary, the adjoint operator of the geodesic ray transform on symmetric $m$-tensors*

$$I^*_m = L_m \circ I^* : H^k_\alpha(\partial_+ SM) \to H^k(S^m(T^*M))$$

*is bounded for any integer $k \geq 0$.*

Now we can extend the definition of the geodesic ray transform so that it acts on $(H^k(S^m(T^*M)))^*$ (the dual space is with respect to the $L^2$ inner product) for integers $k \geq 1$. Let $u \in (H^k(S^m(T^*M)))^*$ and $\varphi \in H^k_\alpha(\partial_+ SM)$. We define $I_m u$ in the sense of distributions:

$$(I_m u, \varphi) := (u, I^*_m \varphi). \tag{15}$$

By Proposition 5.3, the right-hand side of (15) is well-defined. We derive the following corollary:

**Corollary 5.4.** *Given $M$, a compact nontrapping manifold with strictly convex boundary, the operator*

$$I_m : (H^k(S^m(T^*M)))^* \to (H^k_\alpha(\partial_+ SM))^*$$

*defined by (15) is bounded.*

Here the dual space $(H^k_\alpha(\partial_+ SM))^*$ is also with respect to the $L^2$ inner product. Note $H^k_0(\partial_+ SM) \subset H^k_\alpha(\partial_+ SM)$; thus $(H^k_\alpha(\partial_+ SM))^* \subset H^{-k}(\partial_+ SM)$. On the other hand, since $C^\infty(S^m(T^*M))$ is dense in $H^k(S^m(T^*M))$ under the $H^k$-norm, it is clear that $H^{-k}_c(S^m(T^*M^{\text{int}})) \subset (H^k(S^m(T^*M)))^*$; we will use the weaker map in the next section:

$$I_m : H^{-k}_c(S^m(T^*M^{\text{int}})) \to H^{-k}(\partial_+ SM). \tag{16}$$

## 6. Proof of Theorem 1.2

Now we are in a position to prove our main theorem. We start by showing that (1), (2) and (3) are equivalent.

*Proof.* (1) ⇒ (2): Since $M$ is simple, given $u \in L^2(S^m_{\text{sol}}(T^*M))$, by Lemma 4.2, there exists $v \in H^{-1}_c(S^m(T^*\widetilde{M}))$ such that $r_M I^*_m I_m v = u$. Then (16) implies the existence of some $\tilde{\varphi} = I_m v \in H^{-1}(\partial_+ S\widetilde{M})$ such that $u = r_M I^*_m \tilde{\varphi}$. For $w \in H^1_0(S^m(T^*M))$, we define the distribution $\varphi$ acting on $I_m(H^1_0(S^m(T^*M)))$ by

$$(\varphi, I_m w) := (\tilde{\varphi}, I_m \tilde{w}) = (I^*_m \tilde{\varphi}, \tilde{w}),$$

where $\tilde{w} \in H^1_0(S^m(T^*\widetilde{M}))$ is the extension of $w$ which is zero outside $M$. We claim that there exists $C > 0$ such that

$$|(\varphi, I_m w)| \leq C \|I_m w\|_{H^1}$$

for all $w \in H_0^1(S^m(T^*M))$. Assuming the claim, note that $I_m w \in H_0^1(\partial_+ SM)$ and by the Hahn–Banach theorem, $\varphi$ can be extended to a bounded linear functional on $H_0^1(\partial_+ SM)$, still denoted by $\varphi$, i.e., $\varphi \in H^{-1}(\partial_+ SM)$. By the definition of $\varphi$,

$$|(\varphi, I_m w)| = |(\tilde{\varphi}, I_m \tilde{w})| \leq C \|I_m \tilde{w}\|_{H^1}.$$

Therefore to prove the claim, it suffices to show that

$$\|I_m \tilde{w}\|_{H^1(\partial_+ S\tilde{M})} \leq C \|I_m w\|_{H^1(\partial_+ SM)} \tag{17}$$

for some $C > 0$.

Assume at this point that inequality (17) holds and let us continue with the proof. Now $\varphi \in H^{-1}(\partial_+ SM)$ is well-defined. Let $w \in H_0^1(S^m(T^*M))$, and let $\tilde{w}$ be the extension of $w$ into $\tilde{M}$ which is zero outside $M$, so $\tilde{w} \in H_0^1(S^m(T^*\tilde{M}))$. Then

$$(r_M I_m^* \tilde{\varphi}, w) = (I_m^* \tilde{\varphi}, \tilde{w}) = (\tilde{\varphi}, I_m \tilde{w}) = (\varphi, I_m w) = (I_m^* \varphi, w).$$

Thus $u = r_M I_m^* \tilde{\varphi} = I_m^* \varphi$. (The choice of $\varphi$ is not unique.)

(2) $\Rightarrow$ (3): Given $u \in L^2(S^m_{\mathrm{sol}}(T^*M))$, by the assumption, there is $\varphi \in H^{-1}(\partial_+ SM)$ such that $u = I_m^* \varphi$. Since $I_m^* = L_m \circ I^*$, we define $f = I^* \varphi$; then $f \in H^{-1}(SM)$ and $u = L_m f$. Furthermore, given $h \in H_0^2(SM)$,

$$(Xf, h) = (f, -Xh) = (I^* \varphi, -Xh) = (\varphi, -I(Xh)) = 0,$$

i.e., $Xf = 0$.

(3) $\Rightarrow$ (1): Assume $I_m u = 0$ for some $u \in C^\infty(S^m_{\mathrm{sol}}(T^*M))$. Then it is well known that there exists $h \in C^\infty(SM)$ with $h|_{\partial SM} = 0$ such that

$$Xh = -\ell_m u.$$

Moreover, by [Sharafutdinov 2002, Lemma 2.3] there exists $p \in C^\infty(S^{m-1}(T^*M))$ with $p|_{\partial M} = 0$ such that $u|_{\partial M} = dp|_{\partial M}$. When $m = 0$, this just means $u|_{\partial M} = 0$. Calculations in local coordinates show that $X(\ell_{m-1} p) = \ell_m dp$. Thus we obtain

$$X(h + \ell_{m-1} p) = -\ell_m (u - dp),$$

with $(h + \ell_{m-1} p)|_{\partial SM} = 0$.

Under the projection $\pi : SM \to M$, the pullback of the unit normal vector $\nu$ to $\partial M$ is the unit normal vector $\mu$ to $\partial SM$, and in local coordinates

$$X = \xi^i \frac{\partial}{\partial x^i} - \Gamma^i_{jk} \xi^j \xi^k \frac{\partial}{\partial \xi^i},$$

where $\Gamma^i_{jk}$ are the Christoffel symbols. By taking the boundary normal coordinates $(x', x^n)$ near $x \in \partial M$ (so $\nu(x) = \mu(x, \xi) = \partial/\partial x^n$), together with the fact that $(h + \ell_{m-1} p)|_{\partial SM} = 0$, we obtain that for $(x, \xi) \in \partial SM$,

$$0 = -\ell_m (u - dp)(x, \xi) = X(h + \ell_{m-1} p)(x, \xi) = \xi^n \partial_{x^n} (h + \ell_{m-1} p)(x, \xi).$$

The first equality comes from the fact $u - dp|_{\partial M} = 0$. Thus $\partial_\mu(h + \ell_{m-1} p)(x, \xi) = 0$ for all $\xi \notin S_x \partial M$. But since $h$ and $p$ are smooth, and the measure of $S_x \partial M$ is zero on $S_x M$, we get $\partial_\mu(h+\ell_{m-1} p)(x, \xi) = 0$ for all $\xi \in S_x M$, so $h + \ell_{m-1} p \in H_0^2(SM)$.

On the other hand, there exists $f \in H^{-1}(SM)$ with $Xf = 0$ such that $u = L_m f$. It follows that

$$0 = (Xf, h + \ell_{m-1} p) = (f, -X(h + \ell_{m-1} p)) = (f, \ell_m(u - dp)) = (L_m f, u - dp) = \|u\|^2,$$

where the last equality comes from the fact that $u$ is orthogonal to $dp$. Thus $u = 0$, which implies the s-injectivity. □

**Remark 6.1.** By carrying out an argument similar to the one of [Stefanov and Uhlmann 2005, Lemma 4.1], one can actually show that there exists $p \in C^\infty(S^{m-1}(T^*M))$ with $p|_{\partial M} = 0$ such that $\partial_\nu^k u|_{\partial M} = \partial_\nu^k dp|_{\partial M}$ for all integers $k \geq 0$. When $m = 0$, this means the boundary jet of $u$ is zero, i.e., $\partial_\nu^k u|_{\partial M} = 0$ for all $k \geq 0$. Note that [Stefanov and Uhlmann 2005] only considers the case that $u$ is a symmetric 2-tensor field, but the proof works for tensors of any rank. On the other hand, given $\partial_\nu^k u|_{\partial M} = \partial_\nu^k dp|_{\partial M}$, one should be able to prove that $h + \ell_{m-1} p \in H_0^{k+2}(SM)$ for all $k \geq 0$, i.e., $h + \ell_{m-1} p$ also has zero boundary jet. However, for our purposes $k = 0$ is enough.

The thing left to prove is the inequality (17). Actually the $H^k$ norms of $I_m w$ and $I_m \tilde{w}$ are equivalent for arbitrary $k \geq 0$, provided that $w$ is in $H_0^k(S^m(T^*M))$. A simple calculation shows that $\|I_m \tilde{w}\|_{L^2}^2 = (\tilde{w}, I_m^* I_m \tilde{w}) = (w, r_M I_m^* I_m \tilde{w}) = (w, I_m^* I_m w) = \|I_m w\|_{L^2}^2$. We assume $\partial \widetilde{M}$ and $\partial M$ are sufficiently close.

**Lemma 6.2.** *Let $M$ be a compact nontrapping manifold with strictly convex boundary. Given $w \in H_0^k(S^m(T^*M))$, $k \geq 1$, let $\tilde{w} \in H_0^k(S^m(T^*\widetilde{M}))$ be the extension of $w$ to $\widetilde{M}$ by zero. Then there exists $C > 1$ such that*

$$\frac{1}{C}\|I_m w\|_{H^k(\partial_+ SM)} \leq \|I_m \tilde{w}\|_{H^k(\partial_+ S\widetilde{M})} \leq C\|I_m w\|_{H^k(\partial_+ SM)}. \tag{18}$$

*Proof.* We only need to show (17), which is half of (18). Since $\partial M$ and $\partial \widetilde{M}$ are close, we can assume the closure of $\widetilde{M}$ is still compact nontrapping with strictly convex boundary. Given a geodesic $\gamma_{x,\xi}$ on $M$ determined by $(x, \xi) \in \partial_+ SM$, we can uniquely extend it to a geodesic $\gamma_{y,\eta}$ on $\widetilde{M}$ determined by $(y, \eta) \in \partial_+ S\widetilde{M}$. It is not difficult to see that the map

$$T : \partial_+ SM \to \partial_+ S\widetilde{M}, \quad \text{with } T(x, \xi) = (y, \eta),$$

is a diffeomorphism from $\partial_+ SM$ onto its image $T(\partial_+ SM)$. On the other hand, by the definition of $\tilde{w}$, $I_m w(x, \xi) = I_m \tilde{w}(T(x, \xi)) = I_m \tilde{w}(y, \eta)$ and $I_m \tilde{w}(y, \eta) = 0$ for $(y, \eta) \in \partial_+ S\widetilde{M} \setminus T(\partial_+ SM)$.

Since $\partial_+ SM$ and $\partial_+ S\widetilde{M}$ are compact, similar to the proofs of Lemmas 5.1 and 5.2, we will work in local charts. Let $U$ be a domain in $\partial_+ S\widetilde{M}$ with local coordinates $(\tilde{z}^1, \ldots, \tilde{z}^{2n-2})$ and $\varphi$ be a smooth function on $\partial_+ S\widetilde{M}$ with $\operatorname{supp} \varphi \subset U$. In the mean time, there is a domain $V$ in $\partial_+ SM$ with local coordinates $(z^1, \ldots, z^{2n-2})$ such that $T^{-1}(U \cap T(\partial_+ SM)) \subset V$, and $\psi$ is a smooth function on $\partial_+ SM$ with $T^{-1}(U \cap T(\partial_+ SM)) \subset \operatorname{supp} \psi \subset V$ and $\psi \equiv 1$ on $T^{-1}(U \cap T(\partial_+ SM))$. We first consider the case $w \in C_c^\infty(S^m(T^*M)^{\text{int}})$ and show that there exists $C > 0$ such that

$$\|\varphi \cdot I_m \tilde{w}\|_{H^k(U)} \leq C\|\psi \cdot I_m w\|_{H^k(V)}.$$

Notice that for $|\alpha| \le k$,

$$D_{\tilde{z}}^{\alpha}[\varphi \cdot I_m \tilde{w}] = \sum_{\beta+\gamma=\alpha} D_{\tilde{z}}^{\gamma}\varphi \cdot D_{\tilde{z}}^{\beta} I_m \tilde{w}.$$

Thus

$$\begin{aligned}
\left\| D_{\tilde{z}}^{\alpha}[\varphi \cdot I_m \tilde{w}] \right\|_{L^2(U)}^2 &\le \sum_{\beta \le \alpha} C_{\beta,\alpha} \int_U |D_{\tilde{z}}^{\beta} I_m \tilde{w}|^2 \, d\tilde{z} \\
&= \sum_{\beta \le \alpha} C_{\beta,\alpha} \int_{U \cap T(\partial_+ SM)} |D_{\tilde{z}}^{\beta} I_m \tilde{w}(\tilde{z})|^2 \, d\tilde{z} \\
&\le \sum_{|\sigma| \le |\alpha|} C_{\sigma,\alpha} \int_{T^{-1}(U \cap T(\partial_+ SM))} \left| D_z^{\sigma} I_m \tilde{w}(T(z)) \right|^2 J \, dz \\
&\le C' \sum_{|\sigma| \le |\alpha|} \int_{T^{-1}(U \cap T(\partial_+ SM))} \left| D_z^{\sigma} (\psi \cdot I_m w)(z) \right|^2 dz \\
&\le C' \sum_{|\sigma| \le |\alpha|} \int_V \left| D_z^{\sigma} (\psi \cdot I_m w)(z) \right|^2 dz \le C \| \psi \cdot I_m w \|_{H^k(V)}^2,
\end{aligned}$$

where $J$ is the Jacobian related to the diffeomorphism $T$. Therefore

$$\| I_m \tilde{w} \|_{H^k(\partial_+ S\tilde{M})} \le C \| I_m w \|_{H^k(\partial_+ SM)}$$

for $w \in C_c^{\infty}(S^m(T^*M)^{\mathrm{int}})$.

Now for $w \in H_0^k(S^m(T^*M))$, there is a sequence $w_k \in C_c^{\infty}(S^m(T^*M)^{\mathrm{int}})$, $k = 1, 2, \ldots$, which converges to $w$ in the $H^k$ norm. Then it is not difficult to see that the sequence $\tilde{w}_k \in C_c^{\infty}(S^m(T^*\tilde{M}))$ converges to $\tilde{w} \in H_0^k(S^m(T^*\tilde{M}))$. By the boundedness of the operator $I_m$, we know $I_m w_k$ and $I_m \tilde{w}_k$ converge to $I_m w$ and $I_m \tilde{w}$ respectively in the $H^k$ norm. This implies that above estimates are valid for any $w \in H_0^k(S^m(T^*M))$. □

The following proposition that holds on compact nontrapping manifolds with strictly convex boundary shows that items (4) and (5) in Theorem 1.2 are equivalent and any of them implies item (1).

**Proposition 6.3.** *Let $M$ be a compact nontrapping Riemannian manifold with strictly convex boundary and let $u \in C^{\infty}(S^m_{\mathrm{sol}}(T^*M))$. The following are equivalent*:

(i) *There exists $\varphi \in C_{\alpha}^{\infty}(\partial_+ SM)$ such that $u = I_m^* \varphi$.*

(ii) *There exists $f \in C^{\infty}(SM)$ satisfying $Xf = 0$ and $u = L_m f$.*

*Either of these two conditions implies s-injectivity of $I_m$.*

*Proof.* (i) $\Rightarrow$ (ii): By the assumption, there is $\varphi \in C_{\alpha}^{\infty}(\partial_+ SM)$ such that $u = I_m^* \varphi = L_m \circ I^* \varphi$. Define $f = I^* \varphi = \varphi^{\sharp} \in C^{\infty}(SM)$ (since $\varphi \in C_{\alpha}^{\infty}(\partial_+ SM)$); then $u = L_m f$. Moreover, it is clear that $Xf = X\varphi^{\sharp} = 0$ by definition.

(ii) $\Rightarrow$ (i): If there exists $f \in C^{\infty}(SM)$ with $Xf = 0$, this implies that $f = I^*(f|_{\partial_+ SM})$. We define $\varphi = f|_{\partial_+ SM} \in C^{\infty}(\partial_+ SM)$. However, since $\varphi^{\sharp} = f \in C^{\infty}(SM)$, we know $\varphi$ actually sits in the space $C_{\alpha}^{\infty}(\partial_+ SM)$. By the assumption, $u = L_m f = L_m \circ I^* \varphi = I_m^* \varphi$.

The argument that shows that any of these conditions imply s-injectivity of $I_m$ is even easier than the proof that (3) implies (1) in Theorem 1.2 since we do not have to worry about paring $Xf$ with an element in $H_0^2(SM)$. Assuming (ii), integration by parts yields right away that

$$0 = (Xf, h) = (f, -Xh) = (f, \ell_m(u)) = (L_m f, u) = \|u\|^2. \qquad \square$$

Finally we show that in Theorem 1.2, item (1) implies item (4):

Since $M$ is simple, given $u \in C^\infty(S^m_{\mathrm{sol}}(T^*M))$, by Lemma 4.3, there exists $v \in C_c^\infty(S^m(T^*\widetilde{M}))$ such that $r_M I_m^* I_m v = u$. Then it is a standard argument that if we define $\varphi = I^*(I_m v)|_{\partial_+ SM}$, then $I_m^* \varphi = u$. Moreover, since $I^*(I_m v)$ is smooth in the interior of $S\widetilde{M}$, we have $\varphi \in C_\alpha^\infty(\partial_+ SM)$.

The proof of Theorem 1.2 is now complete.

## 7. Alternative proof of Corollary 1.3

Before giving the alternative proof, we will explain how the solenoidal condition of a tensor manifests itself at the level of the transport equation. It seems that this basic relation has not appeared before in the literature, although we believe it was known to experts.

As we already pointed out in the Introduction, by considering the vertical Laplacian $\Delta$ on each fiber $S_x M$ of $SM$, we have a natural $L^2$ decomposition $L^2(SM) = \bigoplus_{m\geq 0} H_m(SM)$ into vertical spherical harmonics. We set $\Omega_m := H_m(SM) \cap C^\infty(SM)$. Then a function $u$ belongs to $\Omega_m$ if and only if $\Delta u = m(m+n-2)u$, where $n = \dim M$. The maps

$$\ell_m : C^\infty(S^m(T^*M)) \to \bigoplus_{k=0}^{[m/2]} \Omega_{m-2k}$$

and

$$L_m : \bigoplus_{k=0}^{[m/2]} \Omega_{m-2k} \to C^\infty(S^m(T^*M))$$

are isomorphisms. These maps give natural identification between functions in $\Omega_m$ and *trace-free* symmetric $m$-tensors (for details on this, see [Guillemin and Kazhdan 1980b; Dairbekov and Sharafutdinov 2010; Paternain et al. 2015a]). The geodesic vector field $X$ maps $\Omega_m$ to $\Omega_{m-1} \oplus \Omega_{m+1}$ and hence we can split it as $X = X_+ + X_-$, where $X_\pm : \Omega_m \to \Omega_{m\pm 1}$ and $X_+^* = -X_-$. Note that

$$X\ell_{m-1} = \ell_m d.$$

Given $f \in \bigoplus_{k=0}^{[m/2]} \Omega_{m-2k}$, in general $Xf \in \bigoplus_{k=0}^{[(m+1)/2]} \Omega_{m+1-2k}$. The next simple lemma characterizes the solenoidal condition in terms of $Xf$.

**Lemma 7.1.** *$Xf \in \Omega_{m+1}$ if and only if $L_m f$ is a solenoidal tensor.*

*Proof.* Note that $L_m f$ is solenoidal if and only if $(L_m f, dh) = 0$ for any $h \in C^\infty(S^{m-1}(T^*M))$ with $h|_{\partial M} = 0$. But

$$(L_m f, dh) = (f, \ell_m dh) = (f, X\ell_{m-1} h) = -(Xf, \ell_{m-1} h)$$

and the last term is zero if and only if $(Xf)_{m-2k-1} = 0$ for $0 \le k \le [(m-1)/2]$ since $\ell_{m-1}h \in \bigoplus_{k=0}^{[(m-1)/2]} \Omega_{m-1-2k}$. □

Another way to look at the condition $Xf \in \Omega_{m+1}$ is that the following equations should hold:

$$X_- f_{m-2k} + X_+ f_{m-2k-2} = 0 \quad \text{for } 0 \le k \le [(m-1)/2].$$

**Lemma 7.2.** *The following are equivalent*:

(1) *Given a nonnegative integer $m$ and $a_m \in \Omega_m$ with $X_- a_m = 0$, there exists $w \in C^\infty(SM)$ such that $Xw = 0$ and $w_m = a_m$.*

(2) *Given a nonnegative integer $m$ and $f = \sum_{k=0}^m f_k$ such that $Xf \in \Omega_m \oplus \Omega_{m+1}$, there exists $w \in C^\infty(SM)$ such that $Xw = 0$ and $\sum_{k=0}^m w_k = f$.*

*Proof.* The fact that (2) implies (1) is quite obvious from the fact that $a_m \in \Omega_m$ with $X_- a_m = 0$ implies $Xa_m = X_+ a_m \in \Omega_{m+1}$.

To prove that (1) implies (2) we proceed by induction on $m$. The case $m = 0$ follows right away since $Xf_0 \in \Omega_1$ and $X_- f_0 = 0$.

Suppose the claim holds for $m$ and let $f = \sum_{k=0}^{m+1} f_k$ be given with $Xf \in \Omega_{m+1} \oplus \Omega_{m+2}$. This is equivalent to saying that $X\left(\sum_{k=0}^m f_k\right) \in \Omega_m \oplus \Omega_{m+1}$ and $X_- f_{m+1} + X_+ f_{m-1} = 0$.

By the induction hypothesis, there exists $w \in C^\infty(SM)$ such that $Xw = 0$ and $w_k = f_k$ for all $k \le m$. The equation $Xw = 0$ in degree $m$ is

$$X_- w_{m+1} + X_+ f_{m-1} = 0$$

and thus

$$X_-(f_{m+1} - w_{m+1}) = 0.$$

Using item (1) in the lemma, there exists $w' = \sum_{m+1}^\infty w'_k \in C^\infty(SM)$ such that $Xw' = 0$ and $w'_{m+1} = f_{m+1} - w_{m+1}$. Then $X(w + w') = 0$ and $\sum_{k=0}^{m+1} (w + w')_k = f$ as desired. □

Finally we show:

**Proposition 7.3.** *The following are equivalent*:

(1) *Given a nonnegative integer $m$ and $u \in C^\infty(S^m_{\mathrm{sol}}(T^*M))$, there exists $f \in C^\infty(SM)$ with $Xf = 0$ such that $L_m f = u$.*

(2) *Given a nonnegative integer $m$ and $a_m \in \Omega_m$ with $X_- a_m = 0$, there exists $w \in C^\infty(SM)$ such that $Xw = 0$ and $w_m = a_m$.*

*Proof.* Assume (1) holds. Given $a_m \in \Omega_m$ with $X_- a_m = 0$, we see using Lemma 7.1 that $L_m a_m$ is a solenoidal tensor. Hence there is $f$ such that $Xf = 0$ and $f_m = L_m^{-1} L_m f = a_m$ (note that $L_m f_k = 0$ for $k > m$). Thus (2) holds.

Conversely if (2) holds, then item (2) in Lemma 7.2 holds. Thus there exists $f \in C^\infty(SM)$ such that $Xf = 0$ and $\sum_{k=0}^{[m/2]} f_{m-2k} = L_m^{-1} u$ and (1) holds. □

***Proof of Corollary 1.3.*** On account of Proposition 7.3, it suffices to show that given $a_m \in \Omega_m$ with $X_- a_m = 0$, there exists $w \in C^\infty(SM)$ such that $Xw = 0$ and $w_m = a_m$. What makes this possible in dimension two is [Paternain et al. 2015b, Lemma 5.6], whose content we now explain.

If $(M, g)$ is an oriented Riemannian surface, there is a global orthonormal frame $\{X, X_\perp, V\}$ of $SM$ equipped with the Sasaki metric, where $X$ is the geodesic vector field, $V$ is the vertical vector field and $X_\perp = [X, V]$. We define the Guillemin–Kazhdan operators [1980a]

$$\eta_\pm = \tfrac{1}{2}(X \pm iX_\perp).$$

If $x = (x_1, x_2)$ are oriented isothermal coordinates near some point of $M$, we obtain local coordinates $(x, \theta)$ on $SM$, where $\theta$ is the angle between $\xi$ and $\partial/\partial x_1$. In these coordinates $V = \partial/\partial\theta$ and $\eta_+$ and $\eta_-$ are $\partial$- and $\bar{\partial}$-type operators; see [Paternain et al. 2015a, Appendix B].

For any $m \in \mathbb{Z}$ we define

$$\Lambda_m = \{u \in C^\infty(SM) : Vu = imu\}.$$

In the $(x, \theta)$-coordinates elements of $\Lambda_m$ look locally like $h(x)e^{im\theta}$. Spherical harmonics may be further decomposed as

$$\Omega_0 = \Lambda_0,$$
$$\Omega_m = \Lambda_m \oplus \Lambda_{-m} \quad \text{for } m \geq 1.$$

Any $u \in C^\infty(SM)$ has a decomposition $u = \sum_{m=-\infty}^{\infty} u_m$, where $u_m \in \Lambda_m$. The geodesic vector field decomposes as

$$X = \eta_+ + \eta_-,$$

where $\eta_\pm : \Lambda_m \to \Lambda_{m\pm1}$. If $m \geq 1$, the action of $X_\pm$ on $\Omega_m$ is given by

$$X_\pm(e_m + e_{-m}) = \eta_\pm e_m + \eta_\mp e_{-m}, \quad e_j \in \Lambda_j,$$

and for $m = 0$, we have $X_+|_{\Omega_0} = \eta_+ + \eta_-$ and $X_-|_{\Omega_0} = 0$.

With these preliminaries out of the way, [Paternain et al. 2015b, Lemma 5.6] says that given $f \in \Lambda_m$, there is a smooth $w \in C^\infty(SM)$ with $Xw = 0$ and $w_m = f$. For $m = 0$, this gives the desired result right away.

Given $a_m \in \Omega_m$ with $X_- a_m = 0$ and $m \geq 1$, we write $a_m = e_m + e_{-m}$ with $e_j \in \Lambda_j$. Then $\eta_- e_m + \eta_+ e_{-m} = 0$. Consider now smooth $p, q$ with $Xp = Xq = 0$ and $p_m = e_m$ and $q_{-m} = e_{-m}$. Then

$$w = \sum_{-\infty}^{-m} q_k + \sum_{m}^{\infty} p_k$$

satisfies $Xw = 0$ and $w_m = a_m$. □

## Acknowledgements

We are very grateful to Mikko Salo and Gunther Uhlmann for several discussions related to this paper; our main result is motivated by a question raised in [Paternain et al. 2015a, Remark 11.7]. We are also very

grateful to Boris Kruglikov and Vladimir Matveev for extensive discussions related to Killing tensors and for the input provided in [Kruglikov and Matveev 2016]. Finally we would like to thank Colin Guillarmou and Plamen Stefanov for invaluable comments on the first draft. In particular, Stefanov provided us with the proof of Lemma 3.5 and Guillarmou with the first part of Remark 4.4. This research was supported by EPSRC grant EP/M023842/1.

## References

[Anikonov and Romanov 1997] Y. E. Anikonov and V. G. Romanov, "On uniqueness of determination of a form of first degree by its integrals along geodesics", *J. Inverse Ill-Posed Probl.* **5**:6 (1997), 487–490. MR Zbl

[Dairbekov and Sharafutdinov 2010] N. S. Dairbekov and V. A. Sharafutdinov, "Conformal Killing symmetric tensor fields on Riemannian manifolds", *Mat. Tr.* **13**:1 (2010), 85–145. In Russian; translated in *Siberian Advances in Mathematics* **21**:1 (2011), 1–41. MR Zbl

[Dairbekov and Uhlmann 2010] N. Dairbekov and G. Uhlmann, "Reconstructing the metric and magnetic field from the scattering relation", *Inverse Probl. Imaging* **4**:3 (2010), 397–409. MR Zbl

[Delay 2012] E. Delay, "Smooth compactly supported solutions of some underdetermined elliptic PDE, with gluing applications", *Comm. Partial Differential Equations* **37**:10 (2012), 1689–1716. MR Zbl

[Guillarmou 2014] C. Guillarmou, "Invariant distributions and X-ray transform for Anosov flows", preprint, 2014. To appear in *J. Differential Geom.* arXiv

[Guillemin and Kazhdan 1980a] V. Guillemin and D. Kazhdan, "Some inverse spectral results for negatively curved 2-manifolds", *Topology* **19**:3 (1980), 301–312. MR Zbl

[Guillemin and Kazhdan 1980b] V. Guillemin and D. Kazhdan, "Some inverse spectral results for negatively curved $n$-manifolds", pp. 153–180 in *Geometry of the Laplace operator* (Honolulu, HI, 1979), Proc. Sympos. Pure Math. **36**, Amer. Math. Soc., Providence, RI, 1980. MR Zbl

[Han 2011] Q. Han, *A basic course in partial differential equations*, Graduate Studies in Mathematics **120**, American Mathematical Society, Providence, RI, 2011. MR Zbl

[Helgason 1999] S. Helgason, *The Radon transform*, 2nd ed., Progress in Mathematics **5**, Birkhäuser, Boston, 1999. MR Zbl

[Kato et al. 2000] T. Kato, M. Mitrea, G. Ponce, and M. Taylor, "Extension and representation of divergence-free vector fields on bounded domains", *Math. Res. Lett.* **7**:5–6 (2000), 643–650. MR Zbl

[Kruglikov and Matveev 2016] B. Kruglikov and V. S. Matveev, "The geodesic flow of a generic metric does not admit nontrivial integrals polynomial in momenta", *Nonlinearity* **29**:6 (2016), 1755–1768. MR Zbl

[McLean 2000] W. McLean, *Strongly elliptic systems and boundary integral equations*, Cambridge University Press, 2000. MR Zbl

[Muhometov 1977] R. G. Muhometov, "The reconstruction problem of a two-dimensional Riemannian metric, and integral geometry", *Dokl. Akad. Nauk SSSR* **232**:1 (1977), 32–35. In Russian. MR

[Paternain et al. 2012] G. P. Paternain, M. Salo, and G. Uhlmann, "The attenuated ray transform for connections and Higgs fields", *Geom. Funct. Anal.* **22**:5 (2012), 1460–1489. MR Zbl

[Paternain et al. 2013] G. P. Paternain, M. Salo, and G. Uhlmann, "Tensor tomography on surfaces", *Invent. Math.* **193**:1 (2013), 229–247. MR Zbl

[Paternain et al. 2014] G. P. Paternain, M. Salo, and G. Uhlmann, "Spectral rigidity and invariant distributions on Anosov surfaces", *J. Differential Geom.* **98**:1 (2014), 147–181. MR Zbl

[Paternain et al. 2015a] G. P. Paternain, M. Salo, and G. Uhlmann, "Invariant distributions, Beurling transforms and tensor tomography in higher dimensions", *Math. Ann.* **363**:1–2 (2015), 305–362. MR Zbl

[Paternain et al. 2015b] G. P. Paternain, M. Salo, and G. Uhlmann, "On the range of the attenuated ray transform for unitary connections", *Int. Math. Res. Not.* **2015**:4 (2015), 873–897. MR Zbl

[Pestov and Sharafutdinov 1988] L. N. Pestov and V. A. Sharafutdinov, "Integral geometry of tensor fields on a manifold of negative curvature", *Sibirsk. Mat. Zh.* **29**:3 (1988), 114–130. In Russian; translated in *Siberian Mathematical Journal* **29**:3 (1988), 427–441. MR Zbl

[Pestov and Uhlmann 2005] L. Pestov and G. Uhlmann, "Two dimensional compact simple Riemannian manifolds are boundary distance rigid", *Ann. of Math.* (2) **161**:2 (2005), 1093–1110. MR Zbl

[Salo and Uhlmann 2011] M. Salo and G. Uhlmann, "The attenuated ray transform on simple surfaces", *J. Differential Geom.* **88**:1 (2011), 161–187. MR Zbl

[Sharafutdinov 1994] V. A. Sharafutdinov, *Integral geometry of tensor fields*, VSP, Utrecht, 1994. MR Zbl

[Sharafutdinov 1999] V. Sharafutdinov, "Ray transform on Riemannian manifolds", lecture notes, University of Oulu, 1999, available at http://math.nsc.ru/∼sharafutdinov/files/Lectures.pdf.

[Sharafutdinov 2002] V. A. Sharafutdinov, "An integral geometry problem in a nonconvex domain", *Sibirsk. Mat. Zh.* **43**:6 (2002), 1430–1442. In Russian; translated in *Siberian Math. J.* **43**:6 (2002), 1159–1168. MR Zbl

[Sharafutdinov et al. 2005] V. Sharafutdinov, M. Skokan, and G. Uhlmann, "Regularity of ghosts in tensor tomography", *J. Geom. Anal.* **15**:3 (2005), 499–542. MR

[Stefanov and Uhlmann 2004] P. Stefanov and G. Uhlmann, "Stability estimates for the X-ray transform of tensor fields and boundary rigidity", *Duke Math. J.* **123**:3 (2004), 445–467. MR Zbl

[Stefanov and Uhlmann 2005] P. Stefanov and G. Uhlmann, "Boundary rigidity and stability for generic simple metrics", *J. Amer. Math. Soc.* **18**:4 (2005), 975–1003. MR Zbl

[Stefanov and Uhlmann 2008] P. Stefanov and G. Uhlmann, "Integral geometry on tensor fields on a class of non-simple Riemannian manifolds", *Amer. J. Math.* **130**:1 (2008), 239–268. MR Zbl

[Stefanov et al. 2014] P. Stefanov, G. Uhlmann, and A. Vasy, "Inverting the local geodesic X-ray transform on tensors", preprint, 2014. To appear in *J. Anal. Math.* arXiv

[Uhlmann and Vasy 2016] G. Uhlmann and A. Vasy, "The inverse problem for the local geodesic ray transform", *Invent. Math.* **205**:1 (2016), 83–120. With an appendix by H. Zhou. MR Zbl

Received 18 Nov 2015. Revised 28 Jul 2016. Accepted 12 Sep 2016.

GABRIEL P. PATERNAIN: g.p.paternain@dpmms.cam.ac.uk
*Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB3 0WB, United Kingdom*

HANMING ZHOU: hz318@dpmms.cam.ac.uk
*Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB3 0WB, United Kingdom*

ANALYSIS AND PDE
Vol. 9, No. 8, 2016
dx.doi.org/10.2140/apde.2016.9.1931

# MULTIPLE VECTOR-VALUED INEQUALITIES VIA THE HELICOIDAL METHOD

CRISTINA BENEA AND CAMIL MUSCALU

We develop a new method of proving vector-valued estimates in harmonic analysis, which we call "the helicoidal method". As a consequence of it, we are able to give affirmative answers to several questions that have been circulating for some time. In particular, we show that the tensor product $\mathrm{BHT} \otimes \Pi$ between the bilinear Hilbert transform BHT and a paraproduct $\Pi$ satisfies the same $L^p$ estimates as the BHT itself, solving completely a problem introduced by Muscalu et al. (*Acta Math.* **193**:2 (2004), 269–296). Then, we prove that for "locally $L^2$ exponents" the corresponding vector-valued $\overrightarrow{\mathrm{BHT}}$ satisfies (again) the same $L^p$ estimates as the BHT itself. Before the present work there was not even a single example of such exponents.

Finally, we prove a biparameter Leibniz rule in mixed norm $L^p$ spaces, answering a question of Kenig in nonlinear dispersive PDE.

## 1. Introduction

Vector-valued estimates for classical Calderón–Zygmund operators are known from the work of Burkholder [1983], Benedek, Calderón and Panzone [Benedek et al. 1962], Rubio de Francia, Ruiz and Torrea [Rubio de Francia et al. 1986], to mention a few. A customary way of proving such vector-valued estimates is through weighted norm inequalities and extrapolation, as explained in [García-Cuerva and Rubio de Francia 1985]. Initially, the vector-valued approach unified the existing theory for maximal operators, square functions, and singular integrals. Later on, the setting was generalized to Banach spaces which have the *unconditional martingale difference* property, and it was shown by Bourgain [1986] that this is in fact a necessary condition for this theory.

For bilinear operators, however, the theory is far from being fully understood, even in the scalar case. In this paper, we study vector-valued estimates for the bilinear Hilbert transform and for paraproducts. Our initial motivation was an AKNS system-related problem, which can be reduced to understanding a Rubio de Francia operator for iterated Fourier integrals. Because of the specific nature of this question, our general approach is concrete, rather than abstract. As much as possible, the present article aims to be self-contained.

Central to time-frequency analysis is the bilinear Hilbert transform operator, defined by

$$\mathrm{BHT}(f,g)(x) = \text{p.v.} \int_{\mathbb{R}} f(x-t)g(x+t)\,\frac{dt}{t}.$$

Muscalu is a member of the Simion Stoilow Institute of Mathematics of the Romanian Academy.

*MSC2010:* 42A45, 42B15, 42B25, 42B37.
*Keywords:* vector-valued estimates for singular and multilinear operators, tensor products in mixed norms, Leibniz rule, AKNS systems.

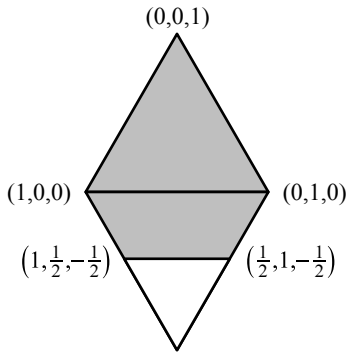



**Figure 1.** Range for BHT operator.

This operator was first introduced by Calderón, in connection with his work on the Cauchy integral on Lipschitz curves. $L^p$ estimates for BHT were proved nearly thirty years later, by M. Lacey and C. Thiele, without establishing the optimality of the range.

**Theorem 1** [Lacey and Thiele 1999]. BHT *is a bounded bilinear operator from* $L^p \times L^q$ *into* $L^s$ *for any* $1 < p, q \le \infty$, $0 < s < \infty$, *satisfying* $\frac{1}{p} + \frac{1}{q} = \frac{1}{s}$ *and* $\frac{2}{3} < s < \infty$.

The range of the operator Range(BHT) consists of the set of triples $(p, q, s)$ satisfying the conditions above. The question that remains open is whether the bilinear Hilbert transform is bounded also for $s \in \left(\frac{1}{2}, \frac{2}{3}\right]$. The Hölder-type condition $\frac{1}{p}+\frac{1}{q}=\frac{1}{s}$ reflects the scaling invariance of the operator, and it can be reformulated as $\frac{1}{p}+\frac{1}{q}+\frac{1}{s'}=1$, where $s'$ is the conjugate exponent of $s$. Thus $(p, q, s) \in$ Range(BHT) if $\left(\frac{1}{p}, \frac{1}{q}, \frac{1}{s'}\right)$ lies in the plane $\{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 1\}$, and is contained inside the convex hull of the points

$$(0, 0, 1), \quad (1, 0, 0), \quad \left(1, \tfrac{1}{2}, -\tfrac{1}{2}\right), \quad \left(\tfrac{1}{2}, 1, -\tfrac{1}{2}\right), \quad (0, 1, 0)$$

(see Figure 1). Regarded as a bilinear multiplier operator, BHT becomes equivalent to

$$(f, g) \mapsto \int_{\xi<\eta} \hat{f}(\xi)\, \hat{g}(\eta)\, e^{2\pi i x(\xi+\eta)}\, d\xi\, d\eta. \tag{1}$$

The method of the proof, which breaks down when $\frac{1}{p} + \frac{1}{q} \ge \frac{3}{2}$, consists of approximating BHT by a *model operator* obtained through a Whitney decomposition of the frequency region $\{\xi < \eta\}$. In essence, this model operator is a superposition of "almost orthogonal" objects of a lower complexity, called *discretized paraproducts*.

Paraproducts play an important role on their own, especially in the analysis of PDE. A *paraproduct* is an expression of the form

$$(f, g) \mapsto \int_{\mathbb{R}} \int_{\mathbb{R}} f(x-t)\, g(x-s)\, k(s, t)\, ds\, dt, \tag{2}$$

where $k(s,t)$ is a Calderón–Zygmund kernel in the plane $\mathbb{R}^2$. Alternatively, a paraproduct can be regarded as a bilinear multiplier operator

$$(f,g)\mapsto \int_{\mathbb{R}^2} m(\xi,\eta)\,\hat{f}(\xi)\,\hat{g}(\eta)\,e^{2\pi i x(\xi+\eta)}\,d\xi\,d\eta,$$

where $m$ is a classical Marcinkiewicz–Mikhlin–Hörmander multiplier in two variables, sufficiently smooth away from the origin. The singularity of the multiplier $m$ consists of one point: $(\xi,\eta)=(0,0)$. On the other hand, we can see from (1) that the BHT multiplier is singular along the line $\xi=\eta$.

We have the following result on paraproducts:

**Theorem 2** [Meyer and Coifman 1997]. *Any bilinear multiplier operator associated to a symbol $m(\xi,\eta)$ satisfying $|\partial^\alpha m(\xi,\eta)|\lesssim|(\xi,\eta)|^{-\alpha}$ for sufficiently many multi-indices $\alpha$, maps $L^p(\mathbb{R})\times L^q(\mathbb{R})$ into $L^s(\mathbb{R})$ provided that $1<p,q\le\infty$, $\frac{1}{2}<s<\infty$, and $\frac{1}{p}+\frac{1}{q}=\frac{1}{s}$.*

Following the presentation in [Muscalu and Schlag 2013], any bilinear operator of this form can be essentially written as a finite sum of paraproducts of the form

$$(f,g)\mapsto \sum_k \big((f*\psi_k)\cdot(g*\psi_k)\big)*\varphi_k(x)=\sum_k P_k(Q_k f\cdot Q_k g), \tag{I}$$

$$(f,g)\mapsto \sum_k \big((f*\varphi_k)\cdot(g*\psi_k)\big)*\psi_k(x)=\sum_k Q_k(P_k f\cdot Q_k g), \tag{II}$$

$$(f,g)\mapsto \sum_k \big((f*\psi_k)\cdot(g*\varphi_k)\big)*\psi_k(x)=\sum_k Q_k(Q_k f\cdot P_k g). \tag{III}$$

From now on, a *paraproduct* will designate any of the expressions (I), (II) or (III), and will be denoted by $\Pi(f,g)$. Here $\psi_k(x)=2^k\psi(2^k x)$, $\varphi_k(x)=2^k\varphi(2^k x)$, $\hat{\varphi}(\xi)\equiv 1$ on $\left[-\frac{1}{2},\frac{1}{2}\right]$ and is supported on $[-1,1]$ and $\hat{\psi}(\xi)=\hat{\varphi}(\xi/2)-\hat{\varphi}(\xi)$. The $\{Q_k\}_k$ represent Littlewood–Paley projections onto the frequency $|\xi|\sim 2^k$, while $\{P_k\}_k$ are convolution operators associated with dyadic dilations of a nice bump function of integral 1.

A classical application of Theorem 2 is the Leibniz rule

$$\|D^\alpha(f\cdot g)\|_s\lesssim\|D^\alpha f\|_{p_1}\|g\|_{q_1}+\|f\|_{p_2}\|D^\alpha g\|_{q_2}, \tag{3}$$

which holds for any $\alpha>0$, as long as $\frac{1}{p_i}+\frac{1}{q_i}=\frac{1}{s}$, $1<p_i,q_i\le\infty$, and $1/(1+\alpha)<s<\infty$. In particular, if $s\ge 1$, which is the case in most applications, the Leibniz rule holds for any $\alpha>0$.

For functions on $\mathbb{R}^2$, with (fractional) partial derivatives in both variables, a corresponding Leibniz rule is

$$\begin{aligned}\big\|D_1^\alpha D_2^\beta(f\cdot g)\big\|_s&\\ \lesssim\|D_1^\alpha D_2^\beta f\|_{p_1}\|g\|_{q_1}&+\|f\|_{p_2}\|D_1^\alpha D_2^\beta g\|_{q_2}+\|D_1^\alpha f\|_{p_3}\|D_2^\beta g\|_{q_3}+\|D_2^\beta f\|_{p_4}\|D_1^\alpha g\|_{q_4}.\end{aligned} \tag{4}$$

The proof of the above inequality relies on discrete biparameter paraproducts $\Pi\otimes\Pi$, which are expressions of the form

$$\sum_{k,l}\big((f*(\varphi_k\otimes\psi_l))\cdot(g*(\psi_k\otimes\varphi_l))\big)*\psi_k\otimes\psi_l(x,y). \tag{5}$$

Muscalu, Pipher, Thiele, and Tao proved the following theorem:

**Theorem 3** [Muscalu et al. 2004a]. *$\Pi\otimes\Pi$ is a bounded operator from $L^p(\mathbb{R}^2)\times L^q(\mathbb{R}^2)$ into $L^s(\mathbb{R}^2)$ provided that $1<p,q\le\infty$, $\frac{1}{p}+\frac{1}{q}=\frac{1}{s}$, and $0<s<\infty$.*

This further implies that (4) is true whenever

$$\frac{1}{p_i}+\frac{1}{q_i}=\frac{1}{s},\quad 1<p_i,q_i\le\infty,\quad\text{and}\quad \max\Big(\frac{1}{1+\alpha},\frac{1}{1+\beta}\Big)<r<\infty.$$

If $r\ge 1$ the last condition is redundant, so (4) holds for any $\alpha,\beta>0$.

Related to this, Carlos Kenig asked the following question, which has been circulating for some time:

**Question 1.** *Assuming that $1\le s_1,s_2<\infty$, and $\alpha,\beta>0$, is there a Leibniz rule for mixed norm $L^p$ spaces of the form*

$$\begin{aligned}\big\|D_1^\alpha D_2^\beta(f\cdot g)\big\|_{L_x^{s_1}L_y^{s_2}}\lesssim{}&\|D_1^\alpha D_2^\beta f\|_{L_x^{p_1}L_y^{p_2}}\|g\|_{L_x^{q_1}L_y^{q_2}}+\|f\|_{L_x^{p_3}L_y^{p_4}}\|D_1^\alpha D_2^\beta g\|_{L_x^{q_3}L_y^{q_4}}\\&+\|D_1^\alpha f\|_{L_x^{p_5}L_y^{p_6}}\|D_2^\beta g\|_{L_x^{q_5}L_y^{q_6}}+\|D_2^\beta f\|_{L_x^{p_7}L_y^{p_8}}\|D_1^\alpha g\|_{L_x^{q_7}L_y^{q_8}}?\end{aligned}$$

Here the mixed norms are defined by

$$\|f\|_{L_x^pL_y^q}:=\big\|\|f\|_{L_y^q}\big\|_{L_x^p}:=\bigg(\int_{\mathbb{R}}\bigg(\int_{\mathbb{R}}|f(x,y)|^q\,dy\bigg)^{\frac{p}{q}}dx\bigg)^{\frac{1}{p}}.\tag{6}$$

A result of a similar type appeared in [Kenig et al. 1993], as an important tool in establishing local well-posedness for the generalized Korteweg–de Vries equation. This is a dispersive, nonlinear equation given by

$$\begin{cases}\dfrac{\partial u}{\partial t}+\dfrac{\partial^3 u}{\partial x^3}+u^k\dfrac{\partial u}{\partial x}=0,\quad t,x\in\mathbb{R},\ k\in\mathbb{Z}^+,\\ u(x,0)=u_0(x).\end{cases}\tag{7}$$

In order to prove existence, the authors use the contraction principle, but to be able to do so, they need to construct a suitable Banach space. The norm of the Banach space involves mixed $L^p$ norms of fractional derivatives in the first variable $D_1^\alpha$, and the Leibniz rule employed in this paper is

$$\big\|D_1^\alpha(f\cdot g)-f\cdot D_1^\alpha g-D_1^\alpha f\cdot g\big\|_{L_x^pL_t^q}\lesssim C\|D_1^{\alpha_1}f\|_{L_x^{p_1}L_t^{q_1}}\|D_1^{\alpha_2}g\|_{L_x^{p_2}L_t^{q_2}}.\tag{8}$$

Here $\alpha\in(0,1)$, $\alpha_1+\alpha_2=\alpha$ and $\frac{1}{p_1}+\frac{1}{p_2}=\frac{1}{p}$, $\frac{1}{q_1}+\frac{1}{q_2}=\frac{1}{q}$. Also, $p,p_1,p_2,q,q_1,q_2\in(1,\infty)$, but one can allow $q_1=\infty$ if $\alpha_1=0$.

The fractional derivatives appear as a consequence of the smoothness requirement on the initial data: $u_0$ is assumed to be in some Sobolev space $H^\alpha(\mathbb{R})$, where $\alpha$ depends on the value of $k$ in (7).

Question 1 is an extension of (8), and we managed to provide an answer by proving estimates for $\Pi\otimes\Pi$ in $L^p$ spaces with mixed norms.

Biparameter bilinear operators were first studied in [Journé 1985], where he introduced a new way of generalizing Calderón–Zygmund operators on product spaces. More exactly, in that work he proved that "bicommutators of Calderón–Coifman-type" are bounded, which translates to "$\Pi\otimes\Pi$ maps $L^2(\mathbb{R}^2)\times L^\infty(\mathbb{R}^2)$ into $L^2(\mathbb{R}^2)$". The full range of estimates for $\Pi\otimes\Pi$ was established in [Muscalu et al. 2004a],

where was also noticed that BHT $\otimes$ BHT does not satisfy any $L^p$ estimates. What remained undecided for some time was the following question:

**Question 2.** *Does the tensor product* BHT $\otimes \Pi$ *satisfy any* $L^p$ *estimates*? *Would it be possible to prove it satisfies the same estimates as the* BHT *itself*?

Some significant progress in answering this question was made by Silva [2014]. It was showed that BHT $\otimes \Pi$ maps $L^p \times L^q$ into $L^s$ under the constraints that $\frac{1}{p} + \frac{2}{q} < 2$ and $\frac{1}{q} + \frac{2}{p} < 2$. Our helicoidal method allows us to remove these restrictions, proving in this way that BHT $\otimes \Pi$ satisfies indeed the same $L^p$ estimates as BHT.

As it turned out, the study of Question 1 and Question 2 is related to proving (sometimes multiple) vector-valued inequalities for $\Pi$ and BHT. Let $\vec{r} = (r_1, r_2, r)$ be a tuple so that $1 < r_1, r_2 \le \infty,\ 1 \le r < \infty$ and $\frac{1}{r_1} + \frac{1}{r_2} = \frac{1}{r}$. We say that an inequality of the type

$$\left\| \left( \sum_k |\mathrm{BHT}(f_k, g_k)|^r \right)^{\frac{1}{r}} \right\|_s \lesssim \left\| \left( \sum_k |f_k|^{r_1} \right)^{r_1} \right\|_p \left\| \left( \sum_k |g_k|^{r_2} \right)^{r_2} \right\|_q \tag{9}$$

represents $L^p$ estimates for vector-valued BHT, corresponding to the exponent $\vec{r}$; in short, we have $L^p$ estimates for $\overrightarrow{\mathrm{BHT}}_{\vec{r}}$.

Some $L^p$ estimates for vector-valued BHT have been proved recently by Silva [2014], provided $r \in \left(\frac{4}{3}, 4\right)$. UMD-valued extensions for the quartile operator (the Fourier–Walsh analogue of BHT) were studied by Hytönen, Lacey and Parissis [Hytönen et al. 2013]. Their results, transferred to the $L^p$ setting, hold under the same constraint that $r \in \left(\frac{4}{3}, 4\right)$. Moreover, through this method it is impossible to obtain vector-valued extensions when $L^1$ or $L^\infty$ spaces are involved, as these are not UMD spaces. A similar abstract approach was taken in [Di Plinio and Ou 2015], where Banach-valued estimates for paraproducts were proved.

In spite of these results, some important questions remained unsettled:

**Question 3.** *Are there any exponents* $\vec{r}$ *as before for which the corresponding vector-valued* $\overrightarrow{\mathrm{BHT}}_{\vec{r}}$ *satisfy the same* $L^p$ *estimates as the* BHT *itself*?

As the question suggests, until the present work, there was not even a single example of such an exponent. We show that whenever $\vec{r}$ is in the "local $\ell^2$ range" (that is, $0 \le \frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r'} \le \frac{1}{2}$), $\overrightarrow{\mathrm{BHT}}_{\vec{r}}$ satisfies the same $L^p$ estimates as the BHT operator. Moreover, whenever $2 \le p, q \le \infty$, we show $L^p$ estimates exist for any exponent $\vec{r} = (r_1, r_2, r)$.

To summarize, the main task of the present work is to give affirmative answers to Question 1, Question 2, and Question 3 described above. In what follows, we will present our main results, sometimes in a more general setting.

**Theorem 4.** *For any* $\alpha, \beta > 0$,

$$\begin{aligned}\left\| D_1^\alpha D_2^\beta (f \cdot g) \right\|_{L_x^{s_1} L_y^{s_2}} \lesssim{}& \|D_1^\alpha D_2^\beta f\|_{L_x^{p_1} L_y^{p_2}} \|g\|_{L_x^{q_1} L_y^{q_2}} + \|f\|_{L_x^{p_3} L_y^{p_4}} \|D_1^\alpha D_2^\beta g\|_{L_x^{q_3} L_y^{q_4}} \\ &+ \|D_1^\alpha f\|_{L_x^{p_5} L_y^{p_6}} \|D_2^\beta g\|_{L_x^{q_5} L_y^{q_6}} + \|D_2^\beta f\|_{L_x^{p_7} L_y^{p_8}} \|D_1^\alpha g\|_{L_x^{q_7} L_y^{q_8}}\end{aligned}$$

*whenever* $1 < p_j, q_j \le \infty,\ \frac{1}{2} < s_1 < \infty,\ 1 \le s_2 < \infty$, *with* $\frac{1}{1+\alpha} < s_1 < \infty$, *and the indices satisfy the natural Hölder-type conditions.*

This answers Question 1 in the affirmative. Of course, one may wonder if Theorem 4 holds in arbitrary dimensions. As the careful reader will notice, our methods allow for such a generalization, with the outer-most Lebesgue exponent possibly less than 1, if all the indices $p_i, q_i$ involved are *strictly* between 1 and $\infty$. However, in applications $L^\infty$ norms appear, so it will be of interest to have a more general theorem for $1 < p_i, q_i \leq \infty$. Although we cannot obtain this result in this paper due to some delicate technical issues, we plan to return to this problem sometime in the future.

An $n$-dimensional version of a Leibniz rule was presented in [Torres and Ward 2015] for indices that are again strictly between 1 and $\infty$:

$$\|D_2^\beta(f\cdot g)\|_{L_x^{s_1}L_y^{s_2}(\mathbb{R}\times\mathbb{R}^n)} \lesssim \|D_2^\beta f\|_{L_x^{p_1}L_y^{p_2}(\mathbb{R}\times\mathbb{R}^n)}\|g\|_{L_x^{q_1}L_y^{q_2}(\mathbb{R}\times\mathbb{R}^n)} + \|f\|_{L_x^{p_1}L_y^{p_2}(\mathbb{R}\times\mathbb{R}^n)}\|D_2^\beta g\|_{L_x^{q_1}L_y^{q_2}(\mathbb{R}\times\mathbb{R}^n)}.$$

This can be regarded as an $n$-dimensional generalization of (8), and it is simpler than our variant of the Leibniz rule because it doesn't require a multiparameter analysis.

Our Theorem 4 is a consequence, modulo technical but "classical" complications, of the following result:

**Theorem 5** (mixed norm estimates for paraproducts on the bidisc). *Let* $1 < p_j, q_j \leq \infty$, $\frac{1}{2} < s_1 < \infty$, $1 \leq s_2 < \infty$, *so that* $\frac{1}{p_j} + \frac{1}{q_j} = \frac{1}{s_j}$, $1 \leq j \leq 2$. *Then*

$$\big\|\Pi\otimes\Pi(f,g)\big\|_{L_x^{s_1}L_y^{s_2}} \lesssim \|f\|_{L_x^{p_1}L_y^{p_2}}\|g\|_{L_x^{q_1}L_y^{q_2}}.$$

The above theorem provides $L^p$ estimates for $\Pi\otimes\Pi$ in mixed norm $L^p$ spaces. Through our methods, we can also recover the results from [Muscalu et al. 2006a], stating that $\Pi\otimes\cdots\otimes\Pi$ maps $L^p(\mathbb{R}^n)\times L^q(\mathbb{R}^n)$ into $L^s(\mathbb{R}^n)$ whenever $1 < p, q \leq \infty$, $\frac{1}{2} < s < \infty$ and $\frac{1}{p} + \frac{1}{q} = \frac{1}{s}$. Moreover, we answer Question 2 by proving that $\mathrm{BHT}\otimes\Pi$ and $\mathrm{BHT}\otimes\Pi^{\otimes n}$ satisfy the same $L^p$ estimates as BHT:

**Theorem 6.** *For any* $p, q, r$ *with* $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, *with* $1 < p, q \leq \infty$ *and* $\frac{2}{3} < r < \infty$,

$$\big\|\mathrm{BHT}\otimes\Pi\otimes\cdots\otimes\Pi(f,g)\big\|_{L^r(\mathbb{R}^{n+1})} \lesssim \|f\|_{L^p(\mathbb{R}^{n+1})}\|g\|_{L^q(\mathbb{R}^{n+1})}.$$

*The same is true for* $\Pi\otimes\cdots\otimes\Pi\otimes\mathrm{BHT}\otimes\Pi\otimes\cdots\otimes\Pi$.

For $n \geq 2$, no such results were known previously, and furthermore, a new approach was necessary for $n \geq 3$. This will be explained later in part (3) of the Remark on page 1939.

Some mixed norm $L^p$ estimates for $\Pi^{\otimes d_1}\otimes\mathrm{BHT}\otimes\Pi^{\otimes d_2}$ can also be proved (see Section 5.1). For $\Pi\otimes\mathrm{BHT}$, they are similar to [Di Plinio and Ou 2015] in the case $n = 1$. We recently learned that in [loc. cit.] mixed norm estimates for $\Pi\otimes\Pi$, close to our Theorem 5, are also obtained.

In proving the results mentioned above, multiple vector-valued extensions for BHT and $\Pi$ play a very important role. Given a totally $\sigma$-finite measure space $(\mathcal{W}, \Sigma, \mu)$, and $f, g : \mathbb{R}\times\mathcal{W}\to\mathbb{C}$, we define

$$\mathrm{BHT}(f,g)(x,w) := \text{p.v.}\int_{\mathbb{R}} f(x-t,w)g(x+t,w)\frac{dt}{t}.$$

Note that for a fixed value $w\in\mathcal{W}$, we have $\mathrm{BHT}(f,g)(x,w) = \mathrm{BHT}(f_w,g_w)(x)$, where $f_w(x) = f(x,w)$.

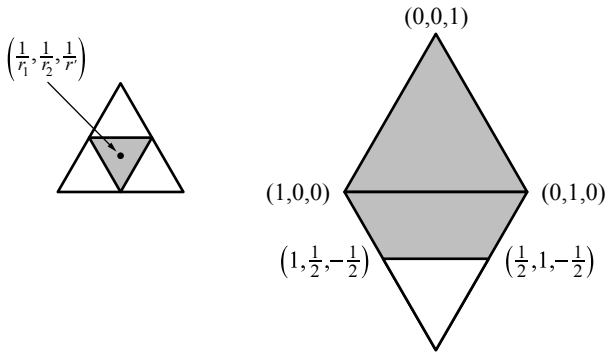



**Figure 2.** Range for vector-valued BHT when $\frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r'} \leq \frac{1}{2}$.

**Theorem 7.** *For any triple $(r_1, r_2, r)$ with $1 < r_1, r_2 \leq \infty$, $1 \leq r < \infty$ and so that $\frac{1}{r_1} + \frac{1}{r_2} = \frac{1}{r}$, there exists a nonempty set $\mathcal{D}_{r_1,r_2,r}$ of triples $(p, q, s)$ satisfying $\frac{1}{p} + \frac{1}{q} = \frac{1}{s}$ for which*

$$\mathrm{BHT} : L^p(\mathbb{R}; L^{r_1}(\mathcal{W}, \mu)) \times L^q(\mathbb{R}; L^{r_2}(\mathcal{W}, \mu)) \to L^s(\mathbb{R}; L^r(\mathcal{W}, \mu)).$$

*This means that there exists a constant $C$ so that*

$$\big\| \|\mathrm{BHT}(f, g)\|_{L^r(\mathcal{W},\mu)} \big\|_{L^s(\mathbb{R})} \leq C \big\| \|f\|_{L^{r_1}(\mathcal{W},\mu)} \big\|_{L^p(\mathbb{R})} \big\| \|g\|_{L^{r_2}(\mathcal{W},\mu)} \big\|_{L^q(\mathbb{R})}.$$

*Depending on the values of $r_1, r_2, r'$, we can give an explicit characterization of $\mathcal{D}_{r_1,r_2,r}$, as follows*:

(i) *If $\frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r'} \leq \frac{1}{2}$, then $\mathcal{D}_{r_1,r_2,r} = \mathrm{Range(BHT)}$.*

(ii) *If $\frac{1}{r_2}, \frac{1}{r'} \leq \frac{1}{2}$ and $\frac{1}{r_1} > \frac{1}{2}$, then $\mathcal{D}_{r_1,r_2,r}$ corresponds to the tuples $(p, q, s) \in \mathrm{Range(BHT)}$ for which $0 \leq \frac{1}{q} < \frac{3}{2} - \frac{1}{r_1}$.*

(iii) *If $\frac{1}{r_1}, \frac{1}{r'} \leq \frac{1}{2}$ and $\frac{1}{r_2} > \frac{1}{2}$, then the range of exponents is similar to the one in* (ii), *with the roles of $r_1$ and $r_2$ interchanged. That is, $\mathcal{D}_{r_1,r_2,r}$ consists of tuples $(p, q, s) \in \mathrm{Range(BHT)}$ for which $0 \leq \frac{1}{p} < \frac{3}{2} - \frac{1}{r_2}$.*

(iv) *If $\frac{1}{r_1}, \frac{1}{r_2} \leq \frac{1}{2}$ and $\frac{1}{r'} > \frac{1}{2}$, then $\mathcal{D}_{r_1,r_2,r}$ corresponds to the tuples $(p, q, s) \in \mathrm{Range(BHT)}$ for which $0 \leq \frac{1}{p}, \frac{1}{q} < \frac{1}{2} + \frac{1}{r}$ and $-\frac{1}{r} < \frac{1}{s'} < 1$.*

See Figures 2–4 for the ranges of BHT in the cases above.

We emphasize that whenever $(p, q, s)$ are such that $0 \leq \frac{1}{p}, \frac{1}{q} \leq \frac{1}{2}$ (and consequently $1 \leq s < \infty$), vector-valued estimates exist for any tuple $(r_1, r_2, r)$. These are the first examples of tuples $(p, q, s)$ which allow for any $\overrightarrow{\mathrm{BHT}}_{\vec{r}}$ extension.

Theorem 7 can be further generalized to multiple vector-valued inequalities. For an $n$-tuple $P = (p_1, \ldots, p_n)$, the mixed $L^P$ norm on the product space

$$(\mathcal{W}, \Sigma, \mu) = \left( \prod_{j=1}^{n} \mathcal{W}_j, \prod_{j=1}^{n} \Sigma_j, \prod_{j=1}^{n} \mu_j \right)$$

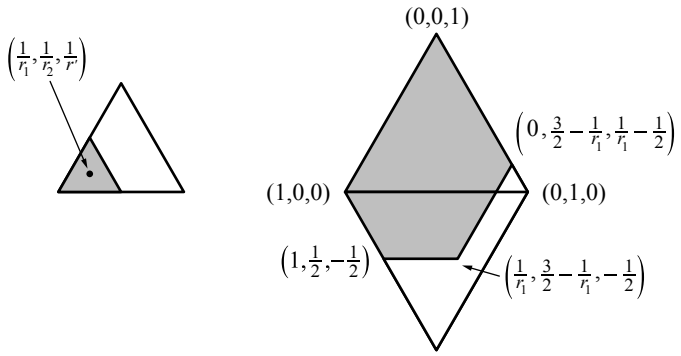



**Figure 3.** Range for vector-valued BHT when $\frac{1}{r_1} > \frac{1}{2}$.

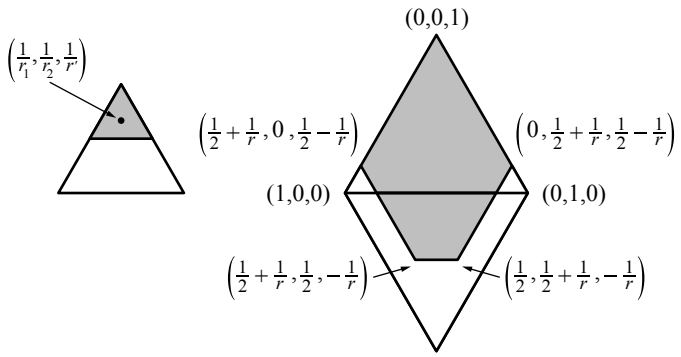



**Figure 4.** Range for vector-valued BHT when $\frac{1}{r'} > \frac{1}{2}$.

is defined as

$$\|f\|_P := \left(\int_{\mathcal{W}_1} \cdots \left(\int_{\mathcal{W}_n} |f(w_1,\dots,w_n)|^{p_n}\, d\mu_n(w_n)\right)^{\frac{p_{n-1}}{p_n}} \cdots d\mu_1(w_1)\right)^{\frac{1}{p_1}}.$$

Consider the tuples $R_1 = (r_1^1,\dots,r_1^n)$, $R_2 = (r_2^1,\dots,r_2^n)$ and $R = (r^1,\dots,r^n)$ satisfying for every $1 \le j \le n$,

$$1 < r_1^j, r_2^j \le \infty, \quad 1 \le r^j < \infty, \quad \frac{1}{r_1^j} + \frac{1}{r_2^j} = \frac{1}{r^j}$$

(from now on, this will be written as $1 < R_1, R_2 \le \infty$, $1 \le R < \infty$, and $\frac{1}{R_1} + \frac{1}{R_2} = \frac{1}{R}$). Then we have the following multiple vector-valued result:

**Theorem 8.** *Let $R_1, R_2$ and $R$ be as above. If the tuples $R_1, R_2, R$ satisfy the condition $(r_1^j, r_2^j, r^j) \in \mathcal{D}_{r_1^{j+1}, r_2^{j+1}, r^{j+1}}$ for every $1 \le j \le n-1$, then there exists a set $\mathcal{D}_{R_1,R_2,R}$ of triples $(p,q,s)$ for which*

$$\mathrm{BHT} : L^p(\mathbb{R}; L^{R_1}(\mathcal{W},\mu)) \times L^q(\mathbb{R}; L^{R_2}(\mathcal{W},\mu)) \to L^s(\mathbb{R}; L^R(\mathcal{W},\mu)).$$

*In addition, $\mathcal{D}_{R_1,R_2,R} = \mathcal{D}_{r_1^1, r_2^1, r^1}$.*

**Remark.** (1) The vector spaces $L^r(\mathcal{W}_j, \Sigma_j, \mu_j)$ can be both discrete $\ell^r$ spaces or the Euclidean $L^r(\mathbb{R})$ spaces. For our applications, they are going to be either of these.

(2) If the exponents $R_1 = (r_1^1, \dots, r_1^n)$, $R_2 = (r_2^1, \dots, r_2^n)$ and $R = (r^1, \dots, r^n)$ are in the "local $L^2$" range, then the multiple vector-valued inequalities hold for any $(p, q, s) \in \text{Range(BHT)}$. As particular cases, we mention

$$\text{BHT} : L^p(\ell^2(\ell^\infty)) \times L^q(\ell^\infty(\ell^2)) \to L^s(\ell^2(\ell^2)),$$
$$\text{BHT} : L^p(\ell^2(\ell^\infty)) \times L^q(\ell^2(\ell^2)) \to L^s(\ell^1(\ell^2))$$

for any $(p, q, s) \in \text{Range(BHT)}$.

Also, for proving an equivalent of Theorem 6 in mixed norm spaces, we need the more complex version

$$\text{BHT} : L_x^{p_1}(L_y^{p_2}(\ell^\infty(\ell^2))) \times L_x^{q_1}(L_y^{q_2}(\ell^2(\ell^2))) \to L_x^{s_1}(L_y^{s_2}(\ell^2(\ell^1))).$$

(3) As mentioned earlier, multiple vector-valued estimates for BHT play an important role in estimating $\text{BHT} \otimes \Pi^{\otimes n}$. In the case $n = 1$, one can obtain estimates for $\text{BHT} \otimes \Pi$ in the Banach range by using duality and vector-valued inequalities of the type

$$\text{BHT} : L^p(\ell^2) \times L^q(\ell^\infty) \to L^s(\ell^2) \quad \text{and} \quad \text{BHT} : L^p(\ell^\infty) \times L^q(\ell^2) \to L^s(\ell^2).$$

However, $\ell^1$-valued estimates cannot be avoided for $n \geq 3$, for example, if $\Pi \otimes \Pi \otimes \Pi$ has the form

$$\Pi \otimes \Pi \otimes \Pi(f, g)(x, y, z) = \sum_{k,l,m} Q_k^1 Q_l^2 P_m^3 \big(P_k^1 Q_l^2 Q_m^3 f \cdot Q_k^1 P_l^2 Q_m^3\big)(x, y, z).$$

This is in part the novelty of our approach in Theorem 6, and it contrasts with the situation of classical Calderón–Zygmund operators, where $\ell^1$-valued estimates cannot be expected.

(4) The optimality of the range in Theorem 7 or that in Theorem 8 remains without answer, for now. Since we use in our proofs the model operator for BHT, the obstructions appearing are similar to those in [Lacey and Thiele 1999]. These are described in the constraint $C(r_1, r_2, r')$ on page 1954.

Equally important are multiple vector-valued inequalities for paraproducts, as they are essential in proving Theorem 4.

**Theorem 9.** *For any tuples* $R_1 = (r_1^1, \dots, r_1^n)$, $R_2 = (r_2^1, \dots, r_2^n)$ *and* $R = (r^1, \dots, r^n)$ *satisfying componentwise* $1 < R_1, R_2 \leq \infty$, $1 \leq R < \infty$, *and* $\frac{1}{R_1} + \frac{1}{R_2} = \frac{1}{R}$,

$$\Pi : L^p(\mathbb{R}; L^{R_1}(\mathcal{W}, \mu)) \times L^q(\mathbb{R}; L^{R_2}(\mathcal{W}, \mu)) \to L^s(\mathbb{R}; L^R(\mathcal{W}, \mu)),$$

*provided* $1 < p, q \leq \infty$, $\frac{1}{2} < s < \infty$, *and* $\frac{1}{p} + \frac{1}{q} = \frac{1}{s}$.

In other words, vector-valued estimates for paraproducts exist within the same range as that of scalar paraproducts. This is also the case with classical Calderón–Zygmund operators.

***Original motivation.*** We now describe the previously mentioned Rubio de Francia operator for iterated Fourier integrals, and the context where it appeared. AKNS systems are systems of differential equations of the form

$$u' = i\lambda Du + Au, \tag{10}$$

where $u = [u_1, \dots, u_n]^t$ is a vector-valued function defined on $\mathbb{R}$, $D$ is a diagonal $n \times n$ matrix with real and distinct entries $d_1, d_2, \dots, d_n$, and $A = (a_{jk}(\cdot))_{j,k=1}^n$ is a matrix-valued function defined on $\mathbb{R}$ and such that $a_{jj} \equiv 0$ for all $1 \le j \le n$.

Then one would like to prove that the solutions $u_j^\lambda$ (which depend on $\lambda$ as well) are bounded "for all times"; that is,

$$\|u_j^\lambda\|_\infty < \infty \quad \text{for a.e. } \lambda \text{ and all } 1 \le j \le n. \tag{11}$$

We want to have such an estimate under the weakest possible assumptions, so we only require the entries of the potential matrix $A$ to be integrable in some $L^p$ spaces:

$$a_{jk}(\cdot) \in L^{p_{jk}}(\mathbb{R}) \quad \text{for all } 1 \le j, k \le n,\ j \ne k.$$

In the case of an upper triangular matrix $A$, whose entries are functions $g_k \in L^{p_k}$, the solutions $u_j(t)$ at a fixed time $t$ are a finite sum of expressions of the form

$$C \int_{x_1 < \cdots < x_m < t} g_1(x_1) \cdots g_m(x_m) e^{i\lambda(\alpha_1 x_1 + \cdots + \alpha_m x_m)}\, dx_1 \cdots dx_m.$$

Here $m \le n$ and $\alpha_k \ne 0$ for all $k$, as a consequence of $d_1 \ne \cdots \ne d_n$. Hence the problem (11) reduces to estimating

$$\widetilde{C}_m^\alpha(g_1, g_2, \dots, g_m)(\lambda) := \sup_t \left| \int_{x_1 < \cdots < x_m < t} g_1(x_1) \cdots g_m(x_m) e^{i\lambda(\alpha_1 x_1 + \cdots + \alpha_m x_m)}\, dx_1 \cdots dx_m \right|.$$

It was proved by Christ and Kiselv [2001a; 2001b] that $\widetilde{C}_m^\alpha$ is a bounded operator:

$$\left\| \widetilde{C}_m^\alpha(g_1, \dots, g_m) \right\|_{s_m} \lesssim \prod_{k=1}^m \|g_k\|_{p_k}$$

for all $1 \le p_k < 2$ such that $\frac{1}{s_m} = \frac{1}{p_1'} + \cdots + \frac{1}{p_m'}$.

On the other hand, if the entries of the matrix $A$ are $L^2$ functions, the previous expression becomes equivalent to

$$\sup_t \left| \int_{x_1 < \cdots < x_m < t} \hat{f}_1(x_1) \cdots \hat{f}_m(x_m) e^{i\lambda(\alpha_1 x_1 + \cdots + \alpha_m x_m)}\, dx_1 \cdots dx_m \right|, \tag{12}$$

denoted $C_m^\alpha(f_1, \dots, f_m)(\lambda)$. For $m = 1$, this is exactly the Carleson operator, while $m = 2$ corresponds to the bi-Carleson operator of [Muscalu et al. 2006b], both of which are known to be bounded operators (with the remark that for the bi-Carleson, the $\alpha_k$ need to satisfy some nondegeneracy condition):

$$\|C_2^\alpha(h_1, h_2)\|_{s_2} \lesssim \|h_1\|_{p_1} \|h_2\|_{p_2}$$

for $1 < p_1, p_2 \le \infty$, $\frac{1}{s_2} = \frac{1}{p_1} + \frac{1}{p_2}$, and $\frac{2}{3} < s_2 < \infty$.

Moreover, if instead of considering the sup in the expression (12), we look at the limiting behavior $\lim_{t\to\infty} u_j(t)$, then we encounter iterated Fourier integrals, for example, the BHT operator as seen in (1), or the bi-est operator of [Muscalu et al. 2004b]:

$$\int_{\xi_1<\xi_2<\xi_3} \hat{f}_1(\xi_1)\,\hat{f}_2(\xi_2)\,\hat{f}_3(\xi_3)\,e^{2\pi i x(\xi_1+\xi_2+\xi_3)}\,d\xi_1\,d\xi_2\,d\xi_3.$$

Now we consider the following mixed problem: The matrix $A$ is the sum of a lower triangular matrix with entries $\hat{f}_k \in L^2$, and an upper triangular matrix with entries $g_k \in L^{p_k}$, where $1 \le p_k < 2$. Using Picard iteration, the solutions $u_j(t)$ can be expressed as a series of terms of the form

$$C\int_R \hat{f}_{11}(\xi_{11})\cdots\hat{f}_{1m_1}(\xi_{1m_1})g_{21}(x_{21})\cdots g_{2n_2}(x_{2n_2})\cdots\hat{f}_{l1}(\xi_{l1})\cdots\hat{f}_{lm_l}(\xi_{lm_l})\,dx\,d\xi,$$

where $R = \{\xi_{11} < \cdots < \xi_{1m_1} < x_{21} < \cdots < x_{2n_2} < \cdots < \xi_{l1} < \cdots < \xi_{lm_l} < t\}$.

The simplest of these operators, where the sup is dropped, is given by

$$M(f_1, f_2, g)(\xi) = \int_{x_1<x_2<x_3} \hat{f}_1(x_1)\,\hat{f}_2(x_2)g(x_3)\,e^{2\pi i\xi(x_1+x_2+x_3)}\,dx_1\,dx_2\,dx_3, \tag{13}$$

where $f_1 \in L^{p_1}$, $f_2 \in L^{p_2}$, $1 < p_1, p_2 < \infty$, and $g \in L^p$ with $1 < p < 2$. The techniques from [Christ and Kiselev 1998; 2001a; 2001b], akin to those used by Paley [1931], are based on a dyadic filtration associated to one of the functions. This involves a structure on $\mathbb{R}$ similar to that of the dyadic mesh: on every level of the filtration, one has a partition of $\mathbb{R}$, and passing to the next level of the filtration means refining the previous partition. We want to use $g$ in order to obtain this structure and for simplicity we assume $\|g\|_p = 1$. Define the function

$$\varphi(x) = \int_{-\infty}^{x} |g(y)|^p\,dy.$$

Its image is the unit interval $[0, 1]$, and the filtration will consist of preimages through $\varphi$ of the collection $\mathcal{D}$ of dyadic intervals in $[0, 1]$. Because $\varphi$ is increasing, whenever $x_2 < x_3$ we have $0 \le \varphi(x_2) \le \varphi(x_3) \le 1$. Hence there exists a unique dyadic interval $\omega \subset [0, 1]$ such that $\varphi(x_2)$ is contained in the left half of $\omega$, which we denote $\omega_L$, while $\varphi(x_3)$ is contained in the right half $\omega_R$. To simplify notation, we identify $\varphi^{-1}(\omega)$ with $\omega$.

Then the operator $M$ can be written as

$$\begin{aligned}\sum_{\omega\in\mathcal{D}} \int_{\substack{x_1<x_2\\ x_2\in\omega_L, x_3\in\omega_R}} &\hat{f}_1(x_1)\,\hat{f}_2(x_2)g(x_3)\,e^{2\pi i\xi(x_1+x_2+x_3)}\,dx_1\,dx_2\,dx_3\\ &= \sum_{\omega}\int_{\substack{x_1<x_2\\ x_1,x_2\in\omega_L, x_3\in\omega_R}} \hat{f}_1(x_1)\,\hat{f}_2(x_2)g(x_3)\,e^{2\pi i\xi(x_1+x_2+x_3)}\,dx_1\,dx_2\,dx_3 \qquad (14)\\ &\quad+\sum_{\omega}\int_{\substack{x_1<L(\omega_L)\\ x_2\in\omega_L, x_3\in\omega_R}} \hat{f}_1(x_1)\,\hat{f}_2(x_2)g(x_3)\,e^{2\pi i\xi(x_1+x_2+x_3)}\,dx_1\,dx_2\,dx_3. \qquad (15)\end{aligned}$$

Here $L(\omega_L)$ denotes the left endpoint of the interval $\omega_L$. We call the operators in (14) and (15) $M_1$ and $M_2$ respectively. The first term $M_1$ accounts for the occurrence of arbitrary intervals (they are in fact

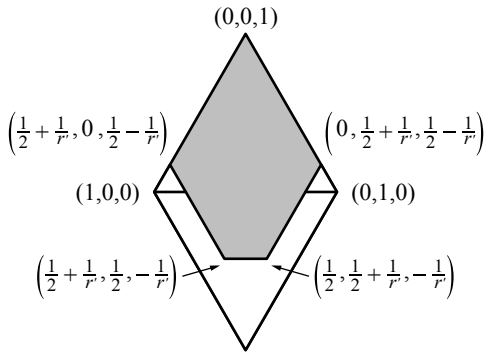



**Figure 5.** Range for $T_r$ operator for $1 \le r \le 2$.

$\varphi^{-1}(\omega_L)$), and this combined with Hölder's inequality motivates the operator

$$T_r(f,g)(x) = \Big( \sum_{k=1}^{N} \Big| \int_{a_k<\xi_1<\xi_2<b_k} \hat{f}(\xi_1)\,\hat{g}(\xi_2)\, e^{2\pi i x(\xi_1+\xi_2)}\, d\xi_1\, d\xi_2 \Big|^r \Big)^{\frac{1}{r}}. \tag{16}$$

We have the following result:

**Theorem 10.** *If* $1 \le r \le 2$, *then*

$$\|T_r(f,g)\|_s \lesssim \|f\|_p \|g\|_q$$

*whenever* $\frac{1}{p}+\frac{1}{q}=\frac{1}{s}$, *and* $p,q,s$ *satisfy*

$$0 \le \tfrac{1}{p}, \tfrac{1}{q} < \tfrac{1}{2}+\tfrac{1}{r}, \qquad -\tfrac{1}{r'} < \tfrac{1}{s'} < 1.$$

*On the other hand, if* $r \ge 2$, *then* $T_r$ *is a bounded operator with the same range as the* BHT *operator*; *see Figure 5.*

In Section 7 we will show how both $M_1$ and $M_2$ are bounded operators:

**Theorem 11.** *The operators* $M_1$ *and* $M_2$ *satisfy the following*:

$$M_1 : L^{p_1} \times L^{p_2} \times L^{p} \to L^{q} \quad \textit{provided } 1<p<2 \textit{ and } \tfrac{1}{p_1}+\tfrac{1}{p_2}+\tfrac{1}{p'}=\tfrac{1}{q},$$

*while*

$$M_2 : L^{p_1} \times L^{p_2} \times L^{p} \to L^{q} \quad \textit{provided } 1<p<2,\ \tfrac{1}{p_2}+\tfrac{1}{p'}<1 \textit{ and } \tfrac{1}{p_1}+\tfrac{1}{p_2}+\tfrac{1}{p'}=\tfrac{1}{q}.$$

*Hence* $M = M_1 + M_2$ *is a bounded operator from* $L^{p_1} \times L^{p_2} \times L^{p} \to L^{q}$ *provided* $1<p<2$, $\frac{1}{p_2}+\frac{1}{p'}<1$ *and* $\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p'}=\frac{1}{q}$.

However, as Robert Kesler [2015] noticed, the boundedness of the operator $M$ can also be proved by making use of a vector-valued extension for the "linear" operator $\mathrm{BHT}(f_1,\cdot)$. The constraint for the exponents is given by $\frac{1}{p_2}+\frac{1}{p'}<1$. So even if $M$ splits as $M = M_1 + M_2$ and the range of $M_1$ is larger, one gets the same range for $M$ through both methods.

Because the intervals $\{[a_k, b_k]\}_k$ are disjoint and arbitrary, we refer to $T_r$ as a *bilinear Rubio de Francia operator for iterated Fourier integrals*. Recall that *Rubio de Francia's square function* is the operator

$$f \mapsto \mathrm{RF}(f)(x) := \Big( \sum_{k=1}^{N} \Big| \int_{I_k} \hat{f}(\xi) e^{2\pi i \xi x}\, d\xi \Big|^2 \Big)^{\frac{1}{2}} = \Big( \sum_{k=1}^{N} |P_{I_k} f(x)|^2 \Big)^{\frac{1}{2}},$$

where $\{I_k = [a_k, b_k]\}_{1 \le k \le N}$ is a family of disjoint intervals, and $P_I(f)$ denotes the Fourier projection of $f$ onto the interval $I$. Using vector-valued singular integrals theory, Rubio de Francia [1985] proved the boundedness of the RF operator on $L^p$ for $p \ge 2$. Interpolating this result with estimates for Carleson's operator [1966], one gets more generally that the operator

$$\mathrm{RF}_\nu(f)(x) := \Big( \sum_{k=1}^{N} |P_{I_k} f(x)|^\nu \Big)^{\frac{1}{\nu}}$$

is bounded on $L^p$, as long as $\frac{1}{p} + \frac{1}{\nu} < 1$.

In the particular case of a *lacunary* family of intervals (that is, $I_k = [2^{k-1}, 2^k]$ and $k \in \mathbb{Z}$), the above operator corresponds to a Littlewood–Paley square function with sharp cutoffs, which is bounded on $L^p(\mathbb{R})$ for any $1 < p < \infty$. Even more, the $L^p$ norm of the square function is comparable to the $L^p$ norm of the initial function:

$$C_p^{-1} \| f \|_p \le \Big\| \Big( \sum_{k \in \mathbb{Z}} \Big| \int_{\mathbb{R}} \mathbf{1}_{\{2^{k-1} \le \xi < 2^k\}} \hat{f}(\xi) e^{2\pi i x \xi}\, d\xi \Big|^2 \Big)^{\frac{1}{2}} \Big\|_p \le C_p \| f \|_p.$$

Rubio de Francia's theorem addresses the boundedness of a square function associated to an arbitrary family of intervals, and in this sense it is optimal: in the case $\nu = 2$, the condition $p \ge 2$ is necessary, while for $\nu > 2$, we need the strict inequality $\nu > p'$.

Returning to our operator $T_r$, note that it can also be regarded as a vector-valued bilinear Hilbert transform

$$T_r(f, g)(x) = \Big( \sum_k \big| \mathrm{BHT}(P_{I_k} f, P_{I_k} g)(x) \big|^r \Big)^{\frac{1}{r}},$$

because the multiplier of the BHT operator is equivalent to $\mathbf{1}_{\{\xi_1 < \xi_2\}}$, as seen in (1).

Using solely Khintchine's inequality, it was proved in [Grafakos and Li 2006] that

$$\Big\| \Big( \sum_k |\mathrm{BHT}(f_k, g_k)|^2 \Big)^{\frac{1}{2}} \Big\|_s \lesssim \Big\| \Big( \sum_k |f_k|^2 \Big)^{\frac{1}{2}} \Big\|_p \Big\| \Big( \sum_k |g_k|^2 \Big)^{\frac{1}{2}} \Big\|_q.$$

This implies the boundedness of $T_r$ for $r \ge 2$, $p, q \ge 2$. But this is a very limited range, and in order to obtain estimates in the case $p < 2$ or $q < 2$, one needs the full power of vector-valued extensions.

We note that our estimates for the operator $T_r$ are sharp, in the sense that the same estimates are satisfied by

$$(f, g) \mapsto \Big( \sum_k \big| P_{I_k} f(x) \cdot P_{I_k} g(x) \big|^r \Big)^{\frac{1}{r}}. \tag{17}$$

In (17), $\mathrm{BHT}(P_{I_k} f, P_{I_k} g)$ is replaced by the product of the functions $P_{I_k} f \cdot P_{I_k} g$. In general, the best one can hope for a bilinear Fourier multiplier operator is that it satisfies the same $L^p$ estimates as the product $(f, g) \mapsto f \cdot g$, and this is the case for $T_r$.

Moreover, in the special case of lacunary dyadic intervals, for any $1 \le r < \infty$, we have that

$$(f, g) \mapsto \Big( \sum_k \Big| \int_{2^k < \xi < \eta < 2^{k+1}} \hat{f}(\xi)\, \hat{g}(\eta)\, e^{2\pi i x(\xi+\eta)}\, d\xi\, d\eta \Big|^r \Big)^{\frac{1}{r}}$$

is a bounded operator from $L^p \times L^q$ to $L^s$ for any $(p, q, s) \in \mathrm{Range}(\mathrm{BHT})$. The cases $p = \infty$ and $q = \infty$ cannot be obtained directly, but follow by duality.

Our initial proof of Theorem 10 did not involve vector-valued bilinear Hilbert transform operators, but it was built around localizations of BHT, in conjunction with several stopping times. Afterwards we realized that this method is suitable for other general situations, which eventually led to the development of the helicoidal method. This applies to paraproducts, BHT, the Carleson operator, the Rubio de Francia operator, etc. In the study of the $T_r$ operator, the stopping times were dictated by level sets of linear Rubio de Francia operators: $\mathrm{RF}_{r_1}(f)$ and $\mathrm{RF}_{r_2}(g)$. For the vector-valued BHT, the three stopping times that are used for estimating the trilinear form are dictated by level sets of

$$\Big( \sum_k |f_k|^{r_1} \Big)^{\frac{1}{r_1}}, \quad \Big( \sum_k |g_k|^{r_2} \Big)^{\frac{1}{r_2}} \quad \text{and} \quad \Big( \sum_k |h_k|^{r'} \Big)^{\frac{1}{r'}}.$$

The method of the proof is described in more detail in Section 2.5.

Lastly, we want to point out an interesting connection with another open problem in time-frequency analysis: the boundedness of the Hilbert transform along vector fields. More exactly, if $v : \mathbb{R}^2 \to \mathbb{R}^2$ is a nonvanishing measurable vector field, then one defines the Hilbert transform along $v$ as

$$H_v f(x, y) = \mathrm{p.v.} \int_{\mathbb{R}} f\big((x, y) - t \cdot v(x, y)\big) \frac{dt}{t}.$$

It was conjectured by Stein that $H_v$ is a bounded operator on $L^2$ whenever $v$ is Lipschitz. Some partial results in this direction are known in the case of a one-variable vector field. M. Bateman and C. Thiele [2013] proved the $L^p$ boundedness of $H_v$ for $\frac{3}{2} < p < \infty$ and provided that $v(x, y) = v(x, 0)$.

The proof makes use of the Littlewood–Paley square function in the second variable and restrictions to certain fixed sets $G$ and $H$, together with single annulus estimates for $H_v$ from [Bateman 2013]. In the special case when $f(x, y) = g(x)h(y)$, estimates for the variational Carleson from [Oberlin et al. 2012] yield the same result whenever $p > \frac{4}{3}$. It is still not known if this can be extended to general functions $f(x, y)$, or whether one can push the lower bound for $p$ below $\frac{4}{3}$.

Silva [2014] uses ideas similar to the ones described above, obtaining in this way vector-valued extensions for BHT whenever $\frac{4}{3} < r < 4$. Our methods allow us to prove that vector-valued extensions exist for any $1 \le r < \infty$ (in fact, for any triple $(r_1, r_2, r)$). It would be interesting to understand whether the localization argument that we are employing can be transferred to the study of the Hilbert transform along vector fields.

Besides having sharp estimates for the local version of the operator, the structure of the intervals chosen through the triple stopping time can play a role in itself. The collections of intervals constitute a maximal covering for the level sets of certain maximal operators, and for that reason, they form a *sparse* collection of intervals (in the sense of [Lerner 2013]). From here, weighted estimates can be deduced, and a similar approach was carried out in [Culiuc et al. 2016].

The rest of the paper is organized as follows: in Section 2 we recall some definitions and results regarding multilinear operators. The helicoidal method is described in detail in Section 2.5. Multiple vector-valued extensions for BHT are presented in Section 3, and those for paraproducts in Section 4. Following in Section 5 are the estimates for $\mathrm{BHT} \otimes \Pi^{\otimes n}$. The Leibniz rules are a modification of mixed norm $L^p$ estimates for $\Pi \otimes \Pi$ and are discussed in Section 6. The Rubio de Francia theorem for iterated Fourier integrals and its application to the AKNS system problem appear in Section 7.

## 2. Some classical results on the bilinear Hilbert transform

In this paper we use Chapter 6 of [Muscalu and Schlag 2013] as a black box, but we recall a few definitions and results to ease the reading of the presentation. Essential here are the notions of *size* and *energy*, which are quantities associated to certain subsets of the phase-frequency space.

**Notation.** For any interval $I \subset \mathbb{R}$, define

$$\tilde{\chi}_I(x) := \left(1 + \frac{\operatorname{dist}(x, I)}{|I|}\right)^{-100}.$$

The mesh of dyadic intervals is denoted by $\mathcal{D}$.

**Definition 12.** A *tile* is a rectangle $P = I_P \times \omega_P$ with the property that $I_P, \omega_P \in \mathcal{D}$ or $\omega_P$ is in a shifted variant of $\mathcal{D}$. We define a *tritile* to be a tuple $P = (P_1, P_2, P_3)$ where each $P_i$ is a tile as defined above and the spatial intervals are the same: $I_{P_i} = I_P$ for all $1 \le i \le 3$.

**Definition 13** (order relation). Given two tiles $P$ and $P'$, we say $P' < P$ if $I_{P'} \subsetneq I_P$ and $\omega_P \subset 3\omega_{P'}$, and $P' \le P$ if $P' < P$ or $P' = P$. Also, $P' \lesssim P$ if $I_{P'} \subset I_P$ and $\omega_P \subseteq 100\omega_{P'}$, and $P' \lesssim' P$ if $P' \lesssim P$ but $P' \not\leq P$.

**Definition 14.** A collection $\mathbb{P}$ of tritiles is said to have *rank 1* if for any $P, P' \in \mathbb{P}$ the following conditions are satisfied:

- If the tritiles are distinct, i.e., $P \neq P'$, then $P'_j \neq P_j$ for all $1 \le j \le 3$.
- If $\omega_{P_{j_0}} = \omega_{P'_{j_0}}$ for some $j_0$, then $\omega_{P_j} = \omega_{P'_j}$ for all $1 \le j \le 3$.
- If $P'_{j_0} \le P_{j_0}$ for some $j_0$, then $P'_j \lesssim P_j$ for all $1 \le j \le 3$.
- If in addition to $P'_{j_0} \le P_{j_0}$ one also assumes $|I_{P'}| \ll |I_P|$, then $P'_j \lesssim' P_j$ for all $j \neq j_0$.

**Definition 15.** Let $\mathbb{P}$ be a sparse rank 1 collection of tritiles, and let $1 \le j \le 3$. A subcollection $T$ of $\mathbb{P}$ is called a *$j$-tree* if and only if there exists a tritile $P_T$ (called the *top* of the tree) such that $P_j \le P_{T,j}$ for all $P \in T$. We write $I_T$ for $I_{P_T}$ and $\omega_{T_j}$ for $\omega_{P_T,j}$ and we say $T$ is a *tree* if it is a $j$-tree for some $1 \le j \le 3$.

**Definition 16.** Let $1 \le i \le 3$. A finite sequence of trees $T_1, \dots, T_M$ is said to be a chain of strongly $i$-disjoint trees if and only if

(i) $P_i \neq P_i'$ for every $P \in T_{l_1}$ and $P' \in T_{l_2}$, with $l_1 \neq l_2$;

(ii) whenever $P \in T_{l_1}$ and $P' \in T_{l_2}$ with $l_1 \neq l_2$ are such that $2\omega_{P_i} \cap 2\omega_{P_i'} \neq \varnothing$, then if $|\omega_{P_i}| < |\omega_{P_i'}|$, one has $I_{P'} \cap I_{T_{l_1}} = \varnothing$, and if $|\omega_{P_i'}| < |\omega_{P_i}|$, one has $I_P \cap I_{T_{l_2}} = \varnothing$.

(iii) whenever $P \in T_{l_1}$ and $P' \in T_{l_2}$ with $l_1 < l_2$ are such that $2\omega_{P_i} \cap 2\omega_{P_i'} \neq \varnothing$ and $|\omega_{P_i}| = |\omega_{P_i'}|$, then $I_{P'} \cap I_{T_{l_1}} = \varnothing$.

**Definition 17.** Let $P$ be a tile. A *wave packet* on $P$ is a smooth function $\phi_P$ which has Fourier support inside $\frac{9}{10}\omega_P$ and is $L^2$-adapted to $I_P$ in the sense that

$$|\phi_P^{(l)}(x)| \le C_{l,M} \frac{1}{|I_P|^{\frac{1}{2}+l}} \left(1 + \frac{\operatorname{dist}(x, I_P)}{|I_P|}\right)^{-M} \tag{18}$$

for sufficiently many derivatives $l$ and any $M > 0$.

**2.1.** ***Model operator for*** **BHT.** A discretized model operator for BHT is given by

$$\mathrm{BHT}_{\mathbb{P}}(f,g)(x) = \sum_{P \in \mathbb{P}} \frac{1}{|I_P|^{\frac{1}{2}}} \langle f, \phi_{P_1}^1 \rangle \langle g, \phi_{P_2}^2 \rangle \phi_{P_3}^3(x), \tag{19}$$

where the family $\mathbb{P}$ of tritiles is sparse and has rank 1, while $(\phi_{P_j}^j)_{P \in \mathbb{P}}$ are wave packets associated to the tiles $P_j$. In some sense, the bilinear Hilbert transform is the canonical example of such an operator. Above we also included the definitions of *trees* and *chains of strongly disjoint trees* because they are essential in understanding such singular bilinear operators.

The model operator from (19) was introduced in [Lacey and Thiele 1999], and the bilinear Hilbert transform itself can be represented as an average of such shifted model operators. The detailed reduction can be found in [Muscalu and Schlag 2013, Chapter 6]. As a consequence, the boundedness of the bilinear Hilbert transform within Range(BHT) can be deduced from similar estimates for the model operator. Similarly, estimates for vector-valued and for the localized bilinear Hilbert transform will follow once we prove their equivalents for the model operator, and we will not insist on the exact distinction between the two.

It is worth mentioning however, that the model operator fails to be bounded for $s \le \frac{2}{3}$, leaving undecided the boundedness of the bilinear Hilbert transform itself for $\frac{1}{2} < s \le \frac{2}{3}$.

Bilinear operators are often studied with the use of the associated trilinear form. In the case of the (model operator for the) BHT operator, the trilinear form is given by

$$\Lambda_{\mathrm{BHT};\mathbb{P}}(f,g,h) = \sum_{P \in \mathbb{P}} \frac{1}{|I_P|^{\frac{1}{2}}} \langle f, \phi_{P_1}^1 \rangle \langle g, \phi_{P_2}^2 \rangle \langle h, \phi_{P_3}^3 \rangle. \tag{20}$$

**Definition 18.** If $\mathbb{P}$ is a collection of tritiles and $I_0$ is a dyadic interval, we denote by $\mathbb{P}(I_0)$ the tiles $P$ in $\mathbb{P}$ whose spatial interval $I_P$ is contained in $I_0$:

$$\mathbb{P}(I_0) := \{P \in \mathbb{P} : I_P \subseteq I_0\}.$$

**Definition 19.** Let $\mathbb{P}$ be a finite collection of tritiles, let $j \in \{1,2,3\}$, and let $f$ be an arbitrary function. We define the *size* of the sequence $\langle f, \phi^j_{P_j}\rangle_P$ by

$$\operatorname{size}\big(\langle f, \phi^j_{P_j}\rangle_P\big) := \sup_{T\subseteq\mathbb{P}} \bigg(\frac{1}{|I_T|}\sum_{P\in T} |\langle f, \phi^j_{P_j}\rangle|^2\bigg)^{\frac{1}{2}}, \tag{21}$$

where $T$ ranges over all trees in $\mathbb{P}$ that are $i$-trees for some $i \neq j$.

**Lemma 20** [Muscalu and Schlag 2013, Lemma 6.13]. *Let $j \in \{1,2,3\}$ and let $E$ be a set of finite measure. Then for every $|f| \le \mathbf{1}_E$ one has*

$$\operatorname{size}\big(\langle f, \phi^j_{P_j}\rangle_P\big) \lesssim \sup_{P\in\mathbb{P}} \frac{1}{|I_P|}\int_E \tilde{\chi}^M_{I_P}\,dx$$

*for all $M > 0$, with implicit constants depending on $M$.*

Thanks to Lemma 20, which is a consequence of the John–Nirenberg inequality, we can work with the simpler "sizes"

$$\operatorname{size} f \sim \sup_{P\in\mathbb{P}} \frac{1}{|I_P|}\int_{\mathbb{R}} |f|\cdot\tilde{\chi}^M_{I_P}\,dx,$$

where $M$ is some large number to be chosen later.

We will also need a size that behaves well with respect to localization. In the formula above we consider the supremum over the spacial intervals $I_P$ of the collection $\mathbb{P}$. In our proofs, we will need to compare $\operatorname{size}_{\mathbb{P}(I_0)} f$ and $(1/|I_0|)\int_{\mathbb{R}}|f|\cdot\tilde{\chi}_{I_0}\,dx$, so the following definition is natural:

**Definition 21.** If $I_0$ is a fixed dyadic interval, then we define

$$\widetilde{\operatorname{size}}_{\mathbb{P}(I_0)} f := \sup_{\substack{J\subseteq 3I_0\\ \exists P\in\mathbb{P}(I_0), I_P\subseteq J}} \frac{1}{|J|}\int_{\mathbb{R}} |f|\cdot\tilde{\chi}^M_J\,dx. \tag{22}$$

We note that for any function $f$,

$$\operatorname{size}_{\mathbb{P}(I_0)} f \le \widetilde{\operatorname{size}}_{\mathbb{P}(I_0)} f.$$

**Definition 22.** Let $\mathbb{P}$ be a finite collection of tritiles, $j \in \{1,2,3\}$ and let $f$ be a fixed function. We define the energy of the sequence $\langle f, \phi^j_{P_j}\rangle_P$ by

$$\operatorname{energy}\big(\langle f, \phi^j_{P_j}\rangle_P\big) := \sup_{n\in\mathbb{Z}} 2^n \sup_{\mathbb{T}}\bigg(\sum_{T\in\mathbb{T}} |I_T|\bigg)^{\frac{1}{2}}, \tag{23}$$

where $\mathbb{T}$ ranges over all chains of strongly $j$-disjoint trees in $\mathbb{P}$ (which are $i$-trees for some $i \neq j$) having the property that

$$\bigg(\sum_{P\in T} |\langle f, \phi^j_{P_j}\rangle|^2\bigg)^{\frac{1}{2}} \ge 2^n |I_T|^{\frac{1}{2}}$$

for all $T \in \mathbb{T}$ and such that

$$\bigg(\sum_{P\in T'} |\langle f, \phi^j_{P_j}\rangle|^2\bigg)^{\frac{1}{2}} \le 2^{n+1} |I_{T'}|^{\frac{1}{2}}$$

for all subtrees $T' \subseteq T \in \mathbb{T}$.

We have the following estimates for the trilinear form and energy:

**Proposition 23** [Muscalu and Schlag 2013, Proposition 6.12]. *Let $\mathbb{P}$ be a finite collection of tritiles. Then*

$$\Lambda_{\mathrm{BHT};\mathbb{P}}(f_1, f_2, f_3) \lesssim \prod_{j=1}^{3} \big(\mathrm{size}(\langle f_j, \phi^j_{P_j}\rangle_P)\big)^{\theta_j} \big(\mathrm{energy}(\langle f_j, \phi^j_{P_j}\rangle_P)\big)^{1-\theta_j}$$

*for any $0 \le \theta_1, \theta_2, \theta_3 < 1$ with $\theta_1 + \theta_2 + \theta_3 = 1$; the implicit constants depend on the $\theta_j$ but are independent of the other parameters.*

**Lemma 24** [Muscalu and Schlag 2013, Lemma 6.14]. *Let $j \in \{1, 2, 3\}$ and $f \in L^2(\mathbb{R})$. Then*

$$\mathrm{energy}\big(\langle f, \phi^j_{P_j}\rangle_P\big) \lesssim \|f\|_2.$$

However, for our specific problem we need more accurate estimates for the localized trilinear form. This will follow in Sections 2.4 and 3.1.

**2.2.** ***Interpolation.*** Since this is a fundamental tool in harmonic analysis, we recall a few facts about interpolation methods. We adapt the results from [Thiele 2006] and emphasize how the constants change through interpolation. In our applications, we need to keep track of the constants. Many of the proofs in the following sections are iterative, and the operatorial norm obtained after interpolation becomes a "size" on the subsequent step of the induction. We recall a few definitions and results, but we will be mainly using their generalization to Banach spaces.

**Definition 25.** For a subset $E \subset \mathbb{R}$ of finite measure, define

$$X(E) = \{f : |f| \le \mathbf{1}_E \text{ a.e.}\}.$$

We will denote by $V$ the linear span of all $X(E)$, which plays an important role because it is a dense subspace of all $L^p$ spaces for $1 \le p < \infty$.

**Definition 26.** A tuple $\alpha = (\alpha_1, \dots, \alpha_n)$ is called *admissible* if for all $1 \le i \le n$,

$$-\infty < \alpha_i < 1 \quad \text{and} \quad \alpha_1 + \cdots + \alpha_n = 1,$$

and there is at most one index $j_0$ so that $\alpha_{j_0} < 0$. We call an index *good* if $\alpha_i > 0$ and *bad* if $\alpha_i \le 0$.

**Definition 27.** A multilinear form $\Lambda : V \times \cdots \times V \to \mathbb{C}$ is of restricted type $\alpha = (\alpha_1, \dots, \alpha_n)$ with $0 \le \alpha_i \le 1$ if there exists a constant $C$ (possibly depending on $\alpha$) such that for each tuple $E = (E_1, \dots, E_n)$ of measurable subsets of $\mathbb{R}$ and for each tuple $f = (f_1, \dots, f_n)$ with $f_j \in X(E_j)$, we have

$$\big|\Lambda(f_1, \dots, f_n)\big| \le C \prod_j |E_j|^{\alpha_j}.$$

**Theorem 28** (similar to [Thiele 2006, Theorem 3.2]). *Let $\beta = (\beta_1, \dots, \beta_n)$ be a tuple of real numbers such that $\sum_j \beta_j = 1$ and $\beta_j > 0$ for all $j$. Assume $\Lambda$ is of restricted type $\alpha$ for all $\alpha$ in a neighborhood of $\beta$ satisfying $\sum_j \alpha_j = 1$, with constant $C(\alpha)$ depending continuously on $\alpha$. Then $\Lambda$ is of strong type $\beta$ with constant $C(\beta)$:*

$$\big|\Lambda(f_1, \dots, f_n)\big| \le C(\beta) \prod_{j=1}^{n} \|f_j\|_{\frac{1}{\beta_j}} \quad \text{for all } f_j \in V.$$

For multilinear operators, it often happens that the target space is an $L^p$ space with $0 < p < 1$. This is not a Banach space, but we can conclude the desired outcome by interpolating weak-$L^q$ estimates for $q$ in a neighborhood of $p$. Additionally, $L^{q,\infty}$ norms are dualized in the following way:

**Lemma 29** [Muscalu and Schlag 2013, Lemma 2.5]. *Let $0 < r \le 1$, and $A > 0$. Then the following statements are equivalent*:

(i) $\|f\|_{r,\infty} \le A$.

(ii) *For every set $E$ with $0 < |E| < \infty$, there exists a major subset $E' \subseteq E$ (i.e., $|E'| \ge |E|/2$) so that $|\langle f, \mathbf{1}_{E'}\rangle| \lesssim A|E|^{\frac{1}{r'}}$, where $\frac{1}{r} + \frac{1}{r'} = 1$. (Note that for $r \ne 1$, we have $r'$ is a negative number.)*

**Definition 30.** Let $\alpha$ be an $n$-tuple of real numbers and assume $\alpha_j \le 1$ for all $j$. An $n$-linear form $\Lambda$ is called of *generalized restricted type* $\alpha$ if there is a constant $C$ (possibly depending on $\alpha$) such that for all tuples $E = (E_1, \dots, E_n)$, there is an index $j_0$ and a major subset $E'_{j_0} \subseteq E_{j_0}$ so that for all tuples $f = (f_1, \dots, f_n)$ with $f_j \in X(E_j)$ for $j \ne j_0$ and $f_{j_0} \in X(E'_{j_0})$,

$$\big|\Lambda(f_1, \dots, f_n)\big| \le C \prod_{j=1}^{n} |E_j|^{\alpha_j}. \tag{24}$$

If a tuple $\alpha = (\alpha_1, \dots, \alpha_n)$ is good, then generalized restricted-type estimates coincide with restricted-type estimates:

**Proposition 31** (similar to [Thiele 2006, Lemma 3.6]). *If $\alpha = (\alpha_1, \dots, \alpha_n)$ is a good tuple, and $\Lambda$ is of generalized restricted type $\alpha$ with constant $C(\alpha)$ and the major subset corresponds to the index $j_0$, then $\Lambda$ is of restricted type $\alpha$ with constant $C(\alpha)/(1 - 2^{-j_0})$.*

**Theorem 32** [Thiele 2006, Theorem 3.8]. *Assume*

$$\Lambda = \big\langle T(f_1, \dots, f_{n-1}), f_n \big\rangle$$

*is of generalized restricted type $\beta$, where $\sum_j \beta_j = 1$. Assume $\beta_k > 0$ for $1 \le k \le n-1$ and $\beta_n \le 0$. Assume $\Lambda$ is also of generalized restricted type $\alpha$ with constant $C(\alpha)$ (continuously depending on $\alpha$) for all $\alpha$ in a neighborhood of $\beta$ satisfying $\sum_j \alpha_j = 1$. Then the multilinear operator $T$ satisfies*

$$\big\|T(f_1, \dots, f_{n-1})\big\|_{\frac{1}{(1-\beta_n)}} \le C(\beta) \prod_{j=1}^{n-1} \|f_j\|_{\frac{1}{\beta_j}}. \tag{25}$$

**2.3.** ***Interpolation for Banach-valued functions.*** The Banach space interpolation theory is very similar to the scalar version, the difference consisting in replacing the norm $|\cdot|$ on $\mathbb{C}$ by $\|\cdot\|_X$ on a Banach space $X$.

We say that $F \in L^p(\mathbb{R}; X)$ provided

$$\|F\|_{L^p(\mathbb{R};X)} := \left(\int_{\mathbb{R}} \|F(x)\|_X^p \, dx\right)^{\frac{1}{p}} < \infty.$$

The question of integrability of $F(x)$ is reduced to the Lebesgue integrability of $x \mapsto \|F(x)\|_X$. The set of vector-valued step functions is dense in $L^p(\mathbb{R}; X)$ and for this reason, similarly to the scalar case, it

will be enough to deal with function in

$$\big\{F : \|F(x)\|_X \le \mathbf{1}_E(x) \text{ a.e. } E \subset \mathbb{R} \text{ subset of finite measure}\big\}.$$

The linear span of such sets will be denoted $V_X$.

The multilinear form associated with an operator is obtained through dualization. More exactly,

$$\|F\|_{L^p(\mathbb{R};X)} := \sup_{\|G\|_{L^{p'}(\mathbb{R};X^*)}\le 1} \Big| \int_{\mathbb{R}} \langle G(x), F(x)\rangle \, dx \Big|$$

whenever $1 \le p < \infty$.

We will deal with a vector-valued multilinear (or multisublinear) operator of the form

$$\vec{T} : L^{p_1}(\mathbb{R}; X_1) \times \cdots \times L^{p_{n-1}}(\mathbb{R}; X_{n-1}) \to L^{p_n}(\mathbb{R}; X_n).$$

The multilinear form associated with this operator, $\Lambda : V_{X_1} \times \cdots \times V_{X_{n-1}} \times V_{X_n^*} \to \mathbb{C}$, is given by

$$\Lambda(F_1, \dots, F_{n-1}, F_n) = \int_{\mathbb{R}} \big\langle \vec{T}(F_1, \dots, F_{n-1})(x), F_n(x) \big\rangle \, dx.$$

The definitions and proofs from the scalar case are adaptable to the vector-valued situation. For completeness, we present them here, adapting the equivalent statements from [Thiele 2006].

**Definition 33.** A tuple $\alpha = (\alpha_1, \dots, \alpha_n)$ is called admissible if $\alpha_1 + \cdots + \alpha_n = 1$, $\alpha_1, \dots, \alpha_n < 1$ and for at most one index $j_0$ we have $\alpha_{j_0} < 0$.

A multisublinear form $\Lambda$ as above is of restricted type $\alpha = (\alpha_1, \dots, \alpha_n)$ for a good admissible tuple $\alpha$ if there exists a constant $C$ so that for each tuple $E = (E_1, \dots, E_n)$ of measurable subsets of $\mathbb{R}$, and for each tuple $F = (F_1, \dots, F_n)$ with $\|F_j\|_X \le \mathbf{1}_{E_j}$, we have

$$\big|\Lambda(F_1, \dots, F_n)\big| \le C |E_1|^{\alpha_1} \cdots |E_n|^{\alpha_n}.$$

**Proposition 34** (equivalent of [Thiele 2006, Theorem 3.2]). *Let $\beta = (\beta_1, \dots, \beta_n)$ be an admissible tuple of real numbers such that $\beta_j > 0$ for all $j$. Assume that $\Lambda$ is of restricted type $\alpha$ for all admissible tuples $\alpha$ in a neighborhood of $\beta$. Then there is a constant $C$ such that for all $F_j \in V_{X_j}$,*

$$\big|\Lambda(F_1, \dots, F_n)\big| \le C \|F_1\|_{L^{1/\beta_1}(\mathbb{R};X_1)} \cdots \|F_n\|_{L^{1/\beta_n}(\mathbb{R};X_n)}.$$

**Definition 35.** Let $\alpha$ be an admissible tuple; the $n$-sublinear form $\Lambda$ is of *generalized restricted type $\alpha$* if there is a constant $C$ such that for all tuples $E = (E_1, \dots, E_n)$ there is an index $j_0$ and a major subset $E'_{j_0}$ of $E_{j_0}$ (that is, $|E'_{j_0}| \ge |E_{j_0}|/2$) such that for all tuples $F = (F_1, \dots, F_n)$ with $\|F_j\|_{X_j} \le \mathbf{1}_{E_j}$ for $j \ne j_0$, and $\|F_{j_0}\|_{X_{j_0}} \le \mathbf{1}_{E'_{j_0}}$, we have

$$\big|\Lambda(F_1, \dots, F_n)\big| \le C \prod_j |E_j|^{\alpha_j}.$$

**Proposition 36.** *If $\Lambda$ is of generalized restricted type $\alpha = (\alpha_1, \dots, \alpha_n)$, and $\alpha_j > 0$ for all $j$, then $\Lambda$ is of restricted type $\alpha$.*

On the other hand, if one of the indices $\alpha_j$ is $\leq 0$, the generalized restricted-type implies only weak-$L^p$ estimates. This works in the case when the multisublinear form is given by

$$\Lambda(F_1,\dots,F_n)=\int_{\mathbb{R}}\big\langle \vec{T}(F_1,\dots,F_{n-1})(x),F_n(x)\big\rangle\,dx, \tag{26}$$

and corresponds to an operator $\vec{T}$ defined on $V_{X_1}\times\cdots\times V_{X_{n-1}}$ and taking values in $V_{X_n}$.

**Proposition 37.** *Let $\Lambda$ be a multisublinear form as in (26), and $\alpha=(\alpha_1,\dots,\alpha_n)$ an admissible tuple with $\alpha_n\leq 0$. Assuming that $\Lambda$ is of generalized restricted type $\alpha$, we have*

$$\lambda\big|\{x:\big\|\vec{T}(F_1,\dots,F_{n-1})(x)\big\|_{X_n}>\lambda\}\big|^{\frac{1}{1-\alpha_n}}\leq A\prod_{j=1}^{n-1}|E_j|^{\alpha_j}$$

*for all tuples $F=(F_1,\dots,F_{n-1})$ with $\|f_j\|_{X_j}\leq \mathbf{1}_{E_j}$.*

**Proposition 38.** *Assume $\Lambda$ is of generalized restricted type $\beta$, where $\beta$ is an admissible tuple with $\beta_n\leq 0$. Assume $\Lambda$ is also of generalized restricted type $\alpha$ for all admissible tuples $\alpha$ in a neighborhood of $\beta$. Then $\vec{T}$ satisfies*

$$\big\|\vec{T}(F_1,\dots,F_{n-1})\big\|_{L^{1/(1-\beta_n)}(\mathbb{R};X_n)}\leq C\prod_{j=1}^{n-1}\|F_j\|_{L^{1/\beta_j}(\mathbb{R};X_j)}. \tag{27}$$

The proofs of the last two propositions follow exactly the same ideas as those corresponding to the scalar case, with very minor differences.

**2.4.** ***A few technical lemmas.*** In this section, we present a few results that will be useful later on for estimating a trilinear form associated to a collection $\mathbb{P}$ of tritiles well-localized in space: $I_P\subset I_0$ for all $P\in\mathbb{P}$.

**Lemma 39.** *If $I_0$ is a fixed dyadic interval, $k\in\mathbb{Z}^+$, and $f$ is a function such that*

$$2^{k-1}\leq\frac{\operatorname{dist}(\operatorname{supp} f,I_0)}{|I_0|}\leq 2^k,$$

*then*

$$\operatorname{energy}_{\mathbb{P}(I_0)}f\lesssim 2^{Mk}\|f\|_2.$$

*Proof.* Following Definition 22, there exists a collection $\mathbb{T}$ of $j$-disjoint trees $T\in\mathbb{T}\subseteq\mathbb{P}(I_0)$, so that

$$(\operatorname{energy}_{\mathbb{P}(I_0)}f)^2\sim\sum_{T\in\mathbb{T}}\sum_{P\in T}\big|\langle f,\phi_{P_j}\rangle\big|^2.$$

We define $\mathcal{T}:=\bigcup_{T\in\mathbb{T}}\bigcup_{P\in T}P$, the collection of all tiles in $\mathbb{T}$, and estimate the right-hand side of the expression above:

$$\sum_{T\in\mathbb{T}}\sum_{P\in T}\big|\langle f,\phi_{P_j}\rangle\big|^2\lesssim\sum_{m\geq 0}\sum_{\substack{I\subseteq I_0\\|I|=2^{-m}|I_0|}}\sum_{\substack{P\in\mathcal{T}\\I_P=I}}\big|\langle f,\phi_{P_j}\rangle\big|^2.$$

The collection of tiles $P \in \mathcal{T}$ with $I_P = I$ for a fixed interval $I$ are all disjoint in frequency. In fact, since they are of the same scale, they are translations of some fixed tile and hence

$$\sum_{\substack{P\in\mathcal{T}\\ I_P=I}} \big|\langle f,\phi_{P_j}\rangle\big|^2 \lesssim \int_{\mathbb{R}} |f(x)|^2 \left(1+\frac{\operatorname{dist}(x,I)}{|I|}\right)^{-2M} dx.$$

This implies

$$\begin{aligned}\sum_{T\in\mathbb{T}}\sum_{P\in T}\big|\langle f,\phi_{P_j}\rangle\big|^2 &\lesssim \sum_{m\geq 0}\sum_{\substack{I\subseteq I_0\\ |I|=2^{-m}|I_0|}} \int_{\mathbb{R}} |f(x)|^2 \left(1+\frac{\operatorname{dist}(x,I)}{|I|}\right)^{-2M} dx\\ &\lesssim \sum_{m\geq 0}\sum_{\substack{I\subseteq I_0\\ |I|=2^{-m}|I_0|}} \|f\|_2^2\, 2^{-2kM}\left(\frac{|I_0|}{|I|}\right)^{-2M}\\ &\lesssim \|f\|_2^2\, 2^{-2kM}\sum_{m\geq 0} 2^{-mM}\\ &\lesssim \|f\|_2^2\, 2^{-2kM}.\end{aligned}$$

□

On the other hand, if $f$ is supported inside $5I_0$, we know from Lemma 24, that $\operatorname{energy}_{\mathbb{P}(I_0)} f \lesssim \|f\|_2$.

Since the collection $\mathbb{P}(I_0)$ is localized in space on the interval $I_0$, we have the following estimate for the trilinear form $\Lambda_{\mathrm{BHT};\mathbb{P}(I_0)}$:

**Lemma 40** (refinement of [Muscalu and Schlag 2013, Proposition 6.12]). *The trilinear form $\Lambda_{\mathrm{BHT};\mathbb{P}(I_0)}$ satisfies*

$$\begin{aligned}&\big|\Lambda_{\mathrm{BHT};\mathbb{P}(I_0)}(f,g,h)\big|\\ &\quad\lesssim (\operatorname{size}_{\mathbb{P}(I_0)} f)^{\theta_1}(\operatorname{size}_{\mathbb{P}(I_0)} g)^{\theta_2}(\operatorname{size}_{\mathbb{P}(I_0)} h)^{\theta_3}\|f\cdot\tilde{\chi}_{I_0}\|_2^{1-\theta_1}\|g\cdot\tilde{\chi}_{I_0}\|_2^{1-\theta_2}\|h\cdot\tilde{\chi}_{I_0}\|_2^{1-\theta_3}\end{aligned} \tag{28}$$

*for any $0\leq\theta_1,\theta_2,\theta_3<1$, with $\theta_1+\theta_2+\theta_3=1$; the implicit constants depend on the $\theta_j$, but are independent of the other parameters.*

*Proof.* For any $l\geq 1$, we define $\mathcal{I}_l := 2^{l+1}I_0\setminus 2^l I_0$, and $\mathcal{I}_0 := 2I_0$. In this way, for any $x\in\mathcal{I}_l$, $1+\operatorname{dist}(x,I_0)/|I_0|\sim 2^l$.

We will be using the following decompositions:

$$f := \sum_{k_1\geq 0} f_{k_1} := \sum_{k_1\geq 0} f\cdot\mathbf{1}_{\mathcal{I}_{k_1}}, \tag{29}$$

and similarly,

$$g := \sum_{k_2\geq 0} g_{k_2} := \sum_{k_2\geq 0} g\cdot\mathbf{1}_{\mathcal{I}_{k_2}},\quad h := \sum_{k_3\geq 0} h_{k_3} := \sum_{k_3\geq 0} h\cdot\mathbf{1}_{\mathcal{I}_{k_3}}.$$

From Proposition 23, the trilinear form can be estimated by

$$\begin{aligned}\left|\Lambda_{\mathrm{BHT};\mathbb{P}(I_0)}(f,g,h)\right| &\lesssim \sum_{k_1,k_2,k_3} \left|\Lambda_{\mathrm{BHT};\mathbb{P}(I_0)}(f_{k_1},g_{k_2},h_{k_3})\right| \\ &\lesssim \sum_{k_1,k_2,k_3} (\mathrm{size}_{\mathbb{P}(I_0)}\, f_{k_1})^{\theta_1} (\mathrm{size}_{\mathbb{P}(I_0)}\, g_{k_2})^{\theta_2} (\mathrm{size}_{\mathbb{P}(I_0)}\, h_{k_3})^{\theta_3} \\ &\qquad (\mathrm{energy}_{\mathbb{P}(I_0)}\, f_{k_1})^{1-\theta_1} (\mathrm{energy}_{\mathbb{P}(I_0)}\, g_{k_2})^{1-\theta_2} (\mathrm{energy}_{\mathbb{P}(I_0)}\, h_{k_3})^{1-\theta_3}.\end{aligned}$$

We will only employ the extra decay in the energy; for the size, we have simply

$$\mathrm{size}_{\mathbb{P}(I_0)}\, f_{k_1} \lesssim \mathrm{size}_{\mathbb{P}(I_0)}\, f$$

uniformly in $k_1$.

On the other hand, since $f_{k_1}$ is supported on $\mathcal{I}_{k_1}$, Lemma 39 implies

$$\mathrm{energy}_{\mathbb{P}(I_0)}\, f_{k_1} \lesssim 2^{-k_1 M} \|f_{k_1}\|_2.$$

Hence we obtain

$$\begin{aligned}\left|\Lambda_{\mathrm{BHT};\mathbb{P}(I_0)}(f,g,h)\right| &\lesssim (\mathrm{size}_{\mathbb{P}(I_0)}\, f)^{\theta_1} (\mathrm{size}_{\mathbb{P}(I_0)}\, g)^{\theta_2} (\mathrm{size}_{\mathbb{P}(I_0)}\, h)^{\theta_3} \\ &\quad \cdot \sum_{k_1,k_2,k_3} (2^{-k_1 M}\|f_{k_1}\|_2)^{1-\theta_1} (2^{-k_2 M}\|g_{k_2}\|_2)^{1-\theta_2} (2^{-k_1 M}\|h_{k_3}\|_2)^{1-\theta_3}.\end{aligned}$$

The expressions in the last line are summable, via Hölder's inequality; more exactly, since $\theta_j < 1$,

$$\begin{aligned}\sum_{k_1\geq 0} 2^{-k_1 M \frac{1-\theta_1}{2}} \left(2^{-k_1 \frac{M}{2(1-\theta_1)}}\|f_{k_1}\|_2\right)^{1-\theta_1} &\lesssim \left(\sum_{k_1} 2^{-k_1 M \frac{1-\theta_1}{1+\theta_1}}\right)^{\frac{1+\theta_1}{2}} \left(\sum_{k_1} 2^{-k_1 \frac{M}{1-\theta_1}}\|f_{k_1}\|_2^2\right)^{\frac{1-\theta_1}{2}} \\ &\lesssim \|f\cdot\tilde{\chi}_{I_0}\|_2^{1-\theta_1}\end{aligned}$$

for $M$ sufficiently large. Note the implicit constants will depend on $\theta_1$ only. This proves inequality (28). □

**2.5.** ***The helicoidal method.*** With the intention of bringing to light the ideas behind our proofs, we present the main strategy in a simplified setting. Unfortunately, we cannot avoid the specific terminology, but one should think of the *sizes* as being averages, while the *energies* are $L^2$ quantities that reflect orthogonality. For estimating the norms $\|\mathrm{BHT}(f,g)\|_s$, we use interpolation results for the trilinear form $\Lambda_{\mathrm{BHT}}(f,g,h) = \langle \mathrm{BHT}(f,g), h\rangle$. In what follows, $\Lambda_{I_0}(f,g,h)$ denotes a space localization of $\Lambda_{\mathrm{BHT}}(f,g,h)$ to the fixed interval $I_0$. More specifically, it is the form associated to a model operator of BHT as in (19), where the spatial intervals of the tiles lie inside the fixed dyadic interval $I_0$. Similarly, $\Lambda^n_{I_0}(f,g,h)$ denotes a space localization of the corresponding trilinear form in the multiple vector-valued setting.

The helicoidal method is an iterated induction procedure suitable for proving vector-valued estimates for linear and multilinear operators. We describe the main ideas in the case of the BHT operator, and later on we will indicate the equivalent statements for paraproducts and the Carleson operator. At the heart of our argument lies the following induction statement:

**Induction statement.** Let $n \geq 0$. We fix $I_0$ a dyadic interval, and $F, G, H'$ subsets of $\mathbb{R}$ of finite measure. Let $R_1 = (r_1^1, \dots, r_1^n)$, $R_2 = (r_2^1, \dots, r_2^n)$ and $R' = ((r')^1, \dots, (r')^n)$ be $n$-tuples so that $\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R'} = 1$, while $f, g$ and $h$ are vector-valued functions satisfying

$$\|f(x)\|_{L^{R_1}(\mathcal{W},\mu)} \leq \mathbf{1}_F(x), \quad \|g(x)\|_{L^{R_2}(\mathcal{W},\mu)} \leq \mathbf{1}_G(x) \quad \text{and} \quad \|h(x)\|_{L^{R'}(\mathcal{W},\mu)} \leq \mathbf{1}_{H'}(x).$$

Then we have the following estimate $\mathcal{P}(n)$ for the trilinear form $\Lambda^n_{I_0}$:

$$\big|\Lambda^n_{I_0}(f,g,h)\big| \lesssim (\widetilde{\text{size}}_{I_0}\mathbf{1}_F)^{\frac{1}{2}+\frac{\theta_1}{2}-\epsilon}\,(\widetilde{\text{size}}_{I_0}\mathbf{1}_G)^{\frac{1}{2}+\frac{\theta_2}{2}-\epsilon}\,(\widetilde{\text{size}}_{I_0}\mathbf{1}_{H'})^{\frac{1}{2}+\frac{\theta_3}{2}-\epsilon}\,|I_0|$$

for every $0 \leq \theta_1, \theta_2, \theta_3 < 1,\ \theta_1 + \theta_2 + \theta_3 = 1$, satisfying an extra condition $C(R_1, R_2, R')$.

In the local $L^2$ case the condition $C(R_1, R_2, R')$ is satisfied automatically: that is, the $\mathcal{P}(n)$ statement is true for all $0 \leq \theta_1, \theta_2, \theta_3$ as above. This condition is the main obstruction in obtaining for $\overrightarrow{\text{BHT}}_{\vec{r}}$ the same range of $L^p$ estimates as that of the scalar BHT; in (37) we point out the source of this constraint. Now we present the proofs of the induction statements $\mathcal{P}(0)$ and $\mathcal{P}(n) \Rightarrow \mathcal{P}(n+1)$. Also, for the reader's convenience, we include the $\mathcal{P}(0) \Rightarrow \mathcal{P}(1)$ step.

As we will see later on, the fact that $\mathcal{P}(n)$ implies our Theorems 7 and 8 is based on a standard triple stopping time argument, involving the above localized sizes.

***Check $\mathcal{P}(0)$:*** This is the scalar BHT case, with $|f| \leq \mathbf{1}_F$, $|g| \leq \mathbf{1}_G$ and $|h| \leq \mathbf{1}_{H'}$. This situation is well understood, and we have from Proposition 23:

$$|\Lambda_{I_0}(f,g,h)| \lesssim (\widetilde{\text{size}}_{I_0} f)^{\theta_1}\,(\widetilde{\text{size}}_{I_0} g)^{\theta_2}\,(\widetilde{\text{size}}_{I_0} h)^{\theta_3}\,(\text{energy}_{I_0} f)^{1-\theta_1}\,(\text{energy}_{I_0} g)^{1-\theta_2}\,(\text{energy}_{I_0} h)^{1-\theta_3}$$

for any $0 \leq \theta_1, \theta_2, \theta_3 < 1$ such that $\theta_1 + \theta_2 + \theta_3 = 1$.

Since we are considering a localized model of BHT, where all the tiles have their spatial intervals $I_P$ lying in $I_0$, one can refine Lemma 20 by replacing $\text{energy}_{I_0} f$ with $\|f \cdot \tilde{\chi}_{I_0}\|_2$. Noticing that

$$\|f \cdot \tilde{\chi}_{I_0}\|_2 \lesssim (\widetilde{\text{size}}_{I_0}\mathbf{1}_F)^{\frac{1}{2}}\,|I_0|^{\frac{1}{2}}$$

and $|I_0|^{\frac{1-\theta_1}{2}}\,|I_0|^{\frac{1-\theta_3}{2}}\,|I_0|^{\frac{1-\theta_3}{2}} = |I_0|$, we obtain the desired $\mathcal{P}(0)$.

***Check $\mathcal{P}(0) \Rightarrow \mathcal{P}(1)$.*** Assume that

$$\left(\sum_k |f_k|^{r_1}\right)^{\frac{1}{r_1}} \leq \mathbf{1}_F, \quad \left(\sum_k |g_k|^{r_2}\right)^{\frac{1}{r_2}} \leq \mathbf{1}_G \quad \text{and} \quad \left(\sum_k |h_k|^{r'}\right)^{\frac{1}{r'}} \leq \mathbf{1}_{H'}. \tag{30}$$

Given that we know $\mathcal{P}(0)$, we will prove $\mathcal{P}(1)$, given by

$$\left|\sum_k \Lambda_{I_0}(f_k, g_k, h_k)\right| \lesssim (\widetilde{\text{size}}_{I_0}\mathbf{1}_F)^{\frac{1}{2}+\frac{\theta_1}{2}-\epsilon}\,(\widetilde{\text{size}}_{I_0}\mathbf{1}_G)^{\frac{1}{2}+\frac{\theta_2}{2}-\epsilon}\,\big(\widetilde{\text{size}}_{I_0}\mathbf{1}_{H'}\big)^{\frac{1}{2}+\frac{\theta_3}{2}-\epsilon}\,|I_0|$$

for any $0 \leq \theta_1, \theta_2, \theta_3 < 1,\ \theta_1 + \theta_2 + \theta_3 = 1$, satisfying the constraint $C(r_1, r_2, r')$, given by

$$\frac{1+\theta_1}{2} - \frac{1}{r_1} > 0, \quad \frac{1+\theta_2}{2} - \frac{1}{r_2} > 0, \quad \frac{1+\theta_3}{2} - \frac{1}{r'} > 0.$$

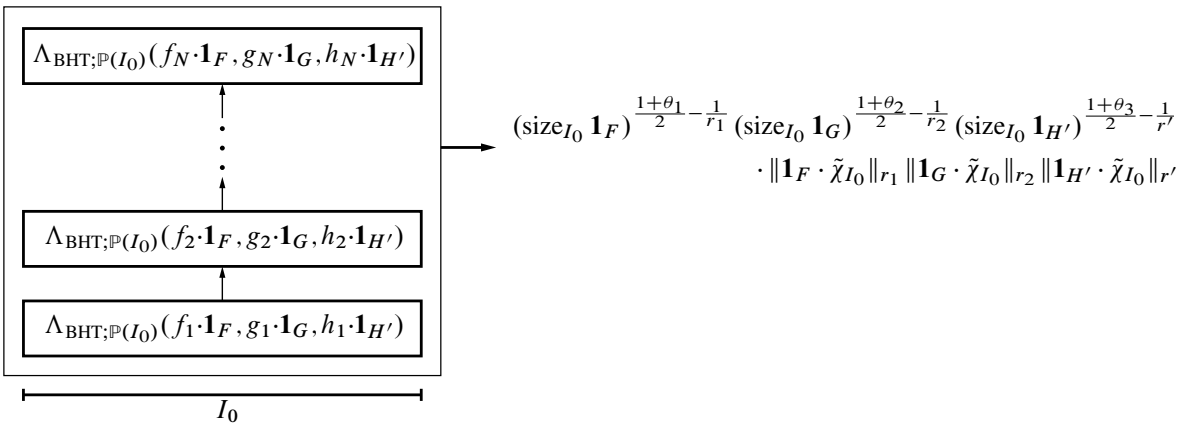



**Figure 6.** Output of the localization process.

Here an intermediate step is necessary in order to get a finer estimate for each $\Lambda_{I_0}(f_k, g_k, h_k)$. That is, we need to prove

$$\Lambda_{I_0}(f_k\cdot\mathbf{1}_F, g_k\cdot\mathbf{1}_G, h_k\cdot\mathbf{1}_{H'}) \lesssim \|\Lambda_{I_0}\|\,\|f_k\cdot\tilde{\chi}_{I_0}\|_{r_1}\|g_k\cdot\tilde{\chi}_{I_0}\|_{r_2}\|h_k\cdot\tilde{\chi}_{I_0}\|_{r'}, \tag{31}$$

where the operatorial norm is given by

$$\|\Lambda_{I_0}\| = (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{r_1}-\epsilon}(\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{r_2}-\epsilon}(\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{r'}-\epsilon}.$$

Once we have such an estimate, we sum in $k$, use Hölder's inequality and (30) to further estimate (31) by

$$\|\Lambda_{I_0}\|\frac{\|\mathbf{1}_F\cdot\tilde{\chi}_{I_0}\|_{r_1}}{|I_0|^{\frac{1}{r_1}}}\frac{\|\mathbf{1}_G\cdot\tilde{\chi}_{I_0}\|_{r_2}}{|I_0|^{\frac{1}{r_2}}}\frac{\|\mathbf{1}_{H'}\cdot\tilde{\chi}_{I_0}\|_{r'}}{|I_0|^{\frac{1}{r'}}}|I_0|.$$

This is illustrated in Figure 6 and it proves $\mathcal{P}(1)$.

The proof of (31) is a slight modification of the proof of the boundedness of the bilinear Hilbert transform. Using interpolation methods, we can assume that $|f_k|\le\mathbf{1}_{E_1}$, $|g_k|\le\mathbf{1}_{E_2}$, $|h_k|\le\mathbf{1}_{E_3}$. So we need to show

$$\Lambda_{I_0}(f_k\cdot\mathbf{1}_F, g_k\cdot\mathbf{1}_G, h_k\cdot\mathbf{1}_{H'}) \lesssim \|\Lambda_{I_0}\|\,|E_1|^{\alpha_1}|E_2|^{\alpha_2}|E_3|^{\alpha_3},$$

where $(\alpha_1,\alpha_2,\alpha_3)$ is an admissible tuple arbitrarily close to $\left(\frac{1}{r_1},\frac{1}{r_2},\frac{1}{r'}\right)$. In order to get the desired expression for $\|\Lambda_{I_0}\|$, we need another stopping time inside $I_0$. This is illustrated in Figure 7.

Let $I\subseteq I_0$ be a subinterval of $I_0$. Now we use $\mathcal{P}(0)$ as follows:

$$\begin{aligned}\left|\Lambda_I(f_k\cdot\mathbf{1}_F, g_k\cdot\mathbf{1}_G, h_k\cdot\mathbf{1}_{H'})\right| &\\ &\lesssim (\widetilde{\mathrm{size}}_I(\mathbf{1}_F\cdot\mathbf{1}_{E_1}))^{\frac{1+\theta_1}{2}-\epsilon}(\widetilde{\mathrm{size}}_I(\mathbf{1}_G\cdot\mathbf{1}_{E_2}))^{\frac{1+\theta_2}{2}-\epsilon}(\widetilde{\mathrm{size}}_I(\mathbf{1}_{H'}\cdot\mathbf{1}_{E_3}))^{\frac{1+\theta_3}{2}-\epsilon}|I|\\ &\lesssim (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\alpha_1-\epsilon}(\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\alpha_2-\epsilon}(\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\alpha_3-\epsilon}\\ &\qquad\cdot(\widetilde{\mathrm{size}}_I\mathbf{1}_{E_1})^{\alpha_1}(\widetilde{\mathrm{size}}_I\mathbf{1}_{E_2})^{\alpha_2}(\widetilde{\mathrm{size}}_I\mathbf{1}_{E_3})^{\alpha_3}|I|.\end{aligned}$$

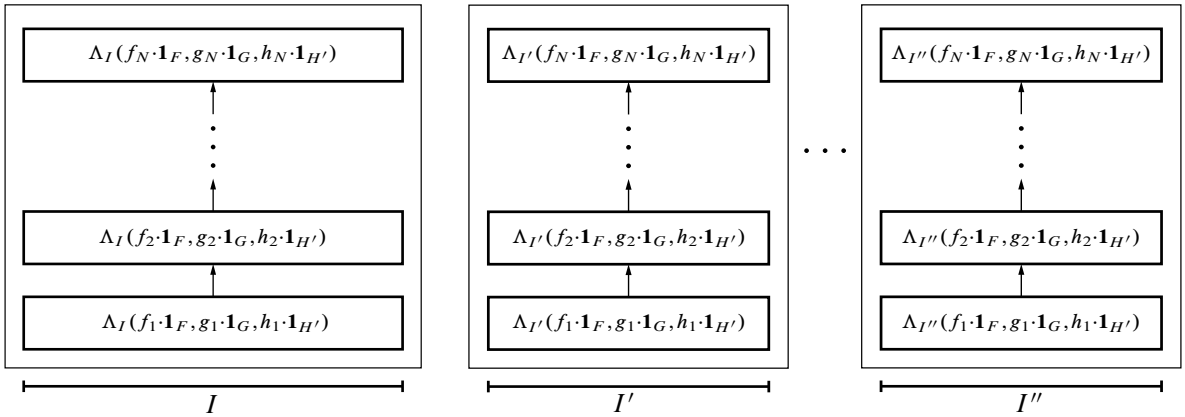



**Figure 7.** Extra stopping time.

In order to obtain the last inequality, we have to make sure that the exponents

$$\frac{1+\theta_1}{2}-\alpha_1-\epsilon, \quad \frac{1+\theta_2}{2}-\alpha_2-\epsilon, \quad \frac{1+\theta_3}{2}-\alpha_3-\epsilon$$

are all positive, which is always the case in the local $L^2$ situation. Since $(\alpha_1,\alpha_2,\alpha_3)$ are arbitrarily close to $\big(\frac{1}{r_1},\frac{1}{r_2},\frac{1}{r'}\big)$, this is the origin of the constraint $C(r_1,r_2,r')$ on page 1954.

Summing over the intervals $I$ given by the alluded to triple stopping time over the corresponding averages, we recover $|E_1|^{\alpha_1}|E_2|^{\alpha_2}|E_3|^{\alpha_3}$. We note that the operatorial norm given by interpolation is

$$(\widetilde{\text{size}}_{I_0}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{r_1}-\tilde{\epsilon}}(\widetilde{\text{size}}_{I_0}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{r_2}-\tilde{\epsilon}}(\widetilde{\text{size}}_{I_0}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{r'}-\tilde{\epsilon}},$$

where $\tilde{\epsilon}$ is slightly larger than the initial $\epsilon$, but the difference between the two is irrelevant.

***Check $\mathcal{P}(n)\Rightarrow\mathcal{P}(n+1)$.*** Lastly, we present the general induction step, in the case of iterated $\ell^p$ spaces. We have multi-indices $\vec{r}_1=(r_1^1,\dots,r_1^n)$, $\vec{r}_2=(r_2^1,\dots,r_2^n)$, $\vec{r'}=((r')^1,\dots,(r')^n)$, and $\|f\|_{\vec{r}_1}\le\mathbf{1}_F$, $\|g\|_{\vec{r}_2}\le\mathbf{1}_G$, $\|h\|_{\vec{r'}}\le\mathbf{1}_{H'}$. Then $\mathcal{P}(n)$ is equivalent to

$$\begin{aligned}\big|\Lambda^n_{I_0}(f,g,h)\big| &= \bigg|\int_{\mathbb{R}}\sum_{\vec{l}}\text{BHT}_{\mathbb{P}(I_0)}(f_{\vec{l}},g_{\vec{l}})(x)\cdot h_{\vec{l}}(x)\,dx\bigg|\\ &\lesssim (\widetilde{\text{size}}_{I_0}\mathbf{1}_F)^{\frac{1}{2}+\frac{\theta_1}{2}-\epsilon}(\widetilde{\text{size}}_{I_0}\mathbf{1}_G)^{\frac{1}{2}+\frac{\theta_2}{2}-\epsilon}(\widetilde{\text{size}}_{I_0}\mathbf{1}_{H'})^{\frac{1}{2}+\frac{\theta_3}{2}-\epsilon}|I_0|, \qquad (32)\end{aligned}$$

whenever $I_0$ is a dyadic interval. For $\mathcal{P}(n+1)$ we consider $n+1$ iterated $\ell^p$ spaces, given by the multi-indices: $\vec{R}_1=(r_1,\vec{r}_1)$, $\vec{R}_2=(r_2,\vec{r}_2)$ and $\vec{R}'=(r',\vec{r'})$, while $f,g$ and $h$ are vector-valued functions satisfying

$$\|f\|_{\vec{R}_1}:=\bigg(\sum_k\|f_k\|_{\vec{r}_1}^{r_1}\bigg)^{\frac{1}{r_1}}\le\mathbf{1}_F, \quad \|g\|_{\vec{R}_2}:=\bigg(\sum_k\|g_k\|_{\vec{r}_2}^{r_2}\bigg)^{\frac{1}{r_2}}\le\mathbf{1}_G, \quad \|h\|_{\vec{R}'}:=\bigg(\sum_k\|h_k\|_{\vec{r'}}^{r'}\bigg)^{\frac{1}{r'}}\le\mathbf{1}_{H'}. \qquad (33)$$

We want a result similar to (32), so we need to estimate

$$\Lambda_{I_0}^{n+1}(f,g,h) := \int_{\mathbb{R}} \sum_k \sum_{\vec{l}} \mathrm{BHT}_{\mathbb{P}(I_0)}(f_{k,\vec{l}}, g_{k,\vec{l}})(x) \cdot h_{k,\vec{l}}(x)\, dx = \sum_k \Lambda_{I_0}^n(f_k, g_k, h_k).$$

We can't directly apply $\mathcal{P}(n)$, and instead we will need the following result, similar to (31):

$$\left|\Lambda_{I_0}^n(f_k, g_k, h_k)\right| \lesssim \|\Lambda_{I_0}^n\| \, \|f_k \cdot \tilde{\chi}_{I_0}\|_{r_1} \|g_k \cdot \tilde{\chi}_{I_0}\|_{r_2} \|h_k \cdot \tilde{\chi}_{I_0}\|_{r'}, \tag{34}$$

where $\|\Lambda_{I_0}^n\| = (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{r_1}-\epsilon} (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{r_2}-\epsilon} (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{r'}-\epsilon}$. Once we have such a result, $\mathcal{P}(n+1)$ follows easily by Hölder, exactly as before.

We will prove (34) by using restricted-type interpolation. Instead of estimating the trilinear form $\Lambda_{I_0}^n$, we will deal with

$$\Lambda_{I_0}^{n,F,G,H'}(f_k, g_k, h_k) := \Lambda_{I_0}(f_k \cdot \mathbf{1}_F, g_k \cdot \mathbf{1}_G, h_k \cdot \mathbf{1}_{H'}). \tag{35}$$

This is natural since condition (33) implies that the functions $f_k$ are supported on $F$, and similarly the functions $g_k$ are supported on $G$ and $h_k$ on $H'$. By interpolation theory, we can assume that

$$\|f_k\|_{\vec{r}_1} \le \mathbf{1}_{E_1}, \quad \|g_k\|_{\vec{r}_2} \le \mathbf{1}_{E_2}, \quad \text{and} \quad \|h_k\|_{\vec{r}'} \le \mathbf{1}_{E_3},$$

and it suffices to prove

$$\left|\Lambda_{I_0}^{n,F,G,H'}(f_k, g_k, h_k)\right| \lesssim \|\Lambda_{I_0}^n\| \, |E_1|^{\alpha_1} |E_2|^{\alpha_2} |E_3|^{\alpha_3} \tag{36}$$

for $(\alpha_1, \alpha_2, \alpha_3)$ in a small neighborhood of $\left(\frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r'}\right)$. Similarly to the case $\mathcal{P}(0) \Rightarrow \mathcal{P}(1)$, we will have a stopping time inside $I_0$, so in fact we need to estimate $\Lambda_I^{n,F,G,H'}(f_k, g_k, h_k)$ for some $I \subseteq I_0$. It is here that we use hypothesis $\mathcal{P}(n)$:

$$\left|\Lambda_I^{n,F,G,H'}(f_k, g_k, h_k)\right| = \left|\Lambda_I^n(f_k \cdot \mathbf{1}_F, g_k \cdot \mathbf{1}_G, h_k \cdot \mathbf{1}_{H'})\right|,$$

with $\|f_k \cdot \mathbf{1}_F\|_{\vec{r}_1} \le \mathbf{1}_{F\cap E_1}$, $\|g_k \cdot \mathbf{1}_G\|_{\vec{r}_2} \le \mathbf{1}_{G\cap E_2}$ and $\|h_k \cdot \mathbf{1}_{H'}\|_{\vec{r}'} \le \mathbf{1}_{H'\cap E_3}$. More precisely,

$$\begin{aligned}
&\left|\Lambda_I^{n,F,G,H'}(f_k, g_k, h_k)\right| \\
&\qquad \lesssim (\widetilde{\mathrm{size}}_I(\mathbf{1}_F \cdot \mathbf{1}_{E_1}))^{\frac{1}{2}+\frac{\theta_1}{2}-\epsilon} (\widetilde{\mathrm{size}}_I(\mathbf{1}_G \cdot \mathbf{1}_{E_2}))^{\frac{1}{2}+\frac{\theta_2}{2}-\epsilon} (\widetilde{\mathrm{size}}_I(\mathbf{1}_{H'} \cdot \mathbf{1}_{E_3}))^{\frac{1}{2}+\frac{\theta_3}{2}-\epsilon} |I| \\
&\qquad \lesssim (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_F)^{\frac{1}{2}+\frac{\theta_1}{2}-\alpha_1-\epsilon} (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_G)^{\frac{1}{2}+\frac{\theta_2}{2}-\alpha_2-\epsilon} (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_{H'})^{\frac{1}{2}+\frac{\theta_3}{2}-\alpha_3-\epsilon} \\
&\qquad\qquad \cdot (\widetilde{\mathrm{size}}_I\mathbf{1}_{E_1})^{\alpha_1} (\widetilde{\mathrm{size}}_I\mathbf{1}_{E_2})^{\alpha_2} (\widetilde{\mathrm{size}}_I\mathbf{1}_{E_3})^{\alpha_3} |I|
\end{aligned}$$

for $(\alpha_1, \alpha_2, \alpha_3)$ in a neighborhood of $\left(\frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r'}\right)$. Due to the stopping time, which is performed with respect to the three sizes, we know the expressions $(\widetilde{\mathrm{size}}_I\mathbf{1}_{E_1})^{\alpha_1}$ add up to $|E_1|^{\alpha_1}$ and it is similar for the sizes of $\mathbf{1}_{E_2}$ and $\mathbf{1}_{E_3}$. Interpolating, we get the desired (36). From the above equation, we can see why the operatorial norm has the form

$$\left\|\Lambda_{I_0}^n\right\| = (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{r_1}-\tilde{\epsilon}} (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{r_2}-\tilde{\epsilon}} (\widetilde{\mathrm{size}}_{I_0}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{r'}-\tilde{\epsilon}}.$$

The $\tilde{\epsilon}$ (which is a slight modification on the $\epsilon$ in the $\mathcal{P}(n)$ statement), appears as an interpolation error; moreover, the conditions

$$\frac{1+\theta_1}{2}-\frac{1}{r_1}>0, \quad \frac{1+\theta_2}{2}-\frac{1}{r_2}>0, \quad \frac{1+\theta_3}{2}-\frac{1}{r'}>0 \tag{37}$$

are necessary, and they imply the constraint $C(R_1, R_2, R')$. This ends the proof of the induction step.

The same method applies in the case of paraproducts. The difference here is that the energies are $L^1$ quantities, and for that reason we don't have any extra assumptions; the range of the multiple vector-valued extensions is the same as that of the paraproducts. The model operator for paraproducts $\Pi$ corresponds to a "rank 0" family of tritiles; that is, once we know the spatial interval $I_P$, there is no other degree of freedom and the frequency intervals are $\big[1/|I_P|, 2/|I_P|\big]$ or $\big[0, 1/|I_P|\big]$. The exact definitions will be introduced in Section 4.

**Induction statement** (paraproducts case). Under the same assumptions as in the induction statement on page 1954, the localized trilinear form for paraproducts satisfies $\mathcal{P}(n)$, given by

$$\big|\Lambda^n_{I_0}(f,g,h)\big| \lesssim (\widetilde{\text{size}}_{I_0}\mathbf{1}_F)^{1-\epsilon}\,(\widetilde{\text{size}}_{I_0}\mathbf{1}_G)^{1-\epsilon}\,(\widetilde{\text{size}}_{I_0}\mathbf{1}_{H'})^{1-\epsilon}\,|I_0|,$$

provided

$$\|f(x)\|_{L^{R_1}(\mathcal{W},\mu)} \le \mathbf{1}_F(x), \quad \|g(x)\|_{L^{R_2}(\mathcal{W},\mu)} \le \mathbf{1}_G(x) \quad \text{and} \quad \|h(x)\|_{L^{R'}(\mathcal{W},\mu)} \le \mathbf{1}_{H'}(x).$$

Finally, we want to point out that the helicoidal method applies equally in the case of (sub)linear operators. One last example is that of the Carleson operator

$$C_{\mathbb{R}} f(x) = \sup_N \left| \int_{\xi<N} \hat{f}(\xi) e^{2\pi i x\xi}\, d\xi \right|$$

for which UMD-valued extensions are already known from the work of Hytönen and Lacey [2013].

Demeter and Silva [2015] gave an alternative proof for $\ell^2$-valued inequalities for the Carleson operator. In fact, they present a new principle, built around ideas from [Bateman and Thiele 2013], for dealing with $\ell^2$-valued inequalities for sublinear operators which are not of Calderón–Zygmund type.

We do not present all the details here, but the essential statement for proving multiple vector-valued inequalities for the Carleson operator, using the helicoidal method, is the following:

**Induction statement** (Carleson operator). Under the same assumptions as in the induction statement on page 1954, the localized bilinear form for the discretized Carleson operator satisfies $\mathcal{P}(n)$, given by

$$\big|\Lambda^n_{C(I_0)}(f,g)\big| \lesssim (\widetilde{\text{size}}_{I_0}\mathbf{1}_F)^{1-\epsilon}\,(\widetilde{\text{size}}_{I_0}\mathbf{1}_G)^{1-\epsilon}\,|I_0|,$$

provided that

$$\|f(x)\|_{L^{R_1}(\mathcal{W},\mu)} \le \mathbf{1}_F(x) \quad \text{and} \quad \|g(x)\|_{L^{R_2}(\mathcal{W},\mu)} \le \mathbf{1}_G(x).$$

Comparing the main statements of the above three examples, we can see from the exponents of the sizes that the range of $L^p$ estimates for the vector-valued Carleson operator and for the vector-valued paraproduct $\Pi$ will coincide with the range of the scalar operator. However, for BHT things are more complicated.

## 3. Multiple vector-valued estimates for BHT

In this section we describe the detailed proof of our Theorems 7 and 8.

**3.1.** ***Estimates for localized*** **BHT.** Here we assume that $F, G$ and $H'$ are fixed subsets of $\mathbb{R}$ of finite measure and $I_0$ is a fixed dyadic interval. We are interested in finding estimates for the bilinear operator

$$\mathrm{BHT}_{I_0}^{F,G,H'}(f,g)(x) := \sum_{P\in\mathbb{P}(I_0)} \frac{1}{|I_P|^{\frac{1}{2}}} \langle f\cdot \mathbf{1}_F, \phi_{P_1}^1\rangle \langle g\cdot \mathbf{1}_G, \phi_{P_2}^2\rangle \phi_{P_3}^3(x)\mathbf{1}_{H'}(x).$$

In doing so, we first study the associated trilinear form

$$\Lambda_{\mathrm{BHT};\mathbb{P}(I_0)}^{F,G,H'}(f,g,h) := \sum_{P\in\mathbb{P}(I_0)} \frac{1}{|I_P|^{\frac{1}{2}}} \langle f\cdot \mathbf{1}_F, \phi_{P_1}^1\rangle \langle g\cdot \mathbf{1}_G, \phi_{P_2}^2\rangle \langle h\cdot \mathbf{1}_{H'}, \phi_{P_3}^3\rangle.$$

While this operator satisfies the same estimates as the bilinear Hilbert transform, the localization to the sets $F, G$ and $H'$, and the restriction to the tiles in $\mathbb{P}(I_0)$ will bring some extra decay. First we prove a result in the "local $L^2$ case", when $\frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r'} < \frac{1}{2}$. In this situation the proof is simpler, because we are employing "energies", which are $L^2$ expressions, and they can easily be related to $L^{r_i}$ averages when $r_i \geq 2$.

**Proposition 41** (the case $r_1, r_2, r' > 2$). *Let $\mathbb{P}$ be a family of tritiles, $I_0$ a dyadic interval and $F, G, H' \subset \mathbb{R}$ sets of finite measure. Then one can find positive numbers $a_1, a_2$ and $a_3$ so that*

$$\begin{aligned}\big|\Lambda_{\mathrm{BHT};\mathbb{P}(I_0)}^{F,G,H'}(f,g,h)\big| &\\ \lesssim (\mathrm{size}_{\mathbb{P}(I_0)}\mathbf{1}_F)^{a_1}(\mathrm{size}_{\mathbb{P}(I_0)}\mathbf{1}_G)^{a_2}(\mathrm{size}_{\mathbb{P}(I_0)}\mathbf{1}_{H'})^{a_3}&\|f\cdot\tilde{\chi}_{I_0}\|_{r_1}\|g\cdot\tilde{\chi}_{I_0}\|_{r_2}\|h\cdot\tilde{\chi}_{I_0}\|_{r'}. \quad (38)\end{aligned}$$

*We can choose $a_j = 1 - \frac{2}{r_j} - \epsilon > 0$ for a very small $\epsilon > 0$.*

*Proof.* In this case we are proving restricted-type estimates by applying directly Proposition 23: let $E_1, E_2, E_3$ be sets of finite measure, and $|f| \leq \mathbf{1}_{E_1}$, $|g| \leq \mathbf{1}_{E_2}$, $|h| \leq \mathbf{1}_{E_3}$. We have

$$\begin{aligned}\Lambda_{\mathrm{BHT}}(f\cdot\mathbf{1}_F, g\cdot\mathbf{1}_G, h\cdot\mathbf{1}_{H'}) \lesssim (\mathrm{size}_{\mathbb{P}(I_0)}(f\cdot\mathbf{1}_F))^{\theta_1}(\mathrm{size}_{\mathbb{P}(I_0)}(g\cdot\mathbf{1}_G))^{\theta_2}(\mathrm{size}_{\mathbb{P}(I_0)}(h\cdot\mathbf{1}_{H'}))^{\theta_3}&\\ \cdot(\mathrm{energy}(f\cdot\mathbf{1}_F))^{1-\theta_1}(\mathrm{energy}(g\cdot\mathbf{1}_G))^{1-\theta_2}(\mathrm{energy}(h\cdot\mathbf{1}_{H'}))^{1-\theta_3} \quad (39)\end{aligned}$$

for any $0 \leq \theta_1, \theta_2, \theta_3 < 1$ such that $\theta_1 + \theta_2 + \theta_3 = 1$. Recall that the sizes can be estimated by

$$\mathrm{size}_{\mathbb{P}(I_0)}(f\cdot\mathbf{1}_F) \lesssim \sup_{P\in\mathbb{P}(I_0)} \frac{1}{|I_P|}\int \mathbf{1}_{E_1}\cdot\mathbf{1}_F\cdot\tilde{\chi}_{I_P}^M\,dx,$$

where $M$ can be chosen as large as we wish. Then we observe that if $E_1$ is supported away from $I_0$, the sizes will decay fast, giving the desired $\|f\cdot\tilde{\chi}_{I_0}\|_{r_1}$ on the right-hand side. It is similar for $E_2$ and $E_3$. For this reason, we can assume that the sets $E_1, E_2, E_3$ are supported on $5I_0$ and then we will need to show only that

$$|\Lambda_{\mathrm{BHT};\mathbb{P}(I_0)}(f\cdot\mathbf{1}_F, g\cdot\mathbf{1}_G, h\cdot\mathbf{1}_{H'})| \lesssim (\mathrm{size}_{\mathbb{P}(I_0)}\mathbf{1}_F)^{a_1}(\mathrm{size}_{\mathbb{P}(I_0)}\mathbf{1}_G)^{a_2}(\mathrm{size}_{\mathbb{P}(I_0)}\mathbf{1}_{H'})^{a_3}\|f\|_{r_1}\|g\|_{r_2}\|h\|_{r'}.$$

We are using the energies precisely for estimating the norms of $f, g$ and $h$, so the sizes are playing the role of a constant here. As we have seen in Lemma 24, the energies are bounded by $L^2$ norms, so from (39), we have

$$\Lambda^{F,G,H'}_{\mathrm{BHT};\mathbb{P}(I_0)}(f,g,h) \lesssim (\mathrm{size}_{\mathbb{P}(I_0)}\, \mathbf{1}_F)^{\theta_1} (\mathrm{size}_{\mathbb{P}(I_0)}\, \mathbf{1}_G)^{\theta_2} (\mathrm{size}_{\mathbb{P}(I_0)}\, \mathbf{1}_{H'})^{\theta_3} |E_1|^{\frac{1-\theta_1}{2}} |E_2|^{\frac{1-\theta_2}{2}} |E_3|^{\frac{1-\theta_3}{2}}.$$

By varying $\theta_1, \theta_2$ and $\theta_3$, we see that these restricted-type estimates are true in a very small neighborhood of $\big(\frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r'}\big)$, and the interpolation, Theorem 28, yields strong-type estimates. Note that the constant in this case is

$$(\mathrm{size}_{\mathbb{P}(I_0)}\, \mathbf{1}_F)^{\theta_1} (\mathrm{size}_{\mathbb{P}(I_0)}\, \mathbf{1}_G)^{\theta_2} (\mathrm{size}_{\mathbb{P}(I_0)}\, \mathbf{1}_{H'})^{\theta_3},$$

which depends on the functions $\mathbf{1}_F, \mathbf{1}_G, \mathbf{1}_{H'}$, the fixed interval $I_0$, the values of $\theta_1, \theta_2$, and $\theta_3$, but not on the functions $f, g, h$. □

Now we deal with the general Banach triangle case, where $\big(\frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r'}\big)$ is an admissible tuple satisfying

$$0 < \tfrac{1}{r_1}, \tfrac{1}{r_2}, \tfrac{1}{r'} < 1.$$

The proof is going to be more complicated because we will need to use the sizes as well for reconstructing the norms of $f, g, h$. In addition, we will also need to use the sizes of $\mathbf{1}_F, \mathbf{1}_G$ and $\mathbf{1}_{H'}$ later on.

**Proposition 42.** *Let $F, G$ and $H'$ be as above and let $\mathbb{P}(I_0)$ be a family of tritiles localized to the dyadic interval $I_0$. Then there exist positive numbers $a_1, a_2$ and $a_3$ so that*

$$\begin{aligned}\big|\Lambda^{F,G,H'}_{\mathrm{BHT};\mathbb{P}(I_0)}(f,g,h)\big| \\ \lesssim (\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_F)^{a_1} (\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_G)^{a_2} (\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_{H'})^{a_3} \|f\cdot\tilde{\chi}_{I_0}\|_{r_1} \|g\cdot\tilde{\chi}_{I_0}\|_{r_2} \|h\cdot\tilde{\chi}_{I_0}\|_{r'},\end{aligned} \tag{40}$$

*where $\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r'} = 1$. In fact, for $\epsilon > 0$ small enough,*

$$a_1 = \frac{1+\theta_1}{2} - \frac{1}{r_1} - \epsilon, \quad a_2 = \frac{1+\theta_2}{2} - \frac{1}{r_2} - \epsilon, \quad a_3 = \frac{1+\theta_3}{2} - \frac{1}{r'} - \epsilon, \tag{41}$$

*where $\theta_1, \theta_2, \theta_3$ are so that $0 \le \theta_1, \theta_2, \theta_3 < 1$, $\theta_1 + \theta_2 + \theta_3 = 1$, and the expressions in (41) are positive.*

*Proof.* In this case, we will use the interpolation, Theorem 32, and for this reason we cannot obtain directly the expression in the right-hand side of (40), which represents *localized* $L^p$ norms. However, as we will see soon, it will be enough to prove that $\Lambda_{\mathrm{BHT};\mathbb{P}(I_0)}$ is of generalized restricted type $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ for $\alpha$ in a small neighborhood of $\big(\frac{1}{r_1}, \frac{1}{r_2}, \frac{1}{r'}\big)$. Then the result in (40) will be a consequence of the fast decay of the wave packets away from $I_0$.

We start with sets of finite measure $E_1, E_2, E_3$ and define $\widetilde{\Omega}$ to be the exceptional set

$$\widetilde{\Omega} := \Big\{x : \mathcal{M}(\mathbf{1}_{E_1}) > C\frac{|E_1|}{|E_3|}\Big\} \cup \Big\{x : \mathcal{M}(\mathbf{1}_{E_2}) > C\frac{|E_2|}{|E_3|}\Big\}.$$

Let $E_3' := E_3 \setminus \widetilde{\Omega}$. We want to prove that (40) holds for any functions $f, g, h$ so that $|f| \le \mathbf{1}_{E_1}$, $|g| \le \mathbf{1}_{E_2}$, and $|h| \le \mathbf{1}_{E_3'}$. For simplicity, we assume that $1 + \mathrm{dist}(I_P, \widetilde{\Omega}^c)/|I_P| \sim 2^d$ for every tile $P \in \mathbb{P}(I_0)$.

Equivalently, we could decompose the collection of tiles into subcollections for which this property holds for all $d \geq 0$. In the end, however, the estimate (40) will be independent of such a decomposition.

With the above assumption, for every $P \in \mathbb{P}(I_0)$, we have

$$\frac{1}{|I_P|}\int_{\mathbb{R}} \mathbf{1}_{E_1}\cdot\mathbf{1}_F\cdot\tilde{\chi}_{I_P}^M\,dx \lesssim 2^d\frac{|E_1|}{|E_3|} \quad \text{and} \quad \frac{1}{|I_P|}\int_{\mathbb{R}} \mathbf{1}_{E_2}\cdot\mathbf{1}_G\cdot\tilde{\chi}_{I_P}^M\,dx \lesssim 2^d\frac{|E_2|}{|E_3|}.$$

This is important because now we can perform a stopping time which will allow us to estimate the "sizes" of the functions $\mathbf{1}_{E_j}$. For each of the functions $\mathbf{1}_F\cdot\mathbf{1}_{E_1}$, $\mathbf{1}_G\cdot\mathbf{1}_{E_2}$ and $\mathbf{1}_{H'}\cdot\mathbf{1}_{E_3'}$, we will be looking for maximal dyadic intervals $J$ which are maximizers for

$$\sup_{\substack{J\subseteq I_0\\ \exists P\in\mathbb{P}(I_0),\, I_P\subseteq J}} \frac{1}{|J|}\int_{\mathbb{R}} \mathbf{1}_{E_1}\cdot\mathbf{1}_F\cdot\tilde{\chi}_J^M\,dx. \tag{42}$$

This is the reason we introduced the new size in Definition 21.

The selection of the intervals and tiles is described in more detail in Section 3.2, so here we only sketch this process.

We start with the largest possible value $2^{-l_1}\lesssim 2^d|E_1|/|E_2|$ and define $\mathcal{I}_{l_1}$ to be the collection of maximal dyadic intervals $I$ with the property that it contains some $I_P\in\mathbb{P}(I_0)$ which is not contained in any of the intervals previously selected, and $I$ also has the property that

$$2^{-l_1-1}\leq\frac{1}{|I|}\int_{\mathbb{R}}\mathbf{1}_{E_1}\cdot\mathbf{1}_F\cdot\tilde{\chi}_I^M\,dx\leq 2^{-l_1}.$$

Then for each $I\in\mathcal{I}_{l_1}$ we find the relevant tiles $P$ with $I_P\subseteq I$, and move them into $\mathbb{P}(I)$. Afterwards we restart the algorithm for the collection $\mathbb{P}(I_0)\setminus\bigcup_{I\in\mathcal{I}_{l_1}}\mathbb{P}(I)$.

The algorithm continues by decreasing $2^{-l_1}$ until all tiles in $\mathbb{P}(I_0)$ are exhausted. In this way, for any $l_1$ and any $I\in\mathcal{I}_{l_1}$, we have $\widetilde{\operatorname{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_1}\cdot\mathbf{1}_F)\sim 2^{-l_1}$. Similarly we define the collections of dyadic intervals $\mathcal{I}_{l_2}$ associated with the functions $\mathbf{1}_{E_2}\cdot\mathbf{1}_G$ as long as $2^{-l_2}\lesssim 2^d|E_2|/|E_3|$.

For the third component, the collections $\mathcal{I}_{l_3}$ are nonempty as long as $2^{-n_3}\lesssim 2^{-\tilde{M}d}$, and in that case, for any $I\in\mathcal{I}_{l_3}$, we have $\widetilde{\operatorname{size}}_{\mathbb{P}(I)}(\mathbf{1}_{H'}\cdot\mathbf{1}_{E_3'})\sim 2^{-n_3}$. The extra decay is due to the fact that $E_3'$ is actually supported on $\widetilde{\Omega}^c$.

Given $l_1,l_2,l_3$ as above, we define $\mathcal{I}^{l_1,l_2,l_3}:=\mathcal{I}_{l_1}\cap\mathcal{I}_{l_2}\cap\mathcal{I}_{l_3}$. This is also going to be a collection of dyadic intervals, and any tile in $\mathbb{P}(I_0)$ will be contained in some $\mathbb{P}(I)$, with $I\in\mathcal{I}^{l_1,l_2,l_3}$. In fact, these collections depend on the parameter $d$ as well, which controls the distance from the exceptional set. We have

$$\mathbb{P}(I_0)=\bigcup_d\bigcup_{l_1,l_2,l_3}\bigcup_{I\in\mathcal{I}_d^{l_1,l_2,l_3}}\mathbb{P}(I),$$

but we suppress the dependency on $d$ in the notation. Thus

$$\Lambda^{F,G,H'}_{\mathrm{BHT};\mathbb{P}(I_0)}(f,g,h)=\sum_{l_1,l_2,l_3}\sum_{I\in\mathcal{I}^{l_1,l_2,l_3}}\Lambda^{F,G,H'}_{\mathrm{BHT};\mathbb{P}(I)}(f,g,h). \tag{43}$$

Every $\Lambda^{F,G,H'}_{\mathrm{BHT};\mathbb{P}(I)}(f,g,h)$ is going to be estimated by Lemma 40:

$$\Lambda^{F,G,H'}_{\mathrm{BHT};\mathbb{P}(I)}(f,g,h) \lesssim (\widetilde{\mathrm{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_1}\cdot\mathbf{1}_F))^{\theta_1}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_2}\cdot\mathbf{1}_G))^{\theta_2}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_3'}\cdot\mathbf{1}_{H'}))^{\theta_3}$$
$$\cdot \|\mathbf{1}_{E_1}\cdot\mathbf{1}_F\cdot\tilde{\chi}_I\|_2^{1-\theta_1}\|\mathbf{1}_{E_2}\cdot\mathbf{1}_G\cdot\tilde{\chi}_I\|_2^{1-\theta_2}\|\mathbf{1}_{E_3'}\cdot\mathbf{1}_{H'}\cdot\tilde{\chi}_I\|_2^{1-\theta_3}.$$

For the particular function $\mathbf{1}_{E_1}\cdot\mathbf{1}_F$ and an interval $I\in\mathcal{I}^{l_1,l_2,l_3}$, we have

$$\left(\int_{\mathbb{R}}\mathbf{1}_{E_1}\cdot\mathbf{1}_F\cdot\tilde{\chi}_I^M\,dx\right)^{\frac{1}{2}} \lesssim 2^{-\frac{l_1}{2}}|I|^{\frac{1}{2}} \lesssim (\widetilde{\mathrm{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_1}\cdot\mathbf{1}_F))^{\frac{1}{2}}|I|^{\frac{1}{2}}.$$

In this way, as long as

$$\frac{1+\theta_1}{2}-\frac{1}{r_1}>0,\quad \frac{1+\theta_2}{2}-\frac{1}{r_2}>0,\quad \frac{1+\theta_3}{2}-\frac{1}{r'}>0, \tag{44}$$

we can estimate $\Lambda^{F,G,H'}_{\mathrm{BHT};\mathbb{P}(I_0)}(f,g,h)$ as

$$\begin{aligned}
&\Lambda^{F,G,H'}_{\mathrm{BHT};\mathbb{P}(I_0)}(f,g,h)\\
&\lesssim \sum_{l_1,l_2,l_3}\sum_{I\in\mathcal{I}^{l_1,l_2,l_3}}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_1}\cdot\mathbf{1}_F))^{\theta_1}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_2}\cdot\mathbf{1}_G))^{\theta_2}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_3'}\cdot\mathbf{1}_{H'}))^{\theta_3}\\
&\qquad\left(\frac{1}{|I|}\int_{\mathbb{R}}\mathbf{1}_{E_1}\cdot\mathbf{1}_F\cdot\tilde{\chi}_I^M\,dx\right)^{\frac{1-\theta_1}{2}}\left(\frac{1}{|I|}\int_{\mathbb{R}}\mathbf{1}_{E_2}\cdot\mathbf{1}_G\cdot\tilde{\chi}_I^M\,dx\right)^{\frac{1-\theta_2}{2}}\left(\frac{1}{|I|}\int_{\mathbb{R}}\mathbf{1}_{E_3'}\cdot\mathbf{1}_{H'}\cdot\tilde{\chi}_I^M\,dx\right)^{\frac{1-\theta_3}{2}}|I|\\
&\lesssim (\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{r_1}}(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{r_2}}(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{r'}-\epsilon}\\
&\qquad\cdot\sum_{l_1,l_2,l_3}\sum_{I\in\mathcal{I}^{l_1,l_2,l_3}}2^{-\frac{l_1}{r_1}}2^{-\frac{l_2}{r_2}}2^{-l_3(\frac{1}{r'}+\epsilon)}|I|.
\end{aligned}\tag{45}$$

The quantity

$$(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{r_1}}(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{r_2}}(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{r'}-\epsilon}$$

is going to represent the operatorial norm $\|\Lambda^{F,G,H'}_{\mathrm{BHT};\mathbb{P}(I_0)}\|$ associated to the trilinear form $\Lambda^{F,G,H}_{\mathrm{BHT};\mathbb{P}(I_0)}$, as seen in (40).

We are left with estimating $\sum_{I\in\mathcal{I}^{l_1,l_2,l_3}}|I|$, which can be realized in three different ways; for example,

$$\sum_{I\in\mathcal{I}^{l_1,l_2,l_3}}|I| \le \sum_{I\in\mathcal{I}_{l_1}}|I| = \left\|\sum_{I\in\mathcal{I}_{l_1}}\mathbf{1}_I\right\|_{1,\infty} \lesssim \left\|\sum_{I\in\mathcal{I}_{l_1}}2^{l_1}\mathcal{M}(\mathbf{1}_{E_1})\cdot\mathbf{1}_I\right\|_{1,\infty} \lesssim 2^{n_1}|E_1|.$$

For this reason, whenever $0\le\alpha_j\le 1$, with $\alpha_1+\alpha_2+\alpha_3=1$, we have

$$\sum_{I\in\mathcal{I}^{l_1,l_2,l_3}}|I| \lesssim (2^{l_1}|E_1|)^{\alpha_1}(2^{l_2}|E_2|)^{\alpha_2}(2^{l_3}|E_3'|)^{\alpha_3}.$$

This yields

$$\begin{aligned}\sum_{l_1,l_2,l_3}\sum_{I\in\mathcal{I}^{l_1,l_2,l_3}} & 2^{-\frac{l_1}{r_1}}2^{-\frac{l_2}{r_2}}2^{-l_3(\frac{1}{r'}+\epsilon)}|I| \\ &\lesssim \sum_{l_1,l_2,l_3} 2^{-l_1(\frac{1}{r_1}-\alpha_1)}2^{-l_2(\frac{1}{r_2}-\alpha_2)}2^{-l_3(\frac{1}{r'}+\epsilon-\alpha_1)}|E_1|^{\alpha_1}|E_2|^{\alpha_2}|E_3|^{\alpha_3} \\ &\lesssim \left(2^d\frac{|E_1|}{|E_3|}\right)^{\frac{1}{r_1}-\alpha_1}\left(2^d\frac{|E_2|}{|E_3|}\right)^{\frac{1}{r_2}-\alpha_2}(2^{-\widetilde{Md}})^{(\frac{1}{r'}+\epsilon-\alpha_3)}|E_1|^{\alpha_1}|E_2|^{\alpha_2}|E_3|^{\alpha_3} \\ &\lesssim 2^{-100d}|E_1|^{\frac{1}{r_1}}|E_2|^{\frac{1}{r_2}}|E_3|^{\frac{1}{r'}}.\end{aligned}$$

Summing over $d$, this proves (40) in the particular case of characteristic functions. Upon interpolating, we lose an $\epsilon$-power of $\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_F$ and $\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_G$ respectively, to get

$$\left|\Lambda^{F,G,H'}_{BHT;\mathbb{P}(I_0)}(f,g,h)\right|\lesssim(\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_F)^{a_1}(\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_G)^{a_2}(\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_{H'})^{a_3}\|f\cdot\tilde{\chi}_{I_0}\|_{r_1}\|g\cdot\tilde{\chi}_{I_0}\|_{r_2}\|h\cdot\tilde{\chi}_{I_0}\|_{r'}.$$

We note that the "weights" $\tilde{\chi}_{I_0}$ will not affect the interpolation process; once we have an inequality that holds for characteristic functions of finite sets, interpolation implies a similar result in full generality.

The exponents $a_1, a_2$ and $a_3$ can be described as

$$a_1=\frac{1+\theta_1}{2}-\frac{1}{r_1}-\epsilon,\quad a_2=\frac{1+\theta_2}{2}-\frac{1}{r_2}-\epsilon,\quad a_3=\frac{1+\theta_3}{2}-\frac{1}{r'}-\epsilon$$

for some sufficiently small $\epsilon$, and for $0\le\theta_1,\theta_2,\theta_3<1$, satisfying $\theta_1+\theta_2+\theta_3=1$, that will be chosen later. □

**Corollary 43** (the case $r=1$). *Let $1<r_1,r_2<\infty$ be such that $\frac{1}{r_1}+\frac{1}{r_2}=1$, and $\theta_1,\theta_2$ satisfy $\frac{1}{2}(1+\theta_1)>\frac{1}{r_1}$ and $\frac{1}{2}(1+\theta_2)>\frac{1}{r_2}$. Then*

$$\begin{aligned}&\|\text{BHT}^{F,G,H'}_{\mathbb{P}(I_0)}(f,g)\|_1 \\ &\quad\lesssim(\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{r_1}-\epsilon}(\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{r_2}-\epsilon}(\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\epsilon}\|f\cdot\tilde{\chi}_{I_0}\|_{r_1}\|g\cdot\tilde{\chi}_{I_0}\|_{r_2}.\end{aligned}$$

*Proof.* A careful inspection of (45) shows that one can choose any triple $(\beta_1,\beta_2,\beta_3)$ with $\beta_1+\beta_2+\beta_3=1$, even with $\beta_3\le 0$, in the place of $\left(\frac{1}{r_1},\frac{1}{r_2},\frac{1}{r'}\right)$. In this case we get

$$\begin{aligned}&\left|\Lambda^{F,G,H'}_{\text{BHT};\mathbb{P}(I_0)}(f,g,h)\right| \\ &\quad\lesssim(\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\beta_1}(\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\beta_2}(\widetilde{\text{size}}_{\mathbb{P}(I_0)}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\epsilon}|E_1|^{\beta_1}|E_2|^{\beta_2}|E_3|^{\beta_3}.\end{aligned}$$

The restrictions are that $\beta_j<\frac{1}{2}(1+\theta_j)$, which works well for very small or negative values of $\beta_3$. Interpolating between tuples $(\beta_1,\beta_2,\beta_3)$ that lie in a small open neighborhood of $\left(\frac{1}{r_1},\frac{1}{r_2},0\right)$, we get the conclusion. In this case, the interpolation is used for estimating the $L^1$ norm of the operator, and not the trilinear form $\Lambda^{F,G,H'}_{\text{BHT};\mathbb{P}(I_0)}$. □

**3.2.** ***Proof of Theorem 7.*** Recall that the vector-valued BHT is defined by

$$\mathrm{BHT}(f,g)(x,w)=\int_{\mathbb{R}} f(x-t,w)g(x+t,w)\frac{dt}{t}=\mathrm{BHT}(f_w,g_w)(x).$$

Then the trilinear form associated with it is

$$\Lambda_{\overrightarrow{\mathrm{BHT}}}(f,g,h)=\int_{\mathbb{R}}\int_{\mathcal{W}}\mathrm{BHT}(f,g)(x,w)h(x,w)\,d\mu(w)\,dx.$$

First we prove generalized restricted-type estimates for $\Lambda_{\overrightarrow{\mathrm{BHT}}}(f,g,h)$, and the general result will follow from the vector-valued interpolation result presented in Proposition 38. Let $F,G$ and $H$ be sets of finite measure. In what follows, we will construct a major subset $H'\subseteq H$ and show

$$|\Lambda_{\overrightarrow{\mathrm{BHT}};\mathbb{P}}(f,g,h)|\lesssim |F|^{\alpha_1}|G|^{\alpha_2}|H|^{\alpha_3} \tag{46}$$

whenever $\|f(x,\cdot)\|_{L^{r_1}(\mathcal{W},\mu)}\leq \mathbf{1}_F(x)$, $\|g(x,\cdot)\|_{L^{r_2}(\mathcal{W},\mu)}\leq \mathbf{1}_G(x)$ and $\|h(x,\cdot)\|_{L^{r'}(\mathcal{W},\mu)}\leq \mathbf{1}_{H'}(x)$. For simplicity, assume $|H|=1$. The exceptional set is defined as

$$\Omega:=\big\{x:\mathcal{M}(\mathbf{1}_F)>C|F|\big\}\cup\big\{x:\mathcal{M}(\mathbf{1}_G)>C|G|\big\}.$$

Because of the $L^1\to L^{1,\infty}$ boundedness of the maximal operator, for a constant $C$ large enough, we have $|\Omega|\ll 1$.

We partition the collection of tritiles according to the scaled distance from the exceptional set

$$\mathbb{P}^d=\left\{P\in\mathbb{P}:1+\frac{\mathrm{dist}(I_P,\Omega^c)}{|I_P|}\sim 2^d\right\}$$

and we will prove estimates equivalent to (46) for the family $\mathbb{P}^d$, with an extra $2^{-10d}$ decay:

$$\big|\Lambda_{\overrightarrow{\mathrm{BHT}};\mathbb{P}^d}(f,g,h)\big|\lesssim 2^{-10d}|F|^{\frac{1}{p}}|G|^{\frac{1}{q}}|H|^{\frac{1}{s'}}. \tag{47}$$

We suppress the $d$-dependency for the moment, but all the subcollections $\mathcal{I}_j^{n_j}$ and $\mathcal{I}^{n_1,n_2,n_3}$ will actually depend on this parameter. At the very end we sum in $d$, and use interpolation, so that the final estimate depends only on the fixed interval $I_0$, and the fixed sets $F,G,H'$.

Now we construct a collection $\{\mathcal{I}_1^{n_1}\}_{n_1\geq\bar{n}_1}$ of relevant dyadic intervals, according to the concentration of $\mathbf{1}_F$:

- Start with $\bar{n}_1$ such that $2^{-\bar{n}_1}\sim 2^d|F|$ and let $\mathbb{P}'_{\bar{n}_1-1}=\mathbb{P}$ (here $\mathbb{P}'_{n_1}$ will play the role of *stock*, or the collection of available tiles).
- Define $\mathcal{I}_1^{\bar{n}_1}$ to be the collection of maximal dyadic intervals $I$ with the property that there exists at least one tile $P\in\mathbb{P}'_{\bar{n}_1}$ with $I_P\subseteq I$ and

$$\frac{1}{|I|}\int \mathbf{1}_F\cdot\tilde{\chi}_I^M\,dx\sim 2^{-\bar{n}_1}. \tag{48}$$

- For every such interval $I$, let $\mathbb{P}_{\bar{n}_1}(I)$ be the collection of tiles $P\in\mathbb{P}'_{\bar{n}_1}$ with the property that $I_P\subseteq I$.
- Set $\mathbb{P}'_{\bar{n}_1}=\mathbb{P}\setminus\bigcup_{I\in\mathcal{I}_1^{\bar{n}_1}}\mathbb{P}_{\bar{n}_1}(I)$.

• Repeat the procedure for all $n_1 \geq \bar{n}_1$. Let $\mathcal{I}_1^{n_1}$ denote the collection of maximal dyadic intervals which contain a time interval $I_P$ for some $P \in \mathbb{P}'_{n_1-1}$ (which was not selected previously) and such that

$$2^{-n_1-1} \leq \frac{1}{|I|} \int \mathbf{1}_F \cdot \tilde{\chi}_I^M \, dx < 2^{-n_1}.$$

• As before, $\mathbb{P}_{n_1}(I) := \{P \in \mathbb{P}'_{n_1} : I_P \subseteq I\}$.

• Set $\mathbb{P}'_{n_1} = \mathbb{P}_{n_1-1} \setminus \bigcup_{I \in \mathcal{I}_1^{n_1}} \mathbb{P}_{n_1}(I)$ and notice that after a finite number of steps, $\mathbb{P}'_{n_1} = \varnothing$.

• Note that we always have $2^{-n_1} \lesssim 2^d |F|$.

For $d$ sufficiently large, the intervals $I_P$ for $P \in \mathbb{P}^d$ are going to be essentially disjoint and the intervals $I \in \mathcal{I}_1^{n_1}$ can be selected in an easier way, but this is not the case, for example, when $d = 0$, which corresponds to $I_P \cap \Omega^c \neq \varnothing$. However, for every $n_1$, the intervals in $\mathcal{I}_1^{n_1}$ are going to be disjoint and this is going to be used later in the proof.

Similarly, $\mathcal{I}_2^{n_2}$ denotes the collection of maximal dyadic intervals $I$ containing at least some $I_P \subseteq I$ for some $P \in \mathbb{P}^d$, and

$$\frac{1}{|I|} \int \mathbf{1}_G \cdot \tilde{\chi}_I^M \, dx \sim 2^{-n_2} \lesssim 2^d |G|.$$

For $\mathbf{1}_{H'}$, let $\mathcal{I}_3^{n_3}$ be the collection of maximal dyadic intervals $I$ containing at least some $I_P$ for some $P \in \mathbb{P}^d$ and such that

$$\frac{1}{|I|} \int \mathbf{1}_{H'} \cdot \tilde{\chi}_I^M \, dx \sim 2^{-n_3} \lesssim 2^{-Md}.$$

We define $\mathcal{I}^{n_1,n_2,n_3} := \mathcal{I}_1^{n_1} \cap \mathcal{I}_2^{n_2} \cap \mathcal{I}_3^{n_3}$, and we further partition $\mathbb{P}^d$ as $\mathbb{P}^d = \bigcup_{n_1,n_2,n_3} \bigcup_{I \in \mathcal{I}^{n_1,n_2,n_3}} \mathbb{P}(I)$.

For $I \in \mathcal{I}_1^{n_1}$, we have $\widetilde{\text{size}}_{\mathbb{P}_{n_1}(I)} \mathbf{1}_F \sim 2^{-n_1}$. When we consider the intersection $I'$ of different intervals in $\mathcal{I}_1^{n_1}, \mathcal{I}_2^{n_2}$ and $\mathcal{I}_3^{n_3}$, all we can say is that $\widetilde{\text{size}}_{\mathbb{P}(I')} \mathbf{1}_F \lesssim 2^{-n_1}$. This fact is the technical obstruction in obtaining vector-valued BHT estimates for any $p, q, s$ in the whole range of BHT.

In a similar way, the relation $(1/|I|) \int_{\mathbb{R}} \mathbf{1}_F \cdot \tilde{\chi}_I^M \, dx \sim 2^{-n_1}$ for $I \in \mathcal{I}_1^{n_1}$ becomes for an interval $I' \in \mathcal{I}_1^{n_1} \cap \mathcal{I}_2^{n_2} \cap \mathcal{I}_3^{n_3}$ an inequality: $(1/|I'|) \int_{\mathbb{R}} \mathbf{1}_F \cdot \tilde{\chi}_{I'}^M \, dx \lesssim 2^{-n_1}$.

The trilinear form in (47) becomes

$$\begin{aligned}
&\sum_{n_1,n_2,n_3} \sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} \Lambda_{\overrightarrow{\text{BHT}};\mathbb{P}(I)}(f,g,h) \\
&\qquad = \sum_{n_1,n_2,n_3} \sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} \int_{\mathbb{R}} \int_{\mathcal{W}} \text{BHT}_{\mathbb{P}(I)}(f_w, g_w)(x) \cdot h_w(x) \, d\mu(w) \, dx \\
&\qquad = \int_{\mathcal{W}} \Bigg( \sum_{n_1,n_2,n_3} \sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} \int_{\mathbb{R}} \text{BHT}_{\mathbb{P}(I)}(f_w \cdot \mathbf{1}_F, g_w \cdot \mathbf{1}_G)(x) \cdot \mathbf{1}_{H'}(x) \cdot h_w(x) \, dx \Bigg) \, d\mu(w).
\end{aligned}$$

Note that the functions $f_w$ are supported on $F$, the $g_w$ on $G$ and the $h_w$ on $H'$, for a.e. $w$. We can apply the localization, Proposition 42, to get

$$\begin{aligned}
&\big|\Lambda^{F,G,H'}_{\text{BHT};\mathbb{P}(I)}(f_w, g_w, h_w)\big| \\
&\qquad \lesssim (\widetilde{\text{size}}_{\mathbb{P}(I)} \mathbf{1}_F)^{a_1} (\widetilde{\text{size}}_{\mathbb{P}(I)} \mathbf{1}_G)^{a_2} (\widetilde{\text{size}}_{\mathbb{P}(I)} \mathbf{1}_{H'})^{a_3} \, \|f_w \cdot \tilde{\chi}_I\|_{r_1} \|g_w \cdot \tilde{\chi}_I\|_{r_2} \|h_w \cdot \tilde{\chi}_I\|_{r'},
\end{aligned}$$

where $\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r'} = 1$.

Recall the expressions for $a_j$ from (41):

$$a_1 = \frac{1+\theta_1}{2} - \frac{1}{r_1} - \epsilon, \quad a_2 = \frac{1+\theta_2}{2} - \frac{1}{r_2} - \epsilon, \quad a_3 = \frac{1+\theta_3}{2} - \frac{1}{r'} - \epsilon,$$

where the only conditions we have on $\theta_1, \theta_2$ and $\theta_3$ are that $\theta_1 + \theta_2 + \theta_3 = 1$ and $a_j > 0$. Using Hölder's inequality, the initial trilinear form can be estimated by

$$\begin{aligned}
&\sum_{n_1,n_2,n_3} \sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} \int_{\mathcal{W}} \big|\Lambda_{\mathrm{BHT};\mathbb{P}(I)}(f_w, g_w, h_w)\big| \\
&\quad\lesssim \sum_{n_1,n_2,n_3} \sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} (\widetilde{\mathrm{size}}_{\mathbb{P}(I)} \mathbf{1}_F)^{a_1} (\widetilde{\mathrm{size}}_{\mathbb{P}(I)} \mathbf{1}_G)^{a_2} (\widetilde{\mathrm{size}}_{\mathbb{P}(I)} \mathbf{1}_{H'})^{a_3} \\
&\qquad\qquad \left(\int_{\mathcal{W}} \|f_w \cdot \tilde{\chi}_I\|_{r_1}^{r_1} \, d\mu(w)\right)^{\frac{1}{r_1}} \left(\int_{\mathcal{W}} \|g_w \cdot \tilde{\chi}_I\|_{r_2}^{r_2} \, d\mu(w)\right)^{\frac{1}{r_2}} \left(\int_{\mathcal{W}} \|h_w \cdot \tilde{\chi}_I\|_{r'}^{r'} \, d\mu(w)\right)^{\frac{1}{r'}} \\
&\quad\lesssim \sum_{n_1,n_2,n_3} \sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} (\widetilde{\mathrm{size}}_{\mathbb{P}(I)} \mathbf{1}_F)^{a_1} (\widetilde{\mathrm{size}}_{\mathbb{P}(I)} \mathbf{1}_G)^{a_2} (\widetilde{\mathrm{size}}_{\mathbb{P}(I)} \mathbf{1}_{H'})^{a_3} \\
&\qquad\qquad \cdot \frac{\|\mathbf{1}_F \cdot \tilde{\chi}_I\|_{r_1}}{|I|^{\frac{1}{r_1}}} \frac{\|\mathbf{1}_G \cdot \tilde{\chi}_I\|_{r_2}}{|I|^{\frac{1}{r_2}}} \frac{\|\mathbf{1}_{H'} \cdot \tilde{\chi}_I\|_{r'}}{|I|^{\frac{1}{r'}}} |I| \\
&\quad\lesssim \sum_{n_1,n_2,n_3} \sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} 2^{-\frac{n_1}{p}} 2^{-\frac{n_2}{q}} 2^{-n_3(a_3+\frac{1}{r'})} |I|.
\end{aligned}$$

In the last inequality we need to assume $\frac{1}{p} \le a_1 + \frac{1}{r_1} = \frac{1}{2}(1+\theta_1)$ and similarly $\frac{1}{q} \le \frac{1}{2}(1+\theta_2)$. We will be summing $|I|$ when $I \in \mathcal{I}^{n_1,n_2,n_3}$. Note that

$$\sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} |I| \le \sum_{I \in \mathcal{I}_1^{n_1}} |I| = \left\| \sum_{I \in \mathcal{I}_1^{n_1}} \mathbf{1}_I \right\|_{1,\infty} \lesssim \left\| \sum_{I \in \mathcal{I}_1^{n_1}} 2^{n_1} \mathcal{M}(\mathbf{1}_F) \cdot \mathbf{1}_I \right\|_{1,\infty} \lesssim 2^{n_1} |F|.$$

Similarly, $\sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} |I| \lesssim 2^{n_2}|G|$ and $\sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} |I| \lesssim 2^{n_3}|H|$ and interpolating these three inequalities we get

$$\sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} |I| \lesssim (2^{n_1}|F|)^{\gamma_1} (2^{n_2}|G|)^{\gamma_2} (2^{n_3}|H|)^{\gamma_3},$$

where $0 \le \gamma_j \le 1$ and $\gamma_1 + \gamma_2 + \gamma_3 = 1$. Finally,

$$\begin{aligned}
\left| \sum_{n_1,n_2,n_3} \sum_{I \in \mathcal{I}^{n_1,n_2,n_3}} \Lambda_{\overrightarrow{\mathrm{BHT}};\mathbb{P}(I)}(f,g,h) \right| &\lesssim \sum_{n_1,n_2,n_3} 2^{-\frac{n_1}{p}} 2^{-\frac{n_2}{q}} 2^{-n_3 \frac{1+\theta_3}{2}} (2^{n_1}|F|)^{\gamma_1} (2^{n_2}|G|)^{\gamma_2} (2^{n_3}|H|)^{\gamma_3} \\
&\lesssim \sum_{n_1,n_2,n_3} 2^{-n_1(\frac{1}{p}-\gamma_1)} 2^{-n_2(\frac{1}{q}-\gamma_2)} 2^{-n_3(\frac{1+\theta_3}{2}-\gamma_3)} |F|^{\gamma_1} |G|^{\gamma_2}.
\end{aligned}$$

The above series converges if we can pick $\gamma_j$ such that

$$\frac{1}{p} > \gamma_1, \quad \frac{1}{q} > \gamma_2 \quad \text{and} \quad \frac{1+\theta_3}{2} > \gamma_3.$$

This will be possible as long as

$$\frac{1}{p}+\frac{1}{q}+\frac{1+\theta_3}{2}>1. \tag{49}$$

If the above conditions are satisfied, we get generalized restricted-type estimates

$$|\Lambda_{\overrightarrow{\mathrm{BHT}}}(f,g,h)|\lesssim |F|^{\frac{1}{p}}|G|^{\frac{1}{q}}.$$

There are four distinct cases:

(i) $\frac{1}{r_1},\frac{1}{r_2},\frac{1}{r'}\le\frac{1}{2}$. In this case, if we pick $\theta_1=\theta_2\sim 0$ and $\theta_3\sim 1$, all the conditions hold and the range of $L^p$ estimates for $\overrightarrow{\mathrm{BHT}}_{\vec{r}}$ is going to be the convex hull of the points

$$(0,0,1),\quad (1,0,0),\quad \big(1,\tfrac12,-\tfrac12\big),\quad \big(\tfrac12,1,-\tfrac12\big),\quad (0,1,0).$$

That is, we get the same range as that of the BHT operator: $p,q>1$, $s>\frac{2}{3}$ and $\frac{1}{p}+\frac{1}{q}=\frac{1}{s}$.

(ii) $\frac{1}{r_2},\frac{1}{r'}\le\frac{1}{2}$ and $\frac{1}{r_1}>\frac{1}{2}$. For the condition $\frac{1}{2}(1+\theta_1)-\frac{1}{r_1}>0$ to hold, we have to choose $\theta_1>\frac{2}{r_1}-1$ and this will imply that the range of the operator, described as a region in the hyperplane $\beta_1+\beta_2+\beta_3=1$, is the convex hull of the points

$$(0,0,1),\quad (1,0,0),\quad \big(1,\tfrac12,-\tfrac12\big),\quad \big(\tfrac{1}{r_1},\tfrac32-\tfrac{1}{r_1},-\tfrac12\big),\quad \big(0,\tfrac32-\tfrac{1}{r_1},\tfrac{1}{r_1}-\tfrac12\big).$$

(iii) $\frac{1}{r_1},\frac{1}{r'}\le\frac{1}{2}$ and $\frac{1}{r_2}>\frac{1}{2}$. Similarly to the previous case, the range of the operator is the convex hull of

$$(0,0,1),\quad (0,1,0),\quad \big(1,\tfrac12,-\tfrac12\big),\quad \big(\tfrac32-\tfrac{1}{r_2},\tfrac{1}{r_2},-\tfrac12\big),\quad \big(\tfrac32-\tfrac{1}{r_2},0,\tfrac{1}{r_2}-\tfrac12\big).$$

(iv) $\frac{1}{r_1},\frac{1}{r_2}\le\frac{1}{2}$ and $\frac{1}{r'}>\frac{1}{2}$. The range is the convex hull of

$$(0,0,1),\quad \big(\tfrac12+\tfrac{1}{r},0,\tfrac12-\tfrac{1}{r}\big),\quad \big(\tfrac12+\tfrac{1}{r},\tfrac12,-\tfrac{1}{r}\big),\quad \big(\tfrac12,\tfrac12+\tfrac{1}{r},-\tfrac{1}{r}\big),\quad \big(0,\tfrac12+\tfrac{1}{r},\tfrac12-\tfrac{1}{r}\big).$$

**3.3.** ***The cases $r=1$ or $r_i=\infty$.*** The proof is similar to the one in the previous Section 3.2. We first consider the case $r=1$. Because the dual space of $L^1(\mathcal{W},\mu)$ is $L^\infty(\mathcal{W},\mu)$, the functions appearing in the trilinear form satisfy

$$\|f(x,\cdot)\|_{L^{r_1}(\mathcal{W},\mu)}\le \mathbf{1}_F(x),\quad \|g(x,\cdot)\|_{L^{r_2}(\mathcal{W},\mu)}\le \mathbf{1}_G(x),\quad \|h(x,\cdot)\|_{L^\infty(\mathcal{W},\mu)}\le \mathbf{1}_{H'}.$$

All the details are identical to the case $r>1$; the restrictions are given by only two inequalities:

$$\frac{1+\theta_1}{2}>\frac{1}{r_1},\quad \frac{1+\theta_2}{2}>\frac{1}{r_2}.$$

In the case $r_1=r_2=2$ and $r=1$, these are automatically satisfied and $\mathcal{D}_{r_1,r_2,r}=\mathrm{Range(BHT)}$.

When $r_1=\infty$, we use the fact that the adjoint $\mathrm{BHT}^{*,1}$ of BHT is a bilinear operator of the same kind, which is bounded from $L^r\times L^{r'}\to L^1$; more precisely,

$$\Lambda_{\mathrm{BHT}}(f_w,g_w,h_w)=\int_{\mathbb{R}}\mathrm{BHT}(f_w,g_w)(x)\cdot h_w(x)\,dx=\int_{\mathbb{R}} f_w(x)\cdot\mathrm{BHT}^{*,1}(g_w,h_w)(x)\,dx.$$

In proving the boundedness of vector-valued BHT via interpolation, we assume

$$\|f(x,\cdot)\|_{L^\infty(\mathcal{W},\mu)}\le \mathbf{1}_F(x),\quad \|g(x,\cdot)\|_{L^{r}(\mathcal{W},\mu)}\le \mathbf{1}_G(x),\quad \|h(x,\cdot)\|_{L^{r'}(\mathcal{W},\mu)}\le \mathbf{1}_{H'}.$$

Then

$$\begin{aligned}&\big|\Lambda_{\mathrm{BHT};\mathbb{P}(I)}(f_w,g_w,h_w)\big|\\&\quad\le \big\|\mathrm{BHT}^{*,1}_{\mathbb{P}(I)}(g_w\cdot\mathbf{1}_G,h_w\cdot\mathbf{1}_{H'})\cdot\mathbf{1}_F\big\|_1\\&\quad\lesssim (\widetilde{\mathrm{size}}_{\mathbb{P}(I)}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\epsilon}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{r}-\epsilon}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{r'}-\epsilon}\|g_w\cdot\tilde{\chi}_I\|_r\|h_w\cdot\tilde{\chi}_I\|_{r'}.\end{aligned}$$

The rest follows as before. Note that in the case $(\infty,2,2)$ we have no constraints on $p,q$, and $s$ except those coming from the original BHT operator itself: indeed, for $\theta_2,\theta_3>0$, we have

$$\frac{1+\theta_2}{2}-\frac{1}{2}>0,\quad \frac{1+\theta_3}{2}-\frac{1}{2}>0.$$

**3.4.** ***Iterated $L^p(\mathcal{W},\mu)$ spaces estimates for* BHT.** Previously, we proved that for any tuple $(r_1,r_2,r)$ with $\frac{1}{r_1}+\frac{1}{r_2}=\frac{1}{r}$, $1\le r<\infty$, and $1<r_1,r_2\le\infty$, we have

$$\mathrm{BHT}: L^p(\mathbb{R};L^{r_1}(\mathcal{W},\mu))\times L^q(\mathbb{R};L^{r_2}(\mathcal{W},\mu))\to L^s(\mathbb{R};L^r(\mathcal{W},\mu))$$

whenever $p,q,r$ are in a certain range $\mathcal{D}_{r_1,r_2,r}$, which can be described in a precise manner. The general ideas for proving multiple vector-valued estimates for BHT (as presented in Theorem 8) via the helicoidal method were described in the Introduction. In this section, we present in more detail the proof in the case of two iterated spaces $\ell^s(\ell^r)$ in order to simplify the notation. First, we prove the following localized vector-valued result:

**Proposition 44.**

$$\Big\|\Big(\sum_{k=1}^N\big|\mathrm{BHT}_{\mathbb{P}(I_0)}(f_k\cdot\mathbf{1}_F,g_k\cdot\mathbf{1}_G)\big|^r\Big)^{\frac{1}{r}}\cdot\mathbf{1}_{H'}\Big\|_s\le\widetilde{C}\Big\|\Big(\sum_{k=1}^N|f_k|^{r_1}\Big)^{\frac{1}{r_1}}\cdot\tilde{\chi}_{I_0}\Big\|_p\Big\|\Big(\sum_{k=1}^N|g_k|^{r_2}\Big)^{\frac{1}{r_2}}\cdot\tilde{\chi}_{I_0}\Big\|_q,$$

*where* $\widetilde{C}=(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{p}-\epsilon}(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{q}-\epsilon}(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{s'}-\epsilon}$.

*Proof.* This is going to be a refinement of the proof of Theorem 7 from the previous section. In constructing the collection of intervals $\mathcal{I}_j^{n_j}$, we note that we only need to select intervals $I$ that are already contained in $I_0$, because all the tiles in $\mathbb{P}(I_0)$ are such that $I_P\subseteq I_0$.

As before, we prove generalized restricted-type estimates, and we assume that the functions have the properties

$$\Big(\sum_k|f_k|^{r_1}\Big)^{\frac{1}{r_1}}\le\mathbf{1}_{E_1},\quad\Big(\sum_k|f_k|^{r_2}\Big)^{\frac{1}{r_2}}\le\mathbf{1}_{E_2},\quad\Big(\sum_k|h_k|^{r'}\Big)^{\frac{1}{r'}}\le\mathbf{1}_{E_3'}.$$

The exceptional set is defined by

$$\widetilde{\Omega}=\Big\{\mathcal{M}(\mathbf{1}_{E_1})>C\frac{|E_1|}{|E_3|}\Big\}\cup\Big\{\mathcal{M}(\mathbf{1}_{E_2})>C\frac{|E_2|}{|E_3|}\Big\},$$

and we assume the tiles to be such that $1+\mathrm{dist}(I_P,\widetilde{\Omega}^c)/|I_P|\sim 2^d$.

For intervals $I \in \mathcal{I}_1^{n_1}$, we have

$$\frac{1}{|I|}\int_{\mathbb{R}} \mathbf{1}_{E_1}\cdot\mathbf{1}_F\cdot\tilde{\chi}_I^M\,dx \sim \widetilde{\text{size}}_{\mathbb{P}_{n_1}(I)}(\mathbf{1}_{E_1}\cdot\mathbf{1}_F) \sim 2^{-n_1} \le 2^d\frac{|E_1|}{|E_3|}.$$

When we consider intervals $I \in \mathcal{I}_1^{n_1}\cap\mathcal{I}_2^{n_2}\cap\mathcal{I}_3^{n_3}$, the above approximations become inequalities. We also need to point out that

$$\widetilde{\text{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_1}\cdot\mathbf{1}_F) \le \widetilde{\text{size}}_{\mathbb{P}(I_0)}(\mathbf{1}_{E_1}\cdot\mathbf{1}_F) \quad\text{and}\quad \frac{1}{|I|}\int_{\mathbb{R}} \mathbf{1}_{E_1}\cdot\mathbf{1}_F\cdot\tilde{\chi}_I^M\,dx \le \widetilde{\text{size}}_{\mathbb{P}(I_0)}(\mathbf{1}_{E_1}\cdot\mathbf{1}_F).$$

Now we add the trilinear forms in order to obtain generalized restricted-type estimates:

$$\begin{aligned}
&\sum_k \left|\Lambda_{\text{BHT};\mathbb{P}(I_0)}(f_k\cdot\mathbf{1}_F, g_k\cdot\mathbf{1}_G, h_k\cdot\mathbf{1}_{H'})\right| \\
&\le \sum_{n_1,n_2,n_3}\sum_{I\in\mathcal{I}^{n_1,n_2,n_3}}\sum_k \left|\Lambda_{\text{BHT};\mathbb{P}(I_0\cap I)}(f_k\cdot\mathbf{1}_F, g_k\cdot\mathbf{1}_G, h_k\cdot\mathbf{1}_{H'})\right| \\
&\lesssim \sum_{n_1,n_2,n_3}\sum_{I\in\mathcal{I}^{n_1,n_2,n_3}} (\widetilde{\text{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_1}\cdot\mathbf{1}_F))^{\frac{1+\theta_1}{2}-\frac{1}{r_1}-\epsilon}(\widetilde{\text{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_2}\cdot\mathbf{1}_G))^{\frac{1+\theta_2}{2}-\frac{1}{r_2}-\epsilon} \\
&\qquad\cdot(\widetilde{\text{size}}_{\mathbb{P}(I)}(\mathbf{1}_{E_3'}\cdot\mathbf{1}_{H'}))^{\frac{1+\theta_3}{2}-\frac{1}{r'}-\epsilon}\frac{\|\mathbf{1}_{E_1}\cdot\mathbf{1}_F\cdot\tilde{\chi}_I\|_{r_1}}{|I|^{\frac{1}{r_1}}}\frac{\|\mathbf{1}_{E_2}\cdot\mathbf{1}_G\cdot\tilde{\chi}_I\|_{r_2}}{|I|^{\frac{1}{r_2}}}\frac{\|\mathbf{1}_{E_3'}\cdot\mathbf{1}_{H'}\cdot\tilde{\chi}_I\|_{r'}}{|I|^{\frac{1}{r'}}}|I|.
\end{aligned}$$

Using the *modified* sizes from Definition 21, this implies

$$\begin{aligned}
&\sum_k \left|\Lambda_{\text{BHT};\mathbb{P}(I_0)}(f_k\cdot\mathbf{1}_F, g_k\cdot\mathbf{1}_G, h_k\cdot\mathbf{1}_{H'})\right| \\
&\quad\lesssim (\widetilde{\text{size}}_{\mathbb{P}(I_0)}(\mathbf{1}_{E_1}\cdot\mathbf{1}_F))^{\frac{1+\theta_1}{2}-\frac{1}{p}-\epsilon}(\widetilde{\text{size}}_{\mathbb{P}(I_0)}(\mathbf{1}_{E_2}\cdot\mathbf{1}_G))^{\frac{1+\theta_2}{2}-\frac{1}{q}-\epsilon}(\widetilde{\text{size}}_{\mathbb{P}(I_0)}(\mathbf{1}_{E_3'}\cdot\mathbf{1}_{H'}))^{\frac{1+\theta_3}{2}-\frac{1}{s'}-\epsilon} \\
&\qquad\cdot\sum_{n_1,n_2,n_3}\sum_{I\in\mathcal{I}^{n_1,n_2,n_3}} 2^{-\frac{n_1}{p}}2^{-\frac{n_2}{q}}2^{-n_3(\frac{1}{s'}+\epsilon)}|I|.
\end{aligned}$$

The last part adds up to something $\lesssim 2^{-\tilde{M}d}|E_1|^{\frac{1}{p}}|E_2|^{\frac{1}{q}}|E_3|^{\frac{1}{s'}}$, which is precisely what we were aiming in the beginning.

The cases when one of $r_1, r_2$ or $r' = \infty$ follow in a similar manner. □

The above proposition is an intermediate step in the proof of $L^p$ estimates for $\overrightarrow{\text{BHT}}_{\vec{R}}$, in the case of two iterated vector spaces, which is presented below.

**Proposition 45.**

$$\left\|\left(\sum_l\left(\sum_k|\text{BHT}(f_{kl},g_{kl})\right|^r\right)^{\frac{s}{r}}\right)^{\frac{1}{s}}\right\|_t \le C\left\|\left(\sum_l\left(\sum_k|f_{kl}|^{r_1}\right)^{\frac{s_1}{r_1}}\right)^{\frac{1}{s_1}}\right\|_p\left\|\left(\sum_l\left(\sum_k|g_{kl}|^{r_r}\right)^{\frac{s_2}{r_2}}\right)^{\frac{1}{s_2}}\right\|_q.$$

*Proof.* Once again, we use generalized restricted-type interpolation; $F, G, H$ are sets of finite measure, with $|H| = 1$. The exceptional set is defined as usual, and $H' = H\setminus\Omega$. The sequences of functions will

be such that

$$\Big(\sum_l\Big(\sum_k |f_{kl}|^{r_1}\Big)^{\frac{s_1}{r_1}}\Big)^{\frac{1}{s_1}}\le \mathbf{1}_F,\quad \Big(\sum_l\Big(\sum_k |g_{kl}|^{r_2}\Big)^{\frac{s_2}{r_2}}\Big)^{\frac{1}{s_2}}\le \mathbf{1}_G,\quad \Big(\sum_l\Big(\sum_k |h_{kl}|^{r'}\Big)^{\frac{s'}{r'}}\Big)^{\frac{1}{s'}}\le \mathbf{1}_{H'}.$$

The collections $\mathcal{I}_j^{n_j}$ are going to be chosen in the same way as in the proof of Theorem 7, depending on the sizes and averages of the characteristic functions $\mathbf{1}_F, \mathbf{1}_G, \mathbf{1}_{H'}$. Proposition 44 yields the following:

$$\sum_k \big|\Lambda_{\mathrm{BHT};\mathbb{P}(I)}(f_{kl},g_{kl},h_{kl})\big|$$

$$\lesssim (\widetilde{\mathrm{size}}_{\mathbb{P}(I)}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{s_1}-\epsilon}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{s_2}-\epsilon}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{s'}-\epsilon} \tag{50}$$

$$\cdot\Big\|\Big(\sum_k |f_{kl}|^{r_1}\Big)^{\frac{1}{r_1}}\tilde{\chi}_I\Big\|_{s_1}\Big\|\Big(\sum_k |g_{kl}|^{r_2}\Big)^{\frac{1}{r_2}}\tilde{\chi}_I\Big\|_{s_2}\Big\|\Big(\sum_k |h_{kl}|^{r'}\Big)^{\frac{1}{r'}}\tilde{\chi}_I\Big\|_{s'}. \tag{51}$$

Then we sum (51) over $l$ as well, and apply Hölder on the triple $(s_1,s_2,s')$. In this way, we recover $\|\mathbf{1}_F\cdot\tilde{\chi}_I\|_{s_1}$, and the corresponding quantities for the second and third entries. We have

$$\Big|\sum_{k,l}\Lambda_{\mathrm{BHT}}(f_{kl},g_{kl},h_{kl})\Big|$$

$$\lesssim \sum_{n_1,n_2,n_3}\sum_{\mathcal{I}^{n_1,n_2,n_3}} (\widetilde{\mathrm{size}}_{\mathbb{P}(I)}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{s_1}-\epsilon}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{s_2}-\epsilon}(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{s'}-\epsilon}$$

$$\cdot\frac{\|\mathbf{1}_F\tilde{\chi}_I\|_{s_1}}{|I|^{\frac{1}{s_1}}}\frac{\|\mathbf{1}_G\tilde{\chi}_I\|_{s_2}}{|I|^{\frac{1}{s_2}}}\frac{\|\mathbf{1}_{H'}\tilde{\chi}_I\|_{s'}}{|I|^{\frac{1}{s'}}}|I|$$

$$\lesssim \sum_{n_1,n_2,n_3}\sum_{\mathcal{I}^{n_1,n_2,n_3}} (\text{“}\widetilde{\mathrm{size}}\text{”}_{\mathbb{P}(I)}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{p}-\epsilon}(\text{“}\widetilde{\mathrm{size}}\text{”}_{\mathbb{P}(I)}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{q}-\epsilon}(\text{“}\widetilde{\mathrm{size}}\text{”}_{\mathbb{P}(I)}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{t'}-\epsilon}$$

$$\cdot 2^{-\frac{n_1}{p}}2^{-\frac{n_2}{q}}2^{-n_3(\frac{-1}{t'}+\epsilon)}|I|.$$

**Remark.** The “sizes” appearing in the line above are not exactly the ones from Definition 19, but the modified ones from Definition 21 . Note that

$$\max\Big(\widetilde{\mathrm{size}}_{\mathbb{P}(I)}\mathbf{1}_F,\frac{1}{|I|}\int_{\mathbb{R}}\mathbf{1}_F\cdot\tilde{\chi}_I^M\,dx\Big)\le \text{“}\widetilde{\mathrm{size}}\text{”}_{\mathbb{P}(I)}\mathbf{1}_F.$$

This is the step where we can prove also the localized version of the statement in Proposition 45. Assuming all the tiles are sitting above an interval $I_0$, we can obtain the same result with operatorial norm

$$(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_F)^{\frac{1+\theta_1}{2}-\frac{1}{p}-\epsilon}(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_G)^{\frac{1+\theta_2}{2}-\frac{1}{q}-\epsilon}(\widetilde{\mathrm{size}}_{\mathbb{P}(I_0)}\mathbf{1}_{H'})^{\frac{1+\theta_3}{2}-\frac{1}{t'}-\epsilon}.$$

The rest of the proof is identical to the simpler vector case of Theorem 7; the quantities on the left-hand side add up to $|F|^{\frac{1}{p}}|G|^{\frac{1}{q}}$, provided

$$\frac{1+\theta_1}{2}>\frac{1}{p},\quad \frac{1+\theta_2}{2}>\frac{1}{q},\quad \frac{1+\theta_3}{2}>\frac{1}{s'}.$$

□

## 4. Similar results for paraproducts: proof of Theorem 9

The paraproduct case is similar to BHT, even though the bilinear Hilbert transform is a much more complicated object. The extra difficulties are hidden in Proposition 23, but we will see from the proof of the vector-valued extensions that the complexity of the paraproduct case is comparable to the "local $L^2$" case for BHT. In both situations, we recover the maximal range for vector-valued estimates.

We will be working with the *discretized paraproduct* of the functions $f$ and $g$, which is defined by

$$\Pi(f,g)(x) = \sum_{I\in\mathcal{I}} \frac{1}{|I|^{\frac{1}{2}}} \langle f, \phi_I^1\rangle \langle g, \phi_I^2\rangle \phi_I^3(x).$$

Here $\mathcal{I}$ is a family of dyadic intervals, and the wave packets $\{\phi_I^j\}_{I\in\mathcal{I}}$ are so that two of the families are *lacunary* ($\phi_I^j$ is a wave packet on $I\times[1/|I|, 2/|I|]$), and the third one is *nonlacunary* ($\phi_I^{j_0}$ is a wave packet on $I\times[0,1/|I|]$). Again, we present the case of $\ell^p$ spaces for simplicity. The operator we are interested in is

$$\vec{\Pi}_r(f,g) := \left(\sum_{k=1}^N \left|\Pi(f_k,g_k)\right|^r\right)^{\frac{1}{r}}.$$

**Remark.** We could alternatively look at operators of the form

$$(f,g)\mapsto \left(\sum_{k=1}^N \left|\Pi_k(f_k,g_k)\right|^r\right)^{\frac{1}{r}},$$

where each paraproduct $\Pi_k$ is associated to a family $\mathcal{I}_k$ of dyadic intervals. The $\Pi_k$ don't need to be precisely the same, but they display a similar behavior. Similarly, for $\overrightarrow{\mathrm{BHT}}$ we could have a "perturbation" $\mathrm{BHT}_w$ for each $w\in\mathcal{W}$, and the method of the proof applies in that case as well.

**4.1.** ***A few results about paraproducts.*** The concepts of *sizes* and *energies* are similar to the corresponding ones for the bilinear Hilbert transform; we don't need to organize the tiles into trees because the family of tiles is of rank 0. We recall some definitions below.

**Definition 46.** Let $\mathcal{I}$ be a family of dyadic intervals. For any $1\le j\le 3$, we define

$$\mathrm{size}_{\mathcal{I}}\big(\langle f,\phi_I^j\rangle_{I\in\mathcal{I}}\big) = \sup_{I\in\mathcal{I}} \frac{|\langle f,\phi_I^j\rangle|}{|I|^{\frac{1}{2}}} \quad \text{if } (\phi_I^j)_I \text{ is nonlacunary}$$

and

$$\mathrm{size}_{\mathcal{I}}\big(\langle f,\phi_I^j\rangle_{I\in\mathcal{I}}\big) = \sup_{I_0\in\mathcal{I}} \frac{1}{|I_0|^{\frac{1}{2}}} \left\| \left(\sum_{\substack{I\in\mathcal{I}\\ I\subseteq I_0}} \frac{|\langle f,\phi_I^j\rangle|^2}{|I|}\cdot \mathbf{1}_I\right)^{\frac{1}{2}}\right\|_{1,\infty} \quad \text{if } (\phi_I^j)_I \text{ is lacunary.}$$

Similarly to the BHT case, energy is defined as

$$\mathrm{energy}_{\mathcal{I}}^j\big(\langle f,\phi_I^j\rangle_{I\in\mathcal{I}}\big) := \sup_{n\in\mathbb{Z}} 2^n \sup_{\mathbb{D}}\left(\sum_{I\in\mathbb{D}} |I|\right),$$

where $\mathbb{D}$ ranges over all collections of disjoint intervals $I_0$ with the property that

$$\frac{|\langle f, \phi_{I_0}^j\rangle|}{|I_0|^{\frac{1}{2}}} \geq 2^n \quad \text{if } (\phi_I^j)_I \text{ is nonlacunary}$$

and

$$\frac{1}{|I_0|}\Big\|\Big(\sum_{\substack{I\in\mathcal{I}\\ I\subseteq I_0}} \frac{|\langle f,\phi_I^j\rangle|^2}{|I|}\cdot \mathbf{1}_I\Big)^{\frac{1}{2}}\Big\|_{1,\infty} \geq 2^n \quad \text{if } (\phi_I^j)_I \text{ is lacunary.}$$

We have estimates similar to Lemmas 20 and 24. However, because we don't need to use orthogonality of trees, the energy becomes an $L^1$ quantity.

**Lemma 47** [Muscalu and Schlag 2013, Lemma 2.13]. *If $F$ is an $L^1$ function and $1 \leq j \leq 3$, then*

$$\operatorname{size}_{\mathcal{I}}^j\big(\langle F,\phi_I^j\rangle_{I\in\mathcal{I}}\big) \lesssim \sup_{I\in\mathcal{I}} \frac{1}{|I|}\int_{\mathbb{R}} |F|\tilde{\chi}_I^M\, dx$$

*for $M > 0$, with implicit constants depending on $M$.*

**Lemma 48** [Muscalu and Schlag 2013, Lemma 2.14]. *If $F$ is an $L^1$ function and $1 \leq j \leq 3$, then*

$$\operatorname{energy}_{\mathcal{I}}^j\big(\langle F,\phi_I^j\rangle_{I\in\mathcal{I}}\big) \lesssim \|F\|_1.$$

**Proposition 49** [Muscalu and Schlag 2013, Proposition 2.12]. *Given a paraproduct $\Pi$ associated with a family $\mathcal{I}$ of intervals,*

$$\begin{aligned}\big|\Lambda_\Pi(f_1,f_2,f_3)\big| &= \Big|\sum_{I\in\mathcal{I}} \frac{1}{|I|^{\frac{1}{2}}}\langle f_1,\phi_I^1\rangle\langle f_2,\phi_I^2\rangle\langle f_3,\phi_I^3\rangle\Big|\\ &\lesssim \prod_{j=1}^3 \big(\operatorname{size}_{\mathcal{I}}^{(j)}(\langle f_j,\phi_I^j\rangle_{I\in\mathcal{I}})\big)^{1-\theta_j}\big(\operatorname{energy}_{\mathcal{I}}^{(j)}(\langle f_j,\phi_I^j\rangle_{I\in\mathcal{I}})\big)^{\theta_j}\end{aligned}$$

*for any $0 \leq \theta_1,\theta_2,\theta_3 < 1$ such that $\theta_1+\theta_2+\theta_3 = 1$, where the implicit constant depends on $\theta_1,\theta_2,\theta_3$ only.*

While the above proposition is the main ingredient, we need "localized" estimates. If $I_0$ is some fixed dyadic interval, then we define

$$\Pi(I_0)(f,g)(x) = \sum_{\substack{I\in\mathcal{I}\\ I\subseteq I_0}} \frac{1}{|I|^{\frac{1}{2}}}\langle f,\phi_I^1\rangle\langle g,\phi_I^2\rangle\phi_I^3(x).$$

Here again we need some localization results which play the role of Proposition 42 and Corollary 43 from the BHT case.

The trilinear form associated to the localized paraproduct is given by

$$\Lambda_{\Pi(I_0)}^{F,G,H'}(f,g,h) := \Lambda_{\Pi(I_0)}(f\cdot\mathbf{1}_F, g\cdot\mathbf{1}_G, h\cdot\mathbf{1}_{H'}).$$

**Proposition 50.** *Let $I_0$ be a fixed dyadic interval and $F, G, H' \subset \mathbb{R}$ sets of finite measure. Then there exist some positive numbers $0 \leq a_1, a_2, a_3 < 1$ so that*

$$\left|\Lambda^{F,G,H'}_{\Pi(I_0)}(f,g,h)\right| \lesssim (\widetilde{\text{size}}_{\mathcal{I}(I_0)}\mathbf{1}_F)^{a_1} (\widetilde{\text{size}}_{\mathcal{I}(I_0)}\mathbf{1}_G)^{a_2} (\widetilde{\text{size}}_{\mathcal{I}(I_0)}\mathbf{1}_{H'})^{a_3} \|f\cdot\tilde{\chi}_{I_0}\|_{r_1} \|g\cdot\tilde{\chi}_{I_0}\|_{r_2} \|h\cdot\tilde{\chi}_{I_0}\|_{r'}$$

*whenever $\frac{1}{r_1}+\frac{1}{r_2}+\frac{1}{r'}=1$, and $1<r_1,r_2,r'<\infty$. Here $a_j = 1-\frac{1}{r_j}-\epsilon$.*

*Proof.* The idea of the proof is very similar to that of Proposition 41. Restricted-type estimates are proved by performing a triple stopping time and then the result follows by interpolation. We leave the routine details to the reader. □

The case $r=1$ is obtained through interpolation of restricted-type estimates only. This comes in contrast with the $r=1$ case for BHT, where generalized restricted-type interpolation is necessary. More exactly, for the BHT operator, in order to conclude estimates for $\left(\frac{1}{r_1},\frac{1}{r_2},0\right)$, one needs to interpolate between good ($\beta_i>0$) and bad ($\beta_3<0$) tuples $\beta=(\beta_1,\beta_2,\beta_3)$.

**Proposition 51.** *If $H'$ is a fixed set of finite measure,*

$$\left|\Lambda_{\Pi(I_0)}(f,g,\mathbf{1}_{H'})\right| \lesssim \widetilde{\text{size}}_{\mathcal{I}(I_0)}\mathbf{1}_{H'} \|f\cdot\tilde{\chi}_{I_0}\|_p \|g\cdot\tilde{\chi}_{I_0}\|_q \tag{52}$$

*whenever $\frac{1}{p}+\frac{1}{q}=1$, and $1<p,q<\infty$.*

*Proof.* In this case $\Lambda_{\Pi(I_0)}(f,g,\mathbf{1}_{H'})$ becomes a bilinear form with respect to the first two entries. Because of the decay of $\tilde{\chi}_{I_0}$, it will be sufficient to prove the proposition in the case $\operatorname{supp} f,g \subseteq 5I_0$. By Theorem 28, it will be enough to show restricted-type estimates for the bilinear form

$$(f,g)\mapsto \Lambda_{\Pi(I_0)}(f,g,\mathbf{1}_{H'}).$$

Let $F$ and $G$ be sets of finite measure and $|f|\leq \mathbf{1}_F$ and $|g|\leq \mathbf{1}_G$. Using Proposition 49 with $\theta_3=0$ and estimating $\widetilde{\text{size}}_{\mathcal{I}(I_0)}f \lesssim 1$ and $\widetilde{\text{size}}_{\mathcal{I}(I_0)}g\lesssim 1$, we get

$$\left|\Lambda_{\Pi(I_0)}(f,g,\mathbf{1}_{H'})\right| \lesssim \widetilde{\text{size}}_{\mathcal{I}(I_0)}\mathbf{1}_{H'}\, |F|^{\theta_1}|G|^{\theta_2},$$

where $\theta_1+\theta_2=1$ and $0<\theta_1,\theta_2<1$. This proves restricted-type estimates in a small neighborhood of $\left(\frac{1}{p},\frac{1}{q}\right)$. □

**4.2.** ***Proof of Theorem 8: a particular case.*** We will be using vector-valued interpolation theorems, as usual. Hence, we fix sets of finite measure $F, G$ and $H$ and we assume $|H|=1$. Let $f=\{f_k\}_k$ and $g=\{g_k\}_k$, with $(\sum_k |f_k|^{r_1})^{\frac{1}{r_1}} \leq \mathbf{1}_F$ and $(\sum_k |g_k|^{r_2})^{\frac{1}{r_2}}\leq \mathbf{1}_G$.

The exceptional set will be

$$\widetilde{\Omega} := \left\{x : \mathcal{M}(\mathbf{1}_F)(x) > C|F|\right\} \cup \left\{x : \mathcal{M}(\mathbf{1}_G)(x) > C|G|\right\}$$

and $H' = H\setminus\widetilde{\Omega}$. We have a sequence of functions $\{h_k\}_k$ with $\left(\sum_k |h_k|^{r'}\right)^{\frac{1}{r'}}\leq \mathbf{1}_{H'}$.

For every $d\geq 0$,

$$\mathcal{I}^d := \left\{ I\in\mathcal{I} : 1+\frac{\operatorname{dist}(I,\Omega^c)}{|I|}\sim 2^d\right\}.$$

When estimating paraproducts associated to the collection $\mathcal{I}^d$, we get an extra $2^{-10d}$ decay and thus the $d$-dependency of the paraproducts can be assumed to be implicit. As before, for each of the sets $F, G$ and $H'$ we define collections of disjoint maximal intervals $\mathcal{J}_1^{n_1}$, $\mathcal{J}_2^{n_2}$ and $\mathcal{J}_3^{n_3}$ respectively. For example, if $I \in \mathcal{J}_1^{n_1}$, then

$$2^{-n_1-1} \leq \frac{1}{|I|}\int_{\mathbb{R}} \mathbf{1}_F \cdot \tilde{\chi}_I \, dx \leq 2^{-n_1} \lesssim |F|.$$

Returning to the operator $\vec{\Pi}_r$, we have for the associated multilinear form

$$\Big|\sum_k \Lambda_{\Pi}(f_k, g_k, h_k)\Big| \leq \sum_{n_1,n_2,n_3} \sum_{I_0 \in \mathcal{J}^{n_1,n_2,n_3}} \sum_k \big|\Lambda_{\Pi(I_0)}(f_k, g_k, h_k)\big|.$$

Now we use the localization results of Proposition 50 to estimate the above expression by

$$\sum_{n_1,n_2,n_3} \sum_{I_0 \in \mathcal{J}^{n_1,n_2,n_3}} \sum_{k=1}^{n} (\widetilde{\text{size}}_{\mathcal{I}(I_0)} \mathbf{1}_F)^{b_1} (\widetilde{\text{size}}_{\mathcal{I}(I_0)} \mathbf{1}_G)^{b_2} (\widetilde{\text{size}}_{\mathcal{I}(I_0)} \mathbf{1}_{H'})^{b_3} \cdot \|f_k \cdot \tilde{\chi}_{I_0}\|_{r_1} \|g_k \cdot \tilde{\chi}_{I_0}\|_{r_2} \|h_k \cdot \tilde{\chi}_{I_0}\|_{r'}$$

$$\lesssim \sum_{n_1,n_2,n_3} \sum_{I_0 \in \mathcal{J}^{n_1,n_2,n_3}} (\widetilde{\text{size}}_{\mathcal{I}(I_0)} \mathbf{1}_F)^{b_1} (\widetilde{\text{size}}_{\mathcal{I}(I_0)} \mathbf{1}_G)^{b_2} (\widetilde{\text{size}}_{\mathcal{I}(I_0)} \mathbf{1}_{H'})^{b_3} \frac{\|\mathbf{1}_F \cdot \tilde{\chi}_{I_0}\|_{r_1}}{|I_0|^{\frac{1}{r_1}}} \frac{\|\mathbf{1}_G \cdot \tilde{\chi}_{I_0}\|_{r_2}}{|I_0|^{\frac{1}{r_2}}} \frac{\|\mathbf{1}_{H'} \cdot \tilde{\chi}_{I_0}\|_{r'}}{|I_0|^{\frac{1}{r'}}} |I_0|.$$

Here we choose some $0 \leq b_j \leq a_j$, which we can do because the sizes are subunitary. Whenever $0 \leq \gamma_j \leq 1$ are so that $\gamma_1 + \gamma_2 + \gamma_3 = 1$,

$$\sum_{I_0 \in \mathcal{J}^{n_1,n_2,n_3}} |I_0| \lesssim (2^{n_1}|F|)^{\gamma_1} (2^{n_2}|G|)^{\gamma_2} (2^{n_3}|H|)^{\gamma_3}.$$

Adding all the pieces together we have

$$\Big|\sum_k \Lambda_{\Pi}(f_k, g_k, h_k)\Big| \lesssim \sum_{n_1,n_2,n_3} 2^{-n_1(b_1+\frac{1}{p}-\gamma_1)} 2^{-n_2(b_2+\frac{1}{q}-\gamma_2)} 2^{-n_3(b_3+\frac{1}{r'}-\gamma_3)} |F|^{\gamma_1} |G|^{\gamma_2}$$
$$\lesssim |F|^{\frac{1}{p}} |G|^{\frac{1}{q}}.$$

Of course, the last inequality is true provided we can choose $\gamma_1, \gamma_2, \gamma_3$ so that the series converges. Choosing the $\theta_j$ and $\alpha_j$ carefully, one can prove that the restricted weak-type estimates hold arbitrarily close to the points

$$(0, 0, 1), \quad (1, 0, 0), \quad (0, 1, 0), \quad (1, 1, -1).$$

Then the general result follows by interpolation.

**Remark.** With a few adjustments, the proof is valid in the case $r = 1$ as well.

## 5. Tensor products BHT $\otimes \Pi^{\otimes n}$

In this section, we will prove the boundedness of the tensor product

$$\mathrm{BHT} \otimes \Pi^{\otimes n} = \mathrm{BHT} \otimes \Pi \otimes \cdots \otimes \Pi : L^p(\mathbb{R}^{n+1}) \times L^q(\mathbb{R}^{n+1}) \to L^r(\mathbb{R}^{n+1})$$

whenever $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, with $\frac{2}{3} < r < \infty$, $1 \le p, q < \infty$.

If $T_1 : L^p(\mathbb{R}^{n_1}) \times L^q(\mathbb{R}^{n_1}) \to L^r(\mathbb{R}^{n_1})$ and $T_2 : L^p(\mathbb{R}^{n_2}) \times L^q(\mathbb{R}^{n_2}) \to L^r(\mathbb{R}^{n_2})$ are two bilinear operators, then the tensor product

$$T_1 \otimes T_2 : L^p(\mathbb{R}^{n_1+n_2}) \times L^q(\mathbb{R}^{n_1+n_2}) \to L^r(\mathbb{R}^{n_1+n_2})$$

will act as $T_1$ in the first variable and as $T_2$ in the second variable. In our case, the operators are given by singular multipliers, and in this situation we can give a characterization of the tensor product. Assume

$$T_1(f,g)(x) = \int_{\mathbb{R}^{2n_1}} \hat{f}(\xi_1)\hat{g}(\xi_2) m_1(\xi_1,\xi_2) e^{2\pi i x(\xi_1+\xi_2)}\, d\xi_1\, d\xi_2$$

and similarly

$$T_2(f,g)(y) = \int_{\mathbb{R}^{2n_2}} \hat{f}(\eta_1)\hat{g}(\eta_2) m_2(\eta_1,\eta_2) e^{2\pi i y(\eta_1+\eta_2)}\, d\eta_1\, d\eta_2.$$

Then the multiplier of the tensor product is precisely $m_1(\xi_1,\xi_2)\cdot m_2(\eta_1,\eta_2)$:

$$\begin{aligned} &T_1 \otimes T_2(f,g)(x,y) \\ &\qquad = \int \hat{f}(\xi_1,\eta_1)\hat{g}(\xi_2,\eta_2) m_1(\xi_1,\xi_2) m_2(\eta_1,\eta_2) e^{2\pi i x(\xi_1+\xi_2)} e^{2\pi i y(\eta_1+\eta_2)}\, d\xi_1\, d\xi_2\, d\eta_1\, d\eta_2. \end{aligned}$$

The multiplier associated with BHT is $\mathrm{sgn}(\xi_1 - \xi_2)$, while the multiplier of a paraproduct of two functions on the real line is a classical Marcinkiewicz–Mikhlin–Hörmander multiplier $m(\xi_1,\xi_2)$, smooth away from the origin, satisfying the condition $|\partial^\alpha m(\xi)| \lesssim |\xi|^{-|\alpha|}$ for sufficiently many multi-indices $\alpha$. The decay in $m$ and a Fourier series decomposition allows one to approximate the multiplier by a finite number of sums of the form

$$\sum_k \hat{\varphi}_k(\xi_1)\hat{\psi}_k(\xi_2)\hat{\psi}_k(\xi_1+\xi_2), \quad \sum_k \hat{\psi}_k(\xi_1)\hat{\varphi}_k(\xi_2)\hat{\psi}_k(\xi_1+\xi_2) \quad \text{or} \quad \sum_k \hat{\psi}_k(\xi_1)\hat{\psi}_k(\xi_2)\hat{\varphi}_k(\xi_1+\xi_2).$$

Recall that $Q_k$ is the Littlewood–Paley projection onto $\{|\xi| \sim 2^k\}$ (which is really the convolution with $\psi_k(\cdot)$), and $P_k$ is the projection onto $\{|\xi| \le 2^k\}$, corresponding to the convolution with $\varphi_k$. Then we can regard paraproducts as being expressions of the form

$$\sum_k Q_k(P_k f \cdot Q_k g)(x,y), \quad \sum_k Q_k(Q_k f \cdot P_k g)(x,y) \quad \text{or} \quad \sum_k P_k(Q_k f \cdot Q_k g)(x,y). \tag{53}$$

It is important in the following proofs that the outermost functions $\hat{\varphi}_k(\xi_1+\xi_2)$ and $\hat{\psi}_k(\xi_1+\xi_2)$ are identically equal to 1 on the supports of $\hat{\psi}_k(\xi_1)\cdot\hat{\psi}_k(\xi_2)$ and $\hat{\psi}_k(\xi_1)\cdot\hat{\varphi}_k(\xi_2)$ respectively. This can always be achieved with the price of an extra decomposition.

**Proposition 52.** *Let $T_m : L^p(\mathbb{R}^n) \times L^q(\mathbb{R}^n) \to L^r(\mathbb{R}^n)$ be a bilinear operator with smooth symbol $m$, and $\Pi : L^p(\mathbb{R}) \times L^q(\mathbb{R}) \to L^r(\mathbb{R})$ a paraproduct as described above.*

(1) *If $\Pi$ is given by $\sum_k Q_k(P_k f \cdot Q_k g)(x, y)$, then*

$$(T_m \otimes \Pi)(f, g)(x, y) = \sum_k Q_k^2\big(T_m(P_k^y f, Q_k^y g)\big)(x) = \sum_k T_m(P_k^y f, Q_k^y g)(x).$$

(2) *If $\Pi$ is given by $\sum_k P_k(Q_k f \cdot Q_k g)(x, y)$, then*

$$(T_m \otimes \Pi)(f, g)(x, y) = \sum_k P_k^2\big(T_m(Q_k^y f, Q_k^y g)\big)(x) = \sum_k T_m(Q_k^y f, Q_k^y g)(x).$$

*Here we need to explain the notation: $Q_k^2$ denotes the projection onto $|\xi_2| \sim 2^k$ in the second variable, and $P_k^y f$ is a function of $x$ only, with the variable $y$ fixed. The exact formulas are*

$$P_k^y f(x) = \int_{\mathbb{R}} \varphi_k(s) f(x, y-s)\, ds, \quad P_k^2 f(x, y) = \int_{\mathbb{R}} \varphi_k(s) f(x, y-s)\, ds,$$
$$Q_k^y f(x) = \int_{\mathbb{R}} \psi_k(s) f(x, y-s)\, ds, \quad Q_k^2 f(x, y) = \int_{\mathbb{R}} \psi_k(s) f(x, y-s)\, ds.$$

*Proof.* The proof is a series of direct computations, and we only present the case (1):

$$\begin{aligned}
&(T_m \otimes \Pi)(f, g)(x, y)\\
&\quad = \int_{\mathbb{R}^{2n+2}} \hat f(\xi_1, \eta_1)\, \hat g(\xi_2, \eta_2) m(\xi_1, \xi_2)\\
&\qquad\qquad \Big(\sum_k \hat\varphi_k(\eta_1)\, \hat\psi_k(\eta_2)\, \hat\psi_k(\eta_1 + \eta_2)\Big) e^{2\pi i x(\xi_1+\xi_2)} e^{2\pi i y(\eta_1+\eta_2)}\, d\xi\, d\eta\\
&\quad = \sum_k \int_{\mathbb{R}^{2n+2}} \hat f(\xi_1, \eta_1)\, \hat g(\xi_2, \eta_2) m(\xi_1, \xi_2)\, \hat\varphi_k(\eta_1)\, \hat\psi_k(\eta_2)\\
&\qquad\qquad \Big(\int_{\mathbb{R}} \psi_k(s) e^{-2\pi i s(\eta_1+\eta_2)}\, ds\Big) e^{2\pi i x(\xi_1+\xi_2)} e^{2\pi i y(\eta_1+\eta_2)}\, d\xi\, d\eta\\
&\quad = \sum_k \int_{\mathbb{R}} \psi_k(s)\big(T_m(P_k^{y-s} f, Q_k^{y-s} g)(x)\big)\, ds\\
&\quad = \sum_k Q_k^2 T_m(P_k^y f, Q_k^y g)(x).
\end{aligned}$$

□

A final ingredient that we will need in the proof of Theorem 6 is the following lemma, which appears in [Ruan 2010]:

**Lemma 53.** *Let $f \in \mathcal{S}(\mathbb{R}^n)$, and $1 \le l \le n$, and $\{i_1, \dots, i_l\} \subset \{1, \dots, n\}$. Then*

$$\|f\|_{L^p} \lesssim \Big\| \Big( \sum_{k_1, \dots, k_l} |Q_{k_1}^{i_1} \cdots Q_{k_l}^{i_l} f|^2 \Big)^{\frac{1}{2}} \Big\|_{L^p}$$

*for any $0 < p < \infty$.*

Lemma 53 above states that the $L^p$ norm of $f$ is bounded by the $L^p$ norm of a square function associated with the variables $x_{i_1}, \dots, x_{i_l}$, even when $0 < p \le 1$. In the case $p > 1$, it is well known that the two norms are equivalent. When $p < 1$, the proof makes use of multiparameter Hardy spaces.

**5.1.** ***Proof of Theorem 6.*** We start with the proof in the case $\mathrm{BHT} \otimes \Pi$, in order to make the presentation clear.

(a) Assume that $\Pi(f,g) = \sum_k Q_k(P_k f \cdot Q_k g)$. Then Proposition 52 implies that $\mathrm{BHT}\otimes\Pi(f,g)(x,y) = \sum_k Q_k^2 \mathrm{BHT}(P_k^y f, Q_k^y g)(x)$. Lemma 53 yields

$$\|\mathrm{BHT}\otimes\Pi\|_{L^s(\mathbb{R}^2)} \lesssim \left\| \left( \sum_k |Q_k^2 \mathrm{BHT}(P_k^y f, Q_k^y g)|^2 \right)^{\frac{1}{2}} \right\|_{L^s(\mathbb{R}^2)}.$$

For the paraproducts that we are considering, $Q_k(P_k f \cdot Q_k g)(y) = P_k f(y) \cdot Q_k g(y)$, so we need to estimate

$$\left\| \left( \sum_k |\mathrm{BHT}(P_k^y f, Q_k^y g)|^2 \right)^{\frac{1}{2}} \right\|_{L^s(\mathbb{R}^2)}.$$

We first estimate the $L^s$ norm of $x \mapsto \left(\sum_k |\mathrm{BHT}(P_k^y f, Q_k^y g)(x)|^2\right)^{\frac{1}{2}}$, and Fubini will imply the desired result for $\mathrm{BHT}\otimes\Pi$. Here we use the vector-valued extension for the bilinear Hilbert transform

$$\mathrm{BHT} : L^p(\ell^\infty) \times L^q(\ell^2) \to L^s(\ell^2),$$

which holds whenever $(p,q,s) \in \mathrm{Range(BHT)}$. More exactly,

$$\begin{aligned}
\|\mathrm{BHT}\otimes\Pi\|_{L^s(\mathbb{R}^2)} &\lesssim \left\| \left\| \left( \sum_k |\mathrm{BHT}(P_k^y f, Q_k^y g)(x)|^2 \right)^{\frac{1}{2}} \right\|_{L_x^s} \right\|_{L_y^s} \\
&\lesssim \left\| \left\| \sup_k |P_k^y f| \right\|_{L_x^p} \left\| \left( \sum_k |Q_k^y g|^2 \right)^{\frac{1}{2}} \right\|_{L_x^q} \right\|_{L_y^s} \\
&\lesssim \left\| \left\| \sup_k |P_k^y f| \right\|_{L_x^p} \right\|_{L_y^p} \left\| \left\| \left( \sum_k |Q_k^y g|^2 \right)^{\frac{1}{2}} \right\|_{L_x^q} \right\|_{L_y^q} \\
&\lesssim \|f\|_p \|g\|_q.
\end{aligned}$$

To get the conclusion, we are using Fubini again, and the boundedness of the maximal and square function operators.

(b) The case $\Pi(f,g) = \sum_k P_k(Q_k f, Q_k g)$ is more direct, but the ideas are similar. The functions $\varphi$ in the paraproduct definition are such that $\Pi(f,g) = \sum_k (Q_k f \cdot Q_k g)$, so we have

$$\mathrm{BHT}\otimes\Pi(f,g)(x,y) = \sum_k \mathrm{BHT}(Q_k^y f, Q_k^y g)(x).$$

Now we use the vector-valued extension $\mathrm{BHT} : L^p(\ell^2) \times L^q(\ell^2) \to L^s(\ell^1)$ (which is well-defined for any $(p, q, s) \in \mathrm{Range}(\mathrm{BHT})$) together with Fubini and the boundedness of the square function to get

$$\begin{aligned}\|\mathrm{BHT} \otimes \Pi\|_{L^s(\mathbb{R}^2)} &\lesssim \Big\| \Big\| \sum_k |\mathrm{BHT}(Q_k^y f, Q_k^y g)(x)| \Big\|_{L_x^s} \Big\|_{L_y^s} \\ &\lesssim \Big\| \Big\| \Big( \sum_k |Q_k^y f|^2 \Big)^{\frac{1}{2}} \Big\|_{L_x^p} \Big\| \Big( \sum_k |Q_k^y g|^2 \Big)^{\frac{1}{2}} \Big\|_{L_x^q} \Big\|_{L_y^s} \\ &\lesssim \|f\|_p \|g\|_q.\end{aligned}$$

The general case of Theorem 6 is similar, but slightly more technical. We present it below for completeness. The paraproducts can be of three types, as seen in (53). This generates a partition of $\{1, \dots, n\}$ into three subsets of indices $\mathcal{I}_1, \mathcal{I}_2$ and $\mathcal{I}_3$ so that if $k \in \mathcal{I}_1$, then

$$\Pi(f, g)(y) = \sum_k Q_k(P_k f \cdot Q_k g)(y),$$

and similarly for $\mathcal{I}_2$ and $\mathcal{I}_3$.

Because the projections on different coordinates commute, i.e., $Q_k^i P_l^j = P_l^j Q_k^i$ and $Q_k^i Q_l^j = Q_l^j Q_k^i$, we can assume

$$\mathcal{I}_1 = \{1, \dots, l\}, \quad \mathcal{I}_2 = \{l+1, \dots, l+d\}, \quad \mathcal{I}_3 = \{l+d+1, \dots, n\}.$$

Of course, we allow the possibility that one or even two of these sets of indices are empty. With this assumption, Proposition 52 applied iteratively yields

$$\begin{aligned}&\mathrm{BHT} \otimes \Pi \otimes \cdots \otimes \Pi(f, g)(x, y_1, \dots, y_n) \\ &= \sum_{k_1, \dots, k_n} Q_{k_1}^1 \cdots Q_{k_l}^l Q_{k_{l+1}}^{l+1} \cdots Q_{k_{l+d}}^{l+d} P_{k_{l+d+1}}^{l+d+1} \cdots P_{k_n}^n \circ \\ &\qquad \mathrm{BHT}\big(P_{k_1}^{y_1} \cdots P_{k_l}^{y_l} Q_{k_{l+1}}^{y_{l+1}} \cdots Q_{k_n}^{y_n} f, Q_{k_1}^{y_1} \cdots Q_{k_l}^{y_l} P_{k_{l+1}} \cdots P_{k_{l+d}}^{y_{l+d}} Q_{k_{l+d+1}}^{y_{l+d+1}} \cdots Q_{k_n}^{y_n} g\big)(x).\end{aligned}$$

The outer-most expressions $Q_{k_1}^1 \cdots Q_{k_l}^l Q_{k_{l+1}}^{l+1} \cdots Q_{k_{l+d}}^{l+d} P_{k_{l+d+1}}^{l+d+1} \cdots P_{k_n}^n$ are extremely important. Expressions of the type $P_k$ will be associated with $\ell^1$ norms, and the $Q_k$ with $\ell^2$ norms and square functions. Here we want to apply Lemma 53, so we need to deal with the $Q_k$ functions first. Once we do this, we can estimate the $L^r$ norm of $\mathrm{BHT} \otimes \Pi \otimes \cdots \Pi(f, g)$ by

$$\begin{aligned}&\Big\| \Big( \sum_{k_1, \dots, k_{l+d}} \Big| \sum_{k_{l+d+1}, \dots, k_n} P_{k_{l+d+1}}^{l+d+1} \cdots P_{k_n}^n \mathrm{BHT}\big(P_{k_1}^{y_1} \cdots Q_{k_{l+1}}^{y_{l+1}} \cdots f, Q_{k_1}^{y_1} \cdots P_{k_{l+1}}^{y_{l+1}} \cdots Q_{k_{l+d+1}}^{y_{l+d+1}} \cdots g\big) \Big|^2 \Big)^{\frac{1}{2}} \Big\|_r \\ &\quad \lesssim \Big\| \Big( \sum_{k_1, \dots, k_{l+d}} \Big| \sum_{k_{l+d+1}, \dots, k_n} \big|\mathrm{BHT}\big(P_{k_1}^{y_1} \cdots Q_{k_{l+1}}^{y_{l+1}} \cdots f, Q_{k_1}^{y_1} \cdots P_{k_{l+1}}^{y_{l+1}} \cdots Q_{k_{l+d+1}}^{y_{l+d+1}} \cdots g\big)\big| \Big|^2 \Big)^{\frac{1}{2}} \Big\|_r \\ &\quad \lesssim \|f\|_p \|g\|_q.\end{aligned}$$

For the last part we used the following vector-valued estimates for the BHT:

$$L^p\big(\underbrace{\ell^\infty(\cdots(\ell^\infty}_{l}\underbrace{(\ell^2(\cdots(\ell^2}_{d}\underbrace{(\ell^2(\cdots(\ell^2))}_{n-l-d}\cdots\big)\times L^q\big(\underbrace{\ell^2(\cdots(\ell^2}_{l}\underbrace{(\ell^\infty(\cdots(\ell^\infty}_{d}\underbrace{(\ell^2(\cdots(\ell^2))}_{n-l-d}\cdots\big)$$
$$\mapsto L^s\big(\underbrace{\ell^2(\cdots(\ell^2}_{l}\underbrace{(\ell^2(\cdots(\ell^2}_{d}\underbrace{(\ell^1(\cdots(\ell^1))}_{n-l-d}\cdots\big)$$

together with the boundedness of the maximal operator and square function.

Similarly, we can obtain estimates for $\Pi^{\otimes^{d_1}}\otimes \mathrm{BHT}\otimes\Pi^{\otimes^{d_2}}$ within the same range as that of BHT. Some partial results in mixed norm $L^p$ spaces can be obtained too, but the general case, for arbitrary values of $d_1$ and $d_2$ remains open. We present a few particular cases that illustrate the main ideas, without being too technical.

(i) Here, we prove mixed norm $L^p$ estimates for $\Pi_1\otimes\mathrm{BHT}\otimes\Pi_3$, where $\Pi_1=\sum_k Q_k^1(P_k^1\cdot Q_k^1)$, $\Pi_3=\sum_l Q_l^3(Q_l^3\cdot P_l^3)$, and the exponents $p_j, q_j$ are in $[2,\infty)$. We note that

$$\Pi_1\otimes\mathrm{BHT}\otimes\Pi_3(f,g)(x,y,z)=\sum_{k,l} Q_k^1Q_l^3\mathrm{BHT}(P_k^xQ_l^zf, Q_k^xP_l^zg)(y),$$

and we want to estimate the above expression in the space $\|\cdot\|_{L_x^{s_1}L_y^{s_2}L_z^{s_3}}$. The key observation is that whenever $1<s_2,s_3<\infty$,

$$\Big\|\sum_{k,l} Q_k^1Q_l^3F(x,y,z)\Big\|_{L_x^{s_1}L_y^{s_2}L_z^{s_3}}\lesssim\Big\|\Big(\sum_{k,l}|Q_k^1Q_l^3F(x,y,z)|^2\Big)^{\frac12}\Big\|_{L_x^{s_1}L_y^{s_2}L_z^{s_3}},\tag{54}$$

which is a Banach-valued equivalent of Lemma 53. This result, for $s_1>1$, can be found in [Fernandez 1987; Rubio de Francia et al. 1986], and it follows from the boundedness of Calderón–Zygmund operators (the dual of the square function is such an operator) on $L^p$ spaces with mixed norms. The proof in the case $s_1\le 1$ is a Banach space adaptation of the proof of Lemma 53. Given the special properties of the $Q_k^1$ and $Q_l^3$ operators, we obtain

$$\|\Pi_1\otimes\mathrm{BHT}\otimes\Pi_3(f,g)\|_{L_x^{s_1}L_y^{s_2}L_z^{s_3}}\lesssim\Big\|\Big(\sum_{k,l}|\mathrm{BHT}(P_k^xQ_l^zf,Q_k^xP_l^zg)(y)|^2\Big)^{\frac12}\Big\|_{L_x^{s_1}L_y^{s_2}L_z^{s_3}}.$$

The multiple vector-valued estimates

$$\mathrm{BHT}: L_y^{p_2}(L_z^{p_3}(\ell^\infty(\ell^2))))\times L_y^{q_2}(L_z^{q_3}(\ell^2(\ell^\infty))))\to L_y^{s_2}(L_z^{s_3}(\ell^2(\ell^2)))),$$

which exist in the local $L^2$ case at least, together with Hölder's inequality imply

$$\begin{aligned}\|\Pi_1\otimes\mathrm{BHT}\otimes\Pi_3(f,g)\|_{L_x^{s_1}L_y^{s_2}L_z^{s_3}}&\\ \lesssim\Big\|\sup_k\Big(\sum_l|P_k^xQ_l^zf(y)|^2\Big)^{\frac12}\Big\|_{L_x^{p_1}L_y^{p_2}L_z^{p_3}}&\Big\|\Big(\sum_k|\sup_l|Q_k^xP_l^zg(y)||^2\Big)^{\frac12}\Big\|_{L_x^{q_1}L_y^{q_2}L_z^{q_3}}\\ \lesssim\|f\|_{L_x^{p_1}L_y^{p_2}L_z^{p_3}}&\|g\|_{L_x^{q_1}L_y^{q_2}L_z^{q_3}}.\end{aligned}$$

The last inequality follows again from Banach-valued extensions of convolution operators. Since our proof makes use of multiple vector-valued estimates for BHT, we cannot obtain mixed norm $L^p$ estimates for all the exponents in the Banach range. From the above example, one can see that besides the constraints imposed by the square functions and maximal operators, we also need $(p_3, q_3, s_3) \in \mathcal{D}_{p_2,q_2,s_2}$.

(ii) If $d_1 = 0$ and $d_2 = 1$, we have

$$\mathrm{BHT} \otimes \Pi : L_x^{p_1} L_y^{p_2} \times L_x^{q_1} L_y^{q_2} \to L_x^{s_1} L_y^{s_2}$$

whenever $1 < p_2, q_2, s_2 < \infty$, $1 < p_1, q_1 \le \infty$, $\frac{2}{3} < s_1 < \infty$ and $(p_2, q_2, s_2) \in \mathcal{D}_{p_1,q_1,s_1}$.

(iii) If $d_1 = 1$ and $d_2 = 0$, we have

$$\Pi \otimes \mathrm{BHT} : L_x^{p_1} L_y^{p_2} \times L_x^{q_1} L_y^{q_2} \to L_x^{s_1} L_y^{s_2}$$

whenever $1 < p_2, q_2, s_2 < \infty$, $1 < p_1, q_1 \le \infty$, $\frac{1}{2} < s_1 < \infty$. Since the "target" spaces (that is, inner spaces in the mixed norms) are strictly between 1 and $\infty$, the outer $L^\infty$ cases (that is, $p_1 = \infty$ or $q_1 = \infty$) follow easily from similar estimates on the adjoints.

We note that mixed norm estimates for $\Pi \otimes \mathrm{BHT}$ appear also in [Di Plinio and Ou 2015], where all the inner spaces involved are $L^p$ spaces with $1 < p < \infty$ (in our notation, that means $1 < p_2, q_2, s_2 < \infty$).

## 6. Leibniz rules: Theorem 4

Now we present some ideas behind the proof of Theorem 4. Littlewood–Paley projections play an important role when dealing with derivatives:

$$\begin{aligned} D_1^\alpha D_2^\beta (f \cdot g)(x,y) &= \sum_{k,l} \big[(f * \varphi_k \otimes \varphi_l) \cdot (g * \psi_k \otimes \psi_l)\big] * (D_1^\alpha \psi_k \otimes D_2^\beta \psi_l)(x,y) \\ &= \sum_{k,l} \big[(f * \varphi_k \otimes \varphi_l) \cdot (g * \psi_k \otimes \psi_l)\big] * (2^{k\alpha} \widetilde{\psi}_k \otimes 2^{l\beta} \widetilde{\psi}_l)(x,y), \end{aligned}$$

where

$$\widehat{\widetilde{\psi}}_k(\xi) = \frac{|\xi|^\alpha}{2^{k\alpha}} \hat{\psi}_k(\xi) \quad \text{and} \quad \widehat{\widetilde{\psi}}_l(\eta) = \frac{|\eta|^\beta}{2^{l\beta}} \hat{\psi}_l(\eta).$$

Then one can move the $2^{k\alpha}$ inside, and couple it with the $\psi_k$ because $2^{k\alpha} \psi_k(x) = D^\alpha \widetilde{\widetilde{\psi}}_k(x)$. Here

$$\widehat{\widetilde{\widetilde{\psi}}}_k(\xi) = \frac{2^{k\alpha}}{|\xi|^\alpha} \hat{\psi}_k(\xi).$$

In this way, we obtain $D_1^\alpha D_2^\beta (f \cdot g) = \widetilde{\Pi} \otimes \widetilde{\Pi}(f, D_1^\alpha D_2^\beta g) +$ eight other similar terms. We can estimate $\Pi \otimes \Pi$ in $L^p$ spaces with mixed norms, as long as the "outside" functions $\widehat{\psi}_k$ and $\widehat{\varphi}_k$ are constantly equal to 1 on $2^{k-2} \le |\xi| \le 2^{k+2}$ and $|\xi| \le 2^{k+2}$ respectively. The operators $\widetilde{\Pi}$ are slightly different, but using Fourier series we can write $\widetilde{\Pi}(F, G)$ as

$$(F,G) \mapsto \sum_{n \in \mathbb{Z}} c_n \sum_{k,l} \big[F * (\varphi_k \otimes \varphi_l) \cdot G * (\widetilde{\widetilde{\psi}}_k \otimes \widetilde{\widetilde{\psi}}_l)\big] * \psi_k \otimes \widetilde{\psi}_{l,n}(x,y).$$

Here the coefficients satisfy $|c_n| \lesssim n^{-M}$, and $\psi_{k,n}(x) = \psi_k(x + 2^{-k}n)$. Now notice that the right-hand side above becomes

$$\sum_n c_n \sum_l Q_l^2 \widetilde{\Pi}(P_{l,n}^y F, \widetilde{\widetilde{Q}}_{l,n}^y G)(x),$$

which is a superposition of $\Pi \otimes \Pi$ operators.

The proof of the Leibniz rule follows from

(1) (multiple) vector-valued estimates for the paraproduct

$$\widetilde{\Pi}(f,g) = \sum_l \big[(f * \varphi_l) \cdot (g * \widetilde{\widetilde{\psi}}_l)\big] * \widetilde{\psi}_l,$$

(2) the boundedness of the shifted maximal and square functions:

$$\Big\| \sup_l |f * \varphi_{l,n}| \Big\|_p \lesssim \log\langle n\rangle \, \|f\|_p, \quad \Big\| \Big( \sum_l |f * \widetilde{\widetilde{\psi}}_{l,n}|^2 \Big)^{\frac{1}{2}} \Big\|_p \lesssim \log\langle n\rangle \, \|f\|_p.$$

Returning to the Leibniz rules, we have for $s_1, s_2 \geq 1$,

$$\begin{aligned}
\Big\| \|D_1^\alpha D_2^\beta (f,g)\|_{L_y^{s_2}} \Big\|_{L_x^{s_1}} &\leq \sum_n |c_n| \Big\| \Big\| \sum_l Q_l^2 \widetilde{\Pi}(P_{l,n}^y F, \widetilde{\widetilde{Q}}_{l,n}^y G) \Big\|_{L_y^{s_2}} \Big\|_{L_x^{s_1}} \\
&\lesssim \sum_n |c_n| \Big\| \Big\| \Big( \sum_l |\widetilde{\Pi}^{\beta_1,\beta_2}(P_{l,n}^y F, \widetilde{\widetilde{Q}}_{l,n}^y G)|^2 \Big)^{\frac{1}{2}} \Big\|_{L_y^{s_2}} \Big\|_{L_x^{s_1}} \\
&\lesssim \sum_n |c_n| \Big\| \|\sup_l |P_{l,n}^y F|\|_{L_y^{p_2}} \Big\|_{L_x^{p_1}} \Big\| \Big\| \Big( \sum_l |\widetilde{\widetilde{Q}}_{l,n}^y G|^2 \Big)^{\frac{1}{2}} \Big\|_{L_y^{q_2}} \Big\|_{L_x^{q_1}} \\
&\lesssim \|f\|_{L_x^{p_1} L_y^{p_2}} \|D_1^\alpha D_2^\beta g\|_{L_x^{q_1} L_y^{q_2}}.
\end{aligned}$$

Here we used the vector-valued estimates

$$\widetilde{\Pi} : L_x^{p_1}(L_y^{p_2}(\ell^\infty)) \times L_x^{q_1}(L_y^{q_2}(\ell^2)) \to L_x^{s_1}(L_y^{s_2}(\ell^2)),$$

as well as the boundedness of the square function and maximal operator. We note that the square function is in the $y$-variable, and for that reason at first we cannot allow $p_2 = \infty$ or $q_2 = \infty$. However, this obstruction can be removed by using duality.

The same proof works in the case $\frac{1}{2} < s_1 < 1$, if $1 < p_2, q_2 < \infty$. In this case, we use the subadditivity of $\|\cdot\|_{s_1}^{s_1}$. The case $\frac{1}{2} < s_1 < 1$ and $p_2 = \infty$ requires a slightly different reasoning, and can be deduced from the corresponding mixed norm estimates for $\Pi \otimes \Pi$. This will be presented at the end of this section.

A slightly more difficult case of the Leibniz rule is when one of the last components is a $\varphi$-type function:

$$\begin{aligned}
D_1^\alpha D_2^\beta (f \cdot g)(x,y) &= \sum_{k,l} \big[(f * \psi_k \otimes \varphi_l) \cdot (g * \psi_k \otimes \psi_l)\big] * (D_1^\alpha \varphi_k \otimes D_2^\beta \psi_l)(x,y) \\
&= \sum_{k,l} \big[(f * \psi_k \otimes \varphi_l) \cdot (g * \psi_k \otimes \psi_l)\big] * (2^{k\alpha} \widetilde{\varphi}_k \otimes 2^{l\beta} \widetilde{\psi}_l)(x,y).
\end{aligned}$$

In this case

$$\widehat{\widetilde{\varphi}}_k(\xi) = \frac{|\xi|^\alpha}{2^{k\alpha}}\hat{\varphi}_k(\xi),$$

but $\widetilde{\varphi}$ doesn't behave as nicely as $\widetilde{\psi}$; since $\widehat{\widetilde{\varphi}}$ is not smooth at the origin, the decay in $\widetilde{\varphi}$ is much slower:

$$|\widetilde{\varphi}(x)| \leq \frac{1}{(1+|x|)^{1+\alpha}}.$$

We use a Fourier series decomposition of $\widehat{\widetilde{\varphi}}_k$ on its support

$$\widehat{\widetilde{\varphi}}_k(\xi) = \sum_{n\in\mathbb{Z}} c_n e^{\frac{2\pi i n\xi}{2^k}} \cdot \hat{\varphi}_k(\xi), \quad \text{where } c_n = \frac{1}{2^k}\int_{\mathbb{R}} \widehat{\widetilde{\varphi}}_k(\xi) e^{-\frac{2\pi i n\xi}{2^k}}\, d\xi.$$

In this case we only have $|c_n| \leq 1/(1+|n|)^{1+\alpha}$, but this is enough for the coefficients to sum up, if $s_1 > 1/(1+\alpha)$. Since $s_2 \geq 1$, we will not have a similar issue when doing the decomposition in the second variable.

Following the same line of ideas, the problem reduces to estimating

$$\sum_n c_n \sum_k P_k^1 \widetilde{\Pi}\big(\widetilde{\widetilde{Q}}_{k,n}^x F, Q_{k,n}^x G\big)(y),$$

and it would imply "mixed square functions" estimates of the form

$$\left\| \left( \sum_n |Q_{k,n}^x G|^2 \right)^{\frac{1}{2}} \right\|_{L_x^{q_1} L_y^{q_2}}.$$

This is bounded as long as $1 < q_1, q_2 < \infty$, and in order to recover the case $p_i = \infty$ or $q_i = \infty$ we want to make sure that the square functions are in the innermost variable, which is $y$. So we need a decomposition of $\widetilde{\psi}_l$, as before. Also, we will need vector-valued estimates for the "generalized paraproduct"

$$(f,g) \mapsto \sum_k (f * \psi_k \cdot g * \psi_k) * \widetilde{\varphi}_k,$$

where the last component $\widetilde{\varphi}$ has slow decay. The vector spaces involved are $(\ell^2, \ell^\infty, \ell^2)$ or $(\ell^2, \ell^2, \ell^1)$, and such estimates can be proved using ideas similar to those in Section 4, modulo standard technical difficulties, as discussed in [Muscalu and Schlag 2013].

We now present the proof of the mixed norm estimates for the biparameter paraproducts:

*Proof of Theorem 5.* Since the other cases are very similar, we can assume that $\Pi_y$, the paraproduct acting on the variable $y$, is of the form

$$\Pi_y(\cdot,\cdot) = \sum_k Q_k\big(P_k(\cdot), Q_k(\cdot)\big).$$

Then we can write $\Pi\otimes\Pi$ as $\Pi\otimes\Pi(f,g)(x,y)=\sum_k Q_k^2\Pi(P_k^y,Q_k^y)(x)$. Then we have

$$\Big\|\Big\|\sum_k Q_k^2\Pi(P_k^y,Q_k^y)(x)\Big\|_{L_y^{s_2}}\Big\|_{L_x^{s_1}}\lesssim\Big\|\Big\|\Big(\sum_k|\Pi(P_k^y,Q_k^y)(x)|^2\Big)^{\frac12}\Big\|_{L_y^{s_2}}\Big\|_{L_x^{s_1}}$$
$$\lesssim\Big\|\|\sup_k|P_k^y f(x)|\|_{L_y^{p_2}}\Big\|_{L_x^{p_1}}\Big\|\Big\|\Big(\sum_k|Q_k^y g(x)|^2\Big)^{\frac12}\Big\|_{L_y^{q_2}}\Big\|_{L_x^{q_1}}.$$

In the above inequality we used the multiple vector-valued estimate

$$\Pi_x: L_x^{p_1}(L_y^{p_2}(\ell^\infty))\times L_x^{q_1}(L_y^{q_2}(\ell^2))\to L_x^{s_1}(L_y^{s_2}(\ell^2)),$$

which is a consequence of Theorem 9.

Now we focus on the case $p_2=\infty, 1<q_2=q<\infty$, since $q_2=\infty$ is symmetric. We want to prove that

$$\Pi\otimes\Pi: L_x^{p_1}L_y^\infty\times L_x^{q_1}L_y^q\to L_x^{s_1}L_y^q,$$

by using Banach-valued restricted-type interpolation. That is, for any sets of finite measure $F,G,H$, we can find a major subset $H'\subseteq H$, and we will prove that

$$\Big|\int_{\mathbb{R}^2}\Pi\otimes\Pi(f,g)(x,y)h(x,y)\,dx\,dy\Big|\lesssim|F|^{\alpha_1}|G|^{\alpha_2}|H|^{\alpha_3}\tag{55}$$

for any functions $f,g$ and $h$ satisfying

$$\|f(x,\cdot)\|_{L_y^\infty}\le\mathbf{1}_F(x),\quad \|g(x,\cdot)\|_{L_y^q}\le\mathbf{1}_G(x),\quad \|h(x,\cdot)\|_{L_y^{q'}}\le\mathbf{1}_{H'}(x),$$

and $(\alpha_1,\alpha_2,\alpha_3)$ any tuple satisfying $\alpha_1+\alpha_2+\alpha_3=1$, situated in the neighborhood of $\big(\frac{1}{p_1},\frac{1}{p_2},\frac{1}{p'}\big)$.

A triple stopping time similar to the one appearing in the proof of Theorem 7 will allow us to recover any exterior $L_x^{p_j}$ norms, while the interior norms are fixed: $L_y^\infty, L_y^q, L_y^q$.

We will consider *localizations* of the paraproduct acting on the $x$-variable. More exactly, the following estimate, the proof of which is a combination of Proposition 50 and $L^p$ estimates for $\Pi\otimes\Pi$, is key:

If $I_0$ is a fixed dyadic interval, then $\Pi_{I_0}^{F,G,H'}\otimes\Pi: L_x^\infty L_y^\infty\times L_x^q L_y^q\to L_x^q L_y^q$ with operatorial norm

$$\|\Pi_{I_0}^{F,G,H'}\otimes\Pi\|_{L_x^\infty L_y^\infty\times L_x^q L_y^q\to L_x^q L_y^q}=\|(\Pi_{I_0}^{F,G,H'}\otimes\Pi)^{*,1}\|_{L_x^{q'}L_y^{q'}\times L_x^q L_y^q\to L_x^1 L_y^1}.$$

The latter is bounded above by

$$\|(\Pi_{I_0}^{F,G,H'}\otimes\Pi)^{*,1}\|_{L_x^{q'}L_y^{q'}\times L_x^q L_y^q\to L_x^1 L_y^1}\lesssim(\widetilde{\text{size}}_{I_0}\mathbf{1}_{H'})^{\frac1q-\epsilon}(\widetilde{\text{size}}_{I_0}\mathbf{1}_G)^{\frac{1}{q'}-\epsilon}(\widetilde{\text{size}}_{I_0}\mathbf{1}_F)^{1-\epsilon},$$

which is a consequence of the localized multiple vector-valued estimates that always appear in the iterative step of the helicoidal method.

More exactly, we have

$$\begin{aligned}\big|\Pi_{I_0}^{F,G,H}\otimes\Pi(f,g)(x,y)h(x,y)\,dx\,dy\big|&\\ \lesssim (\widetilde{\text{size}}_{I_0}\mathbf{1}_{H'})^{\frac{1}{q}-\epsilon}(\widetilde{\text{size}}_{I_0}\mathbf{1}_{G})^{\frac{1}{q'}-\epsilon}(\widetilde{\text{size}}_{I_0}\mathbf{1}_{F})^{1-\epsilon}&\\ \big\|\|h(x,\cdot)\|_{L_y^{q'}}\cdot\tilde{\chi}_{I_0}\big\|_{L_x^{q'}}\big\|\|g(x,\cdot)\|_{L_y^{q}}\cdot\tilde{\chi}_{I_0}\big\|_{L_x^{q}}\|f(\cdot,\cdot)\|_{L_x^\infty L_y^\infty}.&\end{aligned}$$

This implies, after performing the usual stopping times, that

$$\begin{aligned}\left|\int_{\mathbb{R}^2}(\Pi\otimes\Pi)(f,g)(x,y)h(x,y)\,dx\,dy\right|&\lesssim\sum_{n_1,n_2,n_3}\sum_{I_0}\left|\int_{\mathbb{R}^2}(\Pi_{I_0}^{F,G,H'}\otimes\Pi)(f,g)(x,y)h(x,y)\,dx\,dy\right|\\ &\lesssim\sum_{n_1,n_2,n_3}\sum_{I_0}(\widetilde{\text{size}}_{I_0}\mathbf{1}_{F})^{1-\epsilon}(\widetilde{\text{size}}_{I_0}\mathbf{1}_{G})^{1-\epsilon}(\widetilde{\text{size}}_{I_0}\mathbf{1}_{H'})^{1-\epsilon}|I_0|.\end{aligned}$$

From here, the desired $L^p$ estimates follow almost immediately. □

## 7. Rubio de Francia theorem for iterated Fourier integrals

We end by answering the initial question that motivated the study of vector-valued BHT. More exactly, we prove Theorem 10, which is a consequence of Theorem 7, with $r_1, r_2$ chosen carefully so that $\frac{1}{r_1}+\frac{1}{r_2}=\frac{1}{r}$.

*Proof of Theorem 10.* We start with the case $r\geq 2$; this follows from Theorem 7:

$$\left\|\left(\sum_k|\text{BHT}(P_{I_k}f,P_{I_k}g)(x)|^2\right)^{\frac{1}{2}}\right\|_s\lesssim\left\|\left(\sum_k|P_{I_k}f|^{r_1}\right)^{\frac{1}{r_1}}\right\|_p\left\|\left(\sum_k|P_{I_k}g|^{r_2}\right)^{\frac{1}{r_2}}\right\|_q \tag{56}$$

for any $1<p,q<\infty$, $\frac{2}{3}<s<\infty$.

This is implied by Rubio de Francia's theorem, if one can find $r_1$ and $r_2$ with $\frac{1}{r_1}+\frac{1}{r_2}=\frac{1}{2}$ and

$$\frac{1}{p}<\frac{1}{r_1'},\quad \frac{1}{q}<\frac{1}{r_2'}.$$

This is possible as long as $\frac{1}{s}=\frac{1}{p}+\frac{1}{q}<\frac{1}{r_1'}+\frac{1}{r_2'}=\frac{3}{2}$, which coincides with the condition that we have for the range of BHT.

The case $1\leq r<2$ is similar; for $p,q$, and $s$ as above, one needs to find $r_1$ and $r_2\geq 2$ so that

$$2-\frac{1}{r}=\frac{1}{r_1'}+\frac{1}{r_2'}>\frac{1}{p}+\frac{1}{q}.$$

Note that $\frac{1}{p}<\frac{1}{r_1'}=1-\frac{1}{r}+\frac{1}{r_2}\leq\frac{1}{r'}+\frac{1}{2}$, and similarly for $q$. Because of this restriction, the operator $T_r$ is bounded as long as admissible triple $\big(\frac{1}{p},\frac{1}{q},\frac{1}{s'}\big)$ is in the convex hull of the points

$$(0,0,1),\quad \big(\tfrac{1}{2}+\tfrac{1}{r'},\tfrac{1}{2},-\tfrac{1}{r'}\big),\quad \big(\tfrac{1}{2},\tfrac{1}{2}+\tfrac{1}{r'},-\tfrac{1}{r'}\big),\quad \big(\tfrac{1}{2}+\tfrac{1}{r'},0,\tfrac{1}{2}-\tfrac{1}{r'}\big),\quad \big(0,\tfrac{1}{2}+\tfrac{1}{r'},\tfrac{1}{2}-\tfrac{1}{r'}\big).\quad \square$$

**Remark.** An alternative way of proving the boundedness of $T_r$ within the range mentioned in Theorem 10 is by interpolating between

$$L^{p_1} \times L^{q_1} \to L^{s_1}(\ell^2) \quad \text{with } p_1, q_1, s_1 \text{ in the range of the BHT operator, and} \tag{57}$$

$$L^{p_2} \times L^{q_2} \to L^{s_2}(\ell^1) \quad \text{with } p_2, q_2 > 1,\ s_2 \geq 1. \tag{58}$$

**7.1.** ***Boundedness of operators $M_1$ and $M_2$.*** In what follows we prove the boundedness of operators $M_1$ and $M_2$ presented in (14) and (15):

$$M_1(f_1, f_2, g)(\xi) = \sum_{\omega} \int_{\substack{x_1<x_2 \\ x_1,x_2\in\omega_L, x_3\in\omega_R}} \hat{f_1}(x_1)\hat{f_2}(x_2)g(x_3)e^{2\pi i\xi(x_1+x_2+x_3)}\,dx_1\,dx_2\,dx_3$$

and

$$M_2(f_1, f_2, g)(\xi) = \sum_{\omega} \int_{\substack{x_1<L(\omega_L) \\ x_2\in\omega_L, x_3\in\omega_R}} \hat{f_1}(x_1)\hat{f_2}(x_2)g(x_3)e^{2\pi i\xi(x_1+x_2+x_3)}\,dx_1\,dx_2\,dx_3.$$

For both operators, we are going to use the triangle inequality in $L^r$, the target space for operators $M_1$ and $M_2$. However, if $r < 1$, this inequality is not available anymore for the quasinorm $\|\cdot\|_r$ and instead we use the triangle inequality for $\|\cdot\|_r^r$. This is the only difference between the Banach and quasi-Banach case, and for simplicity we assume $r \geq 1$. Also, as previously stated, we assume $\|g\|_p = 1$.

**Proposition 54.** *Let $1 < p < 2$ and $\frac{1}{r} = \frac{1}{s} + \frac{1}{p'} = \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p'}$. Then*

$$\big\|M_1(f_1, f_2, g)\big\|_r \lesssim \|f_1\|_{p_1}\|f_2\|_{p_2}\|g\|_p.$$

*Proof.* Recall that $\omega \in \mathcal{D}$ is the mesh of dyadic intervals contained in $[0, 1]$, and we identify them with their preimage: $\omega \sim \varphi^{-1}(\omega)$. We rewrite $M_1$ as

$$M_1(f_1, f_2, g)(\xi) = \sum_{\omega} \mathrm{BHT}(P_{\omega_L} f_1, P_{\omega_L} f_2)(\xi) \cdot \widehat{g \cdot \mathbf{1}_{\omega_R}}(\xi).$$

Then

$$\begin{aligned}
\big\|M_1(f_1, f_2, g)\big\|_r &\lesssim \sum_{k\geq 0} \Big\| \sum_{|\omega|=2^{-k}} \mathrm{BHT}(P_{\omega_L} f_1, P_{\omega_L} f_2) \cdot \widehat{g \cdot \mathbf{1}_{\omega_R}} \Big\|_r \\
&\lesssim \sum_{k\geq 0} \Big\| \Big( \sum_{|\omega|=2^{-k}} \big|\mathrm{BHT}(P_{\omega_L} f_1, P_{\omega_L} f_2)\big|^p \Big)^{\frac{1}{p}} \Big( \sum_{|\omega|=2^{-k}} \big|\widehat{g \cdot \mathbf{1}_{\omega_R}}\big|^{p'} \Big)^{\frac{1}{p'}} \Big\|_r \\
&\lesssim \sum_{k\geq 0} \Big\| \Big( \sum_{|\omega|=2^{-k}} \big|\mathrm{BHT}(P_{\omega_L} f_1, P_{\omega_L} f_2)\big|^p \Big)^{\frac{1}{p}} \Big\|_s \Big( \sum_{|\omega|=2^{-k}} \|\widehat{g \cdot \mathbf{1}_{\omega_R}}\|_{p'}^{p'} \Big)^{\frac{1}{p'}}.
\end{aligned}$$

We estimate $\|\widehat{g \cdot \mathbf{1}_{\omega_R}}\|_{p'} \lesssim \|g \cdot \mathbf{1}_{\omega_R}\|_p = 2^{-\frac{k}{p}}$ using the Hausdorff–Young theorem. Also, there are $2^k$ dyadic intervals of length $2^{-k}$ in $[0, 1]$ and because of this

$$\big\|M_1(f_1, f_2, g)\big\|_r \lesssim \sum_{k\geq 0} 2^{-k(\frac{1}{p}-\frac{1}{p'})} \Big\| \Big( \sum_{|\omega|=2^{-k}} |\mathrm{BHT}(P_{\omega_L} f_1, P_{\omega_L} f_2)|^p \Big)^{\frac{1}{p}} \Big\|_s.$$

If we estimate the last term using the operator $T_p$ directly, we will not obtain the full range stated above, as there will appear extra constraints of the type

$$\frac{1}{p_1}+\frac{1}{p}<\frac{3}{2},\quad \frac{1}{p_2}+\frac{1}{p}<\frac{3}{2}.$$

Instead, using Hölder and the fact that $1<p<2$, we have

$$\big\|\mathrm{BHT}(P_{\omega_L}f_1,P_{\omega_L}f_2)\big\|_{\ell^p(\omega)}\le\big\|\mathrm{BHT}(P_{\omega_L}f_1,P_{\omega_L}f_2)\big\|_{\ell^2(\omega)}2^{k(\frac{1}{p}-\frac{1}{2})}.$$

Using the boundedness of $T_2$, we have $\|M_1(f_1,f_2,g)\|_r\lesssim\sum_{k\ge0}2^{-k(\frac{1}{2}-\frac{1}{p'})}\|f_1\|_{p_1}\|f_2\|_{p_2}$. □

**Proposition 55.** *Let* $1<p<2$ *and* $\frac{1}{r}=\frac{1}{s}+\frac{1}{p'}=\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p'}$. *Then*

$$\big\|M_2(f_1,f_2,g)\big\|_r\lesssim\|f_1\|_{p_1}\|f_2\|_{p_2}\|g\|_p,$$

*provided* $\frac{1}{p_2}+\frac{1}{p'}<1$.

*Proof.* First, we remark that

$$\big|M_2(f_1,f_2,g)(\xi)\big|\le\sum_{\omega}|Cf_1(\xi)||P_{\omega_L}f_2(\xi)|\widehat{|g\omega_R|}(\xi),$$

where $C$ is the Carleson operator, bounded on $L^p$ whenever $1<p<\infty$. From here on the estimates are similar to those in Proposition 54, but instead of the bilinear operator $T_r(f,g)$ we will have to use the more restrictive Rubio de Francia operator $\mathrm{RF}_\nu$:

$$\begin{aligned}\big\|M_2(f_1,f_2,g)\big\|_r&\le\sum_{k\ge0}\Big\|Cf_1\Big(\sum_{|\omega|=2^{-k}}|P_{\omega_L}f_2|^p\Big)^{\frac{1}{p}}\Big(\sum_{|\omega|=2^{-k}}|\widehat{g\cdot\mathbf{1}_{\omega_R}}|^{p'}\Big)^{\frac{1}{p'}}\Big\|_r\\&\le\sum_{k\ge0}\|Cf_1\|_{p_1}\Big\|\Big(\sum_{|\omega|=2^{-k}}|P_{\omega_L}f_2|^p\Big)^{\frac{1}{p}}\Big\|_{p_2}\Big(\sum_{|\omega|=2^{-k}}\|\widehat{g\cdot\mathbf{1}_{\omega_R}}\|_{p'}^{p'}\Big)^{\frac{1}{p'}}\\&\le\sum_{k\ge0}2^{k(\frac{1}{p}-\frac{1}{\nu})}\|Cf_1\|_{p_1}\Big\|\Big(\sum_{|\omega|=2^{-k}}|P_{\omega_L}f_2|^\nu\Big)^{\frac{1}{\nu}}\Big\|_{p_2}\Big(\sum_{|\omega|=2^{-k}}\|\widehat{g\cdot\mathbf{1}_{\omega_R}}\|_{p'}^{p'}\Big)^{\frac{1}{p'}}\\&\le\sum_{k\ge0}2^{-k(\frac{1}{\nu}-\frac{1}{p'})}\|f_1\|_{p_1}\|\mathrm{RF}_\nu(f_2)\|_{p_2}.\end{aligned}$$

If $p_2\ge2$,we can take $\nu=2$ and there are no other restrictions. In the case $p_2<2$, Rubio de Francia requires $\frac{1}{\nu}+\frac{1}{p_2}<1$. This and the condition $\frac{1}{\nu}-\frac{1}{p'}>0$ (so that the geometric series above is finite) can be summarized as $\frac{1}{p_2}+\frac{1}{p'}<1$. □

## Acknowledgements

Benea was partially supported by NSF grant DMS 1500262 and ERC project FAnFArE no. 637510; Muscalu was partially supported by NSF grant DMS 1500262.

## Note added in proof

We recently improved Theorems 4 and 5, allowing for the exponent $s_2$ to be $< 1$. This is a consequence of new multiple quasi-Banach valued inequalities for $\Pi$. In [Benea and Muscalu 2016], we also prove multiple quasi-Banach valued inequalities for the bilinear Hilbert transform operator, extending also Theorem 7.

## References


[Bateman 2013] M. Bateman, "Single annulus $L^p$ estimates for Hilbert transforms along vector fields", *Rev. Mat. Iberoam.* **29**:3 (2013), 1021–1069. MR 3090145 Zbl 1283.42019

[Bateman and Thiele 2013] M. Bateman and C. Thiele, "$L^p$ estimates for the Hilbert transforms along a one-variable vector field", *Anal. PDE* **6**:7 (2013), 1577–1600. MR 3148061 Zbl 1285.42014

[Benea and Muscalu 2016] C. Benea and C. Muscalu, "Quasi-Banach valued inequalities via the helicoidal method", preprint, 2016. arXiv 1609.01090

[Benedek et al. 1962] A. Benedek, A.-P. Calderón, and R. Panzone, "Convolution operators on Banach space valued functions", *Proc. Nat. Acad. Sci. U.S.A.* **48** (1962), 356–365. MR 0133653 Zbl 0103.33402

[Bourgain 1986] J. Bourgain, "Vector-valued singular integrals and the $H^1$-BMO duality", pp. 1–19 in *Probability theory and harmonic analysis* (Cleveland, Ohio, 1983), Monogr. Textbooks Pure Appl. Math. **98**, Dekker, New York, 1986. MR 830227 Zbl 0602.42015

[Burkholder 1983] D. L. Burkholder, "A geometric condition that implies the existence of certain singular integrals of Banach-space-valued functions", pp. 270–286 in *Conference on harmonic analysis in honor of Antoni Zygmund, Vol. I* (Chicago, IL, 1981), Wadsworth, Belmont, CA, 1983. MR 730072

[Carleson 1966] L. Carleson, "On convergence and growth of partial sums of Fourier series", *Acta Math.* **116** (1966), 135–157. MR 0199631 Zbl 0144.06402

[Christ and Kiselev 1998] M. Christ and A. Kiselev, "Absolutely continuous spectrum for one-dimensional Schrödinger operators with slowly decaying potentials: some optimal results", *J. Amer. Math. Soc.* **11**:4 (1998), 771–797. MR 1621882 Zbl 0899.34051

[Christ and Kiselev 2001a] M. Christ and A. Kiselev, "Maximal functions associated to filtrations", *J. Funct. Anal.* **179**:2 (2001), 409–425. MR 1809116 Zbl 0974.47025

[Christ and Kiselev 2001b] M. Christ and A. Kiselev, "WKB asymptotic behavior of almost all generalized eigenfunctions for one-dimensional Schrödinger operators with slowly decaying potentials", *J. Funct. Anal.* **179**:2 (2001), 426–447. MR 1809117 Zbl 0985.34078

[Culiuc et al. 2016] A. Culiuc, F. Di Plinio, and Y. Ou, "Domination of multilinear singular integrals by positive sparse forms", 2016. arXiv 1603.05317

[Demeter and Silva 2015] C. Demeter and P. Silva, "Some new light on a few classical results", *Colloq. Math.* **140**:1 (2015), 129–147. MR 3352435 Zbl 1325.42010

[Di Plinio and Ou 2015] F. Di Plinio and Y. Ou, "Banach-valued multilinear singular integrals", preprint, 2015. arXiv 1506.05827

[Fernandez 1987] D. L. Fernandez, "Vector-valued singular integral operators on $L^p$-spaces with mixed norms and applications", *Pacific J. Math.* **129**:2 (1987), 257–275. MR 909030 Zbl 0634.42014

[García-Cuerva and Rubio de Francia 1985] J. García-Cuerva and J. L. Rubio de Francia, *Weighted norm inequalities and related topics*, North-Holland Mathematics Studies **116**, North-Holland, Amsterdam, 1985. MR 807149 Zbl 0578.46046

[Grafakos and Li 2006] L. Grafakos and X. Li, "The disc as a bilinear multiplier", *Amer. J. Math.* **128**:1 (2006), 91–119. MR 2197068 Zbl 1143.42015

[Hytönen and Lacey 2013] T. P. Hytönen and M. T. Lacey, "Pointwise convergence of vector-valued Fourier series", *Math. Ann.* **357**:4 (2013), 1329–1361. MR 3124934 Zbl 1285.42005

[Hytönen et al. 2013] T. P. Hytönen, M. T. Lacey, and I. Parissis, "The vector valued quartile operator", *Collect. Math.* **64**:3 (2013), 427–454. MR 3084406 Zbl 1288.42008

[Journé 1985] J.-L. Journé, "Calderón–Zygmund operators on product spaces", *Rev. Mat. Iberoamericana* **1**:3 (1985), 55–91. MR 836284 Zbl 0634.42015

[Kenig et al. 1993] C. E. Kenig, G. Ponce, and L. Vega, "Well-posedness and scattering results for the generalized Korteweg–de Vries equation via the contraction principle", *Comm. Pure Appl. Math.* **46**:4 (1993), 527–620. MR 1211741 Zbl 0808.35128

[Kesler 2015] R. Kesler, "Mixed estimates for degenerate multi-linear operators associated to simplexes", *J. Math. Anal. Appl.* **424**:1 (2015), 344–360. MR 3286565 Zbl 1305.47031

[Lacey and Thiele 1999] M. Lacey and C. Thiele, "On Calderón's conjecture", *Ann. of Math.* (2) **149**:2 (1999), 475–496. MR 1689336 Zbl 0934.42012

[Lerner 2013] A. K. Lerner, "A simple proof of the $A_2$ conjecture", *Int. Math. Res. Not.* **2013**:14 (2013), 3159–3170. MR 3085756 Zbl 1318.42018

[Meyer and Coifman 1997] Y. Meyer and R. Coifman, *Wavelets: Calderón–Zygmund and multilinear operators*, Cambridge Studies in Advanced Mathematics **48**, Cambridge University Press, 1997. MR 1456993 Zbl 0916.42023

[Muscalu and Schlag 2013] C. Muscalu and W. Schlag, *Classical and multilinear harmonic analysis. Vol. II*, Cambridge Studies in Advanced Mathematics **138**, Cambridge University Press, 2013. MR 3052499 Zbl 1282.42001

[Muscalu et al. 2004a] C. Muscalu, J. Pipher, T. Tao, and C. Thiele, "Bi-parameter paraproducts", *Acta Math.* **193**:2 (2004), 269–296. MR 2134868 Zbl 1087.42016

[Muscalu et al. 2004b] C. Muscalu, T. Tao, and C. Thiele, "$L^p$ estimates for the biest, II: The Fourier case", *Math. Ann.* **329**:3 (2004), 427–461. MR 2127985 Zbl 1073.42010

[Muscalu et al. 2006a] C. Muscalu, J. Pipher, T. Tao, and C. Thiele, "Multi-parameter paraproducts", *Rev. Mat. Iberoam.* **22**:3 (2006), 963–976. MR 2320408 Zbl 1114.42005

[Muscalu et al. 2006b] C. Muscalu, T. Tao, and C. Thiele, "The bi-Carleson operator", *Geom. Funct. Anal.* **16**:1 (2006), 230–277. MR 2221256 Zbl 1130.42004

[Oberlin et al. 2012] R. Oberlin, A. Seeger, T. Tao, C. Thiele, and J. Wright, "A variation norm Carleson theorem", *J. Eur. Math. Soc.* (*JEMS*) **14**:2 (2012), 421–464. MR 2881301 Zbl 1246.42016

[Paley 1931] R. E. A. C. Paley, "Some theorems on orthonormal functions", *Studia Math.* **3** (1931), 226–238. Zbl 0003.35201

[Ruan 2010] Z. Ruan, "Multi-parameter Hardy spaces via discrete Littlewood–Paley theory", *Anal. Theory Appl.* **26**:2 (2010), 122–139. MR 2653706 Zbl 1224.42039

[Rubio de Francia 1985] J. L. Rubio de Francia, "A Littlewood–Paley inequality for arbitrary intervals", *Rev. Mat. Iberoamericana* **1**:2 (1985), 1–14. MR 850681 Zbl 0611.42005

[Rubio de Francia et al. 1986] J. L. Rubio de Francia, F. J. Ruiz, and J. L. Torrea, "Calderón–Zygmund theory for operator-valued kernels", *Adv. in Math.* **62**:1 (1986), 7–48. MR 859252 Zbl 0627.42008

[Silva 2014] P. Silva, "Vector-valued inequalities for families of bilinear Hilbert transforms and applications to bi-parameter problems", *J. Lond. Math. Soc.* (2) **90**:3 (2014), 695–724. MR 3291796 Zbl 1311.42021

[Thiele 2006] C. Thiele, *Wave packet analysis*, CBMS Regional Conference Series in Mathematics **105**, American Mathematical Society, Providence, RI, 2006. MR 2199086 Zbl 1101.42001

[Torres and Ward 2015] R. H. Torres and E. L. Ward, "Leibniz's rule, sampling and wavelets on mixed Lebesgue spaces", *J. Fourier Anal. Appl.* **21**:5 (2015), 1053–1076. MR 3393695 Zbl 06493292

Received 20 Jan 2016. Revised 12 Jun 2016. Accepted 12 Jul 2016.

CRISTINA BENEA: cristina.benea@univ-nantes.fr
*Université de Nantes, Laboratoire de Mathématiques, Nantes 44322, France*

CAMIL MUSCALU: camil@math.cornell.edu
*Department of Mathematics, Cornell University, Ithaca, NY 14853, United States*

ANALYSIS AND PDE
Vol. 9, No. 8, 2016
dx.doi.org/10.2140/apde.2016.9.1989

# STRUCTURE OF MODULAR INVARIANT SUBALGEBRAS IN FREE ARAKI–WOODS FACTORS

RÉMI BOUTONNET AND CYRIL HOUDAYER

We show that any amenable von Neumann subalgebra of any free Araki–Woods factor that is globally invariant under the modular automorphism group of the free quasifree state is necessarily contained in the almost periodic free summand.

## 1. Introduction

*Free Araki–Woods factors* were introduced in [Shlyakhtenko 1997]. In the framework of Voiculescu's free probability theory, they can be regarded as the type III counterparts of free group factors using the free Gaussian functor [Voiculescu 1985; Voiculescu et al. 1992]. Following Shlyakhtenko, to any orthogonal representation $U : \mathbb{R} \curvearrowright H_{\mathbb{R}}$ on a real Hilbert space, one associates the *free Araki–Woods* von Neumann algebra $\Gamma(H_{\mathbb{R}}, U)''$. The von Neumann algebra $\Gamma(H_{\mathbb{R}}, U)''$ comes equipped with a unique *free quasifree state* $\varphi_U$ which is always normal and faithful (see Section 2 for a detailed construction). We have $\Gamma(H_{\mathbb{R}}, U)'' \cong \mathrm{L}(\boldsymbol{F}_{\dim(H_{\mathbb{R}})})$ when $U = 1_{H_{\mathbb{R}}}$ and $\Gamma(H_{\mathbb{R}}, U)''$ is a full type III factor when $U \neq 1_{H_{\mathbb{R}}}$.

Let $U : \mathbb{R} \curvearrowright H_{\mathbb{R}}$ be any orthogonal representation. Using Zorn's lemma, we may decompose $H_{\mathbb{R}} = H_{\mathbb{R}}^{\mathrm{ap}} \oplus H_{\mathbb{R}}^{\mathrm{wm}}$ and $U = U^{\mathrm{wm}} \oplus U^{\mathrm{ap}}$, where $U^{\mathrm{ap}} : \mathbb{R} \curvearrowright H_{\mathbb{R}}^{\mathrm{ap}}$ is the *almost periodic*, and $U^{\mathrm{wm}} : \mathbb{R} \curvearrowright H_{\mathbb{R}}^{\mathrm{wm}}$ the *weakly mixing*, subrepresentation of $U : \mathbb{R} \curvearrowright H_{\mathbb{R}}$. Write $M = \Gamma(H_{\mathbb{R}}, U)''$, $N = \Gamma(H_{\mathbb{R}}^{\mathrm{ap}}, U^{\mathrm{ap}})''$ and $P = \Gamma(H_{\mathbb{R}}^{\mathrm{wm}}, U^{\mathrm{wm}})''$, so that we have the *free product* splitting

$$(M, \varphi_U) = (N, \varphi_{U^{\mathrm{ap}}}) * (P, \varphi_{U^{\mathrm{wm}}}).$$

Our main result provides a general structural decomposition for any von Neumann subalgebra $Q \subset M$ that is globally invariant under the modular automorphism group $\sigma^{\varphi_U}$ and shows that when $Q$ is also assumed to be amenable then $Q$ sits inside $N$. It generalizes Theorem C of [Houdayer and Raum 2015] to *arbitrary* free Araki–Woods factors.

**Main Theorem.** *Keep the same notation as above. Let $Q \subset M$ be any unital von Neumann subalgebra that is globally invariant under the modular automorphism group $\sigma^{\varphi_U}$. Then there exists a unique central projection $z \in \mathcal{Z}(Q) \subset M^{\varphi_U} = N^{\varphi_{U^{\mathrm{ap}}}}$ such that*

- *$Qz$ is amenable and $Qz \subset zNz$, and*
- *$Qz^{\perp}$ has no nonzero amenable direct summand and $(Q' \cap M^{\omega})z^{\perp} = (Q' \cap M)z^{\perp}$ is atomic for any nonprincipal ultrafilter $\omega \in \beta(\boldsymbol{N}) \setminus \boldsymbol{N}$.*

*MSC2010:* 46L10, 46L54, 46L36.
*Keywords:* free Araki–Woods factors, Popa's asymptotic orthogonality property, type III factors, ultraproduct von Neumann algebras.

*In particular, for any unital amenable von Neumann subalgebra $Q \subset M$ that is globally invariant under the modular automorphism group $\sigma^{\varphi_U}$, we have $Q \subset N$.*

Our main theorem should be compared to [Houdayer 2014b, Theorem D], which provides a similar result for crossed product $\mathrm{II}_1$ factors arising from free Bogoljubov actions of amenable groups.

The core of our argument is Theorem 3.1 which generalizes [Houdayer and Raum 2015, Theorem 4.3] to arbitrary free Araki–Woods factors. Let us point out that Theorem 3.1 is reminiscent of Popa's asymptotic orthogonality property in free group factors [Popa 1983] which is based on the study of central sequences in the ultraproduct framework. Unlike other results on this theme [Houdayer 2014b; 2015; Houdayer and Ueda 2016], we do not assume here that the subalgebra $Q \subset M$ has a diffuse intersection with the free summand $N$ of the free product splitting $(M, \varphi_U) = (N, \varphi_{U^{\mathrm{ap}}}) * (P, \varphi_{U^{\mathrm{wm}}})$, and so we cannot exploit commutation relations of $Q$-central sequences with elements in $N$. Instead, we use the facts that $Q$ admits central sequences that are invariant under the modular automorphism group $\sigma^{\varphi_U^\omega}$ of the ultraproduct state $\varphi_U^\omega$ and that the modular automorphism group $\sigma^{\varphi_U}$ is weakly mixing on $P$.

## 2. Preliminaries

For any von Neumann algebra $M$, we denote by $\mathcal{Z}(M)$ the center of $M$, by $\mathcal{U}(M)$ the group of unitaries in $M$, by $\mathrm{Ball}(M)$ the unit ball of $M$ with respect to the uniform norm and by $(M, \mathrm{L}^2(M), J, \mathrm{L}^2(M)_+)$ the standard form of $M$. We say that an inclusion of von Neumann algebras $P \subset M$ is *with expectation* if there exists a faithful normal conditional expectation $\mathrm{E}_P : M \to P$. All the von Neumann algebras we consider in this paper are always assumed to be $\sigma$-finite.

Let $M$ be any $\sigma$-finite von Neumann algebra with predual $M_*$ and $\varphi \in M_*$ any faithful state. We write $\|x\|_\varphi = \varphi(x^*x)^{1/2}$ for all $x \in M$. Recall that on $\mathrm{Ball}(M)$, the topology given by $\|\cdot\|_\varphi$ coincides with the $\sigma$-strong topology. Denote by $\xi_\varphi \in \mathrm{L}^2(M)_+$ the unique representing vector of $\varphi$. The mapping $M \to \mathrm{L}^2(M) : x \mapsto x\xi_\varphi$ defines an embedding with dense image such that $\|x\|_\varphi = \|x\xi_\varphi\|_{\mathrm{L}^2(M)}$ for all $x \in M$. We denote by $\sigma^\varphi$ the modular automorphism group of the state $\varphi$. The *centralizer* $M^\varphi$ of the state $\varphi$ is by definition the fixed point algebra of $(M, \sigma^\varphi)$.

Recall from [Houdayer 2014a, Section 2.1] that two subspaces $E, F \subset H$ of a Hilbert space are said to be *$\varepsilon$-orthogonal* for some $0 \leq \varepsilon \leq 1$ if $|\langle \xi, \eta \rangle| \leq \varepsilon \|\xi\| \|\eta\|$ for all $\xi \in E$ and all $\eta \in F$. We then simply write $E \perp_\varepsilon F$.

***Ultraproduct von Neumann algebras.*** Let $M$ be any $\sigma$-finite von Neumann algebra and $\omega \in \beta(\boldsymbol{N}) \setminus \boldsymbol{N}$ any nonprincipal ultrafilter. Define

$$\mathcal{I}_\omega(M) = \{(x_n)_n \in \ell^\infty(M) : x_n \to 0 \ *\text{-strongly as } n \to \omega\},$$
$$\mathcal{M}^\omega(M) = \{(x_n)_n \in \ell^\infty(M) : (x_n)_n \, \mathcal{I}_\omega(M) \subset \mathcal{I}_\omega(M) \text{ and } \mathcal{I}_\omega(M) \, (x_n)_n \subset \mathcal{I}_\omega(M)\}.$$

The *multiplier algebra* $\mathcal{M}^\omega(M)$ is a $\mathrm{C}^*$-algebra and $\mathcal{I}_\omega(M) \subset \mathcal{M}^\omega(M)$ is a norm closed two-sided ideal. Following [Ocneanu 1985, §5.1], we define the *ultraproduct von Neumann algebra* $M^\omega$ by $M^\omega := \mathcal{M}^\omega(M)/\mathcal{I}_\omega(M)$, which is indeed known to be a von Neumann algebra. We denote the image of $(x_n)_n \in \mathcal{M}^\omega(M)$ by $(x_n)^\omega \in M^\omega$.

For every $x \in M$, the constant sequence $(x)_n$ lies in the multiplier algebra $\mathcal{M}^\omega(M)$. We then identify $M$ with $(M+\mathcal{I}_\omega(M))/\mathcal{I}_\omega(M)$ and regard $M \subset M^\omega$ as a von Neumann subalgebra. The map

$$\mathrm{E}_\omega : M^\omega \to M, \quad (x_n)^\omega \mapsto \sigma\text{-weak} \lim_{n\to\omega} x_n$$

is a faithful normal conditional expectation. For every faithful state $\varphi \in M_*$, the formula $\varphi^\omega := \varphi \circ \mathrm{E}_\omega$ defines a faithful normal state on $M^\omega$. Observe that $\varphi^\omega((x_n)^\omega) = \lim_{n\to\omega} \varphi(x_n)$ for all $(x_n)^\omega \in M^\omega$.

Let $Q \subset M$ be any von Neumann subalgebra with faithful normal conditional expectation $\mathrm{E}_Q : M \to Q$. Choose a faithful state $\varphi \in M_*$ in such a way that $\varphi = \varphi \circ \mathrm{E}_Q$. We have $\ell^\infty(Q) \subset \ell^\infty(M)$, $\mathcal{I}_\omega(Q) \subset \mathcal{I}_\omega(M)$ and $\mathcal{M}^\omega(Q) \subset \mathcal{M}^\omega(M)$. We then identify $Q^\omega = \mathcal{M}^\omega(Q)/\mathcal{I}_\omega(Q)$ with $(\mathcal{M}^\omega(Q)+\mathcal{I}_\omega(M))/\mathcal{I}_\omega(M)$ and may regard $Q^\omega \subset M^\omega$ as a von Neumann subalgebra. Observe that the norm $\|\cdot\|_{(\varphi|_Q)^\omega}$ on $Q^\omega$ is the restriction of the norm $\|\cdot\|_{\varphi^\omega}$ to $Q^\omega$. Observe moreover that $(\mathrm{E}_Q(x_n))_n \in \mathcal{I}_\omega(Q)$ for all $(x_n)_n \in \mathcal{I}_\omega(M)$ and $(\mathrm{E}_Q(x_n))_n \in \mathcal{M}^\omega(Q)$ for all $(x_n)_n \in \mathcal{M}^\omega(M)$. Therefore, the mapping $\mathrm{E}_{Q^\omega} : M^\omega \to Q^\omega$ given by $(x_n)^\omega \mapsto (\mathrm{E}_Q(x_n))^\omega$ is a well-defined conditional expectation satisfying $\varphi^\omega \circ \mathrm{E}_{Q^\omega} = \varphi^\omega$. Hence, $\mathrm{E}_{Q^\omega} : M^\omega \to Q^\omega$ is a faithful normal conditional expectation. For more on ultraproduct von Neumann algebras, we refer the reader to [Ando and Haagerup 2014; Ocneanu 1985].

***Free Araki–Woods factors.*** Let $H_\mathbb{R}$ be any real Hilbert space and $U : \mathbb{R} \curvearrowright H_\mathbb{R}$ any orthogonal representation. Denote by $H = H_\mathbb{R} \otimes_\mathbb{R} \mathbb{C} = H_\mathbb{R} \oplus \mathrm{i}H_\mathbb{R}$ the complexified Hilbert space, by $I : H \to H : \xi + \mathrm{i}\eta \mapsto \xi - \mathrm{i}\eta$ the canonical anti-unitary involution on $H$ and by $A$ the infinitesimal generator of $U : \mathbb{R} \curvearrowright H$, that is, $U_t = A^{\mathrm{i}t}$ for all $t \in \mathbb{R}$. Moreover, we have $IAI = A^{-1}$. Observe that $j : H_\mathbb{R} \to H : \zeta \mapsto (2/(A^{-1}+1))^{1/2}\zeta$ defines an isometric embedding of $H_\mathbb{R}$ into $H$. Put $K_\mathbb{R} := j(H_\mathbb{R})$. It is easy to see that $K_\mathbb{R} \cap \mathrm{i}K_\mathbb{R} = \{0\}$ and that $K_\mathbb{R} + \mathrm{i}K_\mathbb{R}$ is dense in $H$. Write $T = IA^{-1/2}$. Then $T$ is a conjugate-linear closed invertible operator on $H$ satisfying $T = T^{-1}$ and $T^*T = A^{-1}$. Such an operator is called an *involution* on $H$. Moreover, we have $\mathrm{dom}(T) = \mathrm{dom}(A^{-1/2})$ and $K_\mathbb{R} = \{\xi \in \mathrm{dom}(T) : T\xi = \xi\}$. In what follows, we simply write

$$\overline{\xi + \mathrm{i}\eta} := T(\xi + \mathrm{i}\eta) = \xi - \mathrm{i}\eta, \quad \forall \xi, \eta \in K_\mathbb{R}.$$

We introduce the *full Fock space* of $H$:

$$\mathcal{F}(H) = \mathbb{C}\Omega \oplus \bigoplus_{n=1}^{\infty} H^{\otimes n}.$$

The unit vector $\Omega$ is known as the *vacuum vector*. For all $\xi \in H$, we define the *left creation* operator $\ell(\xi) : \mathcal{F}(H) \to \mathcal{F}(H)$ by

$$\begin{cases} \ell(\xi)\Omega = \xi, \\ \ell(\xi)(\xi_1 \otimes \cdots \otimes \xi_n) = \xi \otimes \xi_1 \otimes \cdots \otimes \xi_n. \end{cases}$$

We have $\|\ell(\xi)\|_\infty = \|\xi\|$, and $\ell(\xi)$ is an isometry if $\|\xi\| = 1$. For all $\xi \in K_\mathbb{R}$, put $W(\xi) := \ell(\xi) + \ell(\xi)^*$. The crucial result of Voiculescu [Voiculescu et al. 1992, Lemma 2.6.3] is that the distribution of the self-adjoint operator $W(\xi)$ with respect to the vector state $\varphi_U = \langle \cdot\, \Omega, \Omega \rangle$ is the semicircular law of Wigner supported on the interval $[-\|\xi\|, \|\xi\|]$.

**Definition 2.1** [Shlyakhtenko 1997]. Let $H_{\mathbb{R}}$ be any real Hilbert space and $U : \mathbb{R} \curvearrowright H_{\mathbb{R}}$ any orthogonal representation. The *free Araki–Woods* von Neumann algebra associated with $U : \mathbb{R} \curvearrowright H_{\mathbb{R}}$ is defined by

$$\Gamma(H_{\mathbb{R}}, U)'' := \{W(\xi) : \xi \in K_{\mathbb{R}}\}''.$$

We denote by $\Gamma(H_{\mathbb{R}}, U)$ the unital C*-algebra generated by 1 and by all the elements $W(\xi)$ for $\xi \in K_{\mathbb{R}}$.

The vector state $\varphi_U = \langle \cdot\, \Omega, \Omega\rangle$ is called the *free quasifree state* and is faithful on $\Gamma(H_{\mathbb{R}}, U)''$. Let $\xi, \eta \in K_{\mathbb{R}}$ and write $\zeta = \xi + \mathrm{i}\eta$. Put

$$W(\zeta) := W(\xi) + \mathrm{i}W(\eta) = \ell(\zeta) + \ell(\bar{\zeta})^*.$$

Note that the modular automorphism group $\sigma^{\varphi_U}$ of the free quasifree state $\varphi_U$ is given by $\sigma_t^{\varphi_U} = \mathrm{Ad}(\mathcal{F}(U_t))$, where $\mathcal{F}(U_t) = 1_{\mathbb{C}\Omega} \oplus \bigoplus_{n\geq 1} U_t^{\otimes n}$. In particular, it satisfies

$$\sigma_t^{\varphi_U}(W(\zeta)) = W(U_t\zeta), \quad \forall \zeta \in K_{\mathbb{R}} + \mathrm{i}K_{\mathbb{R}}, \forall t \in \mathbb{R}.$$

It is easy to see that for all $n \geq 1$ and all $\zeta_1, \ldots, \zeta_n \in K_{\mathbb{R}} + \mathrm{i}K_{\mathbb{R}}$, $\zeta_1 \otimes \cdots \otimes \zeta_n \in \Gamma(H_{\mathbb{R}}, U)''\Omega$. When $\zeta_1, \ldots, \zeta_n$ are all nonzero, we denote by $W(\zeta_1 \otimes \cdots \otimes \zeta_n) \in \Gamma(H_{\mathbb{R}}, U)''$ the unique element such that

$$\zeta_1 \otimes \cdots \otimes \zeta_n = W(\zeta_1 \otimes \cdots \otimes \zeta_n)\Omega.$$

Such an element is called a *reduced word*. By [Houdayer and Raum 2015, Proposition 2.1(i)] (see also [Houdayer 2014a, Proposition 2.4]), the reduced word $W(\zeta_1 \otimes \cdots \otimes \zeta_n)$ satisfies the *Wick formula* given by

$$W(\zeta_1 \otimes \cdots \otimes \zeta_n) = \sum_{k=0}^{n} \ell(\zeta_1)\cdots\ell(\zeta_k)\ell(\bar{\zeta}_{k+1})^*\cdots\ell(\bar{\zeta}_n)^*.$$

Note that since inner products are assumed to be linear in the first variable, for all $\xi, \eta \in H$ we have $\ell(\xi)^*\ell(\eta) = \overline{\langle \xi, \eta\rangle}1 = \langle \eta, \xi\rangle 1$. In particular, the Wick formula from [Houdayer and Raum 2015, Proposition 2.1(ii)] is

$$\begin{aligned} &W(\xi_1 \otimes \cdots \otimes \xi_r)W(\eta_1 \otimes \cdots \otimes \eta_s) \\ &\qquad = W(\xi_1 \otimes \cdots \otimes \xi_r \otimes \eta_1 \otimes \cdots \otimes \eta_s) + \overline{\langle \bar{\xi}_r, \eta_1\rangle}\, W(\xi_1 \otimes \cdots \otimes \xi_{r-1})W(\eta_2 \otimes \cdots \otimes \eta_s) \end{aligned}$$

for all $\xi_1, \ldots, \xi_r, \eta_1, \ldots, \eta_s \in K_{\mathbb{R}} + \mathrm{i}K_{\mathbb{R}}$. We repeatedly use this fact in the next section. We refer to [Houdayer and Raum 2015, Section 2] for further details.

## 3. Asymptotic orthogonality property in free Araki–Woods factors

Let $U : \mathbb{R} \curvearrowright H_{\mathbb{R}}$ be any orthogonal representation. By Zorn's lemma, we may decompose $H_{\mathbb{R}} = H_{\mathbb{R}}^{\mathrm{ap}} \oplus H_{\mathbb{R}}^{\mathrm{wm}}$ and $U = U^{\mathrm{wm}} \oplus U^{\mathrm{ap}}$, where $U^{\mathrm{ap}} : \mathbb{R} \curvearrowright H_{\mathbb{R}}^{\mathrm{ap}}$ is the *almost periodic*, and $U^{\mathrm{wm}} : \mathbb{R} \curvearrowright H_{\mathbb{R}}^{\mathrm{wm}}$ the *weakly mixing*, subrepresentation of $U : \mathbb{R} \curvearrowright H_{\mathbb{R}}$. Write $M = \Gamma(H_{\mathbb{R}}, U)''$, $N = \Gamma(H_{\mathbb{R}}^{\mathrm{ap}}, U^{\mathrm{ap}})''$ and $P = \Gamma(H_{\mathbb{R}}^{\mathrm{wm}}, U_t^{\mathrm{wm}})''$, so that

$$(M, \varphi_U) = (N, \varphi_{U^{\mathrm{ap}}}) * (P, \varphi_{U^{\mathrm{wm}}}).$$

For notational convenience, we simply write $\varphi := \varphi_U$.

The main result of this section, Theorem 3.1 below, strengthens and generalizes [Houdayer and Raum 2015, Theorem 4.3].

**Theorem 3.1.** *Keep the same notation as above. Let $\omega \in \beta(\boldsymbol{N}) \setminus \boldsymbol{N}$ be any nonprincipal ultrafilter. For all $a \in M \ominus N$, all $b \in M$ and all $x, y \in (M^\omega)^{\varphi^\omega} \cap (M^\omega \ominus M)$, we have*

$$\varphi^\omega(b^* y^* a x) = 0.$$

*Proof.* Denote, as usual, by $H := H_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$ the complexified Hilbert space and by $U : \mathbb{R} \curvearrowright H$ the corresponding unitary representation. Put $H^{\mathrm{ap}} := H_{\mathbb{R}}^{\mathrm{ap}} \otimes_{\mathbb{R}} \mathbb{C}$ and $H^{\mathrm{wm}} := H_{\mathbb{R}}^{\mathrm{wm}} \otimes_{\mathbb{R}} \mathbb{C}$. Put $K_{\mathbb{R}} := j(H_{\mathbb{R}})$, $K_{\mathbb{R}}^{\mathrm{ap}} = j(H_{\mathbb{R}}^{\mathrm{ap}})$ and $K_{\mathbb{R}}^{\mathrm{wm}} := j(H_{\mathbb{R}}^{\mathrm{wm}})$, where $j$ is the isometric embedding $\xi \in H_{\mathbb{R}} \mapsto (2/(1+A^{-1}))^{1/2}\xi \in H$. Denote by $\mathcal{H} = \mathcal{F}(H)$ the full Fock space of $H$. For every $t \in \mathbb{R}$, put $\kappa_t = 1_{\mathbb{C}\Omega} \oplus \bigoplus_{n\geq 1} U_t^{\otimes n} \in \mathcal{U}(\mathcal{H})$. For every $t \in \mathbb{R}$ and every $x \in M$, we have $\sigma_t^\varphi(x)\Omega = \kappa_t(x\Omega)$. We implicitly identify the full Fock space $\mathcal{F}(H)$ with the standard Hilbert space $\mathrm{L}^2(M)$ and the vacuum vector $\Omega \in \mathcal{H}$ with the canonical representing vector $\xi_\varphi \in \mathrm{L}^2(M)_+$.

Put $K_{\mathrm{an}} := \bigcup_{\lambda>1} \mathbf{1}_{[\lambda^{-1},\lambda]}(A)(K_{\mathbb{R}} + \mathrm{i}K_{\mathbb{R}})$. Observe that $K_{\mathrm{an}} \subset K_{\mathbb{R}} + \mathrm{i}K_{\mathbb{R}}$ is a dense subspace of elements $\eta \in K_{\mathbb{R}} + \mathrm{i}K_{\mathbb{R}}$ for which the map $\mathbb{R} \to K_{\mathbb{R}} + \mathrm{i}K_{\mathbb{R}} : t \mapsto U_t\eta$ extends to a $(K_{\mathbb{R}} + \mathrm{i}K_{\mathbb{R}})$-valued entire analytic function, and that $\overline{K_{\mathrm{an}}} = K_{\mathrm{an}}$. For all $\eta \in K_{\mathrm{an}}$, the element $W(\eta)$ is analytic with respect to the modular automorphism group $\sigma^\varphi$ and we have $\sigma_z^\varphi(W(\eta)) = W(A^{\mathrm{i}z}\eta)$ for all $z \in \mathbb{C}$.

Denote by $\mathcal{W}$ the set of reduced words of the form $W(\xi_1 \otimes \cdots \otimes \xi_n)$ for which $n \geq 1$ and $\xi_1, \ldots, \xi_n \in K_{\mathrm{an}}$. By linearity/density, in order to prove Theorem 3.1, we may assume without loss of generality that $a$ and $b$ are reduced words in $\mathcal{W}$. Since moreover $a \in M \ominus N$, we can assume that at least one of its letters $\xi_i$ lies in $K_{\mathbb{R}}^{\mathrm{wm}} + \mathrm{i}K_{\mathbb{R}}^{\mathrm{wm}}$. More precisely, we can write

$$a = a' W(\xi_1 \otimes \cdots \otimes \xi_p) a'',$$
$$b = b' W(\eta_1 \otimes \cdots \otimes \eta_q) b''$$

with $p \geq 1$, $q \geq 0$ and for reduced words $a', a'', b', b''$ in $N$ with letters in $K_{\mathrm{an}} \cap (K_{\mathbb{R}}^{\mathrm{ap}} + \mathrm{i}K_{\mathbb{R}}^{\mathrm{ap}})$, and for $\xi_2, \ldots, \xi_{p-1}, \eta_2, \ldots, \eta_{q-1} \in K_{\mathrm{an}}$ and $\xi_1, \xi_p, \eta_1, \eta_q \in K_{\mathrm{an}} \cap (K_{\mathbb{R}}^{\mathrm{wm}} + \mathrm{i}K_{\mathbb{R}}^{\mathrm{wm}})$. By convention, when $q = 0$, $W(\eta_1 \otimes \cdots \otimes \eta_q)$ is the trivial word 1, so that $b = b'b''$.

Denote by $L \subset K_{\mathbb{R}}^{\mathrm{wm}} + \mathrm{i}K_{\mathbb{R}}^{\mathrm{wm}}$ the finite dimensional subspace generated by $\xi_1, \xi_p, \eta_1, \eta_q$ and such that $\bar{L} = L$. If $q = 0$, then $L$ is simply the subspace generated by $\xi_1, \xi_p, \bar{\xi}_1, \bar{\xi}_p$. Denote by

- $\mathcal{X}(1, r) \subset \mathcal{H}$ the closed linear subspace generated by all the reduced words of the form $e_1 \otimes \cdots \otimes e_n$ with $r \geq 0$, $n \geq r+1$, $e_1, \ldots, e_r \in K_{\mathbb{R}}^{\mathrm{ap}} + \mathrm{i}K_{\mathbb{R}}^{\mathrm{ap}}$ and $e_{r+1} \in L$;
- $\mathcal{X}(2, r) \subset \mathcal{H}$ the closed linear subspace generated by all the reduced words of the form $e_1 \otimes \cdots \otimes e_n$ with $r \geq 0$, $n \geq r+1$, $e_{n-r} \in L$ and $e_{n-r+1}, \ldots, e_n \in K_{\mathbb{R}}^{\mathrm{ap}} + \mathrm{i}K_{\mathbb{R}}^{\mathrm{ap}}$;
- $\mathcal{Y} \subset \mathcal{H}$ the closed linear subspace generated by all the reduced words of the form $e_1 \otimes \cdots \otimes e_n$ with $n \geq 1$ and $e_1, e_n \in L^\perp$.

When $r = 0$, we simply write $\mathcal{X}_1 := \mathcal{X}(1, 0)$ and $\mathcal{X}_2 := \mathcal{X}(2, 0)$. Observe that we have the orthogonal decomposition

$$\mathcal{H} = \mathbb{C}\Omega \oplus \overline{(\mathcal{X}_1 + \mathcal{X}_2)} \oplus \mathcal{Y}.$$

**Claim 3.2.** *Let $\varepsilon \geq 0$ and $t \in \mathbb{R}$ such that $U_t(L) \perp_{\varepsilon/\dim L} L$. Then for all $i \in \{1,2\}$ and all $r \geq 0$, we have*

$$\kappa_t(\mathcal{X}(i,r)) \perp_\varepsilon \mathcal{X}(i,r).$$

*Proof of Claim 3.2.* Choose an orthonormal basis $(\zeta_1, \ldots, \zeta_{\dim L})$ of $L$. We first prove the claim for $i = 1$. We identify $\mathcal{X}(1,r)$ with $L \otimes ((H^{\mathrm{ap}})^{\otimes r} \otimes \mathcal{H})$ using the unitary defined by

$$\mathcal{V}(1,r) : H \otimes (H^{\otimes r} \otimes \mathcal{H}) \to \mathcal{H} : \zeta \otimes \mu \otimes \nu \mapsto \mu \otimes \zeta \otimes \nu.$$

Observe that $\kappa_t \mathcal{V}(1,r) = \mathcal{V}(1,r)(U_t \otimes (U_t)^{\otimes r} \otimes \kappa_t)$ for every $t \in \mathbb{R}$. Let $\Xi_1, \Xi_2 \in \mathcal{X}(1,r)$ be such that $\Xi_1 = \sum_{i=1}^{\dim L} \zeta_i \otimes \Theta_i^1$ and $\Xi_2 = \sum_{j=1}^{\dim L} \zeta_j \otimes \Theta_j^2$ with $\Theta_i^1, \Theta_j^2 \in (H^{\mathrm{ap}})^{\otimes r} \otimes \mathcal{H}$. We have

$$\kappa_t(\Xi_1) = \sum_{i=1}^{\dim L} U_t(\zeta_i) \otimes \kappa_t(\Theta_i^1),$$

and hence

$$|\langle \kappa_t(\Xi_1), \Xi_2 \rangle| \leq \sum_{i,j=1}^{\dim L} |\langle U_t(\zeta_i), \zeta_j \rangle| \|\Theta_i^1\| \|\Theta_j^2\|.$$

Since $|\langle U_t(\zeta_i), \zeta_j \rangle| \leq \varepsilon/\dim L$, we obtain $|\langle \kappa_t(\Xi_1), \Xi_2 \rangle| \leq \varepsilon \|\Xi_1\| \|\Xi_2\|$ by the Cauchy–Schwarz inequality. The proof of the claim for $i = 2$ is entirely analogous. □

Given a closed subspace $\mathcal{K} \subset \mathcal{H}$, we denote by $P_{\mathcal{K}} : \mathcal{H} \to \mathcal{K}$ the orthogonal projection onto $\mathcal{K}$.

**Claim 3.3.** *Take $z = (z_n)^\omega \in (M^\omega)^{\varphi^\omega}$ and let $w_1, w_2 \in N$ be any elements of the following forms*:

- *Either $w_1 = 1$ or $w_1 = W(\zeta_1 \otimes \cdots \otimes \zeta_r)$ with $r \geq 1$ and $\zeta_1, \ldots, \zeta_r \in K_{\mathrm{an}} \cap (K_{\mathbb{R}}^{\mathrm{ap}} + \mathrm{i} K_{\mathbb{R}}^{\mathrm{ap}})$.*
- *Either $w_2 = 1$ or $w_2 = W(\mu_1 \otimes \cdots \otimes \mu_s)$ with $s \geq 1$ and $\mu_1, \ldots, \mu_s \in K_{\mathrm{an}} \cap (K_{\mathbb{R}}^{\mathrm{ap}} + \mathrm{i} K_{\mathbb{R}}^{\mathrm{ap}})$.*

*Then for all $i \in \{1,2\}$, we have $\lim_{n\to\omega} \|P_{\mathcal{X}_i}(w_1 z_n w_2 \Omega)\| = 0$.*

*Proof of Claim 3.3.* Observe that $w_1 z_n w_2 \Omega = w_1 J \sigma^\varphi_{-\mathrm{i}/2}(w_2^*) J z_n \Omega$. Firstly, we have

$$P_{\mathcal{X}(1,r)}(J \sigma^\varphi_{-\mathrm{i}/2}(w_2^*) J z_n \Omega) = J \sigma^\varphi_{-\mathrm{i}/2}(w_2^*) J P_{\mathcal{X}(1,r)}(z_n \Omega),$$
$$P_{\mathcal{X}(2,s)}(w_1 z_n \Omega) = w_1 P_{\mathcal{X}(2,s)}(z_n \Omega).$$

Secondly, for all $\Xi \in \mathcal{H}$, we have

$$P_{\mathcal{X}_1}(w_1 \Xi) = P_{\mathcal{X}_1}(w_1 P_{\mathcal{X}(1,r)}(\Xi)),$$
$$P_{\mathcal{X}_2}(J \sigma^\varphi_{-\mathrm{i}/2}(w_2^*) J \Xi) = P_{\mathcal{X}_2}(J \sigma^\varphi_{-\mathrm{i}/2}(w_2^*) J P_{\mathcal{X}(2,s)}(\Xi)).$$

This implies that

$$P_{\mathcal{X}_1}(w_1 z_n w_2 \Omega) = P_{\mathcal{X}_1}(w_1 J \sigma^\varphi_{-\mathrm{i}/2}(w_2^*) J P_{\mathcal{X}(1,r)}(z_n \Omega)),$$
$$P_{\mathcal{X}_2}(w_1 z_n w_2 \Omega) = P_{\mathcal{X}_2}(w_1 J \sigma^\varphi_{-\mathrm{i}/2}(w_2^*) J P_{\mathcal{X}(2,s)}(z_n \Omega)),$$

and we are left to show that $\lim_{n\to\omega} \|P_{\mathcal{X}(1,r)}(z_n \Omega)\| = \lim_{n\to\omega} \|P_{\mathcal{X}(2,s)}(z_n \Omega)\| = 0$.

Let $i \in \{1, 2\}$ and $k \in \{r, s\}$. Fix $N \geq 0$. Since the orthogonal representation $U : \mathbb{R} \curvearrowright H_{\mathbb{R}}^{\mathrm{wm}}$ is weakly mixing and $L \subset H^{\mathrm{wm}}$ is a finite dimensional subspace, we may choose inductively $t_1, \ldots, t_N \in \mathbb{R}$ such that $U_{t_{j_1}}(L) \perp_{(N \dim(L))^{-1}} U_{t_{j_2}}(L)$ for all $1 \leq j_1 < j_2 \leq N$. By Claim 3.2, this implies that

$$\kappa_{t_{j_1}}(\mathcal{X}(i,k)) \perp_{1/N} \kappa_{t_{j_2}}(\mathcal{X}(i,k)), \quad \forall 1 \leq j_1 < j_2 \leq N.$$

For all $t \in \mathbb{R}$ and all $n \in \boldsymbol{N}$, we have

$$\begin{aligned} \|P_{\mathcal{X}(i,k)}(z_n\Omega)\|^2 &= \langle P_{\mathcal{X}(i,k)}(z_n\Omega), z_n\Omega\rangle \\ &= \langle \kappa_t(P_{\mathcal{X}(i,k)}(z_n\Omega)), \kappa_t(z_n\Omega)\rangle \qquad (\text{since } \kappa_t \in \mathcal{U}(\mathcal{H})) \\ &= \langle P_{\kappa_t(\mathcal{X}(i,k))}(\kappa_t(z_n\Omega)), \kappa_t(z_n\Omega)\rangle. \end{aligned}$$

By [Ando and Haagerup 2014, Theorem 4.1], for all $t \in \mathbb{R}$, we have $(z_n)^\omega = z = \sigma_t^{\varphi^\omega}(z) = (\sigma_t^\varphi(z_n))^\omega$. This implies that $\lim_{n\to\omega} \|\sigma_t^\varphi(z_n) - z_n\|_\varphi = 0$, and hence $\lim_{n\to\omega} \|\kappa_t(z_n\Omega) - z_n\Omega\| = 0$ for all $t \in \mathbb{R}$. In particular, since the sequence $(z_n\Omega)_n$ is bounded in $\mathcal{H}$, we deduce that for all $t \in \mathbb{R}$,

$$\lim_{n\to\omega} \|P_{\mathcal{X}(i,k)}(z_n\Omega)\|^2 = \lim_{n\to\omega} \langle P_{\kappa_t(\mathcal{X}(i,k))}(z_n\Omega), z_n\Omega\rangle.$$

Applying this equality to our well chosen reals $(t_j)_{1\leq j\leq N}$, taking a convex combination and applying the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} \lim_{n\to\omega} \|P_{\mathcal{X}(i,k)}(z_n\Omega)\|^2 &= \lim_{n\to\omega} \frac{1}{N} \sum_{j=1}^N \langle P_{\kappa_{t_j}(\mathcal{X}(i,k))}(z_n\Omega), z_n\Omega\rangle \\ &= \lim_{n\to\omega} \frac{1}{N} \left\langle \sum_{j=1}^N P_{\kappa_{t_j}(\mathcal{X}(i,k))}(z_n\Omega), z_n\Omega \right\rangle \\ &\leq \lim_{n\to\omega} \frac{1}{N} \left\| \sum_{j=1}^N P_{\kappa_{t_j}(\mathcal{X}(i,k))}(z_n\Omega) \right\| \|z_n\|_\varphi. \end{aligned}$$

Then for all $n \in \boldsymbol{N}$, we have

$$\begin{aligned} \left\| \sum_{j=1}^N P_{\kappa_{t_j}(\mathcal{X}(i,k))}(z_n\Omega) \right\|^2 &= \sum_{j_1,j_2=1}^N \left\langle P_{\kappa_{t_{j_1}}(\mathcal{X}(i,k))}(z_n\Omega), P_{\kappa_{t_{j_2}}(\mathcal{X}(i,k))}(z_n\Omega) \right\rangle \\ &\leq \sum_{j=1}^N \|P_{\kappa_{t_j}(\mathcal{X}(i,k))}(z_n\Omega)\|^2 + \sum_{j_1\neq j_2}^N \frac{\|z_n\|_\varphi^2}{N} \\ &\leq N\|z_n\|_\varphi^2 + N^2 \frac{\|z_n\|_\varphi^2}{N} \\ &= 2N\|z_n\|_\varphi^2. \end{aligned}$$

Altogether, we have obtained the inequality $\lim_{n\to\omega} \|P_{\mathcal{X}(i,k)}(z_n\Omega)\|^2 \leq \sqrt{2}\|z\|_{\varphi^\omega}^2/\sqrt{N}$. As $N$ is arbitrarily large, this finishes the proof of Claim 3.3. The above argument is inspired by [Wen 2016, Lemma 10]. Alternatively, we could have used [Houdayer 2014a, Proposition 2.3]. □

**Claim 3.4.** *The subspaces $W(\xi_1\otimes\cdots\otimes\xi_p)\mathcal{Y}$ and $J\sigma^{\varphi}_{-i/2}(W(\bar\eta_q\otimes\cdots\otimes\bar\eta_1))J\mathcal{Y}$ are orthogonal in $\mathcal{H}$. Here, in the case $q=0$, the vector space $J\sigma^{\varphi}_{-i/2}(W(\bar\eta_q\otimes\cdots\otimes\bar\eta_1))J\mathcal{Y}$ is nothing but $\mathcal{Y}$.*

*Proof of Claim 3.4.* Let $m,n\geq 1$ and $e_1,\ldots,e_m,f_1,\ldots,f_n\in H$ with $e_1,e_m,f_1,f_n\in L^{\perp}$, so that the vectors $e_1\otimes\cdots\otimes e_m$ and $f_1\otimes\cdots\otimes f_n$ belong to $\mathcal{Y}$. Since $\bar\xi_p\perp e_1$, $\bar f_n\perp\eta_1$ and $\xi_1\perp f_1$, we have

$$\begin{aligned}
&\big\langle W(\xi_1\otimes\cdots\otimes\xi_p)(e_1\otimes\cdots\otimes e_m),\, J\sigma^{\varphi}_{-i/2}(W(\bar\eta_q\otimes\cdots\otimes\bar\eta_1))J(f_1\otimes\cdots\otimes f_n)\big\rangle\\
&\qquad=\big\langle W(\xi_1\otimes\cdots\otimes\xi_p)W(e_1\otimes\cdots\otimes e_m)\Omega,\, J\sigma^{\varphi}_{-i/2}(W(\bar\eta_q\otimes\cdots\otimes\bar\eta_1))JW(f_1\otimes\cdots\otimes f_n)\Omega\big\rangle\\
&\qquad=\big\langle W(\xi_1\otimes\cdots\otimes\xi_p)W(e_1\otimes\cdots\otimes e_m)\Omega,\, W(f_1\otimes\cdots\otimes f_n)W(\eta_1\otimes\cdots\otimes\eta_q)\Omega\big\rangle\\
&\qquad=\big\langle W(\xi_1\otimes\cdots\otimes\xi_p\otimes e_1\otimes\cdots\otimes e_m)\Omega,\, W(f_1\otimes\cdots\otimes f_n\otimes\eta_1\otimes\cdots\otimes\eta_q)\Omega\big\rangle\\
&\qquad=\big\langle \xi_1\otimes\cdots\otimes\xi_p\otimes e_1\otimes\cdots\otimes e_m,\, f_1\otimes\cdots\otimes f_n\otimes\eta_1\otimes\cdots\otimes\eta_q\big\rangle\\
&\qquad=0.
\end{aligned}$$

Note that in the case $q=0$, the above calculation still makes sense. Indeed, we have

$$\big\langle W(\xi_1\otimes\cdots\otimes\xi_p)(e_1\otimes\cdots\otimes e_m),\,(f_1\otimes\cdots\otimes f_n)\big\rangle=\big\langle \xi_1\otimes\cdots\otimes\xi_p\otimes e_1\otimes\cdots\otimes e_m,\, f_1\otimes\cdots\otimes f_n\big\rangle=0.$$

Since the linear span of all such reduced words $e_1\otimes\cdots\otimes e_m$ generate $\mathcal{Y}$ (and likewise the span of the words $f_1\otimes\cdots\otimes f_n$), we obtain that the subspaces $W(\xi_1\otimes\cdots\otimes\xi_p)\mathcal{Y}$ and $J\sigma^{\varphi}_{-i/2}(W(\bar\eta_q\otimes\cdots\otimes\bar\eta_1))J\mathcal{Y}$ are orthogonal in $\mathcal{H}$. □

Let $x,y\in(M^{\omega})^{\varphi^{\omega}}\cap(M^{\omega}\ominus M)$. We have

$$\begin{aligned}
\varphi^{\omega}(b^*y^*ax)&=\langle ax\xi_{\varphi^{\omega}},\,yb\xi_{\varphi^{\omega}}\rangle\\
&=\lim_{n\to\omega}\langle ax_n\xi_{\varphi},\,y_nb\xi_{\varphi}\rangle\\
&=\lim_{n\to\omega}\big\langle a'W(\xi_1\otimes\cdots\otimes\xi_p)a''x_n\Omega,\,y_nb'W(\eta_1\otimes\cdots\otimes\eta_q)b''\Omega\big\rangle\\
&=\lim_{n\to\omega}\big\langle W(\xi_1\otimes\cdots\otimes\xi_p)a''x_n\sigma^{\varphi}_{-i}((b'')^*)\Omega,\,J\sigma^{\varphi}_{-i/2}(W(\bar\eta_q\otimes\cdots\otimes\bar\eta_1))J(a')^*y_nb'\Omega\big\rangle.
\end{aligned}$$

Put $z_n=a''x_n\sigma^{\varphi}_{-i}((b'')^*)$ and $z_n'=(a')^*y_nb'$. By Claim 3.3, we have that

$$\lim_{n\to\omega}\|P_{\mathcal{X}_i}(z_n\Omega)\|=\lim_{n\to\omega}\|P_{\mathcal{X}_i}(z_n'\Omega)\|=0,\qquad\forall i\in\{1,2\}.$$

Since moreover $\mathrm{E}_{\omega}(x)=\mathrm{E}_{\omega}(y)=0$, we see that $\lim_{n\to\omega}\|P_{\mathbb{C}\Omega}(z_n\Omega)\|=\lim_{n\to\omega}\|P_{\mathbb{C}\Omega}(z_n'\Omega)\|=0$. Since $\mathcal{H}=\mathbb{C}\Omega\oplus\overline{(\mathcal{X}_1+\mathcal{X}_2)}\oplus\mathcal{Y}$, we obtain

$$\lim_{n\to\omega}\|z_n\Omega-P_{\mathcal{Y}}(z_n\Omega)\|=0\quad\text{and}\quad\lim_{n\to\omega}\|z_n'\Omega-P_{\mathcal{Y}}(z_n'\Omega)\|=0.$$

By Claim 3.4, we finally obtain

$$\begin{aligned}
\varphi^{\omega}(b^*y^*ax)&=\lim_{n\to\omega}\big\langle W(\xi_1\otimes\cdots\otimes\xi_p)z_n\Omega,\,J\sigma^{\varphi}_{-i/2}(W(\bar\eta_q\otimes\cdots\otimes\bar\eta_1))Jz_n'\Omega\big\rangle\\
&=\lim_{n\to\omega}\big\langle W(\xi_1\otimes\cdots\otimes\xi_p)P_{\mathcal{Y}}(z_n\Omega),\,J\sigma^{\varphi}_{-i/2}(W(\bar\eta_q\otimes\cdots\otimes\bar\eta_1))JP_{\mathcal{Y}}(z_n'\Omega)\big\rangle=0.
\end{aligned}$$

This finishes the proof of Theorem 3.1. □

## 4. Proof of the Main Theorem

We start by proving the following intermediate result.

**Theorem 4.1.** *Let $(M,\varphi)=(\Gamma(H_{\mathbb{R}},U)'',\varphi_U)$ be any free Araki–Woods factor endowed with its free quasifree state. Keep the same notation as in the introduction. Let $q\in M^{\varphi}=N^{\varphi_{U^{\mathrm{ap}}}}$ be any nonzero projection. Write $\varphi_q=\varphi(q\cdot q)/\varphi(q)$. Then for any amenable von Neumann subalgebra $Q\subset qMq$ that is globally invariant under the modular automorphism group $\sigma^{\varphi_q}$, we have $Q\subset qNq$.*

*Proof.* We may assume that $Q$ has separable predual. Indeed, let $x\in Q$ be any element and denote by $Q_0\subset Q$ the von Neumann subalgebra generated by $x\in Q$ that is globally invariant under the modular automorphism group $\sigma^{\varphi_q}$. Then $Q_0$ is amenable and has separable predual. Therefore, we may assume without loss of generality that $Q_0=Q$, that is, $Q$ has separable predual.

Special case. We first prove the result when $Q\subset qMq$ is globally invariant under $\sigma^{\varphi_q}$ and is an irreducible subfactor, meaning that $Q'\cap qMq=\mathbb{C}q$.

Let $a\in Q$ be any element. Since $Q$ is amenable and has separable predual, $Q'\cap(qMq)^{\omega}$ is diffuse and so is $Q'\cap((qMq)^{\omega})^{\varphi_q^{\omega}}$ by [Houdayer and Raum 2015, Theorem 2.3]. In particular, there exists a unitary $u\in\mathcal{U}\big(Q'\cap((qMq)^{\omega})^{\varphi_q^{\omega}}\big)$ such that $\varphi_q^{\omega}(u)=0$. Note that $\mathrm{E}_{\omega}(u)\in Q'\cap qMq=\mathbb{C}q$, and hence $\mathrm{E}_{\omega}(u)=\varphi_q^{\omega}(u)=0$, so that $u\in(M^{\omega})^{\varphi^{\omega}}\cap(M^{\omega}\ominus M)$. Theorem 3.1 yields $\varphi^{\omega}(a^*u^*(a-\mathrm{E}_N(a))u)=0$. Since moreover $au=ua$ and $u\in\mathcal{U}\big((qMq)^{\varphi_q^{\omega}}\big)$, we have

$$\begin{aligned}\|a\|_{\varphi}^2=\|au\|_{\varphi^{\omega}}^2&=\varphi^{\omega}(u^*a^*au)=\varphi^{\omega}(a^*u^*au)\\&=\varphi^{\omega}(a^*u^*\,\mathrm{E}_N(a)u)=\varphi^{\omega}(ua^*u^*\,\mathrm{E}_N(a))=\varphi(a^*\,\mathrm{E}_N(a))=\|\,\mathrm{E}_N(a)\|_{\varphi}^2.\end{aligned}$$

This shows that $a=\mathrm{E}_N(a)\in N$.

General case. We next prove the result when $Q\subset qMq$ is any amenable subalgebra globally invariant under $\sigma^{\varphi_q}$.

Denote by $z\in\mathcal{Z}(Q)\subset N^{\varphi}$ the unique central projection such that $Qz$ is atomic and $Q(1-z)$ is diffuse. Since $Qz$ is atomic and globally invariant under the modular automorphism group $\sigma^{\varphi_z}$, we have that $\varphi_z|_{Qz}$ is almost periodic and hence $Qz\subset N$. It remains to prove that $Q(1-z)\subset N$. Cutting down by $1-z$ if necessary, we may assume that $Q$ itself is diffuse.

Since $Q\subset qMq$ is diffuse and with expectation and since $M$ is solid (see [Houdayer and Raum 2015, Theorem A] and [Houdayer and Isono 2016, Theorem 7.1], which does not require separability of the predual), the relative commutant $Q'\cap qMq$ is amenable. Up to replacing $Q$ by $Q\vee Q'\cap qMq$, which is still amenable and globally invariant under the modular automorphism group $\sigma^{\varphi_q}$, we may assume that $Q'\cap qMq=\mathcal{Z}(Q)$. Denote by $(z_n)_n$ a sequence of central projections in $\mathcal{Z}(Q)$ such that $\sum_n z_n=q$, $(Qz_0)'\cap z_0Mz_0=\mathcal{Z}(Q)z_0$ is diffuse and $(Qz_n)'\cap z_nMz_n=\mathbb{C}z_n$ for every $n\geq 1$.

- By the special case above, we know that $Qz_n\subset N$ for all $n\geq 1$.
- Since $\mathcal{Z}(Q)z_0\oplus(1-z_0)N(1-z_0)$ is diffuse and with expectation in $N$, its relative commutant inside $M$ is contained in $N$ by [Houdayer and Ueda 2016, Proposition 2.7(1)]. In particular, $Qz_0\subset N$.

Therefore, we have $Q\subset N$. □

*Proof of the main theorem.* Put $\varphi := \varphi_U$. Denote by $z \in \mathcal{Z}(Q) \subset M^\varphi = N^\varphi$ the unique central projection such that $Qz$ is amenable and $Qz^\perp$ has no nonzero amenable direct summand. By Theorem 4.1, we have $Qz \subset zNz$. Fix any nonprincipal ultrafilter $\omega \in \beta(\boldsymbol{N}) \setminus \boldsymbol{N}$. Then $(Q' \cap M^\omega)z^\perp = (Q' \cap M)z^\perp$ is atomic, by [Houdayer and Raum 2015, Theorem A] (see also [Houdayer and Isono 2016, Theorem 7.1]). □

## Acknowledgments

The present work was done when the authors were visiting the University of California at San Diego (UCSD). They thank Adrian Ioana and the mathematics department at UCSD for their kind hospitality.

Boutonnet is supported by NSF Career Grant DMS 1253402. Houdayer is supported by ERC Starting Grant GAN 637601.

## References

[Ando and Haagerup 2014] H. Ando and U. Haagerup, "Ultraproducts of von Neumann algebras", *J. Funct. Anal.* **266**:12 (2014), 6842–6913. MR Zbl

[Houdayer 2014a] C. Houdayer, "A class of $\mathrm{II}_1$ factors with an exotic abelian maximal amenable subalgebra", *Trans. Amer. Math. Soc.* **366**:7 (2014), 3693–3707. MR Zbl

[Houdayer 2014b] C. Houdayer, "Structure of $\mathrm{II}_1$ factors arising from free Bogoljubov actions of arbitrary groups", *Adv. Math.* **260** (2014), 414–457. MR Zbl

[Houdayer 2015] C. Houdayer, "Gamma stability in free product von Neumann algebras", *Comm. Math. Phys.* **336**:2 (2015), 831–851. MR Zbl

[Houdayer and Isono 2016] C. Houdayer and Y. Isono, "Bi-exact groups, strongly ergodic actions and group measure space type III factors with no central sequence", *Comm. Math. Phys.* **348**:3 (2016), 991–1015. MR

[Houdayer and Raum 2015] C. Houdayer and S. Raum, "Asymptotic structure of free Araki–Woods factors", *Math. Ann.* **363**:1-2 (2015), 237–267. MR Zbl

[Houdayer and Ueda 2016] C. Houdayer and Y. Ueda, "Asymptotic structure of free product von Neumann algebras", *Math. Proc. Cambridge Philos. Soc.* **161**:3 (2016), 489–516. MR

[Ocneanu 1985] A. Ocneanu, *Actions of discrete amenable groups on von Neumann algebras*, Lecture Notes in Mathematics **1138**, Springer, Berlin, 1985. MR Zbl

[Popa 1983] S. Popa, "Maximal injective subalgebras in factors associated with free groups", *Adv. in Math.* **50**:1 (1983), 27–48. MR Zbl

[Shlyakhtenko 1997] D. Shlyakhtenko, "Free quasi-free states", *Pacific J. Math.* **177**:2 (1997), 329–368. MR Zbl

[Voiculescu 1985] D. Voiculescu, "Symmetries of some reduced free product $C^*$-algebras", pp. 556–588 in *Operator algebras and their connections with topology and ergodic theory* (Buşteni, 1983), edited by H. Araki et al., Lecture Notes in Math. **1132**, Springer, Berlin, 1985. MR Zbl

[Voiculescu et al. 1992] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables: a noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups*, CRM Monograph Series **1**, American Mathematical Society, Providence, RI, 1992. MR Zbl

[Wen 2016] C. Wen, "Maximal amenability and disjointness for the radial masa", *J. Funct. Anal.* **270**:2 (2016), 787–801. MR Zbl

Received 24 Feb 2016. Revised 9 Jun 2016. Accepted 3 Oct 2016.

RÉMI BOUTONNET: remi.boutonnet@math.u-bordeaux1.fr
*Institut de Mathématiques de Bordeaux, CNRS, Université Bordeaux I, 33405 Talence, France*

CYRIL HOUDAYER: cyril.houdayer@math.u-psud.fr
*Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, CNRS, 91405 Orsay, France*

ANALYSIS AND PDE
Vol. 9, No. 8, 2016
dx.doi.org/10.2140/apde.2016.9.1999

# FINITE-TIME BLOWUP FOR A SUPERCRITICAL DEFOCUSING NONLINEAR WAVE SYSTEM

TERENCE TAO

We consider the global regularity problem for defocusing nonlinear wave systems

$$\Box u = (\nabla_{\mathbb{R}^m} F)(u)$$

on Minkowski spacetime $\mathbb{R}^{1+d}$ with d'Alembertian $\Box := -\partial_t^2 + \sum_{i=1}^d \partial_{x_i}^2$, where the field $u : \mathbb{R}^{1+d} \to \mathbb{R}^m$ is vector-valued, and $F : \mathbb{R}^m \to \mathbb{R}$ is a smooth potential which is positive and homogeneous of order $p+1$ outside of the unit ball for some $p > 1$. This generalises the scalar defocusing nonlinear wave (NLW) equation, in which $m = 1$ and $F(v) = 1/(p+1)|v|^{p+1}$. It is well known that in the energy-subcritical and energy-critical cases when $d \le 2$ or $d \ge 3$ and $p \le 1 + 4/(d-2)$, one has global existence of smooth solutions from arbitrary smooth initial data $u(0), \partial_t u(0)$, at least for dimensions $d \le 7$. We study the supercritical case where $d = 3$ and $p > 5$. We show that in this case, there exists a smooth potential $F$ for some sufficiently large $m$ (in fact we can take $m = 40$), positive and homogeneous of order $p+1$ outside of the unit ball, and a smooth choice of initial data $u(0), \partial_t u(0)$ for which the solution develops a finite-time singularity. In fact the solution is discretely self-similar in a backwards light cone. The basic strategy is to first select the mass and energy densities of $u$, then $u$ itself, and then finally design the potential $F$ in order to solve the required equation. The Nash embedding theorem is used in the second step, explaining the need to take $m$ relatively large.

## 1. Introduction

Let $\mathbb{R}^m$ be a Euclidean space, with the usual Euclidean norm $v \mapsto \|v\|_{\mathbb{R}^m}$ and Euclidean inner product $v, w \mapsto \langle v, w\rangle_{\mathbb{R}^m}$. A function $F : \mathbb{R}^m \to \mathbb{R}^n$ is said to be *homogeneous of order* $\alpha$ for some real $\alpha$ if we have

$$F(\lambda v) = \lambda^\alpha F(v) \tag{1-1}$$

for all $\lambda > 0$ and $v \in \mathbb{R}^m$. In particular, differentiating this at $\lambda = 1$ we obtain *Euler's identity*

$$\langle v, (\nabla_{\mathbb{R}^m} F)(v)\rangle_{\mathbb{R}^m} = \alpha F(v), \tag{1-2}$$

where $\nabla_{\mathbb{R}^m}$ denotes the gradient in $\mathbb{R}^m$, assuming of course that the gradient $\nabla_{\mathbb{R}^m} F$ of $F$ exists at $v$. When $\alpha$ is not an integer, it is not possible for such homogeneous functions to be smooth at the origin unless they are identically zero (this can be seen by performing a Taylor expansion of $F$ around the origin). To avoid this technical issue, we also introduce the notion of $F$ being *homogeneous of order* $\alpha$ *outside of the unit ball*, by which we mean that (1-1) holds for $\lambda \ge 1$ and $v \in \mathbb{R}^m$ with $\|v\|_{\mathbb{R}^m} \ge 1$.

*MSC2010:* primary 35Q30, 35L71; secondary 35L67.
*Keywords:* nonlinear wave equation, Nash embedding theorem.

Define a *potential* to be a function $F : \mathbb{R}^m \to \mathbb{R}$ that is smooth away from the origin; if $F$ is also smooth at the origin, we call it a *smooth potential*. We say that the potential is *defocusing* if $F$ is positive away from the origin, and *focusing* if $F$ is negative away from the origin. In this paper we consider nonlinear wave systems of the form

$$\Box u = (\nabla_{\mathbb{R}^m} F)(u), \tag{1-3}$$

where the unknown field $u : \mathbb{R}^{1+d} \to \mathbb{R}^m$ is assumed to be smooth, $\Box = \partial^\alpha \partial_\alpha = -\partial_t^2 + \sum_{i=1}^d \partial_{x_i}^2$ is the d'Alembertian operator on Minkowski spacetime

$$\mathbb{R}^{1+d} := \{(t, x_1, \ldots, x_d) : t, x_1, \ldots, x_d \in \mathbb{R}\} = \{(t, x) : t \in \mathbb{R}, x \in \mathbb{R}^d\}$$

with the usual Minkowski metric

$$\eta_{\alpha\beta} x^\alpha x^\beta = -t^2 + x_1^2 + \cdots + x_d^2$$

and the usual Einstein summation, raising, and lowering conventions, $m, d \geq 1$ are integers, and $F : \mathbb{R}^m \to \mathbb{R}$ is a smooth potential. This is a Lagrangian field equation, in the sense that (1-3) is (formally, at least) the Euler–Lagrange equations for the Lagrangian

$$\int_{\mathbb{R}^{1+d}} \tfrac{1}{2} \langle \partial^\alpha u, \partial_\alpha u \rangle_{\mathbb{R}^m} + F(u)\, d\eta.$$

We will restrict attention to potentials $F$ which are homogeneous outside of the unit ball of order $p+1$ for some exponent $p > 1$. The well-studied *nonlinear wave equation* (NLW) corresponds to the case when $m = 1$ and $F(v) = |v|^{p+1}/(p+1)$ (for the defocusing NLW) or $F(v) = -|v|^{p+1}/(p+1)$ (for the focusing NLW), with the caveat that one needs to restrict $p$ to be an odd integer if one wants these potentials to be smooth. Later in the paper we will restrict attention to the physical case $d = 3$, basically to take advantage of a form of the sharp Huygens' principle.

The natural initial value problem to study here is the Cauchy initial value problem, in which one specifies a smooth initial position $u_0 : \mathbb{R}^d \to \mathbb{R}^m$ and initial velocity $u_1 : \mathbb{R}^d \to \mathbb{R}^m$, and asks for a smooth solution $u$ to (1-3) with $u(0, x) = u_0(x)$ and $\partial_t u(0, x) = u_1(x)$. Standard energy methods (see, e.g., [Shatah and Struwe 1998]) show that for any choice of smooth initial data $u_0, u_1 : \mathbb{R}^d \to \mathbb{R}^m$, one can construct a solution $u$ to (1-3) in an open neighbourhood $\Omega$ in $\mathbb{R}^{1+d}$ of the initial time slice $\{(0, x) : x \in \mathbb{R}^d\}$ with this initial data. Furthermore, either such a solution can be extended to be globally defined in $\mathbb{R}^{1+d}$, or else there is a solution $u$ defined on some open neighbourhood $\Omega$ of $\{(0, x) : x \in \mathbb{R}^d\}$ that "blows up" in the sense that it cannot be smoothly continued to some boundary point $(t_*, x_*)$ of $\Omega$. The *global regularity problem* for a given choice of potential $F$ asks if the latter situation does not occur, that is to say that for every choice of smooth initial data there is a smooth global solution. Note that as the equation (1-3) enjoys finite speed of propagation, there is no need to specify any decay hypotheses on the initial data as this will not affect the answer to the global regularity problem.

For focusing potentials $F$, there are well-known blowup examples that show that global regularity fails. For instance, if $m = 1$ and $F : \mathbb{R} \to \mathbb{R}$ is given by

$$F(v) := -\frac{2}{(p-1)^2} |v|^{p+1} \tag{1-4}$$

for all $|v| \geq 1$ (and extended arbitrarily in some smooth fashion to the region $|v| < 1$ while remaining negative away from the origin), then $F$ is a focusing potential that is homogeneous of order $p+1$ outside of the unit ball, and the function $u : \{(t,x) \in \mathbb{R}^{1+d} : 0 < t \leq 1\} \to \mathbb{R}$ defined by

$$u(t,x) := t^{-\frac{2}{p-1}} \tag{1-5}$$

solves (1-3) but blows up at the boundary $t = 0$; applying the time reversal symmetry $(t,x) \mapsto (1-t,x)$, we obtain a counterexample to global regularity for this choice of $F$. We will thus henceforth restrict attention to defocusing potentials $F$, which excludes ODE-type blowup examples (1-5) in which $u(t,x)$ depends only on $t$.

The *energy* (or *Hamiltonian*)

$$E[u(t)] := \int_{\mathbb{R}^d} \tfrac{1}{2}\|\partial_t u(t,x)\|^2_{\mathbb{R}^m} + \tfrac{1}{2}\|\nabla_x u(t,x)\|^2_{\mathbb{R}^m \otimes \mathbb{R}^d} + F(u(t,x))\,dx \tag{1-6}$$

is (formally, at least) conserved by the flow (1-3). A dimensional analysis of this quantity then naturally splits the range of parameters $(d, p)$ into three cases:

- The *energy-subcritical* case when $d \leq 2$, or when $d \geq 3$ and $p < 1 + \frac{4}{d-2}$.
- The *energy-critical* case when $d \geq 3$ and $p = 1 + \frac{4}{d-2}$.
- The *energy-supercritical* case when $d \geq 3$ and $p > 1 + \frac{4}{d-2}$.

In the energy-subcritical and energy-critical cases one has global regularity for any defocusing NLW system, at least when $d \leq 7$; see[1] [Jörgens 1961] for the subcritical case, and [Grillakis 1990; 1992; Struwe 1988; Shatah and Struwe 1998] for the critical case. These results were also extended to the logarithmically supercritical case (in which the potential $F$ grows faster than the energy-critical potential by a logarithmic factor) in [Tao 2007; Roy 2009]. A major ingredient in the proof of global regularity in these cases is the conservation of the energy (1-6), which is nonnegative in the defocusing case. In the energy-critical (and logarithmically supercritical) case, one also takes advantage of Morawetz inequalities such as

$$\int_0^T \int_{\mathbb{R}^d} \frac{F(u(t,x))}{|x|}\,dx\,dt \leq C E[u(0)] \tag{1-7}$$

for any time interval $[0, T]$ on which the solution exists. These bounds can be deduced from the properties of the stress-energy tensor

$$T_{\alpha\beta} := \langle \partial_\alpha u, \partial_\beta u\rangle - \tfrac{1}{2}\eta_{\alpha\beta}(\langle \partial^\gamma u, \partial_\gamma u\rangle_{\mathbb{R}^m} + F(u))$$

and in particular in the divergence-free nature $\partial^\beta T_{\alpha\beta} = 0$ of this tensor.

It thus remains to address the energy-supercritical case for defocusing smooth potentials $F$. In this case it is known that the Cauchy problem is ill-posed in various technical senses at low regularities [Lebeau 2001; 2005; Christ et al. 2003; Brenner and Kumlin 2000; Burq et al. 2007; Ibrahim et al. 2011],

[1] Several of these references restrict attention to the scalar NLW or to three spatial dimensions, but the arguments extend without difficulty to the energy-critical NLW systems considered here in the range $3 \leq d \leq 7$. There are technical difficulties establishing global regularity in extremely high dimension, even when the potential $F$ and all of its derivatives are bounded; see, e.g., [Brenner and von Wahl 1981].

despite the existence of global weak solutions [Segal 1963; Strauss 1989], as well as global smooth solutions from sufficiently small initial data [Lindblad and Sogge 1996] (assuming that $F$ vanishes to sufficiently high order at the origin); see also [Zheng 1991] for a partial regularity result. However, to the author's knowledge, finite-time blowup of smooth solutions has not actually been demonstrated for such equations. The main result of this paper is to establish such a finite-time blowup for at least some choices of defocusing potential $F$ and parameters $d, p, m$:

**Theorem 1.1** (finite-time blowup). *Let $d = 3$, let $p > 1 + \frac{4}{d-2}$, and let*

$$m \geq 2 \max\left( \frac{(d+1)(d+6)}{2}, \frac{(d+1)(d+4)}{2} + 5 \right) + 2$$

*be an integer. Then there exists a defocusing smooth potential $F : \mathbb{R}^m \to \mathbb{R}$ that is homogeneous of order $p+1$ outside of the unit ball, and a smooth choice of initial data $u_0, u_1 : \mathbb{R}^d \to \mathbb{R}^m$ such that there is no global smooth solution $u : \mathbb{R}^{1+d} \to \mathbb{R}^m$ to the nonlinear wave system (1-3) with initial data $u(0) = u_0$, $\partial_t u(0) = u_1$.*

Of course, since $d$ is set equal to 3, the conditions on $p$ and $m$ reduce to $p > 5$ and $m \geq 40$ respectively. However, our restriction to the $d = 3$ case is largely for technical reasons (basically in order to exploit the strong Huygens principle), and we believe the results should extend to higher values of $d$, with the indicated constraints on $d$ and $p$, though we will not pursue this matter here. The rather large value of $m$ is due to our use of the Nash embedding theorem (!) at one stage of the argument. It would of course be greatly desirable to lower the number $m$ of degrees of freedom down to 1, in order to establish blowup for the scalar defocusing supercritical NLW, but our methods crucially need a large value of $m$ in order to ensure that a certain map from a $(1+d)$-dimensional space into the sphere $S^{m-1}$ is embedded, which is where the Nash embedding theorem comes in. Nevertheless, even though Theorem 1.1 does not directly show that the scalar defocusing supercritical NLW exhibits finite-time blowup, it does demonstrate a significant *barrier* to any attempt to prove global regularity for this equation, as such an attempt must necessarily use some special property of the scalar equation that is not shared by the more general system (1-3).

We briefly discuss the methods used to prove Theorem 1.1. The singularity constructed is a discretely self-similar blowup in a backwards light cone; see the reduction to Theorem 2.1 below. In particular, the blowup is "locally of type II" in the sense that scale-invariant norms inside the light cone stay bounded, but not "globally of type II", as a significant amount of energy (as measured using scale-invariant norms) radiates out of the backwards light cone at all scales. This is compatible with the results in [Kenig and Merle 2008; Killip and Visan 2011a; 2011b], which rule out "global" type II blowup, but not "local" type II blowup. It would be natural to seek a *continuously* self-similar smooth blowup solution, but it turns out[2] that these are ruled out; see Proposition 2.2 below. Hence we will not restrict attention to

[2]On the other hand, it is possible to use perturbative methods to create *rough* solutions to (1-3) that are continuously self-similar: see [Planchon 2000; Ribaud and Youssfi 2002]. However, these methods do not seem to be adaptable to generate smooth solutions, and indeed Proposition 2.2 suggests that there are strong obstacles in trying to create such an adaptation. The negative result here also stands in contrast to the situation of high-dimensional wave maps into negatively curved targets, where ODE methods were used in [Cazenave et al. 1998] to construct continuously self-similar blowup examples in seven and higher spatial dimensions.

continuously self-similar solutions. It also turns out to be convenient not to initially restrict attention to spherically symmetric solutions, although we will eventually do so later in the argument.

Traditionally, one thinks of the potential $F$ as being prescribed in advance, and the field $u$ as the unknown to be solved for. However, as we have the freedom to *select* $F$ in Theorem 1.1, it turns out to be more convenient to prescribe $u$ first, and only then design an $F$ for which the equation (1-3) is obeyed. This turns out to be possible as long as the map

$$\theta : (t,x) \mapsto \frac{u(t,x)}{\|u(t,x)\|_{\mathbb{R}^m}}$$

has certain nondegeneracy properties, and if the stress-energy tensor $T_{\alpha\beta}$ (which can be defined purely in terms of $u$) is divergence-free; see the reduction to Theorem 3.2 below. The stress-energy tensor $T_{\alpha\beta}$ (or more precisely, some related fields which we call the *mass density* $M$ and the *energy tensor* $E_{\alpha\beta}$) can be viewed as prescribing the metric geometry of the map $\theta$, and the Nash embedding theorem can then be used to locate a choice of $\theta$ with the desired nondegeneracy properties and the prescribed metric, so long as the fields $M$ and $E_{\alpha\beta}$ obey a number of conditions (one of which relates to the divergence-free nature of the stress-energy tensor, and another to the positive definiteness of the Gram matrix of $u$). This reduces the problem to a certain "semidefinite program" (see Theorem 4.1), in which one now only needs to specify the fields $M$ and $E_{\alpha\beta}$, rather than the original field $u$ or the potential $F$.

It is at this point (after some additional technical reductions in which certain fields are allowed to degenerate to zero) that it finally becomes convenient to make symmetry reductions, working with fields $M$, $E_{\alpha\beta}$ that are both continuously self-similar and spherically symmetric, and assuming that there are no angular components to the energy tensor. In three spatial dimensions, this reduces the divergence-free nature of the stress-energy tensor to a single transport equation for the null energy $e_+$ (which, in terms of the original field $u$, is given in polar coordinates by $e_+ = \frac{1}{2}\|(\partial_t + \partial_r)(ru)\|_H^2$), in terms of a certain "potential energy density" $V$ (which, in terms of the original data $u$ and $F$, is given by $V = rF(u)$); see Theorem 5.4 for a precise statement. The strategy is then to solve for these fields $e_+, V$ first, and then choose all the remaining unknown fields in such a way that the remaining requirements of the semidefinite program are satisfied. This turns out to be possible if the fields $e_+, V$ are chosen to concentrate close to the boundary of the light cone.

## 2. Reduction to discretely self-similar solution

We begin the proof of Theorem 1.1.

We first observe that from finite speed of propagation and the symmetries of the equation, Theorem 1.1 follows from the claim below, in which the solution is restricted to a truncated light cone and is discretely self-similar and the potential is now homogeneous everywhere (not just outside of the unit ball), but no longer required to be smooth. This reduction does not use any of the hypotheses on $m, d, p$.

**Theorem 2.1** (first reduction). *Let $d = 3$, let $p > 1 + \frac{4}{d-2}$, and let*

$$m \geq 2\max\left(\frac{(d+1)(d+6)}{2}, \frac{(d+1)(d+4)}{2} + 5\right) + 2$$

*be an integer. Then there exists a defocusing potential* $F : \mathbb{R}^m \to \mathbb{R}$ *which is homogeneous of order* $p+1$ *and a smooth function* $u : \Gamma_d \to \mathbb{R}^m \backslash \{0\}$ *on the light cone* $\Gamma_d := \{(t,x) \in \mathbb{R}^{1+d} : t > 0; |x| \leq t\}$ *that solves (1-3) on its domain and is nowhere vanishing, and also discretely self-similar in the sense that there exists* $S > 0$ *such that*

$$u(e^S t, e^S x) = e^{-\frac{2}{p-1}S} u(t,x) \tag{2-1}$$

*for all* $(t,x) \in \Gamma_d$.

A key point here is that $u$ is smooth all the way up to the boundary of the light cone $\Gamma_d$, rather than merely being smooth in the interior. The exponent $-\frac{2}{p-1}$ is mandated by dimensional analysis considerations. It would be natural to consider solutions that are continuously self-similar in the sense that (2-1) holds for *all* $S \in \mathbb{R}$, but as we shall shortly see, it will not be possible to generate such solutions in the three-dimensional defocusing setting.

Let us assume Theorem 2.1 for the moment, and show how it implies Theorem 1.1. Let $F$, $S$, $u$ be as in Theorem 2.1. Since $u$ is smooth and nonzero on the compact region $\{(t,x) \in \Gamma_d : e^{-S} \leq t \leq 1\}$, it is bounded from below in this region. By replacing $u$ with $Cu$ and $F$ with $v \mapsto C^2 F(v/C)$ for some large constant $C$, we may thus assume that

$$\|u(t,x)\|_{\mathbb{R}^m} \geq 1$$

whenever $(t,x) \in \Gamma_d$ with $e^{-S} \leq t \leq 1$. Using the discrete self-similarity property (2-1), we then have this bound for all $0 < t \leq 1$; in fact we have a lower bound on $\|u(t,x)\|_{\mathbb{R}^m}$ that goes to infinity as $t \to 0$, ensuring in particular that $u$ has no smooth extension to $(0,0)$.

Using a smooth cutoff function, one can find a smooth defocusing potential $\widetilde{F} : \mathbb{R}^m \to \mathbb{R}$ that agrees with $F$ in the region $\{v \in \mathbb{R}^m : \|v\|_{\mathbb{R}^m} \geq 1\}$. Then $u$ solves (1-3) with this potential in the truncated light cone $\{(t,x) \in \mathbb{R}^{1+d} : 0 < t \leq 1; |x| \leq t\}$ with $F$ replaced by $\widetilde{F}$. Choose smooth initial data $v_0, v_1 : \mathbb{R}^d \to \mathbb{R}^m$ such that

$$v_0(x) = u(1,x)$$

and

$$v_1(x) = -\partial_t u(1,x)$$

for all $|x| \leq 1$ (where we use $|x| := \|x\|_{\mathbb{R}^d}$ to denote the magnitude of $x \in \mathbb{R}^d$); such data exists from standard smooth extension theorems (see, e.g., [Seeley 1964]) since the functions $u(1,x)$, $\partial_t u(1,x)$ are smooth on the closed ball $\{x : |x| \leq 1\}$. Suppose for contradiction that Theorem 1.1 failed (with $F$ replaced by $\widetilde{F}$); then we have a global smooth solution $v : \mathbb{R}^{1+d} \to \mathbb{R}^m$ to (1-3) (for $\widetilde{F}$) with initial data $v(0) = v_0$, $\partial_t v(0) = v_1$. The function $\tilde{u} : (t,x) \mapsto v(1-t,x)$ is then another global smooth solution to (1-3) (for $\widetilde{F}$) such that $\tilde{u}(1,x) = u(1,x)$ and $\partial_t \tilde{u}(1,x) = \partial_t u(1,x)$ for all $|x| \leq 1$. Finite speed of propagation (see, e.g., [Tao 2006, Proposition 3.3]) then shows that $\tilde{u}$ and $u$ agree in the region $\{(t,x) \in \mathbb{R}^{1+d} : 0 < t \leq 1; |x| \leq t\}$; as $\tilde{u}$ is smoothly extendible to $(0,0)$, we know $u$ is also, giving the desired contradiction. This concludes the derivation of Theorem 1.1 from Theorem 2.1.

It remains to prove Theorem 2.1. This will be the focus of the remaining sections of the paper. For now, let us show why continuously self-similar solutions are not available in the defocusing case, at

least for some choices of parameters $d$, $p$. The point will be that continuous self-similarity gives a new monotonicity formula for a certain quantity $f(t,r)$ (measuring a sort of "equipartition of energy") that can be used to derive a contradiction.

**Proposition 2.2** (no self-similar defocusing solutions). *Let $d \geq 3$ and $p > 1$ be such that $\frac{d-3}{2} - \frac{2}{p-1} < 0$, let $m$ be a natural number, and let $F : \mathbb{R}^m \to \mathbb{R}$ be a defocusing potential that is homogeneous of order $p+1$. Then there does not exist a smooth solution $u : \Gamma_d \to \mathbb{R}^m \backslash \{0\}$ to (1-3) that is homogeneous of order $-\frac{2}{p-1}$.*

Note in particular that in the physical case $d = 3$, the condition $\frac{d-3}{2} - \frac{2}{p-1} < 0$ is automatic, and so no self-similar defocusing solutions exist in this case. We do not know if this condition is necessary in the above proposition.

*Proof.* Suppose for contradiction that such a $u$ exists. Equation (1-3) in polar coordinates $(t, r, \omega)$ reads

$$-\partial_{tt} u + \partial_{rr} u + \frac{d-1}{r} \partial_r u + \frac{1}{r^2} \Delta_\omega u = (\nabla F)(u),$$

where $\Delta_\omega$ is the Laplace–Beltrami operator on the sphere $S^{d-1}$. Making the substitution

$$\phi(t, r, \omega) := r^{\frac{d-1}{2}} u(t, r, \omega), \tag{2-2}$$

this becomes

$$-\partial_{tt}\phi + \partial_{rr}\phi - \frac{1}{r^2}\Big(-\Delta_\omega + \frac{(d-1)(d-3)}{4}\Big)\phi = r^{\frac{d-1}{2}} (\nabla F)(r^{-\frac{d-1}{2}} \phi) \tag{2-3}$$

for $r > 0$.

We introduce the scaling vector field $S := t\partial_t + r\partial_r$ and the Lorentz boost $L := r\partial_t + t\partial_r$. Observe that $L$ and $S$ commute with

$$-S^2 + L^2 = (t^2 - r^2)(-\partial_{tt} + \partial_{rr}) \tag{2-4}$$

and thus

$$-\langle S^2\phi, L\phi\rangle_{\mathbb{R}^m} + \langle L^2\phi, L\phi\rangle_{\mathbb{R}^m} = (t^2 - r^2)\langle -\partial_{tt}\phi + \partial_{rr}\phi, L\phi\rangle_{\mathbb{R}^m}.$$

As $u$ is assumed homogeneous of order $-\frac{2}{p-1}$, we know $\phi$ is homogeneous of order $\frac{d-1}{2} - \frac{2}{p-1}$. From Euler's identity (1-2) we thus have $\phi$ an eigenfunction of $S$,

$$S\phi = \Big(\frac{d-1}{2} - \frac{2}{p-1}\Big)\phi,$$

and thus (by the commutativity of $L$ and $S$)

$$\langle L\phi, S^2\phi\rangle_{\mathbb{R}^m} = \langle LS\phi, S\phi\rangle_{\mathbb{R}^m} = \tfrac{1}{2} L \|S\phi\|^2_{\mathbb{R}^m}.$$

We also have

$$\langle L\phi, L^2\phi\rangle_{\mathbb{R}^m} = \tfrac{1}{2} L \|L\phi\|^2_{\mathbb{R}^m}.$$

Putting all of these facts together, we conclude that

$$L\big(-\tfrac{1}{2}\|S\phi\|^2_{\mathbb{R}^m} + \tfrac{1}{2}\|L\phi\|^2_{\mathbb{R}^m}\big) = (t^2 - r^2)\langle -\partial_{tt}\phi + \partial_{rr}\phi, L\phi\rangle_{\mathbb{R}^m}.$$

A computation similar to (2-4) shows that

$$-\|S\phi\|_{\mathbb{R}^m}^2 + \|L\phi\|_{\mathbb{R}^m} = (t^2 - r^2)(-\|\partial_t \phi\|_{\mathbb{R}^m}^2 + \|\partial_r \phi\|_{\mathbb{R}^m}^2).$$

Since $t^2 - r^2$ is annihilated by $L$, we conclude that

$$L\big(-\tfrac{1}{2}\|\partial_t \phi\|_{\mathbb{R}^m}^2 + \tfrac{1}{2}\|\partial_r \phi\|_{\mathbb{R}^m}^2\big) = \langle -\partial_{tt}\phi + \partial_{rr}\phi, \phi\rangle_{\mathbb{R}^m}.$$

By (2-3), the right-hand side is equal to

$$\frac{1}{r^2}\langle -\Delta_\omega \phi, L\phi\rangle_{\mathbb{R}^m} + \frac{(d-1)(d-3)}{4r^2}\langle \phi, L\phi\rangle_{\mathbb{R}^m} + r^{\frac{d-1}{2}}\big\langle (\nabla F)(r^{-\frac{d-1}{2}}\phi), L\phi\big\rangle_{\mathbb{R}^m}.$$

To deal with the angular Laplacian, we integrate over $S^{d-1}$ and then integrate by parts to conclude that

$$\begin{aligned} L\int_{S^{d-1}} \big(-\tfrac{1}{2}\|\partial_t \phi\|_{\mathbb{R}^m}^2 + \tfrac{1}{2}\|\partial_r \phi\|_{\mathbb{R}^m}^2\big)\, d\omega \\ = \int_{S^{d-1}} \frac{1}{2r^2} L\|\nabla_\omega \phi\|_{\mathbb{R}^m \otimes \mathbb{R}^d}^2 + \frac{(d-1)(d-3)}{8r^2} L\|\phi\|_{\mathbb{R}^m}^2 + r^{\frac{d-1}{2}}\big\langle (\nabla F)(r^{-\frac{d-1}{2}}\phi), L\phi\big\rangle_{\mathbb{R}^m}\, d\omega, \end{aligned}$$

where we use the fact that the Lorentz boost $L$ commutes with angular derivatives, and where $d\omega$ denotes surface measure on $S^{d-1}$.

From the chain and product rules, noting that $Lr = t$, we have

$$L\phi = r^{\frac{d-1}{2}} L(r^{-\frac{d-1}{2}}\phi) + \frac{d-1}{2}\frac{t}{r}\phi$$

and thus (using (1-2))

$$\begin{aligned} \big\langle (\nabla F)(r^{-\frac{d-1}{2}}\phi), L\phi\big\rangle_{\mathbb{R}^m} &= r^{\frac{d-1}{2}}\left( LF(r^{-\frac{d-1}{2}}\phi) + \frac{d-1}{2}\frac{t}{r}\big\langle r^{-\frac{d-1}{2}}\phi, (\nabla F)(r^{-\frac{d-1}{2}}\phi)\big\rangle_{\mathbb{R}^m}\right) \\ &= r^{\frac{d-1}{2}}\left( LF(r^{-\frac{d-1}{2}}\phi) + \frac{(d-1)(p+1)}{2}\frac{t}{r} F(r^{-\frac{d-1}{2}}\phi)\right). \end{aligned}$$

Putting all this together, we see that if we introduce the quantity

$$\begin{aligned} f(t,r) := \int_{S^{d-1}} -\frac{1}{2}\|\partial_t \phi\|_{\mathbb{R}^m}^2 + \frac{1}{2}\|\partial_r \phi\|_{\mathbb{R}^m}^2 - \frac{1}{2r^2}\|\nabla_\omega \phi\|_{\mathbb{R}^m \otimes \mathbb{R}^d}^2 \\ - \frac{(d-1)(d-3)}{8r^2}\|\phi\|_{\mathbb{R}^m}^2 - r^{d-1} F(r^{-\frac{d-1}{2}}\phi)\, d\omega \end{aligned}$$

then we have the formula

$$Lf = \int_{S^{d-1}} \frac{t}{r^3}\|\nabla_\omega \phi\|_{\mathbb{R}^m \otimes \mathbb{R}^d}^2 + \frac{(d-1)(d-3)t}{4r^3}\|\phi\|_{\mathbb{R}^m}^2 + \frac{(d-1)(p-1)}{2}\frac{t}{r} r^{d-1} F(r^{-\frac{d-1}{2}}\phi)\, d\omega$$

for any $r > 0$. In particular, $f(\cosh y, \sinh y)$ is a strictly function of $y$ for $y > 0$, since

$$\frac{d}{dy} f(\cosh y, \sinh y) = (Lf)(\cosh y, \sinh y) > 0$$

with the strict positivity coming from the defocusing nature of $F$. On the other hand, when $y \to 0^+$, we see from (2-2) that all the negative integrands in the definition of $f(\cosh y, \sinh y)$ go to zero, and thus

$$\lim_{y \to 0^+} f(\cosh y, \sinh y) \geq 0.$$

Combining these two facts, we conclude in particular that

$$\lim_{y \to +\infty} f(\cosh y, \sinh y) > 0. \tag{2-5}$$

On the other hand, as $\phi$ is homogeneous of order $\frac{d-1}{2} - \frac{2}{p-1}$ and $F$ is homogeneous of order $p+1$, we see that the integrand in the definition of $f(t,r)$ is homogeneous of order $2\big(\frac{d-3}{2} - \frac{2}{p-1}\big)$, which is negative by hypothesis. This implies that $f(\cosh y, \sinh y)$ goes to zero as $y \to +\infty$, contradicting (2-5). □

## 3. Eliminating the potential

We now exploit the freedom to select the defocusing potential $F$ by eliminating it from the equations of motion. To motivate this elimination, let us temporarily make the a priori assumption that we have a solution $u$ to (1-3) in the light cone $\Gamma_d$ from Theorem 2.1 that is nowhere vanishing. Taking the inner product of (1-3) with $u$ and using (1-2) then gives an equation for $F(u)$:

$$F(u) = \frac{1}{p+1} \langle u, \Box u \rangle_{\mathbb{R}^m}. \tag{3-1}$$

In particular, since $F$ is defocusing and $u$ is nowhere vanishing, we have the defocusing property

$$\langle u, \Box u \rangle_{\mathbb{R}^m} > 0 \tag{3-2}$$

throughout $\Gamma_d$. Next, if $\partial_\alpha$ denotes one of the $d+1$ derivative operators $\partial_t, \partial_{x_1}, \ldots, \partial_{x_d}$, we have from the chain rule that

$$\partial_\alpha F(u) = \langle \partial_\alpha u, (\nabla F)(u) \rangle_{\mathbb{R}^m}$$

and hence from (1-3) and (3-1) we have the equation

$$\partial_\alpha \langle u, \Box u \rangle_{\mathbb{R}^m} = (p+1) \langle \partial_\alpha u, \Box u \rangle_{\mathbb{R}^m}. \tag{3-3}$$

**Remark 3.1.** One can rewrite the equation (3-3) in the more familiar form

$$\partial^\beta T_{\alpha\beta} = 0,$$

where $T_{\alpha\beta}$ is the stress-energy tensor

$$T_{\alpha\beta} = \langle \partial_\alpha u, \partial_\beta u \rangle_{\mathbb{R}^m} - \eta_{\alpha\beta} \Big( \tfrac{1}{2} \langle \partial^\gamma u, \partial_\gamma u \rangle_{\mathbb{R}^m} + \frac{1}{p+1} \langle u, \Box u \rangle \Big).$$

Now assume that $u$ obeys the discrete self-similarity hypothesis (2-1). Let $\theta := u / \|u\|_{\mathbb{R}^m}$ denote the direction vector of $u$; then $\theta$ is smooth map from $\Gamma_d$ to the unit sphere $S^{m-1} := \{v \in \mathbb{R}^m : \|v\|_{\mathbb{R}^m} = 1\}$ of $\mathbb{R}^m$. From the discrete self-similarity (2-1) we see that $\theta$ is invariant under the dilation action of the multiplicative group $e^{S\mathbb{Z}} := \{e^{nS} : n \in \mathbb{Z}\}$ on $\Gamma_d$. Thus $\theta$ descends to a smooth map $\tilde{\theta} : \Gamma_d / e^{S\mathbb{Z}} \to S^{m-1}$

on the compact quotient $\Gamma_d/e^{S\mathbb{Z}}$, which is a smooth surface with boundary (diffeomorphic to the product of a $d$-dimensional closed ball and a circle). Under some nondegeneracy hypotheses on this map, we can now eliminate the potential $F$, reducing Theorem 2.1 to the following claim:

**Theorem 3.2** (second reduction). *Let $d = 3$, let $p > 1 + \frac{4}{d-2}$, and let*

$$m \geq 2\max\left(\frac{(d+1)(d+6)}{2}, \frac{(d+1)(d+4)}{2} + 5\right) + 2$$

*be an integer. Then there exists $S > 0$ and a smooth nowhere vanishing function $u : \Gamma_d \to \mathbb{R}^m \backslash \{0\}$ which is discretely self-similar in the sense of (2-1) and obeys the defocusing property (3-2) and the equations (3-3) throughout $\Gamma_d$. Furthermore, the map $\tilde{\theta} : \Gamma_d/\lambda^{\mathbb{Z}} \to S^{m-1}$ defined as above is injective, and immersed in the sense that the $d+1$ derivatives $\partial_\alpha \theta(t,x)$ for $\alpha = 0, \ldots, d$ are linearly independent in $\mathbb{R}^m$ for each $(t,x) \in \Gamma_d$.*

Let us assume Theorem 3.2 for now and see how it implies Theorem 2.1. As in the previous section, our arguments here will not depend on our hypotheses on $m$,$d$, $p$.

Since the map $\tilde{\theta} : \Gamma_d/e^{S\mathbb{Z}} \to S^{m-1}$ is assumed to be injective and immersed, it is a smooth embedding of the set $\Gamma_d/e^{S\mathbb{Z}}$ to $S^{m-1}$, so that $\tilde{\theta}(\Gamma_d/e^{S\mathbb{Z}}) = \theta(\Gamma_d)$ is a smooth manifold with boundary contained in $S^{m-1}$. We define a function $F_0 : \theta(\Gamma_d) \to \mathbb{R}$ by the formula

$$F_0\left(\frac{u(t,x)}{\|u(t,x)\|_{\mathbb{R}^m}}\right) := \frac{1}{(p+1)\|u(t,x)\|_{\mathbb{R}^m}^{p+1}} \langle u(t,x), \Box u(t,x)\rangle_{\mathbb{R}^m} \tag{3-4}$$

for any $(t,x) \in \Gamma_d$. As $\theta$ is injective and $u$ is nowhere vanishing and discretely self-similar, one verifies that $F_0$ is well-defined. As the map $\theta$ is immersed, we also see that $F_0$ is smooth. From (3-2) we see that $F_0$ is positive on $\theta(\Gamma_d)$. Intuitively, $F_0$ is going to be our choice for $F$ on the set $\theta(\Gamma_d)$ (this choice is forced upon us by (3-1) and homogeneity).

We define an auxiliary function $T : \theta(\Gamma_d) \to \mathbb{R}^m$ by the formula

$$T\left(\frac{u(t,x)}{\|u(t,x)\|_{\mathbb{R}^m}}\right) := \frac{1}{\|u(t,x)\|_{\mathbb{R}^m}^{p}} \Box u(t,x) - \frac{1}{\|u(t,x)\|_{\mathbb{R}^m}^{p+2}} \langle u(t,x), \Box u(t,x)\rangle_{\mathbb{R}^m} u(t,x) \tag{3-5}$$

for all $(t,x) \in \Gamma_d$; geometrically, this is the orthogonal projection of $(1/\|u\|_{\mathbb{R}^m}^p)\Box u$ to the tangent plane of $S^m$ at $u/\|u\|_{\mathbb{R}^m}$, and will be our choice for the $S^{m-1}$ gradient

$$(\nabla_{S^{m-1}} F)\left(\frac{u}{\|u\|}\right) = (\nabla_{\mathbb{R}^m} F)\left(\frac{u}{\|u\|_{\mathbb{R}^m}}\right) - \left\langle \frac{u}{\|u\|_{\mathbb{R}^m}}, (\nabla_{\mathbb{R}^m} F)\left(\frac{u}{\|u\|}\right)\right\rangle_{\mathbb{R}^m} \frac{u}{\|u\|_{\mathbb{R}^m}}$$

of $F$ at $u/\|u\|_{\mathbb{R}^m}$.

As $\theta$ is injective and $u$ is nowhere vanishing and discretely self-similar, one verifies as before that $T$ is well-defined, and from the immersed nature of $\theta$ we see that $T$ is smooth. Clearly $T(\omega)$ is also orthogonal to $\omega$ for any $\omega \in \theta(\Gamma_d)$. We also claim that $T$ is an extension of the gradient $\nabla_{\theta(\Gamma_d)} F_0$ of $F_0$ on $\theta(\Gamma_d)$, in the sense that

$$\langle v, \nabla_{\theta(\Gamma_d)} F_0(\omega)\rangle_{\mathbb{R}^m} = \langle v, T(\omega)\rangle_{\mathbb{R}^m} \tag{3-6}$$

for any $\omega \in \theta(\Gamma_d)$ and tangent vectors $v \in T_\omega \theta(\Gamma_d)$ to $\theta(\Gamma_d)$ at $\omega$. To verify (3-6), we write

$$\omega = \frac{u(t,x)}{\|u(t,x)\|_{\mathbb{R}^m}} = \frac{u}{\|u\|}$$

for some $(t,x) \in \Gamma_d$; henceforth we suppress the explicit dependence on $(t,x)$ for brevity. The tangent space to $\theta(\Gamma_d)$ at $\omega$ is spanned by $\partial_\alpha(u/\|u\|)$ for $\partial_\alpha = \partial_t, \partial_{x_1}, \dots, \partial_{x_d}$, so it suffices to show that

$$\left\langle \partial_\alpha \frac{u}{\|u\|}, \nabla_{\theta(\Gamma_d)} F_0(\omega) \right\rangle_{\mathbb{R}^m} = \left\langle \partial_\alpha \frac{u}{\|u\|}, T(\omega) \right\rangle_{\mathbb{R}^m}$$

for each $\partial_\alpha$. But from the chain and product rules and (3-4), (3-3) and (3-5) we have

$$\begin{aligned}
\left\langle \partial_\alpha \frac{u}{\|u\|}, \nabla_{\theta(\Gamma_d)} F_0(\omega) \right\rangle_{\mathbb{R}^m} &= \partial_\alpha F_0\left(\frac{u}{\|u\|}\right) \\
&= \frac{1}{p+1} \partial_\alpha \left( \frac{1}{\|u\|_{\mathbb{R}^m}^{p+1}} \langle u, \Box u \rangle_{\mathbb{R}^m} \right) \\
&= -\frac{\langle u, \partial_\alpha u \rangle_{\mathbb{R}^m}}{\|u\|_{\mathbb{R}^m}^{p+3}} \langle u, \Box u \rangle_{\mathbb{R}^m} + \frac{\partial_\alpha \langle u, \Box u \rangle_{\mathbb{R}^m}}{(p+1)\|u\|_{\mathbb{R}^m}^{p+1}} \\
&= -\frac{\langle u, \partial_\alpha u \rangle_{\mathbb{R}^m}}{\|u\|_{\mathbb{R}^m}^{p+3}} \langle u, \Box u \rangle_{\mathbb{R}^m} + \frac{\langle \partial_\alpha u, \Box u \rangle_{\mathbb{R}^m}}{\|u\|_{\mathbb{R}^m}^{p+1}} \\
&= \left\langle \frac{1}{\|u\|_{\mathbb{R}^m}} \partial_\alpha u, T\left(\frac{u}{\|u\|_{\mathbb{R}^m}}\right) \right\rangle_{\mathbb{R}^m} \\
&= \left\langle \partial_\alpha \frac{u}{\|u\|_{\mathbb{R}^m}}, T\left(\frac{u}{\|u\|_{\mathbb{R}^m}}\right) \right\rangle_{\mathbb{R}^m}
\end{aligned}$$

as desired, where in the final line comes from the orthogonality of $T(u/\|u\|_{\mathbb{R}^m})$ with scalar multiples of $u$.

We now claim that we may find an open neighbourhood $U$ of $\theta(\Gamma_d)$ in $S^{m-1}$ and a smooth extension $F_1 : U \to \mathbb{R}$ of $F_0$, with the property that

$$\nabla_{S^{m-1}} F_1(\omega) = T(\omega) \tag{3-7}$$

for all $\omega \in \theta(\Gamma_d)$. Indeed, we can define

$$F_1(\omega + v) := F_0(\omega) + \langle v, T(\omega) \rangle_{\mathbb{R}^m}$$

for all $\omega \in \theta(\Gamma_d)$ and sufficiently small $v \in \mathbb{R}^m$ orthogonal to the tangent space $T_\omega \theta(\Gamma_d / e^{S\mathbb{Z}})$ with $\omega + v \in S^{m-1}$; one can verify that this is well-defined as a smooth extension of $F_0$ to a sufficiently small normal neighbourhood of $\theta(\Gamma_d)$ with the desired gradient property (3-7) (here we use (3-6) to deal with tangential components of the gradient), and one may smoothly extend this to an open neighbourhood of $\theta(\Gamma_d)$ by Seeley's theorem [1964].

Next, if we extend $F_1$ by zero to all of $S^{m-1}$ and define $F_2 : S^{m-1} \to \mathbb{R}$ to be the function $F_2 := \psi F_1 + (1-\psi)$ for some smooth function $\psi : S^{m-1} \to [0,1]$ supported in $U$ that equals 1 on a neighbourhood of $\theta(\Gamma_d)$, then $F_2$ is a smooth extension of $F_0$ to $S^{m-1}$ that is strictly positive, and which also obeys

(3-7). If we then set $F : \mathbb{R}^m \to \mathbb{R}$ to be the function

$$F(\lambda\omega) := \lambda^{p+1} F_2(\omega)$$

for all $\lambda \geq 0$ and $\omega \in S^{m-1}$, then $F$ is a defocusing potential, homogeneous of order $p+1$, which extends $F_0$, and such that

$$\nabla_{S^{m-1}} F(\omega) = T(\omega)$$

for $\omega \in \theta(\Gamma_d)$. By homogeneity (1-1), the radial derivative $\langle \omega, \nabla_{\mathbb{R}^m} F(\omega)\rangle_{\mathbb{R}^m}$ is

$$(p+1)F(\omega) = (p+1)F_0(\omega)$$

for such $\omega$, and hence for $\omega = u/\|u\|$ by (3-5) and (3-4) we have

$$\begin{aligned}\nabla_{\mathbb{R}^m} F(\omega) &= T(\omega) + (p+1)F_0(\omega)\omega \\ &= \frac{1}{\|u\|^p}\Box u - \frac{\langle u, \Box u\rangle}{\|u\|^{p+2}}u + \frac{p+1}{(p+1)\|u\|^{p+1}}\langle u, \Box u\rangle \frac{u}{\|u\|} \\ &= \frac{1}{\|u\|^p}\Box u;\end{aligned}$$

since $\nabla_{\mathbb{R}^m} F$ is homogeneous of order $p$, this gives (1-3) as required.

It remains to establish Theorem 3.2. This will be the focus of the remaining sections of the paper.

## 4. Eliminating the field

Having eliminated the potential $F$ from the problem, the next step is (perhaps surprisingly) to eliminate the unknown field $u$, replacing it with quadratic data such as the mass density

$$M(t,x) := \|u(t,x)\|^2_{\mathbb{R}^m} \tag{4-1}$$

and the energy tensor

$$E_{\alpha\beta}(t,x) := \langle \partial_\alpha u(t,x), \partial_\beta u(t,x)\rangle. \tag{4-2}$$

If $u$ has the discrete self-similarity property (2-1), then $M$ and $E$ similarly obey the discrete self-similarity properties

$$M(e^S t, e^S x) = e^{-\frac{4}{p-1}S} M(t,x) \tag{4-3}$$

and

$$E_{\alpha\beta}(e^S t, e^S x) = e^{-\frac{2(p+1)}{p-1}S} E_{\alpha\beta}(t,x). \tag{4-4}$$

Next, observe from the product rule that

$$\langle u, \Box u\rangle_{\mathbb{R}^m} = \tfrac{1}{2}\Box M - \eta^{\beta\gamma} E_{\beta\gamma}, \tag{4-5}$$

where $\eta$ is the Minkowski metric. Thus, the defocusing property (3-2) can be rewritten as

$$\tfrac{1}{2}\Box M - \eta^{\alpha\beta} E_{\alpha\beta} > 0. \tag{4-6}$$

In a similar spirit, we have

$$\langle \partial_\alpha u, \Box u\rangle = \partial^\beta E_{\alpha\beta} - \tfrac{1}{2}\partial_\alpha(\eta^{\beta\gamma}E_{\beta\gamma})$$

and hence the equation (3-3) can be expressed in terms of $M$ and $E$ as

$$\partial_\alpha\big(\tfrac{1}{2}\Box M - \eta^{\beta\gamma}E_{\beta\gamma}\big) = (p+1)\big(\partial^\beta E_{\alpha\beta} - \tfrac{1}{2}\partial_\alpha(\eta^{\beta\gamma}E_{\beta\gamma})\big). \tag{4-7}$$

Finally, observe that the $(2+d)\times(2+d)$ Gram matrix

$$\begin{pmatrix} \langle u(t,x), u(t,x)\rangle_{\mathbb{R}^m} & \langle u(t,x), \partial_t u(t,x)\rangle_{\mathbb{R}^m} & \cdots & \langle u(t,x), \partial_{x_d} u(t,x)\rangle_{\mathbb{R}^m} \\ \langle \partial_t u(t,x), u(t,x)\rangle_{\mathbb{R}^m} & \langle \partial_t u(t,x), \partial_t u(t,x)\rangle_{\mathbb{R}^m} & \cdots & \langle \partial_t u(t,x), \partial_{x_d} u(t,x)\rangle_{\mathbb{R}^m} \\ \vdots & \vdots & \ddots & \vdots \\ \langle \partial_{x_d} u(t,x), u(t,x)\rangle_{\mathbb{R}^m} & \langle \partial_{x_d} u(t,x), \partial_t u(t,x)\rangle_{\mathbb{R}^m} & \cdots & \langle \partial_{x_d} u(t,x), \partial_{x_d} u(t,x)\rangle_{\mathbb{R}^m} \end{pmatrix} \tag{4-8}$$

can be expressed in terms of $E, M$ as

$$\begin{pmatrix} M(t,x) & \frac{1}{2}\partial_t M(t,x) & \cdots & \frac{1}{2}\partial_{x_d} M(t,x) \\ \frac{1}{2}\partial_t M(t,x) & E_{00}(t,x) & \cdots & E_{0d}(t,x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}\partial_{x_d} M(t,x) & E_{d0}(t,x) & \cdots & E_{dd}(t,x) \end{pmatrix}. \tag{4-9}$$

In particular, the matrix (4-9) is positive semidefinite for every $t, x$.

It turns out that with the aid of the Nash embedding theorem and our hypothesis that $m$ is large, we can largely reverse the above observations, reducing Theorem 3.2 to the following claim that no longer directly involves the field $u$ (or the range dimension $m$).

**Theorem 4.1** (third reduction). *Let $d = 3$, and let $p > 1 + \frac{4}{d-2}$. Then there exists $S > 0$ and smooth functions $M : \Gamma_d \to \mathbb{R}$ and $E_{\alpha\beta} : \Gamma_d \to \mathbb{R}$ for $\alpha, \beta = 0, \dots, d$ which are discretely self-similar in the sense of (4-3) and (4-4), obey the defocusing property (4-6) and the equation (4-7) on $\Gamma_d$ for all $\alpha = 0, \dots, d$, and such that the matrix (4-9) is strictly positive definite on $\Gamma_d$ (in particular, this forces $M$ to be strictly positive).*

Let us assume Theorem 4.1 for the moment and show Theorem 3.2. Let $d$, $p$, $S$, $M$, $E_{\alpha\beta}$ be as in Theorem 4.1, and let $m$ be as in Theorem 3.2. Our task is to obtain a function $u : \Gamma_d \to \mathbb{R}^m\backslash\{0\}$ obeying all the properties claimed in Theorem 3.2.

The idea is to build $u$ in such a fashion that (4-1) and (4-2) are obeyed. Accordingly, we will use an ansatz

$$u(t,x) := M(t,x)^{\frac{1}{2}}\theta(t,x) \tag{4-10}$$

for some smooth $\theta : \Gamma_d \to S^{m-1}$ to be constructed shortly. As $M$ is strictly positive, such a function $u$ will be smooth on $\Gamma$ and obey (4-1); differentiating, we see that

$$\langle u, \partial_\alpha u\rangle_{\mathbb{R}^m} = \tfrac{1}{2}\partial_\alpha M \tag{4-11}$$

for $\alpha = 0, \dots, d$. If $\theta$ obeys the discrete self-similarity property

$$\theta(e^S t, e^S x) = \theta(t,x) \tag{4-12}$$

then $u$ will obey (2-1). Thus we shall impose (4-12); that is to say we assume that $\theta$ is lifted from a smooth map $\tilde{\theta} : \Gamma_d/e^{S\mathbb{Z}} \to S^{m-1}$.

From the product rule, (4-1) and (4-11) we have (after some calculation)

$$\langle \partial_\alpha \theta, \partial_\beta \theta \rangle_{\mathbb{R}^m} = M^{-1} \langle \partial_\alpha u, \partial_\beta u \rangle_{\mathbb{R}^m} - M^{-2} (\partial_\alpha M) \partial_\beta M.$$

Thus, if we wish for (4-2) to be obeyed, then the $(1+d) \times (1+d)$ Gram matrix

$$(\langle \partial_\alpha \theta, \partial_\beta \theta \rangle_{\mathbb{R}^m})_{\alpha,\beta=0,\dots,d}$$

must be equal to

$$M \big( E_{\alpha\beta} - (\partial_\alpha M) M^{-1} \partial_\beta M \big)_{\alpha,\beta=0,\dots,d}. \tag{4-13}$$

The matrix in (4-13) is a Schur complement of the matrix in (4-9). Since the matrix in (4-9) is assumed to be strictly positive definite, we conclude that (4-13) is also.

If we denote the matrix in (4-13) by $g(t,x)$, then from (4-3) and (4-4) we have the discrete self-similarity property

$$g(e^S t, e^S x) = e^{-2S} g(t,x). \tag{4-14}$$

As $g$ is a positive definite and symmetric $(1+d)\times(1+d)$ matrix, we can view $g$ as a smooth Riemannian metric on $\Gamma_d$. Given that the dilation operator $(t,x) \mapsto (e^S t, e^S x)$ dilates tangent vectors to $\Gamma_d$ by a factor of $e^S$, we see that the metric $g$ is lifted from a smooth Riemannian metric $\tilde{g}$ on the quotient space $\Gamma_d/e^{S\mathbb{Z}}$.

The space $(\Gamma_d/e^{S\mathbb{Z}}, \tilde{g})$ is a smooth compact $(1+d)$-dimensional Riemannian manifold with boundary; it is easy to embed it in a smooth compact $(1+d)$-dimensional Riemannian manifold without boundary (for instance by using the theorems in [Seeley 1964]). Applying the Nash embedding theorem (for instance in the form in [Günther 1991]), we can thus isometrically embed $(\Gamma_d/e^{S\mathbb{Z}}, \tilde{g})$ in a Euclidean space $\mathbb{R}^D$ with

$$D := \max\left( \frac{(d+1)(d+6)}{2}, \frac{(d+1)(d+4)}{2} + 5 \right).$$

The embedded copy of $(\Gamma_d/e^{S\mathbb{Z}}, \tilde{g})$ is compact and is thus contained in a cube $[-R,R]^D$ for some finite $R$. We use a generic[3] linear isometry from $\mathbb{R}^D$ to $\mathbb{R}^{D+1}$ to embed $[-R,R]^D$ to some compact subset of $\mathbb{R}^{D+1}$. The image of this isometry is a generic hyperplane, which can be chosen to avoid the lattice $(1/\sqrt{2D+2})\mathbb{Z}^{D+1}$, and thus we can embed $[-R,R]^D$ isometrically into the torus $\mathbb{R}^{D+1}/\big((1/\sqrt{D+1})\mathbb{Z}^{D+1}\big)$, which is isometric to $(1/\sqrt{D+1})(S^1)^{D+1}$. But from Pythagoras' theorem, $(1/\sqrt{D+1})(S^1)^{D+1}$ is contained in $S^{2D+1}$, which is in turn contained in $S^{m-1}$ by the largeness hypothesis on $m$. Thus we have an isometric embedding $\tilde{\theta} : \Gamma_d/e^{S\mathbb{Z}} \to S^{m-1}$ from $(\Gamma_d/e^{S\mathbb{Z}}, \tilde{g})$ into the round sphere $S^{m-1}$. In particular, $\tilde{\theta}$ is injective and immersed, and lifting $\tilde{\theta}$ back to $\Gamma_d$, we obtain a smooth map $\theta : \Gamma_d \to S^{m-1}$ with Gram matrix (4-13) that is discretely self-similar in the sense of (4-12), so that the function $u$ defined by (4-10) obeys (2-1). Reversing the calculations that led to (4-13), we

[3] We thank Marc Nardmann for this argument, which improved the value of $m$ from our previous argument by a factor of approximately two.

see that the Gram matrix (4-8) of $u$ is given by (4-9). In particular, (4-2) holds. Reversing the derivation of (4-6), we now obtain (3-2), while from reversing the derivation of (4-7), we obtain (3-3). We have now obtained all the required properties claimed by Theorem 3.2, as desired.

It remains to establish Theorem 4.1. This will be the focus of the remaining sections of the paper.

## 5. Reduction to a self-similar (1+1)-dimensional problem

In reducing Theorem 1.1 to Theorem 4.1, we have achieved the somewhat remarkable feat of converting a nonlinear PDE problem to a convex (or positive semidefinite) PDE problem, in that all of the constraints[4] on the remaining unknowns $M$, $E_{\alpha\beta}$ are linear equalities and inequalities, or assertions that certain matrices are positive definite. Among other things, this shows that if one has a given solution $M$, $E_{\alpha\beta}$ to Theorem 4.1, and then one averages that solution over some compact symmetry group that acts on the space of such solutions, then the average will also be a solution to Theorem 4.1. In particular, one can then reduce without any loss of generality to considering solutions that are invariant with respect to that symmetry.

For instance, given that $M$, $E_{\alpha\beta}$ are already discretely self-similar by (4-3) and (4-4), the space of solutions has an action of the compact dilation group $\mathbb{R}^+/e^{S\mathbb{Z}}$, with (the quotient representative of) any real number $\lambda > 0$ acting on $M$, $E_{\alpha\beta}$ by the action

$$(\lambda \cdot M)(t,x) := \frac{1}{\lambda^{\frac{4}{p-1}}} M\Big(\frac{t}{\lambda}, \frac{x}{\lambda}\Big)$$

and

$$(\lambda \cdot E_{\alpha\beta})(t,x) := \frac{1}{\lambda^{\frac{2(p+1)}{p-1}}} E_{\alpha\beta}\Big(\frac{t}{\lambda}, \frac{x}{\lambda}\Big);$$

this is initially an action of the multiplicative group $\mathbb{R}^+$, but descends to an action of $\mathbb{R}^+/e^{S\mathbb{Z}}$ thanks to (4-3) and (4-4). By the preceding discussion, we may restrict without loss of generality to the case when $M$, $E_{\alpha\beta}$ are invariant with respect to this $\mathbb{R}^+/e^{S\mathbb{Z}}$, or equivalently that $M$, $E_{\alpha\beta}$ are homogeneous of order $-\frac{4}{p-1}$ and $\frac{2(p+1)}{p-1}$ respectively. With this restriction, the parameter $S$ no longer plays a role and may be discarded.

**Remark 5.1.** This reduction may seem at first glance to be in conflict with the negative result in Proposition 2.2. However, the requirement that the mass density $M$ and the energy tensor $E_{\alpha\beta}$ be homogeneous is strictly weaker than the hypothesis that the field $u$ itself is homogeneous. For instance, one could imagine a "twisted self-similar" solution in which the homogeneity condition (1-1) on $u$ is replaced with a more general condition of the form

$$u(\lambda t, \lambda x) = \lambda^{-\frac{2}{p-1}} \exp(J \log \lambda) u(t,x)$$

for all $(t,x) \in \Gamma_d$ and $\lambda > 0$, where $J : \mathbb{R}^m \to \mathbb{R}^m$ is a fixed skew-adjoint linear transformation. (To be compatible with (1-3), one would also wish to require that the potential $F$ is invariant with respect to the orthogonal transformations $\exp(sJ)$ for $s \in \mathbb{R}$.) Such solutions $u$ would not be homogeneous, but the associated densities $M$, $E_{\alpha\beta}$ would still be homogeneous of the order specified above.

[4] Compare with the "kernel trick" in machine learning, or with semidefinite relaxation in optimization.

We may similarly apply the above reductions to the orthogonal group $O(d)$, which acts on the scalar field $M$ and on the 2-tensor $E_{\alpha\beta}$ in the usual fashion; thus

$$(UM)(t,x) := M(t,U^{-1}x)$$

and

$$(UE)_{\alpha\beta}(t,x)(Uv)^\alpha(Uv)^\beta = E_{\alpha\beta}(t,U^{-1}x)v^\alpha v^\beta$$

for all $(t,x)\in\Gamma_d$, $U\in O(d)$, and $v\in\mathbb{R}^{1+d}$, where $U$ acts on $\mathbb{R}^{1+d}$ by $(t,x)\mapsto(t,Ux)$. This allows us to reduce to fields $M, E_{\alpha\beta}$ which are $O(d)$-invariant; thus $M$ is spherically symmetric, and $E_{\alpha\beta}$ takes the form[5]

$$E_{00} = E_{tt}, \tag{5-1}$$

$$E_{0i} = E_{i0} = \frac{x_i}{r}E_{tr}, \tag{5-2}$$

$$E_{ij} = \frac{x_i x_j}{r^2}(E_{rr}-E_{\omega\omega}) + \delta_{ij}E_{\omega\omega} \tag{5-3}$$

for $i,j=1,\dots,d$ and some spherically symmetric scalar functions $E_{tt}, E_{tr}, E_{rr}, E_{\omega\omega}$, where $r := |x|$ is the radial variable and $\delta_{ij}$ is the Kronecker delta. Observe that if $E_{tt}, E_{rr}, E_{\omega\omega}:\Gamma_1\to\mathbb{R}$ are smooth even functions and $E_{tr}:\Gamma_1\to\mathbb{R}$ is a smooth odd function on the $(1+1)$-dimensional light cone

$$\Gamma_1 := \{(t,r)\in\mathbb{R}^{1+1} : t>0; -t\le r\le t\}$$

with $E_{rr}-E_{\omega\omega}$ vanishing to second order at $r=0$, then the above equations define a smooth field $E_{\alpha\beta}$ on $\Gamma_d$, which will be homogeneous of order $-\frac{2(p+1)}{p-1}$ if $E_{tt}, E_{tr}, E_{rr}, E_{\omega\omega}$ are.

Using polar coordinates, we have

$$\frac{1}{2}\Box M - \eta^{\beta\gamma}E_{\beta\gamma} = \frac{1}{2}\Big(-\partial_{tt}M + \partial_{rr}M + \frac{d-1}{r}M\Big) - (-E_{tt}+E_{rr}+(d-1)E_{\omega\omega});$$

thus the condition (4-6) is now

$$\frac{1}{2}\Big(-\partial_{tt}M + \partial_{rr}M + \frac{d-1}{r}M\Big) - \big(-E_{tt}+E_{rr}+(d-1)E_{\omega\omega}\big) > 0. \tag{5-4}$$

By rotating $x$ to be of the form $x=re_1$, we see that the matrix (4-9) is conjugate to

$$\begin{pmatrix} M & \frac{1}{2}\partial_t M & \frac{1}{2}\partial_r M & 0 & \cdots & 0\\ \frac{1}{2}\partial_t M & E_{tt} & E_{tr} & 0 & \cdots & 0\\ \frac{1}{2}\partial_r M & E_{tr} & E_{rr} & 0 & \cdots & 0\\ 0 & 0 & 0 & E_{\omega\omega} & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & 0 & \cdots & E_{\omega\omega}\end{pmatrix}$$

[5] To see that $E_{\alpha\beta}$ must be of this form, rotate the spatial variable $x$ to equal $x=re_1$, then use the orthogonal transformation $(x_1,x_2,\dots,x_d)\mapsto(x_1,-x_2,\dots,-x_d)$, which preserves $re_1$, to see that $E_{0i}=E_{1i}=0$ for all $i=2,\dots,d$; further use of orthogonal transformations preserving $re_1$ can be then used to show that $E_{ij}=0$ and $E_{ii}=E_{jj}$ for $2\le i<j\le d$ (basically because the only matrices that commute with all orthogonal transformations are scalar multiples of the identity). This places $E_{\alpha\beta}$ in the desired form in the $x=re_1$ case, and the general case follows from rotation.

so the positive definiteness of (4-9) is equivalent to the positive definiteness of the $3\times 3$ matrix

$$\begin{pmatrix} M & \frac{1}{2}\partial_t M & \frac{1}{2}\partial_r M \\ \frac{1}{2}\partial_t M & E_{tt} & E_{tr} \\ \frac{1}{2}\partial_r M & E_{tr} & E_{rr} \end{pmatrix} \tag{5-5}$$

together with the positivity of $E_{\omega\omega}$. It will be convenient to isolate the $r=0$ case of this condition (in order to degenerate $E_{rr}$ to zero at $r=0$ later in the argument). In this case, the odd functions $\partial_r M$ and $E_{tr}$ vanish, and $E_{rr}$ is equal to $E_{\omega\omega}$, so the condition reduces to the positive definiteness of the $2\times 2$ matrix

$$\begin{pmatrix} M & \frac{1}{2}\partial_t M \\ \frac{1}{2}\partial_t M & E_{tt} \end{pmatrix} \tag{5-6}$$

together with the aforementioned positivity of $E_{\omega\omega}$.

Finally, we turn to the condition (4-7). Again, we can rotate the position $x$ to be of the form $x = re_1$. In the angular cases $\alpha = 2,\dots,d$, both sides of (4-7) automatically vanish, basically because $\partial_\alpha f(re_1) = 0$ for any spherically symmetric $f$ (and because $E_{\alpha\beta}$ vanishes to second order for any $\beta\neq\alpha$). So the only nontrivial cases of (4-7) are $\alpha = 0$ and $\alpha = 1$. Applying (5-1), (5-2), and (5-3), we can write these cases of (4-7) as

$$\begin{aligned}\partial_t\Big[\frac{1}{2}\Big(-\partial_{tt}M + \partial_{rr}M + \frac{d-1}{r}M\Big) - \big(-E_{tt}+E_{rr}+(d-1)E_{\omega\omega}\big)\Big] \\ = (p+1)\Big[-\partial_t E_{tt} + \partial_r E_{tr} + \frac{d-1}{r}E_{tr} - \frac{1}{2}\partial_t\big(-E_{tt}+E_{rr}+(d-1)E_{\omega\omega}\big)\Big]\end{aligned} \tag{5-7}$$

and

$$\begin{aligned}\partial_r\Big[\frac{1}{2}\Big(-\partial_{tt}M + \partial_{rr}M + \frac{d-1}{r}M\Big) - \big(-E_{tt}+E_{rr}+(d-1)E_{\omega\omega}\big)\Big] \\ = (p+1)\Big[-\partial_t E_{tr} + \partial_r E_{rr} + \frac{d-1}{r}(E_{rr}-E_{\omega\omega}) - \frac{1}{2}\partial_r\big(-E_{tt}+E_{rr}+(d-1)E_{\omega\omega}\big)\Big]\end{aligned} \tag{5-8}$$

respectively.

To summarise, we have reduced Theorem 4.1 to

**Theorem 5.2** (fourth reduction). *Let $d=3$, and let $p > 1+\frac{4}{d-2}$. Then there exist smooth even functions $M$, $E_{tt}$, $E_{rr}$, $E_{\omega\omega}:\Gamma_1\to\mathbb{R}$ and a smooth odd function $E_{tr}:\Gamma_1\to\mathbb{R}$, with $M$ homogeneous of order $-\frac{4}{p-1}$ and $E_{tt}$, $E_{tr}$, $E_{rr}$, $E_{\omega\omega}$ homogeneous of order $-\frac{2(p+1)}{p-1}$, and with $E_{rr}-E_{\omega\omega}$ vanishes to second order at $r=0$, obeying the defocusing property (5-4) and the equations (5-7) and (5-8) on $\Gamma_1$, such that*

$$E_{\omega\omega} > 0 \tag{5-9}$$

*and the $3\times 3$ matrix (5-5) is strictly positive definite on $\Gamma_1$ with $r\neq 0$, and the $2\times 2$ matrix (5-6) is positive definite when $r=0$.*

It remains to prove Theorem 5.2. To do so, we make a few technical relaxations. Firstly, we claim that we may relax the strict conditions (5-4) and (5-9) to their nonstrict counterparts

$$\frac{1}{2}\Big(-\partial_{tt}M + \partial_{rr}M + \frac{d-1}{r}M\Big) - (-E_{tt} + E_{rr} + (d-1)E_{\omega\omega}) \geq 0 \tag{5-10}$$

and

$$E_{\omega\omega} \geq 0. \tag{5-11}$$

To see this, suppose that $M, E_{tt}, E_{tr}, E_{rr}, E_{\omega\omega}$ obey the conclusions of Theorem 5.2 with the conditions (5-4), (5-9) replaced by (5-10), (5-11). We let $\varepsilon > 0$ be a small quantity to be chosen later, and define new fields $M^\varepsilon, E^\varepsilon_{tt}, E^\varepsilon_{tr}, E^\varepsilon_{rr}, E^\varepsilon_{\omega\omega}$ by the formulae[6]

$$\begin{aligned}
M^\varepsilon &:= M - c\varepsilon t^{-\frac{4}{p-1}},\\
E^\varepsilon_{tt} &:= E_{tt} - (d+1)\varepsilon t^{-\frac{2(p+1)}{p-1}},\\
E^\varepsilon_{tr} &:= E_{tr},\\
E^\varepsilon_{rr} &:= E_{rr} + \varepsilon t^{-\frac{2(p+1)}{p-1}},\\
E^\varepsilon_{\omega\omega} &:= E_{\omega\omega} + \varepsilon t^{-\frac{2(p+1)}{p-1}},
\end{aligned}$$

where $c$ is the constant such that

$$\frac{c}{2}\frac{4}{p-1}\frac{p+3}{p-1} - (2d+1) = \frac{p+1}{2}.$$

Clearly these new fields $M^\varepsilon, E^\varepsilon_{tt}, E^\varepsilon_{tr}, E^\varepsilon_{rr}, E^\varepsilon_{\omega\omega}$ are still smooth, with $M^\varepsilon, E^\varepsilon_{tt}, E^\varepsilon_{rr}, E^\varepsilon_{\omega\omega}$ even and $E^\varepsilon_{tr}$ odd, with $M^\varepsilon$ homogeneous of order $-\frac{4}{p-1}$ and $E^\varepsilon_{tt}, E^\varepsilon_{tr}, E^\varepsilon_{rr}, E^\varepsilon_{\omega\omega}$ homogeneous of order $-\frac{2(p+1)}{p-1}$, with $E^\varepsilon_{rr} - E^\varepsilon_{\omega\omega}$ vanishing to second order at $r = 0$. A calculation using the definition of $c$ shows that the equations (5-7) and (5-8) continue to be obeyed when the fields $M, E_{tt}, E_{tr}, E_{rr}, E_{\omega\omega}$ are replaced by $M^\varepsilon, E^\varepsilon_{tt}, E^\varepsilon_{tr}, E^\varepsilon_{rr}, E^\varepsilon_{\omega\omega}$. With this replacement, the left-hand side of (5-10) increases by

$$\frac{p+1}{2}\varepsilon t^{-\frac{2(p+1)}{p-1}},$$

and so (5-4) now holds. The remaining task is to show that with these new fields $M^\varepsilon, E^\varepsilon_{tt}, E^\varepsilon_{tr}, E^\varepsilon_{rr}, E^\varepsilon_{\omega\omega}$, (5-5) is positive definite when $r \neq 0$ and (5-6) is positive definite when $r = 0$. By the scale invariance it suffices to verify these latter properties when $t = 1$. The positive definiteness of (5-6) when $r = 0$ then follows by continuity for $\varepsilon$ small enough. For (5-5), we have to take a little care because the condition $r \neq 0$ is noncompact. We need to ensure the positive definiteness of

$$\begin{pmatrix} M - c\varepsilon & \frac{1}{2}\partial_t M + \frac{2c}{p-1}\varepsilon & \frac{1}{2}\partial_r M \\ \frac{1}{2}\partial_t M + \frac{2c}{p-1}\varepsilon & E_{tt} - (d+1)\varepsilon & E_{tr} \\ \frac{1}{2}\partial_r M & E_{tr} & E_{rr} + \varepsilon \end{pmatrix}$$

[6]The ability to freely manipulate the fields $M, E_{tt}, E_{tr}, E_{rr}, E_{\omega\omega}$ in this fashion is a major advantage of the formulation of Theorem 5.2. It would be very difficult to perform analogous manipulations if the original field $u$ or the potential $F$ were still present.

when $t = 1$ and $r \neq 0$ for $\varepsilon$ small enough. Continuity will ensure this if $|r|$ is bounded away from zero (independently of $\varepsilon$), so we may assume that $r$ is in a small neighbourhood of the origin (independent of $\varepsilon$). Given that the above matrix is already positive definite when $\varepsilon = 0$, it suffices by a continuity argument to show that the above matrix has positive determinant for sufficiently small $\varepsilon$; by the hypothesis (5-5) and the fundamental theorem of calculus, it thus suffices to show that

$$\frac{d}{d\varepsilon} \det \begin{pmatrix} M - c\varepsilon & \frac{1}{2}\partial_t M + \frac{2c}{p-1}\varepsilon & \frac{1}{2}\partial_r M \\ \frac{1}{2}\partial_t M + \frac{2c}{p-1}\varepsilon & E_{tt} - (d+1)\varepsilon & E_{tr} \\ \frac{1}{2}\partial_r M & E_{tr} & E_{rr} + \varepsilon \end{pmatrix} > 0$$

for $r$ near zero and sufficiently small $\varepsilon$. But since $\partial_r M, E_{tr}, E_{rr}$ vanish at $r = 0$, we can use cofactor expansion to write the left-hand side as

$$\begin{pmatrix} M(1,0) & \frac{1}{2}\partial_t M(1,0) \\ \frac{1}{2}\partial_t M(1,0) & E_{tt}(1,0) \end{pmatrix} + O(|r|) + O(\varepsilon)$$

and the claim then follows from the hypothesis (5-5). This concludes the relaxation of the conditions (5-4), (5-9) to (5-10), (5-11).

Now that we allow equality in (5-11), we sacrifice some generality by restricting to the special case $E_{\omega\omega} = 0$ (which basically corresponds to considering spherically symmetric blowup solutions). While this gives up some flexibility, this will simplify our calculations a bit as we now only have four fields $M$, $E_{tt}$, $E_{tr}$, $E_{rr}$ to deal with, rather than five.

Until now we have avoided using the hypothesis $d = 3$. Now we will embrace this hypothesis. In Proposition 2.2 it was convenient to make the change of variables $\phi = r^{\frac{d-1}{2}} u = ru$ to eliminate lower-order terms such as $\left(\frac{d-1}{r}\right)\partial_r u$; this change of variables is particularly pleasant in the three-dimensional case as the lower-order term involving the coefficient $\frac{1}{4}(d-1)(d-3)$ vanishes completely (this vanishing is closely tied to the strong Huygens principle in three dimensions). The corresponding change of variables in this setting, aimed at eliminating the lower-order terms $\left(\frac{d-1}{r}\right)E_{tr}$ and $\left(\frac{d-1}{r}\right)E_{rr}$ in (5-7) and (5-8), is to replace the fields $M$, $E_{tt}$, $E_{tr}$, $E_{rr}$ by the fields $\widetilde{M}$, $\widetilde{E}_{tt}$, $\widetilde{E}_{tr}$, $\widetilde{E}_{rr} : \Gamma_1 \to \mathbb{R}^+$ defined by

$$\begin{aligned} \widetilde{M} &:= r^2 M, \\ \widetilde{E}_{tt} &:= r^2 E_{tt}, \\ \widetilde{E}_{tr} &:= r^2 E_{tr} + \tfrac{1}{2} r \partial_t M = r^2 E_{tr} + \frac{1}{2r}\partial_t \widetilde{M}, \\ \widetilde{E}_{rr} &:= r^2 E_{rr} + r\partial_r M + M = r^2 E_{rr} + \frac{1}{r}\partial_r \widetilde{M} - \frac{1}{r^2}\widetilde{M}. \end{aligned}$$

Observe that if $\widetilde{M}$, $\widetilde{E}_{tt}$, $\widetilde{E}_{rr}$ are smooth and even, and $\widetilde{E}_{tr}$ is odd, with $\widetilde{M}$, $\widetilde{E}_{tt}$ vanishing to second order at $r = 0$,

$$\widetilde{E}_{tr} - \frac{1}{2r}\partial_t \widetilde{M}$$

vanishing to third order, and

$$\widetilde{E}_{rr} - \frac{1}{r}\partial_r \widetilde{M} + \frac{1}{r^2}\widetilde{M}$$

to fourth order, then these fields determine smooth fields $M$, $E_{tt}$, $E_{tr}$, $E_{rr}$ with $M$, $E_{tt}$, $E_{rr}$ even, $E_{tr}$ odd, and $E_{rr}$ vanishing to second order at $r = 0$. Furthermore, if $\widetilde{M}$ is homogeneous of order $\frac{2p-6}{p-1}$ and $\widetilde{E}_{tt}, \widetilde{E}_{tr}, \widetilde{E}_{rr}$ are homogeneous of order $-\frac{4}{p-1}$, then we know $M$ will be homogeneous of order $-\frac{4}{p-1}$ and $E_{tt}$, $E_{tr}$, $E_{rr}$ will be homogeneous of order $-\frac{2(p+1)}{p-1}$.

If we introduce the quantity

$$V := \frac{1}{p+1}\Big(\frac{1}{2}(-\partial_{tt}\widetilde{M} + \partial_{rr}\widetilde{M}) + \widetilde{E}_{tt} - \widetilde{E}_{rr}\Big) \tag{5-12}$$

then a brief calculation shows that

$$V = \frac{r^2}{2(p+1)}\Big(\Big(-\partial_{tt}M + \partial_{rr}M + \frac{2}{r}M\Big) - (-E_{tt} + E_{rr})\Big)$$

and so the condition (5-10) is equivalent to

$$V \geq 0. \tag{5-13}$$

The equations (5-7) and (5-8) can now be expressed as

$$\partial_t\Big[\frac{1}{r^2}V\Big] = -\partial_t E_{tt} + \partial_r E_{tr} + \frac{2}{r}E_{tr} - \frac{1}{2}\partial_t(-E_{tt} + E_{rr})$$

and

$$\partial_r\Big[\frac{1}{r^2}V\Big] = -\partial_t E_{tr} + \partial_r E_{rr} + \frac{2}{r}E_{rr} - \frac{1}{2}\partial_r(-E_{tt} + E_{rr}),$$

which rearrange as an energy conservation law

$$\partial_t\Big(\frac{1}{2}E_{tt} + \frac{1}{2}E_{rr} + \frac{1}{r^2}V\Big) = \partial_r E_{tr} + \frac{2}{r}E_{tr}$$

and a momentum conservation law

$$\partial_t E_{tr} = \partial_r\Big(\frac{1}{2}E_{tt} + \frac{1}{2}E_{rr} - \frac{1}{r^2}V\Big) + \frac{2}{r}E_{rr};$$

multiplying these equations by $r^2$ and writing $E_{tt}, E_{tr}, E_{rr}$ in terms of $\widetilde{E}_{tt}, \widetilde{E}_{tr}, \widetilde{E}_{rr}$ and $\widetilde{M}$ one obtains (after some calculation, as well as (5-12) in the case of (5-15)) the slightly simpler equations

$$\partial_t\big(\tfrac{1}{2}\widetilde{E}_{tt} + \tfrac{1}{2}\widetilde{E}_{rr} + V\big) = \partial_r\widetilde{E}_{tr} \tag{5-14}$$

and

$$\partial_t\widetilde{E}_{tr} = \partial_r\big(\tfrac{1}{2}\widetilde{E}_{tt} + \tfrac{1}{2}\widetilde{E}_{rr} - V\big) - \frac{p-1}{r}V. \tag{5-15}$$

The expressions in (5-14) are even, while the expressions in (5-15) are odd. Thus we may combine these equations into a single equation by adding them together, which after some rearranging becomes the transport-type equation

$$(\partial_t - \partial_r)e_+ + (\partial_t + \partial_r)V = -\frac{p-1}{r}V, \tag{5-16}$$

where $e_+$ is the null energy density

$$e_+ := \tfrac{1}{2}\widetilde{E}_{tt} + \tfrac{1}{2}\widetilde{E}_{rr} + \widetilde{E}_{tr}. \tag{5-17}$$

**Remark 5.3.** It may be instructive to derive these equations in the specific context of a solution $u : \Gamma_3 \to \mathbb{R}$ to the scalar defocusing NLW

$$\Box u = |u|^{p-1}u,$$

which in polar coordinates becomes

$$-\partial_{tt}u + \partial_{rr}u + \frac{2}{r}u = |u|^{p-1}u.$$

Making the change of variables $\phi = ru$, this becomes

$$-\partial_{tt}\phi + \partial_{rr}\phi = \frac{|\phi|^{p-1}\phi}{r^{p-1}}.$$

Introducing the null energy

$$e_+ := \tfrac{1}{2}|\partial_t\phi + \partial_r\phi|^2$$

and the potential energy

$$V := \frac{1}{p+1}\frac{|\phi|^{p+1}}{r^{p-1}}$$

as well as the additional densities

$$\widetilde{M} := |\phi|^2, \quad \widetilde{E}_{tt} := |\partial_t\phi|^2, \quad \widetilde{E}_{rr} := |\partial_r\phi|^2, \quad \widetilde{E}_{tr} := \partial_t\phi\partial_r\phi,$$

one can readily verify the identities (5-12), (5-16), and (5-17). It is similar for the other properties of $\widetilde{M}$, $\widetilde{E}_{tt}$, $\widetilde{E}_{rr}$, $\widetilde{E}_{tr}$ identified in this section.

Finally, we translate the positive definiteness of (5-5) (when $r \neq 0$) and (5-6) (when $r = 0$) into conditions involving the fields $\widetilde{M}$, $\widetilde{E}_{tt}$, $\widetilde{E}_{rr}$, $\widetilde{E}_{tr}$. From the identity

$$\begin{pmatrix} \widetilde{M} & \frac{1}{2}\partial_t\widetilde{M} & \frac{1}{2}\partial_r\widetilde{M} \\ \frac{1}{2}\partial_t\widetilde{M} & \widetilde{E}_{tt} & \widetilde{E}_{tr} \\ \frac{1}{2}\partial_r\widetilde{M} & \widetilde{E}_{tr} & \widetilde{E}_{rr} \end{pmatrix} = r^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{r} & 0 & 1 \end{pmatrix} \begin{pmatrix} M & \frac{1}{2}\partial_t M & \frac{1}{2}\partial_r M \\ \frac{1}{2}\partial_t M & E_{tt} & E_{tr} \\ \frac{1}{2}\partial_r M & E_{tr} & E_{rr} \end{pmatrix} \begin{pmatrix} 1 & 0 & \frac{1}{r} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

we see (for $r \neq 0$) that (5-5) is strictly positive definite if and only if the matrix

$$\begin{pmatrix} \widetilde{M} & \frac{1}{2}\partial_t\widetilde{M} & \frac{1}{2}\partial_r\widetilde{M} \\ \frac{1}{2}\partial_t\widetilde{M} & \widetilde{E}_{tt} & \widetilde{E}_{tr} \\ \frac{1}{2}\partial_r\widetilde{M} & \widetilde{E}_{tr} & \widetilde{E}_{rr} \end{pmatrix} \tag{5-18}$$

is strictly positive definite. Now we turn to (5-6) when $r = 0$. By homogeneity, it suffices to verify this condition when $(t, r) = (1, 0)$. From (1-1), we have $\partial_t M(1,0) = -\frac{4}{p-1}M(1,0)$, so the positive definiteness of (5-6) is equivalent to the condition

$$E_{tt}(1,0) > \Big(\frac{2}{p-1}\Big)^2 M(1,0) > 0,$$

which in terms of $\widetilde{E}_{tt}, \widetilde{M}$ becomes

$$\partial_{rr}\widetilde{E}_{tt}(1,0) > \Big(\frac{2}{p-1}\Big)^2 \partial_{rr}\widetilde{M}(1,0) > 0. \tag{5-19}$$

Summarising the above discussion, we now see that Theorem 5.2 is a consequence of the following:

**Theorem 5.4** (fifth reduction). *Let $p > 5$. Then there exist smooth even functions $\widetilde{M}, \widetilde{E}_{tt}, \widetilde{E}_{rr} : \Gamma_1 \to \mathbb{R}$ and a smooth odd function $\widetilde{E}_{tr} : \Gamma_1 \to \mathbb{R}$, with $\widetilde{M}$ homogeneous of order $\frac{2p-6}{p-1}$ and $\widetilde{E}_{tt}$, $\widetilde{E}_{tr}$, $\widetilde{E}_{rr}$ homogeneous of order $-\frac{4}{p-1}$, with $\widetilde{M}$, $\widetilde{E}_{tt}$ vanishing to second order at $r = 0$,*

$$\widetilde{E}_{tr} - \frac{1}{2r}\partial_t \widetilde{M}$$

*vanishing to third order, and*

$$\widetilde{E}_{rr} - \frac{1}{r}\partial_r \widetilde{M} + \frac{1}{r^2}\widetilde{M}$$

*to fourth order. Furthermore, if one defines the fields $V, e_+ : \Gamma_1 \to \mathbb{R}$ by (5-12) and (5-17), we have the weak defocusing property (5-13) and the null transport equation (5-16). Finally, the matrix (5-18) is strictly positive definite for $r \neq 0$, and for $r = 0$ one has the condition (5-19).*

It remains to establish Theorem 5.4. This will be the focus of the final section of the paper.

## 6. Constructing the mass and energy fields

Fix $p > 5$. We will need a large constant $A > 1$ depending only on $p$, and then sufficiently small parameter $\delta > 0$ (depending on $p, A$) to be chosen later. We use the notation $X \lesssim Y$, $Y \gtrsim X$, or $X = O(Y)$ to denote an estimate of the form $|X| \leq CY$, where $C$ can depend on $p$ but is independent of $\delta, A$.

We need to construct smooth fields $\widetilde{M}, \widetilde{E}_{tt}, \widetilde{E}_{rr}, \widetilde{E}_{tr} : \Gamma_1 \to \mathbb{R}$ which generate some further fields $V, e_+ : \Gamma_1 \to \mathbb{R}$, which are all required to obey a certain number of constraints. The problem is rather underdetermined, and so there will be some flexibility in selecting these fields; most of these fields will end up being concentrated in the region $\{(t,r) \in \Gamma_1 : r = (\pm 1 + O(\delta))t\}$ near the boundary of the light cone. Given that the constraint (5-16) only involves the two fields $V$ and $e_+$, it is natural to proceed by constructing $V$ and $e_+$ first. In fact we will proceed as follows.

***Selection of $e_+$ in the left half of the cone.*** We begin by making a choice for the function $e_+ : \Gamma_1 \to \mathbb{R}$ in the left half $\Gamma_1^l := \{(t,r) \in \Gamma_1 : r \leq 0\}$ of the cone. When $t = 1$, we choose $e_+(1,r)$ to be a smooth function with the following properties:

- One has
$$e_+(1,r) = (1+r)^{-\frac{4}{p-1}} \tag{6-1}$$
for $-1 + \delta \leq r \leq 0$.
- One has
$$e_+(1,r) \geq (1+r)^{-\frac{4}{p-1}} \tag{6-2}$$
for $-1 + \frac{1}{2}\delta \leq r \leq -1 + \delta$. Furthermore, one has
$$\int_{-1+\frac{1}{2}\delta}^{-1+\delta} e_+(1,r)\, dr \geq A\delta^{1-\frac{4}{p-1}}. \tag{6-3}$$

- One has

$$\delta^{-\frac{4}{p-1}} \lesssim e_+(1,r) \lesssim A\delta^{-\frac{4}{p-1}} \tag{6-4}$$

and

$$\left|\frac{d}{dr}e_+(1,r)\right| \lesssim A\delta^{-\frac{p+3}{p-1}} \tag{6-5}$$

for $-1 \le r \le -1+\delta$.

Clearly we can find a smooth function $r \to e_+(1,r)$ on $[-1,0]$ with these properties. We then extend $e_+$ to the entire left half $\Gamma_1^l$ of the cone by requiring it to be homogeneous of order $-\frac{4}{p-1}$; thus

$$e_+(t,r) := t^{-\frac{4}{p-1}} e_+\Big(1, \frac{r}{t}\Big). \tag{6-6}$$

In particular, $e_+$ is smooth on this half of the cone, and we have

$$e_+(t,r) = (t+r)^{-\frac{4}{p-1}}$$

for $-(1-\delta)t \le r \le 0$.

The properties (6-1)–(6-5) are largely used to ensure that the potential energy $V$ that we will construct below is nonnegative.

***Selection of $V$ in the left half of the cone.*** Once $e_+$ has been selected on $\Gamma_1^l$, we construct $V$ on $\Gamma_1^l$ by solving (5-16), or more explicitly by the formula

$$V(t,r) := \frac{1}{2|r|^{p-1}} \int_r^0 |s|^{p-1}((\partial_t - \partial_r)e_+)(t-r+s,s)\,ds \tag{6-7}$$

for $-t \le r < 0$. Note that as $(\partial_t - \partial_r)e_+$ vanishes for $-(1-\delta)t < r < 0$, the potential energy $V$ vanishes on this region also, and so one can smoothly extend $V$ to all of $\Gamma_1^l$. It is easy to see that $V$ is homogeneous of order $-\frac{4}{p-1}$. From the fundamental theorem of calculus and the chain rule, we have

$$(\partial_t + \partial_r)(|r|^{p-1}V) = |r|^{p-1}(\partial_t - \partial_r)e_+$$

for $-t \le r < 0$, and hence by the product rule we see that (5-16) is obeyed for $-t \le r < 0$, and hence to all of $\Gamma_1^l$ by smoothness. We have already seen that $V$ vanishes in the region $-(1-\delta)t < r \le 0$. In the region $-t \le r \le -(1-\delta)t$, we have the following estimate and nonnegativity property:

**Proposition 6.1.** *For $-t \le r \le -(1-\delta)t$, we have*

$$0 \le V(t,r) \lesssim At^{-\frac{4}{p-1}}\delta^{\frac{p-5}{p-1}}.$$

We remark that to get the lower bound $V(t,r)$, the supercriticality hypothesis $p > 5$ will be crucial.

*Proof.* By homogeneity we may assume that $t - r = 2$, so that $t = 1 - O(\delta)$ and $r = -1 + O(\delta)$, and it will suffice to show that

$$0 \le V(t,r) \lesssim A\delta^{\frac{p-5}{p-1}}. \tag{6-8}$$

Write $e_+(t,r) = (t+r)^{-\frac{4}{p-1}} + f(t,r)$; then from (6-7) we have

$$V(t,r) = \frac{1}{2|r|^{p-1}} \int_r^0 |s|^{p-1} ((\partial_t - \partial_r) f)(2+s,s)\, ds. \tag{6-9}$$

The function $f$ is homogeneous of order $-\frac{4}{p-1}$; hence by (1-2)

$$(t\partial_t + t\partial_r) f = -\frac{4}{p-1} f.$$

From the identity

$$\partial_t - \partial_r = -\frac{t+r}{t-r}(\partial_t + \partial_r) + \frac{2}{t-r}(t\partial_t + t\partial_r)$$

and the chain rule, we thus have

$$((\partial_t - \partial_r) f)(2+s,s) = -(1+s)\frac{d}{ds} f(2+s,s) - \frac{4}{p-1} f(2+s,s).$$

Inserting this into (6-9) and integrating by parts, we conclude that

$$V(t,r) = \frac{1+r}{2} f(t,r) + \frac{1}{2|r|^{p-1}} \int_r^0 \frac{d}{ds}(|s|^{p-1}(1+s)) f(2+s,s) - |s|^{p-1} \frac{4}{p-1} f(2+s,s)\, ds,$$

which by the product rule is equal to

$$V(t,r) = \frac{1+r}{2} f(t,r) + \frac{1}{2|r|^{p-1}} \int_r^0 |s|^{p-1} \left[ \frac{p-5}{p-1} + \frac{(p-1)(1+s)}{s} \right] f(2+s,s)\, ds. \tag{6-10}$$

Note that $f(2+s,s)$ is only nonzero when $s = -1 + O(\delta)$, in which case it is of size $O(A\delta^{-\frac{4}{p-1}})$ thanks to (6-2) and (6-4). This gives the upper bound in (6-8). Now we turn to the lower bound. First suppose that $-\big(1 - \frac{1}{2}\delta\big)t \le r$; then $f$ is nonnegative in all of its appearances in (6-10). As we are in the supercritical case $p > 5$, the factor

$$\frac{p-5}{p-1} + \frac{(p-1)(1+s)}{s}$$

is positive (indeed it is $\gtrsim 1$) for $\delta$ small enough, and the claim follows in this case.

It remains to consider the case when $-t \le r \le -\big(1 - \frac{1}{2}\delta\big)t$. In this case we can use the lower bound

$$f(t,r) \ge -(t+r)^{-\frac{4}{\delta}}$$

and conclude that the term $\frac{1}{2}(1+r) f(t,r)$ is at least $-O(\delta^{\frac{p-5}{p-1}})$. A similar argument shows that the contribution to (6-10) coming from those $s$ with

$$-(2+s) \le s \le -\big(1 - \tfrac{1}{2}\delta\big)(2+s)$$

is at least $-O(\delta^{\frac{p-5}{p-1}})$. On the other hand, from (6-3), the contribution of those $s$ with

$$s > -\big(1 - \tfrac{1}{2}\delta\big)(2+s)$$

is $\gtrsim (A - O(1))\delta^{\frac{p-5}{p-1}}$. As $A$ is assumed to be large, the claim follows. □

On the support of $V$ in $\Gamma_1^l$, we see from (6-5) and (6-6) that

$$(\partial_t - \partial_r)e_+ = O(At^{-\frac{p+3}{p-1}}\delta^{-\frac{4}{p-1}})$$

and hence by (5-16) and Proposition 6.1,

$$(\partial_t + \partial_r)V = O(At^{-\frac{p+3}{p-1}}\delta^{-\frac{4}{p-1}}). \tag{6-11}$$

***Selection of $V$ in the right half of the cone.*** Once $V$ has been constructed in the left half $\Gamma_1^l$ of the light cone, we extend it to the right half $\Gamma_1^r := \{(t,r) \in \Gamma_1 : r \geq 0\}$ by even extension; thus

$$V(t,r) := V(t,-r)$$

for all $(t,r) \in \Gamma_1^r$. Since $V$ vanished for $-(1-\delta)t \leq r \leq 0$, we see that $V$ is smooth on all of $\Gamma_1$, and vanishing in the interior cone $\{(t,r) \in \Gamma_1 : |r| \leq (1-\delta)t\}$. It also obeys the nonnegativity property (5-13). From reflecting (6-11) and Proposition 6.1 we have the bounds

$$V = O(At^{-\frac{4}{p-1}}\delta^{\frac{p-5}{p-1}}) \tag{6-12}$$

and

$$(\partial_t - \partial_r)V = O(At^{-\frac{p+3}{p-1}}\delta^{-\frac{4}{p-1}}) \tag{6-13}$$

when $(1-\delta)t \leq r \leq t$.

***Selection of $e_+$ in the right half of the cone.*** Thus far, $V$ has been defined on all of $\Gamma_1$, and $e_+$ defined on $\Gamma_1^l$. We now extend $e_+$ to $\Gamma_1^r$ by solving (5-16), or more precisely by setting

$$e_+(t,r) := e_+(t+r,0) + \int_0^r ((\partial_t + \partial_r)V)(t+r-s,s) + \frac{p-1}{s}V(t+r-s,s)\,ds \tag{6-14}$$

for $0 < r \leq t$; note that the integral is well-defined since $V$ vanishes near the time axis. One easily checks that $e_+(t,r) = (t+r)^{-\frac{4}{p-1}}$ for $0 \leq r \leq (1-\delta)t$, and so $e_+$ extends smoothly to all of $\Gamma_1$ and is equal to $(t+r)^{-\frac{4}{p-1}}$ in the interior cone $\{(t,r) \in \Gamma_1 : |r| \leq (1-\delta)t\}$. It is also clear from construction that $e_+$ is homogeneous of order $-\frac{4}{p-1}$. From the fundamental theorem of calculus we see that $e_+$ and $V$ obey (5-16) on $\Gamma_1^r$, and hence on all of $\Gamma_1$. From (6-12) and (6-13) we see that the integrand is of size $O(At^{-\frac{p+3}{p-1}}\delta^{-\frac{4}{p-1}})$ when $r = (1-O(\delta))t$, and vanishes otherwise, which leads (for $\delta$ small enough) to the crude upper and lower bounds

$$t^{-\frac{p+3}{p-1}} \lesssim e_+(t,r) \lesssim t^{-\frac{p+3}{p-1}} \tag{6-15}$$

throughout $\Gamma_1^r$.

***Selection of $e_-$ and $\widetilde{E}_{tr}$.*** We reflect the function $e_+$ around the time axis to create a new function $e_- : \Gamma_1 \to \mathbb{R}$:

$$e_-(t,r) := e_+(t,r).$$

Like $e_+$, the function $e_-$ is smooth and homogeneous of order $-\frac{4}{p-1}$. It equals $(t-r)^{-\frac{4}{p-1}}$ in the interior cone $\{(t,r) \in \Gamma_1 : |r| \leq (1-\delta)t\}$. On $\Gamma_1^l$ it obeys the crude upper and lower bounds

$$t^{-\frac{p+3}{p-1}} \lesssim e_-(t,r) \lesssim t^{-\frac{p+3}{p-1}} \tag{6-16}$$

and in the region $(1-\delta)t \leq r \leq t$ we have the bounds

$$(\delta t)^{-\frac{4}{p-1}} \lesssim e_-(t,r) \lesssim A(\delta t)^{-\frac{4}{p-1}} \tag{6-17}$$

thanks to (6-4).

Recall from (5-17) that the field $e_+$ is intended to ultimately be of the form $\frac{1}{2}\widetilde{E}_{tt} + \frac{1}{2}\widetilde{E}_{rr} + \widetilde{E}_{tr}$. Similarly, $e_-$ is intended to be of the form

$$e_- = \tfrac{1}{2}\widetilde{E}_{tt} + \tfrac{1}{2}\widetilde{E}_{rr} - \widetilde{E}_{tr}. \tag{6-18}$$

Accordingly, we may now define $\widetilde{E}_{tr}$ as

$$\widetilde{E}_{tr} := \frac{e_+ - e_-}{2}. \tag{6-19}$$

This is clearly smooth, odd, and homogeneous of order $-\frac{4}{p-1}$. We also see that the quantity $\widetilde{E}_{tt} + \widetilde{E}_{rr}$ is now specified:

$$\widetilde{E}_{tt} + \widetilde{E}_{rr} = e_+ + e_-. \tag{6-20}$$

We are left with two remaining unknown scalar fields to specify: the mass density $\widetilde{M}$ and the energy equipartition $-\widetilde{E}_{tt} + \widetilde{E}_{rr}$, which determines the fields $\widetilde{E}_{tt}$ and $\widetilde{E}_{rr}$ by (6-20). The requirements needed for Theorem 5.4 that have not already been verified are as follows:

- $\widetilde{M}$ is smooth, even, and homogeneous of order $\frac{2p-6}{p-1}$, and $-\widetilde{E}_{tt} + \widetilde{E}_{rr}$ is smooth, even, and homogeneous of order $-\frac{4}{p-1}$.
- $\widetilde{M}$, $\widetilde{E}_{tt}$ vanish to second order at $r = 0$, $\widetilde{E}_{tr} - \frac{1}{2r}\partial_t \widetilde{M}$ vanishes to third order, and $\widetilde{E}_{rr} - \frac{1}{r}\partial_r \widetilde{M} + \frac{1}{r^2}\widetilde{M}$ to fourth order.
- One has the equations (5-12) and (5-17) (and hence also (6-18)).
- The matrix (5-18) is strictly positive definite for $r \neq 0$, and for $r = 0$ one has the condition (5-19).

As there is only one equation (beyond homogeneity and reflection symmetry) constraining $\widetilde{M}$ and $-\widetilde{E}_{tt} + \widetilde{E}_{rr}$ — namely, (5-12) — the problem of selecting these two fields is underdetermined, and thus subject to a certain amount of arbitrary choices. We will select these fields first in the exterior region $\{(t,r) \in \Gamma : |r| \geq \frac{t}{2}\}$, and then fill in the interior using a different method.

***Selection of $M, -\widetilde{E}_{tt} + \widetilde{E}_{rr}$ away from the time axis.*** In the exterior region $\{(t,r) \in \Gamma : |r| \geq \frac{t}{2}\}$, we shall simply select the field $\widetilde{M}$ to be a small but otherwise rather arbitrary field, and then use (5-12) to determine $-\widetilde{E}_{tt} + \widetilde{E}_{rr}$.

More precisely, let $\widetilde{M}(1,r)$ be a smooth even function on the region $\{r : \frac{1}{2} \leq |r| \leq 1\}$ obeying the following properties:

- For $\frac{1}{2} \leq |r| \leq \frac{3}{4}$, one has

$$\widetilde{M}(1,r) = \delta\big((1+r)^{\frac{2p-6}{p-1}} + (1-r)^{\frac{2p-6}{p-1}}\big). \tag{6-21}$$

  (This condition will not be used directly in this part of the construction, but is needed for compatibility with the next part.)

- For $\frac{1}{2} \leq |r| \leq 1$, one has the bounds

$$\delta \lesssim \widetilde{M}(1,r) \lesssim \delta \tag{6-22}$$

and

$$\frac{d}{dr}\widetilde{M}(1,r),\ \frac{d^2}{dr^2}\widetilde{M}(1,r) = O(\delta). \tag{6-23}$$

It is clear that one can select such a function. We then extend $\widetilde{M}$ to $\{(t,r) \in \Gamma_1 : |r| \geq \frac{t}{2}\}$ by requiring that $\widetilde{M}$ be homogeneous of order $\frac{2p-6}{p-1}$. Then $\widetilde{M}$ is smooth and even, and one has the bounds

$$\delta t^{\frac{2p-6}{p-1}} \lesssim \widetilde{M}(t,r) \lesssim \delta t^{\frac{2p-6}{p-1}}, \tag{6-24}$$

$$\frac{d}{dr}\widetilde{M}(t,r),\ \frac{d}{dt}\widetilde{M}(t,r) = O(\delta t^{\frac{p-5}{p-1}}), \tag{6-25}$$

$$\frac{d^2}{dr^2}\widetilde{M}(t,r),\ \frac{d^2}{dt^2}\widetilde{M}(t,r) = O(\delta t^{-\frac{4}{p-1}}) \tag{6-26}$$

in the region $\{(t,r) \in \Gamma_1 : |r| \geq \frac{t}{2}\}$.

We then define $-\widetilde{E}_{tt} + \widetilde{E}_{rr}$ on this region by enforcing (5-12); thus

$$-\widetilde{E}_{tt} + \widetilde{E}_{rr} := \tfrac{1}{2}(-\partial_{tt}\widetilde{M} + \partial_{rr}\widetilde{M}) - (p+1)V. \tag{6-27}$$

Combining this with (6-20), this defines $\widetilde{E}_{tt}$ and $\widetilde{E}_{rr}$. It is easy to see that these fields are smooth, even and homogeneous of order $-\frac{4}{p-1}$ on $\{(t,r) \in \Gamma_1 : |r| \geq \frac{t}{2}\}$.

We now claim that the matrix (5-18) is strictly positive definite in the region $\{(t,r) \in \Gamma_1 : |r| \geq \frac{t}{2}\}$. By homogeneity and reflection symmetry, it suffices to verify this when $t = 1$ and $\frac{1}{2} \leq r \leq 1$. Using the identity

$$\begin{pmatrix} \widetilde{M} & \frac{1}{2}(\partial_t+\partial_r)\widetilde{M} & \frac{1}{2}(\partial_t-\partial_r)\widetilde{M} \\ \frac{1}{2}(\partial_t+\partial_r)\widetilde{M} & 2e_+ & -\widetilde{E}_{tt}+\widetilde{E}_{rr} \\ \frac{1}{2}(\partial_t-\partial_r)\widetilde{M} & -\widetilde{E}_{tt}+\widetilde{E}_{rr} & 2e_- \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \widetilde{M} & \frac{1}{2}\partial_t\widetilde{M} & \frac{1}{2}\partial_r\widetilde{M} \\ \frac{1}{2}\partial_t\widetilde{M} & \widetilde{E}_{tt} & \widetilde{E}_{tr} \\ \frac{1}{2}\partial_r M & \widetilde{E}_{tr} & \widetilde{E}_{rr} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix},$$

it suffices to show that the matrix

$$\begin{pmatrix} \widetilde{M} & \frac{1}{2}(\partial_t+\partial_r)\widetilde{M} & \frac{1}{2}(\partial_t-\partial_r)\widetilde{M} \\ \frac{1}{2}(\partial_t+\partial_r)\widetilde{M} & 2e_+ & -\widetilde{E}_{tt}+\widetilde{E}_{rr} \\ \frac{1}{2}(\partial_t-\partial_r)\widetilde{M} & -\widetilde{E}_{tt}+\widetilde{E}_{rr} & 2e_- \end{pmatrix}$$

is strictly positive definite.

If $r \leq 1-\delta$, then all off-diagonal terms are $O(\delta)$ thanks to (6-23) and (6-27), while the diagonal terms are $\gtrsim \delta$, $\gtrsim 1$, and $\gtrsim 1$ respectively, and the positive definiteness is easily verified, since the associated quadratic form is at least

$$\gtrsim \delta x_1^2 + x_2^2 + x_3^2 - O(\delta|x_1||x_2|) - O(\delta|x_1||x_3|) - O(\delta|x_2||x_3|),$$

which is easily seen to be positive for $\delta$ small enough. If $r < 1-\delta$, then the off-diagonal terms are $O(\delta)$ in the top row and left column, and $O(A\delta^{\frac{p-5}{p-1}})$ in the bottom right minor by (6-12), while the diagonal

terms are $\gtrsim \delta$, $\gtrsim 1$, and $\gtrsim \delta^{-\frac{4}{p-1}}$ by (6-22), (6-15) and (6-17), so the associated quadratic form is

$$\gtrsim \delta x_1^2 + x_2^2 + \delta^{-\frac{4}{p-1}} x_3^2 - O(\delta |x_1||x_2|) - O(\delta |x_1||x_3|) - O(A\delta^{\frac{p-5}{p-1}} |x_2||x_3|),$$

which is again positive definite (note that $A\delta^{\frac{p-5}{p-1}}$ can be chosen to be much smaller than the geometric mean of $\delta$ and $\delta^{-\frac{4}{p-1}}$).

***Selection of $M$, $\widetilde{E}_{tt}$, $\widetilde{E}_{rr}$ near the time axis.*** Now we restrict attention to the interior region $\Gamma_1^i := \{(t,r) \in \Gamma_1 : |r| \leq \frac{t}{2}\}$; all identities and estimates here are understood to be on this region unless otherwise specified.

We will now reverse the Gram matrix reduction from previous sections, and construct $\widetilde{M}$, $\widetilde{E}_{tt}$, $\widetilde{E}_{rr}$ in $\Gamma_1^i$ from an (infinite-dimensional) vector-valued solution to the (free, (1+1)-dimensional) wave equation. Let $H$ be a Hilbert space and let $t \mapsto f(t)$ be a family of vectors $f(t)$ in $H$ smoothly parmeterised by a parameter $t \in (0, +\infty)$ (so that all derivatives in $t$ exist in the strong sense and are continuous); we will select this family more precisely later. We introduce the smooth vector-valued field $\phi : \Gamma_1^i \to H$ by the formula

$$\phi(t,r) := f(t+r) - f(t-r)$$

and we will define $\widetilde{M}$, $\widetilde{E}_{tt}$, $\widetilde{E}_{rr} : \Gamma_1^i \to \mathbb{R}$ by the formulae

$$\begin{aligned}
\widetilde{M}(t,r) &:= \langle \phi(t,r), \phi(t,r) \rangle_H, \\
\widetilde{E}_{tt}(t,r) &:= \langle \partial_t \phi(t,r), \partial_t \phi(t,r) \rangle_H, \\
\widetilde{E}_{rr}(t,r) &:= \langle \partial_r \phi(t,r), \partial_r \phi(t,r) \rangle_H.
\end{aligned}$$

Since $\phi$ is smooth and odd in $r$, these functions are smooth and even in $r$. If we impose the additional hypothesis that the Gram matrix $\langle f(s), f(t) \rangle_H$ has the scaling symmetry

$$\langle f(\lambda s), f(\lambda t) \rangle_H = \lambda^{\frac{2p-6}{p-1}} \langle f(s), f(t) \rangle_H \tag{6-28}$$

for $s, t, \lambda > 0$, then $M$ will be homogeneous of order $\frac{2p-6}{p-1}$; furthermore, by differentiating (6-28) with respect to both $s$ and $t$ we see that

$$\langle f'(\lambda s), f'(\lambda t) \rangle_H = \lambda^{-\frac{4}{p-1}} \langle f'(s), f'(t) \rangle_H \tag{6-29}$$

(where $f'$ denotes the derivative of $f$) and so $\widetilde{E}_{tt}$, $\widetilde{E}_{rr}$ will be homogeneous of order $-\frac{4}{p-1}$.

Observe that

$$\begin{aligned}
\tfrac{1}{2}\widetilde{E}_{tt} + \tfrac{1}{2}\widetilde{E}_{rr} + \langle \partial_t \phi, \partial_r \phi \rangle_H &= \tfrac{1}{2} \|(\partial_t + \partial_r)\phi\|_H^2 \\
&= 2\|f'(t+r)\|_H^2
\end{aligned}$$

and similarly

$$\tfrac{1}{2}\widetilde{E}_{tt} + \tfrac{1}{2}\widetilde{E}_{rr} - \langle \partial_t \phi, \partial_r \phi \rangle_H = 2\|f'(t-r)\|_H^2.$$

Thus, if we impose the additional normalisation

$$\|f'(1)\|_H = \frac{1}{\sqrt{2}}, \tag{6-30}$$

and hence by (6-29),

$$\|f'(t)\|_H = \frac{1}{\sqrt{2}} t^{-\frac{2}{p-1}}, \tag{6-31}$$

we see from the identities $e_\pm(t,r) = (t \pm r)^{-\frac{4}{p-1}}$ in $\Gamma_1^i$ that

$$\tfrac{1}{2}\widetilde{E}_{tt} + \tfrac{1}{2}\widetilde{E}_{rr} \pm \langle \partial_t \phi, \partial_r \phi \rangle_H = e_\pm.$$

In particular, (6-20) holds, and from (6-19) one has

$$\widetilde{E}_{tr} = \langle \partial_t \phi, \partial_r \phi \rangle_H.$$

We also obtain the equations (5-17) and (6-18).

Next, it is clear that $\phi$ solves the wave equation

$$-\partial_{tt}\phi + \partial_{rr}\phi = 0,$$

so in particular

$$\langle \phi, -\partial_{tt}\phi + \partial_{rr}\phi \rangle_H = 0,$$

which implies in particular (cf. (4-5)) that

$$\tfrac{1}{2}(-\partial_{tt}\widetilde{M} + \partial_{rr}\widetilde{M}) + \widetilde{E}_{tt} - \widetilde{E}_{rr} = 0.$$

Since $V$ vanishes on $\Gamma_1$, we conclude that (5-12) holds.

Next, from differentiating the formula for $\widetilde{M}$, one has

$$\tfrac{1}{2}\partial_t \widetilde{M} = \langle \phi, \partial_t \phi \rangle_H$$

and

$$\tfrac{1}{2}\partial_r \widetilde{M} = \langle \phi, \partial_t \phi \rangle_H$$

and so the quadratic form associated with (5-18) factorises as

$$\|x_1\phi + x_2\partial_t\phi + x_3\partial_r\phi\|_H^2.$$

This is clearly positive semidefinite at least; to make it positive definite for $r \neq 0$, it will suffice to enforce the condition

$$f(s), f(t), f'(s), f'(t) \text{ linearly independent} \tag{6-32}$$

for all distinct $s, t > 0$.

Suppose we assume the long-range orthogonality condition

$$\langle f(s), f(t) \rangle_H = 0 \tag{6-33}$$

whenever $\frac{t}{s} > 1.1$ or $\frac{s}{t} > 1.1$. Then in the region $\{(t,r) \in \Gamma_1^i : |r| \geq \frac{t}{4}\}$ away from the time axis, we have from Pythagoras' theorem that

$$\widetilde{M}(t,r) = \|f(t+r)\|_H^2 + \|f(t-r)\|_H^2.$$

In particular, if we also impose the normalisation

$$\|f(1)\|_H = \sqrt{\delta} \tag{6-34}$$

then (from (6-28)) we have

$$\widetilde{M}(t,r) = \delta\big((t+r)^{\frac{2p-6}{p-1}} + (t-r)^{\frac{2p-6}{p-1}}\big)$$

in the region $\{(t,r) \in \Gamma_1^i : |r| \geq \frac{t}{4}\}$. In particular from (6-21) and homogeneity we see that $\widetilde{M}$ on $\Gamma_1^i$ joins up smoothly with its counterpart in the exterior region $\{(t,r) \in \Gamma_1 : |r| \geq \frac{t}{2}\}$; by (5-12) we see that $-\widetilde{E}_{tt} + \widetilde{E}_{rr}$ does too. By (6-20) and (6-19) we now see that all of the fields $M$, $\widetilde{E}_{tt}$, $\widetilde{E}_{rr}$, $\widetilde{E}_{tr}$ are smooth on all of $\Gamma_1$.

Now we study the vanishing properties of the various fields constructed at $r = 0$ for a fixed value of $t$. From Taylor expansion we have

$$\phi(t,r) = 2rf'(t) + \tfrac{1}{3}r^3 f'''(t) + O(|r|^5)$$

as $r \to 0$ (where the error term denotes a quantity in $H$ of norm $O(|r|^5)$, and the implied constant can depend on $t$ and $\phi$). Furthermore, these asymptotics behave in the expected fashion with respect to differentiation in time or space; thus for instance

$$\begin{aligned}\partial_r\phi(t,r) &= 2f'(t) + r^2 f'''(t) + O(|r|^4),\\ \partial_t\phi(t,r) &= 2rf''(t) + \tfrac{1}{3}r^3 f^{(4)}(t) + O(|r|^5).\end{aligned}$$

Taking inner products, we conclude the asymptotics

$$\begin{aligned}\widetilde{M}(t,r) &= 4r^2\|f'(t)\|_H^2 + \tfrac{4}{3}r^4\langle f'(t), f'''(t)\rangle_H + O(|r|^6),\\ \widetilde{E}_{tt}(t,r) &= 4r^2\|f''(t)\|_H^2 + O(|r|^4),\\ \widetilde{E}_{tr}(t,r) &= 4r\langle f'(t), f''(t)\rangle_H + O(|r|^3),\\ \widetilde{E}_{rr} &= 4\|f'(t)\|_H^2 + 4r^2\langle f'(t), f'''(t)\rangle_H + O(|r|^4).\end{aligned}$$

The asymptotic for $\widetilde{M}$ behaves well with respect to derivatives; thus for instance

$$\begin{aligned}\partial_t\widetilde{M}(t,r) &= 8r^2\langle f'(t), f''(t)\rangle_H + O(|r|^4),\\ \partial_r\widetilde{M}(t,r) &= 8r\|f'(t)\|_H^2 + \tfrac{16}{3}r^3\langle f'(t), f'''(t)\rangle_H + O(|r|^5).\end{aligned}$$

Among other things, this shows (using (6-30)) that the condition (5-19) reduces to

$$\|f''(1)\|_H > \frac{2}{p-1}\frac{1}{\sqrt{2}}. \tag{6-35}$$

It is also clear from these asymptotics that $\widetilde{E}$ and $\widetilde{E}_{tt}$ vanish to second order, and $\widetilde{E}_{tr} - \frac{1}{2r}\partial_t\widetilde{M}$ vanishes to third order; a brief calculation also shows that $\widetilde{E}_{rr} - \frac{1}{r}\partial_r\widetilde{M} + \frac{1}{r^2}\widetilde{M}$ vanishes to fourth order.

To summarise: in order to conclude all the required properties for Theorem 5.4, it suffices to locate a smooth curve $t \mapsto f(t)$ in a Hilbert space $H$ which obeys the hypotheses (6-28), (6-30), (6-32), (6-33), (6-34) and (6-35).

We take the Hilbert space $H$ to be the space $L^2(\mathbb{R})$ of square-integrable real-valued functions on $\mathbb{R}$ with Lebesgue measure. The functions $f(t) \in H$ will take the form

$$f(t)(x) := t^{\frac{p-3}{p-1}} \psi(x - \log t),$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is a bump function whose (closed) support is precisely $[0, 0.01]$ (that is to say, the set $\{\psi \neq 0\}$ is a dense subset of $[0, 0.01]$) depending on $\delta$ and $p$ to be chosen shortly. It is clear from construction that (6-28) and (6-33) hold. The condition (6-34) becomes

$$\int_{\mathbb{R}} \psi(x)^2 \, dx = \delta,$$

while the condition (6-30) becomes

$$\int_{\mathbb{R}} \psi'(x)^2 \, dx = \tfrac{1}{2}.$$

It is easy to see that we can select $\psi$ with closed support precisely $[0, 0.01]$ with both of these normalisations, basically because the Dirichlet form $\langle \phi', \psi' \rangle$ is unbounded on $L^2([0, 0.01])$.

Now we verify the linear independence claim (6-32). We may assume without loss of generality that $s = 1$ and $t > 1$. Then we have a linear dependence between $\psi$ and $\psi'$ in a neighbourhood of 0; since $\psi, \psi'$ vanish to the left of 0, the Picard uniqueness theorem for ODEs then implies that $\psi$ vanishes a little to the right of 0 also, contradicting the hypothesis that $\psi$ has closed support containing 0. This gives (6-32).

A similar argument shows that $f'(1)$ and $f''(1)$ are linearly independent. Squaring and differentiating (6-31) at $t = 1$ gives

$$\langle f'(1), f''(1) \rangle_H = -\frac{2}{p-1} \frac{1}{2}$$

and (6-35) then follows from (6-30) and the Cauchy–Schwarz inequality, using the linear independence to get the strict inequality. This (finally) completes the proof of Theorem 5.4 and hence Theorem 1.1.

## Acknowledgments

The author is supported by NSF grant DMS-1266164 and by a Simons Investigator Award. The author is also indebted to Sergiu Klainerman for suggesting this question, and Yi Jin Gao for corrections.

## References

[Brenner and Kumlin 2000] P. Brenner and P. Kumlin, "On wave equations with supercritical nonlinearities", *Arch. Math. (Basel)* **74**:2 (2000), 129–147. MR 1735230 Zbl 0971.35051

[Brenner and von Wahl 1981] P. Brenner and W. von Wahl, "Global classical solutions of nonlinear wave equations", *Math. Z.* **176**:1 (1981), 87–121. MR 606174 Zbl 0457.35059

[Burq et al. 2007] N. Burq, S. Ibrahim, and P. Gérard, "Instability results for nonlinear Schrödinger and wave equations", preprint, 2007.

[Cazenave et al. 1998] T. Cazenave, J. Shatah, and A. S. Tahvildar-Zadeh, "Harmonic maps of the hyperbolic space and development of singularities in wave maps and Yang–Mills fields", *Ann. Inst. H. Poincaré Phys. Théor.* **68**:3 (1998), 315–349. MR 1622539 Zbl 0918.58074

[Christ et al. 2003] M. Christ, J. Colliander, and T. Tao, "Ill-posedness for nonlinear Schrodinger and wave equations", preprint, 2003. arXiv math/0311048

[Grillakis 1990] M. G. Grillakis, "Regularity and asymptotic behaviour of the wave equation with a critical nonlinearity", *Ann. of Math.* (2) **132**:3 (1990), 485–509. MR 1078267 Zbl 0736.35067

[Grillakis 1992] M. G. Grillakis, "Regularity for the wave equation with a critical nonlinearity", *Comm. Pure Appl. Math.* **45**:6 (1992), 749–774. MR 1162370 Zbl 0785.35065

[Günther 1991] M. Günther, "Isometric embeddings of Riemannian manifolds", pp. 1137–1143 in *Proceedings of the International Congress of Mathematicians, Vol. II* (Kyoto, 1990), Math. Soc. Japan, Tokyo, 1991. MR 1159298 Zbl 0745.53031

[Ibrahim et al. 2011] S. Ibrahim, M. Majdoub, and N. Masmoudi, "Well- and ill-posedness issues for energy supercritical waves", *Anal. PDE* **4**:2 (2011), 341–367. MR 2859857 Zbl 1267.35131

[Jörgens 1961] K. Jörgens, "Das Anfangswertproblem im Grossen für eine Klasse nichtlinearer Wellengleichungen", *Math. Z.* **77** (1961), 295–308. MR 0130462 Zbl 0111.09105

[Kenig and Merle 2008] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation", *Acta Math.* **201**:2 (2008), 147–212. MR 2461508 Zbl 1183.35202

[Killip and Visan 2011a] R. Killip and M. Visan, "The defocusing energy-supercritical nonlinear wave equation in three space dimensions", *Trans. Amer. Math. Soc.* **363**:7 (2011), 3893–3934. MR 2775831 Zbl 1230.35068

[Killip and Visan 2011b] R. Killip and M. Visan, "The radial defocusing energy-supercritical nonlinear wave equation in all space dimensions", *Proc. Amer. Math. Soc.* **139**:5 (2011), 1805–1817. MR 2763767 Zbl 1218.35148

[Lebeau 2001] G. Lebeau, "Non linear optic and supercritical wave equation", *Bull. Soc. Roy. Sci. Liège* **70**:4–6 (2001), 267–306. MR 1904059 Zbl 1034.35137

[Lebeau 2005] G. Lebeau, "Perte de régularité pour les équations d'ondes sur-critiques", *Bull. Soc. Math. France* **133**:1 (2005), 145–157. MR 2145023 Zbl 1071.35020

[Lindblad and Sogge 1996] H. Lindblad and C. D. Sogge, "Long-time existence for small amplitude semilinear wave equations", *Amer. J. Math.* **118**:5 (1996), 1047–1135. MR 1408499 Zbl 0869.35064

[Planchon 2000] F. Planchon, "Self-similar solutions and semi-linear wave equations in Besov spaces", *J. Math. Pures Appl.* (9) **79**:8 (2000), 809–820. MR 1782103 Zbl 0979.35106

[Ribaud and Youssfi 2002] F. Ribaud and A. Youssfi, "Global solutions and self-similar solutions of semilinear wave equation", *Math. Z.* **239**:2 (2002), 231–262. MR 1888223 Zbl 1005.35068

[Roy 2009] T. Roy, "Global existence of smooth solutions of a 3D log-log energy-supercritical wave equation", *Anal. PDE* **2**:3 (2009), 261–280. MR 2603799 Zbl 1195.35222

[Seeley 1964] R. T. Seeley, "Extension of $C^\infty$ functions defined in a half space", *Proc. Amer. Math. Soc.* **15**:4 (1964), 625–626. MR 0165392 Zbl 0127.28403

[Segal 1963] I. E. Segal, "The global Cauchy problem for a relativistic scalar field with power interaction", *Bull. Soc. Math. France* **91** (1963), 129–135. MR 0153967 Zbl 0178.45403

[Shatah and Struwe 1998] J. Shatah and M. Struwe, *Geometric wave equations*, Courant Lecture Notes in Mathematics **2**, New York University, Courant Institute of Mathematical Sciences, New York, 1998. MR 1674843 Zbl 0993.35001

[Strauss 1989] W. A. Strauss, *Nonlinear wave equations*, CBMS Regional Conference Series in Mathematics **73**, American Mathematical Society, Providence, RI, 1989. MR 1032250 Zbl 0714.35003

[Struwe 1988] M. Struwe, "Globally regular solutions to the $u^5$ Klein–Gordon equation", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **15**:3 (1988), 495–513. MR 1015805 Zbl 0728.35072

[Tao 2006] T. Tao, *Nonlinear dispersive equations: local and global analysis*, CBMS Regional Conference Series in Mathematics **106**, American Mathematical Society, Providence, RI, 2006. MR 2233925 Zbl 1106.35001

[Tao 2007] T. Tao, "Global regularity for a logarithmically supercritical defocusing nonlinear wave equation for spherically symmetric data", *J. Hyperbolic Differ. Equ.* **4**:2 (2007), 259–265. MR 2329385 Zbl 1124.35043

[Zheng 1991] Y. X. Zheng, "Concentration in sequences of solutions to the nonlinear Klein–Gordon equation", *Indiana Univ. Math. J.* **40**:1 (1991), 201–235. MR 1101227 Zbl 0714.35079

Received 24 Feb 2016. Revised 6 Aug 2016. Accepted 29 Sep 2016.

TERENCE TAO: tao@math.ucla.edu
*Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, United States*

ANALYSIS AND PDE
Vol. 9, No. 8, 2016
dx.doi.org/10.2140/apde.2016.9.2031

# A LONG $\mathbb{C}^2$ WITHOUT HOLOMORPHIC FUNCTIONS

LUKA BOC THALER AND FRANC FORSTNERIČ

*Dedicated to John Erik Fornæss*

We construct for every integer $n > 1$ a complex manifold of dimension $n$ which is exhausted by an increasing sequence of biholomorphic images of $\mathbb{C}^n$ (i.e., a *long* $\mathbb{C}^n$), but does not admit any nonconstant holomorphic or plurisubharmonic functions. Furthermore, we introduce new holomorphic invariants of a complex manifold $X$, the *stable core* and the *strongly stable core*, which are based on the long-term behavior of hulls of compact sets with respect to an exhaustion of $X$. We show that every compact polynomially convex set $B \subset \mathbb{C}^n$ such that $B = \overline{B^\circ}$ is the strongly stable core of a long $\mathbb{C}^n$; in particular, holomorphically nonequivalent sets give rise to nonequivalent long $\mathbb{C}^n$'s. Furthermore, for every open set $U \subset \mathbb{C}^n$ there exists a long $\mathbb{C}^n$ whose stable core is dense in $U$. It follows that for any $n > 1$ there is a continuum of pairwise nonequivalent long $\mathbb{C}^n$'s with no nonconstant plurisubharmonic functions and no nontrivial holomorphic automorphisms. These results answer several long-standing open problems.

## 1. Introduction

A complex manifold $X$ of dimension $n$ is said to be a *long* $\mathbb{C}^n$ if it is the union of an increasing sequence of domains $X_1 \subset X_2 \subset X_3 \subset \cdots \subset \bigcup_{j=1}^{\infty} X_j = X$ such that each $X_j$ is biholomorphic to the complex Euclidean space $\mathbb{C}^n$. It is immediate that any long $\mathbb{C}$ is biholomorphic to $\mathbb{C}$. However, for $n > 1$, this class of complex manifolds is still very mysterious. The long-standing question, whether there exists a long $\mathbb{C}^n$ which is not biholomorphic to $\mathbb{C}^n$, was answered in 2010 by E. F. Wold [2010], who constructed a long $\mathbb{C}^n$ that is not holomorphically convex, hence not a Stein manifold. Wold's construction is based on his examples of non-Runge Fatou–Bieberbach domains in $\mathbb{C}^n$ (see [Wold 2008]; an exposition of both results can be found in [Forstnerič 2011, Section 4.20]). In spite of these interesting examples, the theory has not been developed since. In particular, it remained unknown whether there exist long $\mathbb{C}^2$'s without nonconstant holomorphic functions, and whether there exist at least two nonequivalent non-Stein long $\mathbb{C}^2$'s.

We begin with the following result, which answers the first question affirmatively.

**Theorem 1.1.** *For every integer $n > 1$ there exists a long $\mathbb{C}^n$ without any nonconstant holomorphic or plurisubharmonic functions.*

Theorem 1.1 is proved in Section 3. It contributes to the line of counterexamples to the classical union problem for Stein manifolds: *is an increasing union of Stein manifolds always Stein?* For domains in $\mathbb{C}^n$ this question was raised by Behnke and Thullen [1934], and an affirmative answer was given in [Behnke and Stein 1939]. Some progress on the general question was made by Stein [1956] and Docquier and

*MSC2010:* 32E10, 32E30, 32H02.
*Keywords:* holomorphic function, Stein manifold, long $\mathbb{C}^n$, Fatou–Bieberbach domain, Chern–Moser normal form.

Grauert [1960]. The first counterexample was given in any dimension $n \geq 3$ by J. E. Fornæss [1976]; he found an increasing union of balls that is not holomorphically convex, hence not Stein. The key ingredient in his proof is a construction of a biholomorphic map $\Phi : \Omega \to \Phi(\Omega) \subset \mathbb{C}^3$ on a bounded neighborhood $\Omega \subset \mathbb{C}^3$ of any compact set $K \subset \mathbb{C}^3$ with nonempty interior such that the polynomial hull of $\Phi(K)$ is not contained in $\Phi(\Omega)$. (A phenomenon of this type was first described by Wermer [1959].) Fornæss and Stout [1977] constructed an increasing union of three-dimensional polydiscs without nonconstant holomorphic functions. Fornæss [1978] gave a counterexample to the union problem in dimension 2. Increasing unions of hyperbolic Stein manifolds were studied further by Fornæss and Sibony [1981] and Fornæss [2004]. Wold [2010] constructed the first example of a non-Stein long $\mathbb{C}^2$. For the connection with Bedford's conjecture, see the survey [Abbondandolo et al. 2013].

Another question that has been asked repeatedly over a long period of time is whether there exist infinitely many nonequivalent long $\mathbb{C}^n$'s for any or all $n > 1$. Up to now, only two different long $\mathbb{C}^2$'s have been known, namely the standard $\mathbb{C}^2$ and a non-Stein long $\mathbb{C}^2$ constructed by Wold [2010]. In dimension $n > 2$ one can get a few more examples by considering Cartesian products of long $\mathbb{C}^k$'s for different values of $k$. In this paper, we introduce new biholomorphic invariants of a complex manifold, the *stable core* and the *strongly stable core* (see Definition 1.5), which allow us to distinguish certain long $\mathbb{C}^n$'s from one another. In our opinion, this is the main new contribution of the paper from the conceptual point of view. With the help of these invariants, we answer the above mentioned question affirmatively by proving the following result.

Recall that a compact subset $B$ of a topological space $X$ is said to be *regular* if it is the closure of its interior, $B = \overline{B^\circ}$.

**Theorem 1.2.** *Let $n > 1$. To every regular compact polynomially convex set $B \subset \mathbb{C}^n$ we can associate a complex manifold $X(B)$, which is a long $\mathbb{C}^n$ containing a biholomorphic copy of $B$, such that every biholomorphic map $\Phi : X(B) \to X(C)$ between two such manifolds takes $B$ onto $C$. In particular, for every holomorphic automorphism $\Phi \in \mathrm{Aut}(X(B))$, the restriction $\Phi|_B$ is an automorphism of $B$. We can choose $X(B)$ such that it has no nonconstant holomorphic or plurisubharmonic functions.*

It follows that the manifold $X(B)$ can be biholomorphic to $X(C)$ only if $B$ is biholomorphic to $C$. Our construction likely gives many nonequivalent long $\mathbb{C}^n$'s associated to the same set $B$. A more precise result is given by Theorem 1.6 below.

By considering the special case when $B$ is the closure of a strongly pseudoconvex domain, Theorem 1.2 shows that the moduli space of long $\mathbb{C}^n$'s contains the moduli space of germs of smooth strongly pseudoconvex real hypersurfaces in $\mathbb{C}^n$. This establishes a surprising connection between long $\mathbb{C}^n$'s and the Cauchy–Riemann geometry. It has been known since Poincaré's paper [1907] that most pairs of smoothly bounded strongly pseudoconvex domains in $\mathbb{C}^n$ are not biholomorphic to each other, at least not by maps extending smoothly to the closed domains. It was shown much later by C. Fefferman [1974] that the latter condition is automatically fulfilled. (For elementary proofs of Fefferman's theorem, see [Pinchuk and Khasanov 1987; Forstnerič 1992].) A complete set of local holomorphic invariants of a strongly pseudoconvex real-analytic hypersurface is provided by the *Chern–Moser normal form*; see [Chern and Moser 1974]. Most such domains have no holomorphic automorphisms other than the identity

map. (For surveys of this topic, see, e.g., [Baouendi et al. 1999; Forstnerič 1993].) Hence, Theorem 1.2 implies the following corollary.

**Corollary 1.3.** *For every $n > 1$ there is a continuum of pairwise nonequivalent long $\mathbb{C}^n$'s with no nonconstant holomorphic or plurisubharmonic functions and no nontrivial holomorphic automorphisms.*

We now describe the new biholomorphic invariants alluded to above, the *stable core* and the *strongly stable core* of a complex manifold. Their definition is based on the stable hull property defined below, which a compact set in a complex manifold may or may not have. Given a pair of compact sets $K \subset L$ in a complex manifold $X$, we write

$$\widehat{K}_{\mathscr{O}(L)} = \{x \in L : |f(x)| \le \sup_K |f| \text{ for all } f \in \mathscr{O}(L)\}, \tag{1-1}$$

where $\mathscr{O}(L)$ is the algebra of holomorphic functions on neighborhoods of $L$.

**Definition 1.4** (the stable hull property). A compact set $K$ in a complex manifold $X$ has the *stable hull property* (SHP) if there exists an exhaustion $K_1 \subset K_2 \subset \cdots \subset \bigcup_{j=1}^{\infty} K_j = X$ by compact sets such that $K \subset K_1$, $K_j \subset K_{j+1}^{\circ}$ for every $j \in \mathbb{N}$, and the increasing sequence of hulls $\widehat{K}_{\mathscr{O}(K_j)}$ stabilizes, i.e., there is a $j_0 \in \mathbb{N}$ such that

$$\widehat{K}_{\mathscr{O}(K_j)} = \widehat{K}_{\mathscr{O}(K_{j_0})} \quad \text{for all } j \ge j_0. \tag{1-2}$$

Obviously, SHP is a biholomorphically invariant property: if a compact set $K \subset X$ satisfies condition (1-2) with respect to some exhaustion $(K_j)_{j\in\mathbb{N}}$ of $X$, then for any biholomorphic map $F : X \to Y$ the set $F(K) \subset Y$ satisfies (1-2) with respect to the exhaustion $L_j = F(K_j)$ of $Y$. What is less obvious, but needed to make this condition useful, is its independence of the choice of the exhaustion; see Lemma 4.1.

**Definition 1.5.** Let $X$ be a complex manifold.

(i) The *stable core* of $X$, denoted $\mathrm{SC}(X)$, is the open set consisting of all points $x \in X$ which admit a compact neighborhood $K \subset X$ with the stable hull property.

(ii) A regular compact set $B \subset X$ is called the *strongly stable core* of $X$, denoted $\mathrm{SSC}(X)$, if $B$ has the stable hull property, but no compact set $K \subset X$ with $K^{\circ} \setminus B \neq \varnothing$ has the stable hull property.

Clearly, the stable core always exists and is a biholomorphic invariant, in the sense that any biholomorphic map $X \to Y$ maps $\mathrm{SC}(X)$ onto $\mathrm{SC}(Y)$. In particular, every holomorphic automorphism of $X$ maps the stable core $\mathrm{SC}(X)$ onto itself. The strongly stable core $\mathrm{SSC}(X)$ need not exist in general; if it does, then its interior equals the stable core $\mathrm{SC}(X)$ and $\mathrm{SSC}(X) = \overline{\mathrm{SC}(X)}$. In (ii), we must restrict attention to *regular* compact sets since otherwise the definition would be ambiguous.

**Theorem 1.6.** *Let $n > 1$.*

(a) *For every regular compact polynomially convex set $B \subset \mathbb{C}^n$ (i.e., $B = \overline{B^{\circ}}$) there exists a long $\mathbb{C}^n$, $X(B)$, which admits no nonconstant plurisubharmonic functions and whose strongly stable core equals $B$: $\mathrm{SSC}(X(B)) = B$.*

(b) *For every open set $U \subset \mathbb{C}^n$ there exists a long $\mathbb{C}^n$, $X$, which admits no nonconstant holomorphic functions and satisfies $\mathrm{SC}(X) \subset U$ and $\overline{U} = \overline{\mathrm{SC}(X)}$.*

In Theorem 1.6 we have identified the sets $B, U \subset \mathbb{C}^n$ with their images in the long $\mathbb{C}^n$, $X = \bigcup_{k=1}^{\infty} X_k$, by identifying $\mathbb{C}^n$ with the first domain $X_1 \subset X$.

Assuming Theorem 1.6, we now prove Theorem 1.2.

*Proof of Theorem 1.2.* Let $B$ be a regular compact polynomially convex set in $\mathbb{C}^n$ for some $n > 1$. By Theorem 1.6 there exists a long $\mathbb{C}^n$, $X = X(B)$, whose strongly stable core is $B$. Assume that $F \in \mathrm{Aut}(X)$. Then $F(B)$ has SHP (see Definition 1.4). Since $B$ is the biggest regular compact subset of $X$ with SHP (see (ii) in Definition 1.5), we have that $\Phi(B) \subset B$. Applying the same argument to the inverse automorphism $\Phi^{-1} \in \mathrm{Aut}(X)$ gives $\Phi^{-1}(B) \subset B$, and hence $B \subset \Phi(B)$. Both properties together imply that $\Phi(B) = B$, and hence $\Phi|_B \in \mathrm{Aut}(B)$.

In the same way, we see that a biholomorphic map $X(B) \to X(C)$ between two long $\mathbb{C}^n$'s, furnished by part (a) in Theorem 1.6, maps $B$ biholomorphically onto $C$. Hence, if $B$ is not biholomorphic to $C$, then $X(B)$ is not biholomorphic to $X(C)$. □

Theorem 1.6 is proved in Section 4. We construct manifolds with these properties by improving the recursive procedure devised by Wold [2008; 2010]. The following key ingredient was introduced in [Wold 2008]; it will henceforth be called the *Wold process* (see Remark 3.2).

Given a compact holomorphically convex set $L \subset \mathbb{C}^* \times \mathbb{C}^{n-1}$ with nonempty interior, there is a holomorphic automorphism $\psi \in \mathrm{Aut}(\mathbb{C}^* \times \mathbb{C}^{n-1})$ such that the polynomial hull $\widehat{\psi(L)}$ of the set $\psi(L)$ intersects the hyperplane $\{0\} \times \mathbb{C}^{n-1}$. By precomposing $\psi$ with a suitably chosen Fatou–Bieberbach map $\theta : \mathbb{C}^n \hookrightarrow \mathbb{C}^* \times \mathbb{C}^{n-1}$, we obtain a Fatou–Bieberbach map $\phi = \psi \circ \theta : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ such that, for a given polynomially convex set $K \subset \mathbb{C}^n$ with nonempty interior, we have that $\widehat{\phi(K)} \not\subset \phi(\mathbb{C}^n)$.

At every step of the recursion we perform the Wold process simultaneously on finitely many pairwise disjoint compact sets $K_1, \ldots, K_m$ in the complement of the given regular polynomially convex set $B \subset \mathbb{C}^n$, chosen such that $\bigcup_{j=1}^{m} K_j \cup B$ is polynomially convex, thereby ensuring that polynomial hulls of their images $\phi(K_j)$ escape from the range of the injective holomorphic map $\phi : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ constructed in the recursive step. At the same time, we ensure that $\phi$ is close to the identity map on a neighborhood of $B$, and hence the image $\phi(B)$ remains polynomially convex. In practice, the sets $K_j$ will be small pairwise disjoint closed balls in the complement of $B$ whose number will increase during the process. We devise the process so that every point in a certain countable dense set $A = \{a_j\}_{j=1}^{\infty} \subset X \setminus B$ is the center of a decreasing sequence of balls whose $\mathscr{O}(X_k)$-hulls escape from each compact set in $X$; hence none of these balls has the stable hull property. This implies that $B$ is the strongly stable core of $X$.

To prove part (b), we modify the recursion by introducing a new small ball $B' \subset U \setminus B$ at every stage. Thus, the set $B$ acquires additional connected components during the recursive process. The sequence of added balls $B_l$ is chosen such that their union is dense in the given open subset $U \subset \mathbb{C}^n$, while the sequence of sets $K_j$ on which the Wold process is performed densely fills the complement $X \setminus \overline{U}$. It follows that the stable core of the limit manifold $X = \bigcup_{k=1}^{\infty} X_k$ is contained in $U$ and is everywhere dense in $U$.

By combining the technique used in the proof of Theorem 1.1 (see Section 3) with those in [Forstnerič 2012, proof of Theorem 1.1], one can easily obtain the following result for holomorphic families of long $\mathbb{C}^n$'s. (Compare with [Forstnerič 2012, Theorem 1.1].) We leave out the details.

**Theorem 1.7.** *Let $Y$ be a Stein manifold, and let $A$ and $B$ be disjoint finite or countable sets in $Y$. For every integer $n > 1$ there exists a complex manifold $X$ of dimension* $\dim Y + n$ *and a surjective holomorphic submersion $\pi : X \to Y$ with the following properties*:

- *the fiber $X_y = \pi^{-1}(y)$ over any point $y \in Y$ is a long $\mathbb{C}^n$;*
- *$X_y$ is biholomorphic to $\mathbb{C}^n$ if $y \in A$;*
- *$X_y$ does not admit any nonconstant plurisubharmonic function if $y \in B$.*

*If the base $Y$ is $\mathbb{C}^p$, then $X$ may be chosen to be a long $\mathbb{C}^{p+n}$.*

Note that one or both of the sets $A$ and $B$ in Theorem 1.7 may be chosen everywhere dense in $Y$. The same result holds if $A$ is a union of at most countably many closed complex subvarieties of $Y$ and the set $B$ is countable.

Several interesting questions on long $\mathbb{C}^n$'s remain open; we record some of them.

**Problem 1.8.** (A) Does there exist a long $\mathbb{C}^2$ which admits a nonconstant holomorphic function, but is not Stein?

(B) To what extent is it possible to prescribe the algebra $\mathscr{O}(X)$ of a long $\mathbb{C}^n$?

(C) Does there exist a long $\mathbb{C}^n$ for any $n > 1$ which is a Stein manifold different from $\mathbb{C}^n$?

(D) Does there exist a long $\mathbb{C}^n$ without nonconstant meromorphic functions?

(E) What can be said about the (non)existence of complex analytic subvarieties of positive dimension in non-Stein long $\mathbb{C}^n$'s?

In dimensions $n > 2$, an affirmative answer to problem (A) is provided by the product $X = \mathbb{C}^p \times X^{n-p}$ for any $p = 1, \dots, n-2$, where $X^{n-p}$ is a long $\mathbb{C}^{n-p}$ without nonconstant holomorphic functions, furnished by Theorem 1.1. Note that $\mathscr{O}(\mathbb{C}^p \times X^{n-p}) \cong \mathscr{O}(\mathbb{C}^p)$ is the algebra of functions coming from the base. Indeed, any example furnished by Theorem 1.7, with $Y = \mathbb{C}^p$ as base ($p \geq 1$) and $B$ dense in $\mathbb{C}^p$, is of this kind.

Regarding question (D), note that the Fatou–Bieberbach maps $\phi_k : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ used in our constructions have rationally convex images, in the sense that for any compact polynomially convex set $K \subset \mathbb{C}^n$ its image $\phi_k(K)$ is a rationally convex set in $\mathbb{C}^n$; this gives rise to nontrivial meromorphic functions on the resulting long $\mathbb{C}^n$'s.

Since every long $\mathbb{C}^n$ is an *Oka manifold* [Lárusson 2010; Forstnerič 2011, Proposition 5.5.6, p. 200], the results of this paper also contribute to our understanding of the class of Oka manifolds, that is, manifolds which are the most natural targets for holomorphic maps from Stein manifolds and reduced Stein spaces.

Note that every long $\mathbb{C}^n$ is a topological cell according to a theorem of Brown [1961]. Furthermore, it was shown by Wold [2010, Theorem 1.2] that, if $X = \bigcup_{k=1}^{\infty} X_k$ is a long $\mathbb{C}^n$ and $(X_k, X_{k+1})$ is a Runge pair for every $k \in \mathbb{N}$, then $X$ is biholomorphic to $\mathbb{C}^n$. Since the Runge property always holds in the $\mathscr{C}^\infty$ category, i.e., for smooth diffeomorphisms of Euclidean spaces, his proof can be adjusted to show that every long $\mathbb{C}^n$ is also diffeomorphic to $\mathbb{R}^{2n}$. Hence, Theorems 1.2 and 1.6 imply the following corollary.

**Corollary 1.9.** *For every $n > 1$ there exists a continuum of pairwise nonequivalent Oka manifolds of complex dimension $n$ which are all diffeomorphic to $\mathbb{R}^{2n}$.*

In Section 5 we show that $\mathbb{C}^n$ for any $n > 1$ can also be represented as an increasing union of non-Runge Fatou–Bieberbach domains.

## 2. Preliminaries

In this section, we introduce the notation and recall the basic ingredients.

We denote by $\mathscr{O}(X)$ the algebra of all holomorphic functions on a complex manifold $X$. For a compact set $K \subset X$, $\mathscr{O}(K)$ stands for the algebra of functions holomorphic in open neighborhoods of $K$ (in the sense of germs on $K$). The $\mathscr{O}(X)$*-convex hull* of $K$ is

$$\widehat{K}_{\mathscr{O}(X)} = \{x \in X : |f(x)| \le \sup_K |f| \text{ for all } f \in \mathscr{O}(X)\}.$$

When $X = \mathbb{C}^n$, the set $\widehat{K} = \widehat{K}_{\mathscr{O}(\mathbb{C}^n)}$ is the *polynomial hull* of $X$. If $\widehat{K}_{\mathscr{O}(X)} = K$, we say that $K$ is *holomorphically convex in* $X$; if $X = \mathbb{C}^n$ then $K$ is polynomially convex. More generally, if $K \subset L$ are compact sets in $X$, we define the hull $\widehat{K}_{\mathscr{O}(L)}$ by (1-1).

Given a point $p \in \mathbb{C}^n$, we denote by $\mathbb{B}(p; r)$ the closed ball of radius $r$ centered at $p$.

We shall frequently use the following basic result; see, e.g., [Stout 1971; 2007] for the first part (which is a simple application of E. Kallin's lemma) and [Forstnerič 1986] for the second part.

**Lemma 2.1.** *Assume that $B \subset \mathbb{C}^n$ is a compact polynomially convex set. For any $p_1, \dots, p_m \in \mathbb{C}^n \setminus B$ and for all sufficiently small numbers $r_1 > 0, \dots, r_m > 0$, the set $\bigcup_{j=1}^m \mathbb{B}(p_j, r_j) \cup B$ is polynomially convex. Furthermore, if $B$ is the closure of a bounded strongly pseudoconvex domain with $\mathscr{C}^2$ boundary, then any sufficiently $\mathscr{C}^2$-small deformation of $B$ in $\mathbb{C}^n$ is still polynomially convex.*

The key ingredient in our proofs is the main result of the Andersén–Lempert theory as formulated by Forstnerič and Rosay [1993, Theorem 1.1]; see Theorem 2.3 below. We use it not only for $\mathbb{C}^n$, but also for $X = \mathbb{C}^* \times \mathbb{C}^{n-1}$. The result holds for any Stein manifold which enjoys the following *density property* introduced by Varolin [2001]. (See also [Forstnerič 2011, Definition 4.10.1].)

**Definition 2.2.** A complex manifold $X$ enjoys the (holomorphic) *density property* if every holomorphic vector field on $X$ can be approximated, uniformly on compacts, by Lie combinations (sums and commutators) of $\mathbb{C}$-complete holomorphic vector fields on $X$.

By [Andersén 1990; Andersén and Lempert 1992], the complex Euclidean space $\mathbb{C}^n$ for $n > 1$ enjoys the density property. More generally, Varolin proved that any complex manifold $X = (\mathbb{C}^*)^k \times \mathbb{C}^l$ with $k + l \ge 2$ and $l \ge 1$ enjoys the density property [Varolin 2001]. For surveys of this subject, see for instance [Forstnerič 2011, Chapter 4; Kaliman and Kutzschebauch 2011].

**Theorem 2.3.** *Let $X$ be a Stein manifold with the density property, and let*

$$\Phi_t : \Omega_0 \to \Omega_t = \Phi_t(\Omega_0) \subset X, \quad t \in [0, 1]$$

*be a smooth isotopy of biholomorphic maps of* $\Omega_0$ *onto Runge domains* $\Omega_t \subset X$ *such that* $\Phi_0 = \mathrm{Id}_{\Omega_0}$. *Then, the map* $\Phi_1 : \Omega_0 \to \Omega_1$ *can be approximated uniformly on compacts in* $\Omega_0$ *by holomorphic automorphisms of* $X$.

This is a version of [Forstnerič and Rosay 1993, Theorem 1.1] in which $\mathbb{C}^n$ is replaced by an arbitrary Stein manifold with the density property; see also [Forstnerič 2011, Theorem 4.10.6]. For a detailed proof of Theorem 2.3, see [Forstnerič and Rosay 1993, Theorem 1.1] for the case $X = \mathbb{C}^n$ and [Ritter 2013, Theorem 8] for the general case (which follows the one in [Forstnerič and Rosay 1993] essentially verbatim).

## 3. Construction of a long $\mathbb{C}^n$ without holomorphic functions

In this section, we prove Theorem 1.1. We begin by recalling the general construction of a long $\mathbb{C}^n$; see [Wold 2010] or [Forstnerič 2011, Section 4.20].

Recall that a *Fatou–Bieberbach map* is an injective holomorphic map $\phi : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ such that $\phi(\mathbb{C}^n) \subsetneq \mathbb{C}^n$; the image $\phi(\mathbb{C}^n)$ of such a map is called a *Fatou–Bieberbach domain*. Every complex manifold $X$ which is a long $\mathbb{C}^n$ is determined by a sequence of Fatou–Bieberbach maps $\phi_k : \mathbb{C}^n \to \mathbb{C}^n$ $(k = 1, 2, 3, \ldots)$. The elements of $X$ are represented by infinite strings $x = (x_i, x_{i+1}, \ldots)$, where $i \in \mathbb{N}$ and for every $k = i, i+1, \ldots$ we have $x_k \in \mathbb{C}^n$ and $x_{k+1} = \phi_k(x_k)$. Another string $y = (y_j, y_{j+1}, \ldots)$ determines the same element of $X$ if and only if one of the following possibilities holds:

- $i = j$ and $x_i = y_i$ (and hence $x_k = y_k$ for all $k > i$);
- $i < j$ and $y_j = \phi_{j-1} \circ \cdots \circ \phi_i(x_i)$;
- $j < i$ and $x_i = \phi_{i-1} \circ \cdots \circ \phi_j(y_j)$.

For each $k \in \mathbb{N}$, let $\psi_k : \mathbb{C}^n \hookrightarrow X$ be the injective map sending $z \in \mathbb{C}^n$ to the equivalence class of the string $(z, \phi_k(z), \phi_{k+1}(\phi_k(z)), \ldots)$. Set $X_k = \psi_k(\mathbb{C}^n)$ and let $\iota_k : X_k \hookrightarrow X_{k+1}$ be the inclusion map induced by left shift $(x_k, x_{k+1}, x_{k+2}, \ldots) \mapsto (x_{k+1}, x_{k+2}, \ldots)$. Then

$$\iota_k \circ \psi_k = \psi_{k+1} \circ \phi_k, \quad k = 1, 2, \ldots. \tag{3-1}$$

Recall that a compact set $L$ in a complex manifold $X$ is said to be *holomorphically contractible* if there exist a neighborhood $U \subset X$ of $L$ and a smooth 1-parameter family of injective holomorphic maps $F_t : U \to U$ $(t \in [0, 1))$ such that $F_0$ is the identity map on $U$, $F_t(L) \subset L$ for every $t \in [0, 1]$, and $\lim_{t\to 1} F_t$ is a constant map $L \mapsto p \in L$.

The first part of the following lemma is the key ingredient in the construction of the sequence $(\phi_k)_{k\in\mathbb{N}}$ determining a long $\mathbb{C}^n$ as in Theorem 1.1. The same construction gives the second part, which we include for future applications. We write $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$.

**Lemma 3.1.** *Let* $K$ *be a compact set with nonempty interior in* $\mathbb{C}^n$ *for some* $n > 1$. *For every point* $a \in \mathbb{C}^n$ *there exists an injective holomorphic map* $\phi : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ *such that the polynomial hull of the set* $\phi(K)$ *contains the point* $\phi(a)$. *More generally, if* $L \subset \mathbb{C}^n$ *is a compact holomorphically contractible set*

*disjoint from $K$ such that $K \cup L$ is polynomially convex, then there exists an injective holomorphic map $\phi : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ such that $\phi(L) \subset \widehat{\phi(K)}$ and $\widehat{\phi(K)} \setminus \phi(\mathbb{C}^n) \neq \varnothing$.*

*Proof.* To simplify the notation, we consider the case $n = 2$; it will be obvious that the same proof applies in any dimension $n \geq 2$. We follow Wold's construction [2008; 2010] up to a certain point, adding a new twist at the end.

Let $M$ be a compact set in $\mathbb{C}^* \times \mathbb{C}$ enjoying the following properties:

(1) $M$ is a disjoint union of two smooth, embedded, totally real discs.

(2) $M$ is holomorphically convex in $\mathbb{C}^* \times \mathbb{C}$.

(3) the polynomial hull $\widehat{M}$ of $M$ contains the origin $(0, 0) \in \mathbb{C}^2$.

A set $M$ with these properties was constructed by Stolzenberg [1966]; it has been reproduced in [Stout 1971, pp. 392–396; Wold 2008, Section 2; Forstnerič 2011, Section 4.20].

Choose a Fatou–Bieberbach map $\theta : \mathbb{C}^2 \hookrightarrow \mathbb{C}^* \times \mathbb{C}$ whose image $\theta(\mathbb{C}^2)$ is Runge in $\mathbb{C}^2$. For example, we may take the basin of an attracting fixed point of a holomorphic automorphism of $\mathbb{C}^2$ which fixes $\{0\} \times \mathbb{C}$; see [Rosay and Rudin 1988] for explicit examples. Replacing the set $K$ by its polynomial hull $\widehat{K}$, we may assume that $K$ is polynomially convex. Since $\theta(\mathbb{C}^2)$ is Runge in $\mathbb{C}^2$, the set $\theta(K)$ is also polynomially convex, and hence $\mathscr{O}(\mathbb{C}^* \times \mathbb{C})$-convex. By [Wold 2008, Lemma 3.2], there exists a holomorphic automorphism $\psi \in \mathrm{Aut}(\mathbb{C}^* \times \mathbb{C})$ such that

$$\psi(M) \subset \theta(K^\circ).$$

The construction of such an automorphism $\psi$ uses Theorem 2.3 applied to the manifold $X = \mathbb{C}^* \times \mathbb{C}$. We include a brief outline.

By shrinking each of the two discs in $M$ within themselves until they become very small and then translating them into $K^\circ$ within $\mathbb{C}^* \times \mathbb{C}$, we find an isotopy of diffeomorphisms $h_t : M = M_0 \to M_t \subset \mathbb{C}^* \times \mathbb{C}$ ($t \in [0, 1]$), where each $M_t = h_t(M)$ is a totally real $\mathscr{O}(\mathbb{C}^* \times \mathbb{C})$-convex submanifold of $\mathbb{C}^* \times \mathbb{C}$, such that $M_1 \subset K^\circ$. Since $\mathbb{C}^* \times \mathbb{C}$ has the holomorphic density property (see [Varolin 2001]), each diffeomorphism $h_t$ can be approximated uniformly on $M$ (and even in the smooth topology on $M$) by holomorphic automorphisms of $\mathbb{C}^* \times \mathbb{C}$. This is done in two steps. First, we approximate $h_t$ by a smooth isotopy of biholomorphic maps $f_t : U_0 \to U_t$ from a neighborhood $U_0$ of $M_0$ onto a neighborhood $U_t$ of $M_t$; this is done as in [Forstnerič and Løw 1997]. Since the submanifold $M_t$ is totally real and $\mathscr{O}(\mathbb{C}^* \times \mathbb{C})$-convex for each $t \in [0, 1]$, we can arrange that the neighborhood $U_t$ is Runge in $\mathbb{C}^* \times \mathbb{C}$ for each $t \in [0, 1]$. Hence, Theorem 2.3 furnishes an automorphism $\psi \in \mathrm{Aut}(\mathbb{C}^* \times \mathbb{C})$ which approximates the diffeomorphism $h_1 : M \to M_1$ sufficiently closely such that $\psi(M) \subset B$. It follows that the injective holomorphic map $\tilde{\phi} = \psi^{-1} \circ \theta : \mathbb{C}^2 \hookrightarrow \mathbb{C}^* \times \mathbb{C}$ satisfies $M \subset \tilde{\phi}(K^\circ)$. Note that $K' := \tilde{\phi}(K)$ is a compact $\mathscr{O}(\mathbb{C}^* \times \mathbb{C})$-convex set which contains $M$ in its interior. Therefore, its polynomial hull $\widehat{K'}$ contains a neighborhood of $\widehat{M}$, and hence a neighborhood $V \subset \mathbb{C}^2$ of the origin $(0, 0) \in \mathbb{C}^2$. We may assume that $\overline{V} \cap K' = \varnothing$.

Let $a \in \mathbb{C}^2$. If $\tilde{\phi}(a) \in \widehat{K'}$, then we take $\phi = \tilde{\phi}$ and we are done. If this is not the case, we choose a point $a' \in V \cap (\mathbb{C}^* \times \mathbb{C})$ and apply Theorem 2.3 to find a holomorphic automorphism $\tau \in \mathrm{Aut}(\mathbb{C}^* \times \mathbb{C})$ which is close to the identity map on $K'$ and satisfies $\tau(\tilde{\phi}(a)) = a'$. Such $\tau$ exists since the union of $K'$ with a

single point of $\mathbb{C}^*\times\mathbb{C}$ is $\mathscr{O}(\mathbb{C}^*\times\mathbb{C})$-convex, so it suffices to apply the cited result to an isotopy of injective holomorphic maps which is the identity near $K'$ and which moves $\tilde{\phi}(a)$ to $a'$ in $\mathbb{C}^*\times\mathbb{C}\setminus K'$. Assuming that $\tau$ is sufficiently close to the identity map on $K'$, we have $M\subset\tau(K')$, and hence $a'\in\widehat{M}\subset\widehat{\tau(K')}$. Clearly, the map $\phi=\tau\circ\tilde{\phi}:\mathbb{C}^2\to\mathbb{C}^*\times\mathbb{C}$ satisfies $\phi(a)\in\widehat{\phi(K)}$. This proves the first part of the lemma.

The second part is proved similarly. Since the set $L':=\theta(L)\subset\mathbb{C}^*\times\mathbb{C}$ is holomorphically contractible and $K'\cup L'$ is $\mathscr{O}(\mathbb{C}^*\times\mathbb{C})$-convex, there exists an automorphism $\tau\in\mathrm{Aut}(\mathbb{C}^*\times\mathbb{C})$ which approximates the identity map on $K'$ and satisfies $\tau(L')\subset V\cap(\mathbb{C}^*\times\mathbb{C})$. (To find such $\tau$, we apply Theorem 2.3 to a smooth isotopy $h_t:U\to h_t(U)\subset\mathbb{C}^*\times\mathbb{C}$ $(t\in[0,1])$ of injective holomorphic maps on a small neighborhood $U\subset\mathbb{C}^*\times\mathbb{C}$ of $K'\cup L'$ such that $h_0$ is the identity on $U$, $h_t$ is the identity near $K'$ for every $t\in[0,1]$, and $h_1(L')\subset V$. On the set $L'$, $h_t$ first squeezes $L'$ within itself almost to a point and then moves it to a position within $V$. Clearly, such an isotopy can be found such that $h_t(K'\cup L')=K_t\cup h_t(L')$ is $\mathscr{O}(\mathbb{C}^*\times\mathbb{C})$-convex for all $t\in[0,1]$.) If $\tau$ is sufficiently close to the identity on $K'$, then the polynomial hull $\widehat{\tau(K')}$ still contains $V$, and hence $\tau(L')\subset V\subset\widehat{\tau(K')}$. The map $\phi=\tau\circ\tilde{\phi}:\mathbb{C}^2\to\mathbb{C}^*\times\mathbb{C}$ then satisfies the desired conclusion. □

*Proof of Theorem 1.1.* Pick a compact set $K\subset\mathbb{C}^n$ with nonempty interior and a countable dense sequence $\{a_j\}_{j\in\mathbb{N}}$ in $\mathbb{C}^n$. Set $K_1=\widehat{K}$. Lemma 3.1 furnishes an injective holomorphic map $\phi_1:\mathbb{C}^n\to\mathbb{C}^n$ such that

$$\phi_1(a_1)\in\widehat{\phi_1(K_1)}=:K_2. \tag{3-2}$$

Applying Lemma 3.1 to the set $K_2$ and the point $\phi_1(a_2)\in\mathbb{C}^n$ gives an injective holomorphic map $\phi_2:\mathbb{C}^n\hookrightarrow\mathbb{C}^n$ such that

$$\phi_2(\phi_1(a_2))\in\widehat{\phi_2(K_2)}=:K_3.$$

From the first step we also have $\phi_1(a_1)\in K_2$, and hence $\phi_2(\phi_1(a_1))\in K_3$.

Continuing inductively, we obtain a sequence $\phi_j:\mathbb{C}^n\hookrightarrow\mathbb{C}^n$ of injective holomorphic maps for $j=1,2,\ldots$ such that, setting $\Phi_k=\phi_k\circ\cdots\circ\phi_1:\mathbb{C}^n\hookrightarrow\mathbb{C}^n$, we have

$$\Phi_k(a_j)\in\widehat{\Phi_k(K)}\quad\text{for all } j=1,\ldots,k. \tag{3-3}$$

In the limit manifold $X=\bigcup_{k=1}^{\infty}X_k$ (the long $\mathbb{C}^n$) determined by the sequence $(\phi_k)_{k=1}^{\infty}$, the $\mathscr{O}(X)$-hull of the initial set $K\subset\mathbb{C}^n=X_1\subset X$ clearly contains the set $\Phi_k(K)\subset X_{k+1}$ for each $k=1,2,\ldots$. (We have identified the $k$-th copy of $\mathbb{C}^n$ in the sequence with its image $\psi_k(\mathbb{C}^n)=X_k\subset X$.) It follows from (3-3) that the hull $\widehat{K}_{\mathscr{O}(X)}$ contains the set $\{a_j\}_{j\in\mathbb{N}}\subset\mathbb{C}^n=X_1$. Since this set is everywhere dense in $\mathbb{C}^n$, every holomorphic function on $X$ is bounded on $X_1=\mathbb{C}^n$, and hence constant. By the identity principle, it follows that the function is constant on all of $X$.

The same argument shows that the plurisubharmonic hull $\widehat{K}_{\mathrm{Psh}(X)}$ of $K$ contains the set $A_1:=\{a_j\}_{j\in\mathbb{N}}\subset\mathbb{C}^n\cong X_1$, and hence every plurisubharmonic function $u\in\mathrm{Psh}(X)$ is bounded from above on $A_1$. Since $A_1$ is dense in $X_1$, it follows that $u$ is bounded from above on $X_1$. (This is obvious if $u$ is continuous; the general case follows by observing that $u$ can be approximated from above, uniformly on compacts in $X_1\cong\mathbb{C}^n$, by continuous plurisubharmonic functions.) It follows from Liouville's theorem for plurisubharmonic functions that $u$ is constant on $X_1$.

In order to ensure that $u$ is constant on each copy $X_k \cong \mathbb{C}^n$ ($k \in \mathbb{N}$) in the given exhaustion of $X$, we modify the construction as follows. After choosing the first Fatou–Bieberbach map $\phi_1 : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ such that $\phi_1(a_1) \in \widehat{\phi_1(K)}$ (see (3-2)), we choose a countable dense set $A_2' = \{a_{2,1}', a_{2,2}', \ldots\}$ in $\mathbb{C}^n \setminus \phi_1(\mathbb{C}^n)$ and set $A_2 = \phi_1(A_1) \cup A_2'$ to get a countable dense set in $X_2 \cong \mathbb{C}^n$. Next, we find a Fatou–Bieberbach map $\phi_2 : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ such that the first two points $\phi_1(a_1), \phi_1(a_2)$ of the set $\phi_1(A_1)$, and also the first point $a_{2,1}'$ of $A_2'$, are mapped by $\phi_2$ into the polynomial hull of $\phi_2(\phi_1(K))$. We continue inductively. At the $k$-th stage of the construction we have chosen a Fatou–Bieberbach map $\phi_k : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$, and we take $A_{k+1} = \phi_k(A_k) \cup A_{k+1}'$, where $A_{k+1}'$ is a countable dense set in $\mathbb{C}^n \setminus \phi_k(A_k)$. In the manifold $X$ we thus get an increasing sequence $A_1 \subset A_2 \subset \cdots$ whose union $A := \bigcup_{k=1}^{\infty} A_k$ is dense in $X$ and such that every point of $A$ ends up in the hull $\widehat{K}_{\mathscr{O}(X_k)} = \widehat{K}_{\mathrm{Psh}(X_k)}$ for all sufficiently big $k \in \mathbb{N}$. (See the proof of Theorem 1.6 for more details in a related context.) Hence, the plurisubharmonic hull $\widehat{K}_{\mathrm{Psh}(X)}$ contains the countable dense subset $A$ of $X$. We conclude as before that any plurisubharmonic function on $X$ is bounded on every $X_k \cong \mathbb{C}^n$, and hence constant. □

**Remark 3.2** (Wold process). The key ingredient in the proof of Lemma 3.1 is the method, introduced by E. F. Wold [2008], of stretching the image of a compact set in $\mathbb{C}^* \times \mathbb{C}^{n-1}$ by an automorphism of $\mathbb{C}^* \times \mathbb{C}^{n-1}$ so that its image swallows a compact set $M$ whose polynomial hull in $\mathbb{C}^n$ intersects the hyperplane $\{0\} \times \mathbb{C}^{n-1}$. This is called the *Wold process*. A recursive application of this method, possibly at several places simultaneously and with additional approximation of the identity map on a certain other compact polynomially convex set (see Lemma 4.3), causes the hulls of the respective sets to reach out of all domains $X_k \cong \mathbb{C}^n$ in the exhaustion of $X$.

## 4. Construction of manifolds $X(B)$

In this section, we construct long $\mathbb{C}^n$'s satisfying Theorems 1.2 and 1.6.

We begin by showing that the stable hull property of a compact set in a complex manifold $X$ (see Definition 1.4) is independent of the choice of exhaustion of $X$ by compact sets.

**Lemma 4.1.** *Let $X = \bigcup_{j=1}^{\infty} K_j$, where $K_j \subset K_{j+1}^\circ$ is a sequence of compact sets. Let $B$ be a compact set in $X$. Assume that there exists an integer $j_0 \in \mathbb{N}$ such that $B \subset K_{j_0}$ and*

$$\widehat{B}_{\mathscr{O}(K_j)} = \widehat{B}_{\mathscr{O}(K_{j_0})} \quad \text{for all } j \geq j_0. \tag{4-1}$$

*Then $B$ satisfies the same condition with respect to any exhaustion of $X$ by an increasing sequence of compact sets.*

*Proof.* Set $C := \widehat{B}_{\mathscr{O}(K_{j_0})}$. Let $(L_l)_{l\in\mathbb{N}}$ be another exhaustion of $X$ by compact sets satisfying $L_l \subset L_{l+1}^\circ$ for all $l \in \mathbb{N}$. Pick an integer $l_0 \in \mathbb{N}$ such that $C \subset L_{l_0}$. Since both sequences $K_j^\circ$ and $L_l^\circ$ exhaust $X$, we can find sequences of integers $j_1 < j_2 < j_3 < \cdots$ and $l_1 < l_2 < l_3 < \cdots$ such that $j_0 \leq j_1$, $l_0 \leq l_1$, and

$$K_{j_0} \subset L_{l_1} \subset K_{j_1} \subset L_{l_2} \subset K_{j_2} \subset L_{l_3} \subset \cdots.$$

From this and (4-1) we obtain

$$C = \widehat{B}_{\mathscr{O}(K_{j_0})} \subset \widehat{B}_{\mathscr{O}(L_{l_1})} \subset \widehat{B}_{\mathscr{O}(K_{j_1})} = C \subset \widehat{B}_{\mathscr{O}(L_{l_2})} \subset \widehat{B}_{\mathscr{O}(K_{j_2})} = C \subset \cdots.$$

It follows that $\widehat{B}_{\mathscr{O}(L_{l_j})} = C$ for all $j \in \mathbb{N}$. Since the sequence of hulls $\widehat{B}_{\mathscr{O}(L_l)}$ is increasing with $l$, we conclude that

$$\widehat{B}_{\mathscr{O}(L_l)} = C \quad \text{for all } l \geq l_1.$$

Hence, $B$ has the stable hull property with respect to the exhaustion $(L_l)_{l\in\mathbb{N}}$ of $X$. □

**Remark 4.2.** If a complex manifold $X$ is exhausted by an increasing sequence of Stein domains $X_1 \subset X_2 \subset \cdots \subset \bigcup_{j=1}^{\infty} X_j = X$ (this holds for example if $X$ is a long $\mathbb{C}^n$ or a short $\mathbb{C}^n$, where the latter term refers to a manifold exhausted by biholomorphic copies of the ball), then we can choose an exhaustion $K_1 \subset K_2 \subset \cdots \subset \bigcup_{j=1}^{\infty} K_j = X$ such that $K_j$ is a compact set contained in $X_j$ and $\widehat{(K_j)}_{\mathscr{O}(X_j)} = K_j$ for every $j \in \mathbb{N}$. If $K$ is a compact set contained in some $K_{j_0}$, then clearly $\widehat{K}_{\mathscr{O}(K_j)} = \widehat{K}_{\mathscr{O}(X_j)}$ for all $j \geq j_0$. In such case, $K$ has the stable hull property if and only if the sequence of hulls $\widehat{K}_{\mathscr{O}(X_j)}$ stabilizes. This notion is especially interesting for a long $\mathbb{C}^n$. Imagining the exhaustion $X_j \cong \mathbb{C}^n$ of $X$ as an increasing sequence of universes, the stable hull property means that $K$ only influences finitely many of these universes in a nontrivial way, while a set without SHP has nontrivial influence on all subsequent universes. □

We shall need the following lemma, which generalizes [Wold 2008, Lemma 3.2].

**Lemma 4.3.** *Let $n > 1$. Assume that $B$ is a compact polynomially convex set in $\mathbb{C}^n$, $K_1, \ldots, K_m$ are pairwise disjoint compact sets with nonempty interiors in $\mathbb{C}^n \setminus B$ such that $B \cup \left(\bigcup_{j=1}^m K_j\right)$ is polynomially convex, and $\beta \subset \mathbb{C}^n \setminus \left(B \cup \left(\bigcup_{j=1}^m K_j\right)\right)$ is a finite set. Then there exists a Fatou–Bieberbach map $\phi : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ satisfying the following conditions*:

(i) $\widehat{\phi(B)} = \phi(B)$;

(ii) $\widehat{\phi(K_j)} \not\subset \phi(\mathbb{C}^n)$ *for all* $j = 1, \ldots, m$;

(iii) $\phi(\beta) \subset \widehat{\phi(K_1)}$.

*Furthermore, we can choose $\phi$ such that $\phi|_B$ is as close as desired to the identity map.*

*Proof.* For simplicity of notation we give the proof for $n = 2$; the same argument applies for any $n \geq 2$.

By enlarging $B$ slightly, we may assume that it is a compact strongly pseudoconvex and polynomially convex domain in $\mathbb{C}^n$. Choose a closed ball $\mathcal{B} \subset \mathbb{C}^2$ containing $B$ in its interior. Let $\Lambda \subset \mathbb{C}^2 \setminus \mathcal{B}$ be an affine complex line. Up to an affine change of coordinates on $\mathbb{C}^2$ we may assume that $\Lambda = \{0\} \times \mathbb{C}$.

As in the proof of Lemma 3.1, we find an injective holomorphic map $\theta_1 : \mathbb{C}^2 \hookrightarrow \mathbb{C}^* \times \mathbb{C}$ whose image is Runge in $\mathbb{C}^2$, and hence the set $\theta_1(\mathcal{B})$ is polynomially convex. Since $\mathcal{B}$ is contractible, we can connect the identity map on $\mathcal{B}$ to $\theta_1|_{\mathcal{B}}$ by an isotopy of biholomorphic maps $h_t : \mathcal{B} \to \mathcal{B}_t$ ($t \in [0, 1]$) with Runge images in $\mathbb{C}^* \times \mathbb{C}$. Theorem 2.3 furnishes an automorphism $\theta_2 \in \mathrm{Aut}(\mathbb{C}^* \times \mathbb{C})$ such that $\theta_2$ approximates $\theta_1^{-1}$ on $\theta_1(\mathcal{B})$. The composition $\theta = \theta_2 \circ \theta_1 : \mathbb{C}^2 \hookrightarrow \mathbb{C}^* \times \mathbb{C}$ is then an injective holomorphic map which is close to the identity on $\mathcal{B}$. Assuming that the approximation is close enough, the set $B' := \theta(B)$ is polynomially convex in view of Lemma 2.1.

Set $K = \bigcup_{j=1}^m K_j$, $K'_j = \theta(K_j)$ for $j = 1, \ldots, m$, and $K' = \theta(K) = \bigcup_{j=1}^m K'_j$. Note that the set $B' \cup K' = \theta(B \cup K) \subset \mathbb{C}^* \times \mathbb{C}$ is $\mathscr{O}(\mathbb{C}^* \times \mathbb{C})$-convex.

Choose $m$ pairwise disjoint copies $M_1, \ldots, M_m \subset (\mathbb{C}^* \times \mathbb{C}) \setminus B'$ of Stolzenberg's [1966] compact set $M$ described in the proof of Lemma 3.1. Explicitly, each set $M_j$ is $\mathscr{O}(\mathbb{C}^* \times \mathbb{C})$-convex and its polynomial

hull $\widehat{M}_j$ intersects the complex line $\{0\}\times\mathbb{C}$ (which lies in the complement of $\theta(\mathbb{C}^2)$). By placing the sets $M_j$ sufficiently far apart and away from $B'$, we may assume that the compact set $B'\cup\left(\bigcup_{j=1}^m M_j\right)$ is $\mathscr{O}(\mathbb{C}^*\times\mathbb{C})$-convex. Pick a slightly bigger compact set $B''\subset\theta(\mathbb{C}^2)$, containing $B'$ in its interior, such that the sets $B''\cup\left(\bigcup_{j=1}^m K_j'\right)$ and $B''\cup\left(\bigcup_{j=1}^m M_j\right)$ are still $\mathscr{O}(\mathbb{C}^*\times\mathbb{C})$-convex.

We claim that for every $\epsilon>0$ there is an automorphism $\psi\in\mathrm{Aut}(\mathbb{C}^*\times\mathbb{C})$ such that

(a) $|\psi(z)-z|<\epsilon$ for all $z\in B''$, and

(b) $\psi(M_j)\subset K_j'$ for $j=1,\dots,m$.

To obtain such a $\psi$, we apply the construction in the proof of Lemma 3.1 to find an isotopy of smooth diffeomorphisms

$$h_t : M=\bigcup_{j=1}^m M_j \to M^t=h_t(M)\subset\mathbb{C}^*\times\mathbb{C},\quad t\in[0,1],$$

such that $h_0=\mathrm{Id}\,|_M$, the set $M^t=\bigcup_{j=1}^m h_t(M_j)$ consists of smooth totally real submanifolds, $B''\cap M^t=\varnothing$ for all $t\in[0,1]$, $B''\cup M^t$ is $\mathscr{O}(\mathbb{C}^*\times\mathbb{C})$-convex for all $t\in[0,1]$, and $h_1(M_j)\subset K_j'^{\circ}$ for $j=1,\dots,m$. It follows that $h_1$ can be approximated uniformly on $M$ by a holomorphic automorphism $\psi\in\mathrm{Aut}(\mathbb{C}^*\times\mathbb{C})$ which at the same time approximates the identity map on $B''$. (For the details in a similar context, see [Forstnerič and Rosay 1993, proof of Theorem 2.3] or [Forstnerič 2011, proof of Corollary 4.12.4].) The injective holomorphic map

$$\phi:=\psi^{-1}\circ\theta:\mathbb{C}^2\hookrightarrow\mathbb{C}^*\times\mathbb{C}$$

then approximates the identity map on a neighborhood of $B$ and satisfies $M_j\subset\phi(K_j)$ for $j=1,\dots,m$. It follows that

$$\widehat{\phi(K_j)}\cap(\{0\}\times\mathbb{C})\neq\varnothing\quad\text{for all } j=1,\dots,m.$$

If the approximation $\psi|_{B''}\approx\mathrm{Id}$ in (a) is close enough, then the set $\phi(B)=\psi^{-1}(B')$ is still polynomially convex by Lemma 2.1. Clearly, $\phi$ satisfies properties (i) and (ii), and property (iii) can be achieved by applying Lemma 3.1. □

*Proof of Theorem 1.6(a).* Let $B$ be the given regular compact polynomially convex set in $\mathbb{C}^n$. To begin the induction, set $B_1:=B\subset\mathbb{C}^n=X_1$ and choose a pair of disjoint countable set

$$A_1=\{a_1^l : l\in\mathbb{N}\}\subset\mathbb{C}^n\setminus B_1,\qquad \overline{A}_1=\mathbb{C}^n\setminus B_1^\circ,$$
$$\Gamma_1=\{\gamma_1^l : l\in\mathbb{N}\}\subset\mathbb{C}^n\setminus(A_1\cup B_1),\quad \overline{\Gamma}_1=\mathbb{C}^n\setminus B_1^\circ.$$

Let $\mathbb{B}(a_1^1,r_1)$ denote the closed ball of radius $r_1$ centered at $a_1^1$. By choosing $r_1>0$ small enough we may ensure that $\mathbb{B}(a_1^1,r_1)\cap B_1=\varnothing$, $\gamma_1^1\notin\mathbb{B}(a_1^1,r_1)\cup B_1$, and the set $\mathbb{B}(a_1^1,r_1)\cup B_1$ is polynomially convex (see Lemma 2.1). Lemma 4.3 furnishes an injective holomorphic map $\phi_1:\mathbb{C}^n\hookrightarrow\mathbb{C}^n$ such that the set $B_2:=\phi_1(B_1)\subset\mathbb{C}^n$ is polynomially convex, while the compact set

$$C_{1,1}^1:=\phi_1(\mathbb{B}(a_1^1,r_1))\subset\mathbb{C}^n$$

satisfies

$$\widehat{C_{1,1}^1}\setminus\phi_1(\mathbb{C}^n)\neq\varnothing\quad\text{and}\quad\phi_1(\gamma_1^1)\in\widehat{C_{1,1}^1}.$$

We proceed recursively. Suppose that for some $k \in \mathbb{N}$ we have found

- injective holomorphic maps $\phi_1, \dots, \phi_k : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$,
- compact polynomially convex sets $B_1, B_2, \dots, B_{k+1}$ in $\mathbb{C}^n$ such that $B_{i+1} = \phi_i(B_i)$ for $i = 1, \dots, k$,
- countable sets $A_1, \dots, A_k \subset \mathbb{C}^n$ such that for every $i = 1, \dots, k$ we have

$$A_i \subset \mathbb{C}^n \setminus B_i, \quad \overline{A}_i = \mathbb{C}^n \setminus B_i^\circ, \quad A_i = \phi_{i-1}(A_{i-1}) \cup \{a_i^l : l \in \mathbb{N}\}$$

  (where we set $A_0 = \varnothing$),

- countable sets $\Gamma_1, \dots, \Gamma_k \subset \mathbb{C}^n$ such that for every $i = 1, \dots, k$ we have

$$\Gamma_i \subset \mathbb{C}^n \setminus A_i \cup B_i, \quad \overline{\Gamma}_i = \mathbb{C}^n \setminus B_i^\circ, \quad \Gamma_i = \phi_{i-1}(\Gamma_{i-1}) \cup \{\gamma_i^l : l \in \mathbb{N}\}$$

  (where we set $\Gamma_0 = \varnothing$), and

- numbers $r_1 > \cdots > r_k > 0$

such that, setting for all $(i, l) \in \mathbb{N}^2$ with $1 \le i + l \le k + 1$

$$b_{k,i}^l := \phi_{k-1} \circ \cdots \circ \phi_i(a_i^l) \in A_k \quad \text{if } (i,l) \ne (k,1), \qquad b_{k,k}^1 := a_k^1,$$
$$\beta_{k,i}^l := \phi_{k-1} \circ \cdots \circ \phi_i(\gamma_i^l) \in \Gamma_k \quad \text{if } (i,l) \ne (k,1), \qquad \beta_{k,k}^1 := \gamma_k^1,$$

the following conditions hold for all pairs $(i, l) \in \mathbb{N}^2$ with $i + l \le k + 1$:

($1_k$) the closed balls $\mathbb{B}(b_{k,i}^l, r_k)$ are pairwise disjoint and contained in $\mathbb{C}^n \setminus B_k$, and

$$\{\beta_{k,i}^l : i + l \le k + 1\} \cap \bigcup_{i+l \le k+1} \mathbb{B}(b_{k,i}^l, r_k) = \varnothing$$

(since $A_k \cap \Gamma_k = \varnothing$, the latter condition holds provided $r_k > 0$ is small enough);

($2_k$) the set $\bigcup_{i+l \le k+1} \mathbb{B}(b_{k,i}^l, r_k) \cup B_k$ is polynomially convex;

($3_k$) the set $(\phi_{k-1} \circ \cdots \circ \phi_i)^{-1}(\mathbb{B}(b_{k,i}^l, r_k))$ is contained in $\mathbb{B}(a_i^l, r_i/2^k)$;

($4_k$) the set $C_{k,i}^l := \phi_k(\mathbb{B}(b_{k,i}^l, r_k))$ satisfies $\widehat{C_{k,i}^l} \setminus \phi_k(\mathbb{C}^n) \ne \varnothing$;

($5_k$) $\{\phi_k(\beta_{k,i}^l) : i + l \le k + 1\} \subset \widehat{C_{k,1}^1}$.

We now explain the inductive step. We begin by adding to $\phi_k(A_k)$ countably many points in $\mathbb{C}^n \setminus (\phi_k(A_k) \cup B_{k+1})$ to get a countable set

$$A_{k+1} = \phi_k(A_k) \cup \{a_{k+1}^l : l \in \mathbb{N}\} \subset \mathbb{C}^n \setminus B_{k+1}$$

such that

$$\overline{A}_{k+1} = \mathbb{C}^n \setminus B_{k+1}^\circ.$$

In the same way, we find the next countable set

$$\Gamma_{k+1} = \phi_k(\Gamma_k) \cup \{\gamma_{k+1}^l : l \in \mathbb{N}\} \subset \mathbb{C}^n \setminus (A_{k+1} \cup B_{k+1})$$

such that

$$\overline{\Gamma}_{k+1} = \mathbb{C}^n \setminus B_{k+1}^\circ.$$

For every pair of indices $(i, l) \in \mathbb{N}^2$ with $i + l \le k + 2$, we set

$$b_{k+1,i}^l := \phi_k \circ \cdots \circ \phi_i(a_i^l) \in A_{k+1} \quad \text{if } (i,l) \ne (k+1,1), \qquad b_{k+1,k+1}^1 := a_{k+1}^1,$$
$$\beta_{k+1,i}^l := \phi_k \circ \cdots \circ \phi_i(\gamma_i^l) \in \Gamma_{k+1} \quad \text{if } (i,l) \ne (k+1,1), \qquad \beta_{k+1,k+1}^1 := \gamma_{k+1}^1.$$

Choose a number $r_{k+1}$ with $0 < r_{k+1} < r_k$ and so small that the following conditions hold for all $(i, l) \in \mathbb{N}^2$ with $i + l \le k + 2$:

$(1_{k+1})$ the closed balls $\mathbb{B}(b_{k+1,i}^l, r_{k+1})$ are pairwise disjoint and contained in $\mathbb{C}^n \setminus B_{k+1}$, and

$$\{\beta_{k+1,i}^l : i + l \le k+2\} \cap \Big( \bigcup_{i+l \le k+2} \mathbb{B}(b_{k+1,i}^l, r_{k+1}) \cup B_{k+1} \Big) = \varnothing;$$

$(2_{k+1})$ the set $\bigcup_{i+l\le k+2} \mathbb{B}(b_{k+1,i}^l, r_{k+1}) \cup B_{k+1}$ is polynomially convex;

$(3_{k+1})$ the set $(\phi_k \circ \cdots \circ \phi_i)^{-1}(\mathbb{B}(b_{k+1,i}^l, r_{k+1}))$ is contained in $\mathbb{B}(a_i^l, r_i/2^{k+1})$.

Lemma 4.3 gives a Fatou–Bieberbach map $\phi_{k+1} : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ such that the compact set $B_{k+2} := \phi_{k+1}(B_{k+1})$ is polynomially convex, while the compact sets

$$C_{k+1,i}^l := \phi_{k+1}(\mathbb{B}(b_{k+1,i}^l, r_{k+1})), \quad i + l \le k + 2$$

satisfy the following conditions:

$(4_{k+1})$ $\widehat{C_{k+1,i}^l} \setminus \phi_{k+1}(\mathbb{C}^n) \ne \varnothing$ for all $(i, l) \in \mathbb{N}^2$ with $i + l \le k + 2$;

$(5_{k+1})$ $\{\phi_{k+1}(\beta_{k+1,i}^l) : i + l \le k + 2\} \subset \widehat{C_{k+1,1}^1}$.

This completes the induction step and the recursion may continue.

Let $X = \bigcup_{k=1}^\infty X_k$ be the long $\mathbb{C}^n$ determined by the sequence $(\phi_k)_{k=1}^\infty$. Since the set $B_k \subset \mathbb{C}^n$ is polynomially convex and $B_{k+1} = \phi_k(B_k)$ for all $k \in \mathbb{N}$, the sequence $(B_k)_{k\in\mathbb{N}}$ determines a subset $B = B_1 \subset X$ such that

$$\widehat{B}_{\mathscr{O}(X_k)} = B \quad \text{for all } k \in \mathbb{N}. \tag{4-2}$$

This means that the initial compact set $B \subset \mathbb{C}^n = X_1$ has the stable hull property in $X$.

By the construction, the countable sets $A_k \subset \mathbb{C}^n \setminus B_k$ satisfy $\phi_k(A_k) \subset A_{k+1}$ for each $k \in \mathbb{N}$, and hence they determine a countable set $A \subset X \setminus B$. Furthermore, since $\overline{A}_k = \mathbb{C}^n \setminus B_k^\circ$ for every $k \in \mathbb{N}$, it follows that $\overline{A} = X \setminus B^\circ$. Similarly, the family $(\Gamma_k)_{k\in\mathbb{N}}$ determines a countable set $\Gamma \subset X \setminus B$ such that $\overline{\Gamma} = X \setminus B^\circ$.

We now show that $B$ is the biggest regular compact set in $X$ with the stable hull property. Note that condition $(4_k)$, together with the fact that each set $C_{k,i}^l$ contains one of the sets $C_{k+1,i'}^{l'}$ in the next generation according to condition $(3_{k+1})$ (and hence it contains one of the sets $C_{k+j,i'}^{l'}$ for every $j = 1, 2, \ldots$), implies

$$\widehat{(C_{k,i}^l)}_{\mathscr{O}(X_{k+j+1})} \setminus X_{k+j} \ne \varnothing \quad \text{for all } j = 0, 1, 2, \ldots. \tag{4-3}$$

Thus, none of the sets $C^l_{k,i}$ has the stable hull property. Our construction ensures that the centers of these sets form a dense sequence in $X \setminus B$, consisting of all points in the set $A$ determined by the family $(A_k)_{k\in\mathbb{N}}$, in which every point appears infinitely often. Furthermore, condition $(3_k)$ shows that every compact set $K \subset X$ with $K^\circ \setminus B \neq \varnothing$ contains one (in fact, infinitely many) of the sets $C^l_{k,i}$. In view of (4-3), it follows that there is an integer $k_0 \in \mathbb{N}$ (depending on $K$) such that

$$\widehat{K}_{\mathscr{O}(X_{k+1})} \not\subset X_k \quad \text{for all } k \geq k_0.$$

This means that $K$ does not have the stable hull property. It follows that the set $B$ is the strongly stable core of $X$.

Finally, condition $(5_k)$ ensures that the $\mathscr{O}(X)$-hull of a compact ball centered at the point $a^1_1 \in A$ contains the countable set $\Gamma \subset X$ determined by the family $\{\Gamma_k\}_{k\in\mathbb{N}}$. Since $\Gamma$ is dense in $X \setminus B$, it follows that the manifold $X$ does not admit any nonconstant plurisubharmonic function. (See the proof of Theorem 1.1 for the details.)

This proves part (a) of Theorem 1.6. □

*Proof of Theorem 1.6(b).* Let $U \subset \mathbb{C}^n$ be an open set. Pick a regular compact polynomially convex set $B$ contained in $U$. We modify the recursion in the proof of part (a) by adding to $B$ a new small closed ball $B' \subset U \setminus B$ at every stage. In this way, we inductively build an increasing sequence $B = B^1 \subset B^2 \subset \cdots \subset U$ of compact polynomially convex sets whose union $\mathcal{B} := \bigcup_{k=1}^\infty B^k \subset U$ is everywhere dense in $U$, and a sequence of Fatou–Bieberbach maps $\phi_k : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ such that, writing

$$B^k_1 = B^k \quad \text{and} \quad B^k_{j+1} = \phi_j(B^k_j) \text{ for all } j, k \in \mathbb{N},$$

the following two conditions hold:

(i) $B^k = B^{k-1} \cup \mathcal{B}^k$ for all $k > 1$, where $\mathcal{B}^k$ is a small closed ball in $U \setminus B^{k-1}$;

(ii) the set $B^k_j$ is polynomially convex for all $j, k \in \mathbb{N}$.

At the $k$-th stage of the construction we have already chosen Fatou–Bieberbach maps $\phi_1, \ldots, \phi_k$, but we can nevertheless achieve condition (ii) for all $j = 1, \ldots, k+1$ by choosing the ball $\mathcal{B}^k$ sufficiently small. Indeed, the image of a small ball by an injective holomorphic map is a small strongly convex domain, and hence the polynomial convexity of the set $B^k_j$ for $j = 1, \ldots, k+1$ follows from Lemma 2.1. For values $j > k+1$, (ii) is achieved by the construction in the proof of Lemma 4.3; indeed, each of the subsequent maps $\phi_{k+1}, \phi_{k+2}, \ldots$ in the sequence preserves polynomial convexity of $B^k_{k+1}$.

By identifying the sets $U$ and $B^k = B^k_1$ (considered as subsets of $\mathbb{C}^n = X_1$) with their images in the limit manifold $X = \bigcup_{k=1}^\infty X_k$, we thus obtain the following analogue of (4-2):

$$\widehat{(B^k)}_{\mathscr{O}(X_j)} = B^k \quad \text{for all } j, k \in \mathbb{N}.$$

This means that each set $B^k$ ($k \in \mathbb{N}$) lies in the stable core $\mathrm{SC}(X)$. Since $\bigcup_{k=1}^\infty B^k$ is dense in $U$ by the construction, we have that $\overline{U} \subset \overline{\mathrm{SC}(X)}$.

On the other hand, writing $U_1 = U$ and $U_{k+1} := \phi_k \circ \cdots \circ \phi_1(U)$ for $k = 1, 2, \ldots$, the balls $\mathbb{B}(b^l_{k,i}, r_k)$ chosen at the $k$-th stage of the construction (see the proof of part (a)) are contained in $\mathbb{C}^n \setminus \overline{U}$ and, as

$k$ increases, they include more and more points from a countable dense set $A \subset X \setminus \overline{U}$, which is built inductively as in the proof of part (a). By performing the Wold process on each of the balls $\mathbb{B}(b_{k,i}^l, r_k)$ (see condition $(4_k)$ above) at every stage, we can ensure that none of the points of $A$ belongs to the stable core $\mathrm{SC}(X)$. Since $\mathrm{SC}(X)$ is an open set by the definition and $\overline{A} = X \setminus U$, we conclude that $\mathrm{SC}(X) \subset U$. We have seen above that $\overline{U} \subset \overline{\mathrm{SC}(X)}$, and hence $\overline{\mathrm{SC}(X)} = \overline{U}$.

It remains to show that $X$ can be chosen such that it does not admit any nonconstant holomorphic function. By the same argument as in the proof of part (a), we can find a countable dense set $\Gamma \subset X \setminus (A \cup \overline{U})$ which is dense in $X \setminus U$ and is contained in the $\mathscr{O}(X)$-hull of a certain compact set in $X \setminus \overline{U}$. It follows that every plurisubharmonic function $f$ on $X$ is bounded above on $\Gamma$, and hence on $\overline{\Gamma} = X \setminus U$. If $\overline{U}$ is compact, the maximum principle implies that $f$ is also bounded on $U$; hence it is bounded on $X$ and therefore constant. If $U$ is not relatively compact then we are unable to make this conclusion. However, we can easily ensure that $X \setminus \overline{U}$ contains a Fatou–Bieberbach domain; indeed, it suffices to choose the first Fatou–Bieberbach map $\phi_1 : \mathbb{C}^n \to \mathbb{C}^n$ in the sequence determining $X$ such that $\mathbb{C}^n \setminus \phi_1(\mathbb{C}^n)$ contains a Fatou–Bieberbach domain $\Omega$. In this case, every holomorphic function $f \in \mathscr{O}(X)$ is bounded on $\Gamma$, and hence on $\Omega$, so it is constant on $\Omega \cong \mathbb{C}^n$. Therefore it is constant on $X$ by the identity principle.

This proves part (b) and hence completes the proof of Theorem 1.6. □

## 5. An exhaustion of $\mathbb{C}^2$ by non-Runge Fatou–Bieberbach domains

In this section, we show the following result (see also [Boc Thaler 2016, Section 4.4]).

**Proposition 5.1.** *Let $n > 1$. There exists an increasing sequence $X_1 \subset X_2 \subset \cdots \subset \bigcup_{k=1}^{\infty} X_k = \mathbb{C}^n$ of Fatou–Bieberbach domains in $\mathbb{C}^n$ which are not Runge in $\mathbb{C}^n$.*

We shall construct such an example by ensuring that all Fatou–Bieberbach maps $\phi_k : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ in the sequence (see Section 3) have non-Runge images, but they approximate the identity map on increasingly large balls centered at the origin. For this purpose, we shall need the following lemma.

**Lemma 5.2.** *Let $B$ and $\mathcal{B}$ be a pair of closed disjoint balls in $\mathbb{C}^n$ $(n > 1)$. For every $\epsilon > 0$ there exists a Fatou–Bieberbach map $\phi : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ satisfying the following conditions*:

(a) $\|\phi|_B - \mathrm{Id}\| < \varepsilon$;

(b) $\|\phi^{-1}|_B - \mathrm{Id}\| < \varepsilon$;

(c) *$\phi(\mathcal{B})$ is not polynomially convex.*

*Proof.* Pick a slightly bigger ball $B'$ containing $B$ in the interior such that $B' \cap \mathcal{B} = \varnothing$. By an affine linear change of coordinates, we may assume that $B' \subset \mathbb{C}^* \times \mathbb{C}^{n-1}$. Choose a Fatou–Bieberbach map $\theta : \mathbb{C}^n \hookrightarrow \mathbb{C}^* \times \mathbb{C}^{n-1}$ such that $\theta|_{B'}$ is close to the identity. (See the proof of Lemma 4.3.) Theorem 2.3 provides a $\psi \in \mathrm{Aut}(\mathbb{C}^* \times \mathbb{C}^{n-1})$ which approximates the identity map on $\theta(B')$ and such that $\psi(\theta(\mathcal{B}))$ is not polynomially convex (in fact, its polynomial hull intersects the hyperplane $\{0\} \times \mathbb{C}^{n-1}$). The composition $\phi = \psi \circ \theta : \mathbb{C}^n \hookrightarrow \mathbb{C}^* \times \mathbb{C}^{n-1}$ then satisfies condition (a) on $B'$, and condition (c). If $\phi$ is sufficiently close to the identity on $B'$, then it also satisfies condition (b) since $B \subset B'^{\circ}$. □

*Proof of Proposition 5.1.* Let $B_k = \mathbb{B}(0,k) \subset \mathbb{C}^n$ denote the closed ball of radius $k$ centered at the origin. Choose an integer $n_1 \in \mathbb{N}$ and a small ball $\mathcal{B}^1$ disjoint from $B_{n_1}$. Let $\phi_1 : \mathbb{C}^n \to \mathbb{C}^n$ be a Fatou–Bieberbach map satisfying the following conditions:

(1) $\|\phi_1 - \mathrm{Id}\| < \varepsilon_1$ on $B_{n_1}$;

(2) $\|\phi_1^{-1} - \mathrm{Id}\| < \varepsilon_1$ on $B_{n_1}$;

(3) $\phi(\mathcal{B}^1)$ is not polynomially convex.

Suppose inductively that for some $k \in \mathbb{N}$ we have already found Fatou–Bieberbach maps $\phi_1, \dots, \phi_k$, integers $n_1 < n_2 < \cdots < n_k$, and balls $\mathcal{B}^j \subset \mathbb{C}^n \setminus B_{n_j}$ for $j = 1, \dots, k$ such that the following conditions hold:

$(1_k)$ $\|\phi_k - \mathrm{Id}\| < \varepsilon_k$ on $B_{n_k}$;

$(2_k)$ $\|\phi_k^{-1} - \mathrm{Id}\| < \varepsilon_k$ on $B_{n_k}$;

$(3_k)$ $\phi_k(\mathcal{B}^k)$ is not polynomially convex.

Choose an integer $n_{k+1} > n_k$ such that

$$\phi_k(B_{n_k+1}) \cup \phi_k(\phi_{k-1}(B_{n_{k-1}+2})) \cup \cdots \cup \phi_k\big(\cdots(\phi_1(B_{n_1+k}))\big) \cup \phi_k(\mathcal{B}^k) \subset B_{n_{k+1}},$$

and pick a ball $\mathcal{B}^{k+1} \subset \mathbb{C}^n \setminus B_{n_{k+1}}$. By Lemma 5.2, there exists a Fatou–Bieberbach map $\phi_{k+1} : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ satisfying the following conditions:

$(1_{k+1})$ $\|\phi_{k+1} - \mathrm{Id}\| < \varepsilon_{k+1}$ on $B_{n_{k+1}}$;

$(2_{k+1})$ $\|\phi_{k+1}^{-1} - \mathrm{Id}\| < \varepsilon_{k+1}$ on $B_{n_{k+1}}$;

$(3_{k+1})$ $\phi_{k+1}(\mathcal{B}^{k+1})$ is not polynomially convex.

This closes the induction step.

Let $X = \bigcup_{k=1}^{\infty} X_k$ be the long $\mathbb{C}^n$ determined by the sequence $(\phi_k)_k$, let $\iota_k : X_k \hookrightarrow X_{k+1}$ denote the inclusion map, and let $\psi_k : \mathbb{C}^n \to X_k \subset X$ denote the biholomorphic map from $\mathbb{C}^n$ onto the $k$-th element of the exhaustion such that

$$\iota_k \circ \psi_k = \psi_{k+1} \circ \phi_k, \quad k = 1, 2, \dots.$$

(See Section 3, in particular (3-1).) By the construction, the sequence $\psi_k(B_{n_k})$ is a Runge exhaustion of $X$. If the sequence $\varepsilon_k > 0$ has been chosen to be summable, then the sequence $\psi_k$ converges on every ball $B_{n_j}$ and the limit map $\Psi = \lim_{k\to\infty} \psi_k : \mathbb{C}^n \to X$ is a biholomorphism (see [Forstnerič 2011, Corollary 4.4.2, p. 115]). In the terminology of Dixon and Esterle [1986, Theorem 5.2], we have that

$$(\psi_k, B_{n_k}) \to (\Psi, \mathbb{C}^n) \quad \text{as } k \to \infty,$$

where $\Psi(\mathbb{C}^n) = X$ and $\Psi$ is biholomorphic. □

**Remark 5.3.** If we only assume that the images of Fatou–Bieberbach maps $\phi_k : \mathbb{C}^n \hookrightarrow \mathbb{C}^n$ contain large enough balls centered at the origin, we get an exhaustion of a long $\mathbb{C}^n$ with Runge images of balls. By [Arosio et al. 2013, Theorem 3.4], such a long $\mathbb{C}^n$ is biholomorphic to a Stein Runge domain in $\mathbb{C}^n$. Therefore, the following problem is closely related to problem (C) stated in the introduction.

(C′) If a long $\mathbb{C}^n$ is exhausted by Runge images of balls, is it necessarily biholomorphic to $\mathbb{C}^n$?

In this connection, we mention that the first author proved in his thesis [Boc Thaler 2016, Theorem IV.15, p. 62] that $\mathbb{C}^n$ is the only Stein manifold with the density property (see Definition 2.2) having an exhaustion by Runge images of balls.

## Acknowledgements

The authors are partially supported by the research program P1-0291 and the grant J1-5432 from ARRS, Republic of Slovenia. A part of the work on this paper was done while F. Forstnerič was visiting the Erwin Schrödinger Institute in Vienna, Austria, and the Faculty of Science, University of Granada, Spain, in November 2015. He wishes to thank both institutions for the invitation and excellent working conditions. The authors express their gratitude to the anonymous referees for their valuable remarks which helped us to improve and streamline the presentation.

## References

[Abbondandolo et al. 2013] A. Abbondandolo, L. Arosio, J. E. Fornæss, P. Majer, H. Peters, J. Raissy, and L. Vivas, "A survey on non-autonomous basins in several complex variables", 2013. arXiv 1311.3835

[Andersén 1990] E. Andersén, "Volume-preserving automorphisms of $\mathbf{C}^n$", *Complex Variables Theory Appl.* **14**:1-4 (1990), 223–235. MR 1048723 Zbl 0705.58008

[Andersén and Lempert 1992] E. Andersén and L. Lempert, "On the group of holomorphic automorphisms of $\mathbf{C}^n$", *Invent. Math.* **110**:2 (1992), 371–388. MR 1185588 Zbl 0770.32015

[Arosio et al. 2013] L. Arosio, F. Bracci, and E. F. Wold, "Solving the Loewner PDE in complete hyperbolic starlike domains of $\mathbb{C}^N$", *Adv. Math.* **242** (2013), 209–216. MR 3055993 Zbl 1283.30054

[Baouendi et al. 1999] M. S. Baouendi, P. Ebenfelt, and L. P. Rothschild, *Real submanifolds in complex space and their mappings*, Princeton Mathematical Series **47**, Princeton University Press, 1999. MR 1668103 Zbl 0944.32040

[Behnke and Stein 1939] H. Behnke and K. Stein, "Konvergente Folgen von Regularitätsbereichen und die Meromorphiekonvexität", *Math. Ann.* **116**:1 (1939), 204–216. MR 1513225 Zbl 64.0322.03

[Behnke and Thullen 1934] H. Behnke and P. Thullen, "Zur Theorie der Funktionen mehrerer komplexer Veränderlichen", *Math. Ann.* **109**:1 (1934), 313–323. MR 1512896 Zbl 0008.36504

[Boc Thaler 2016] L. Boc Thaler, *Fatou components*, Ph.D. thesis, University of Ljubljana, 2016, http://www.matknjiz.si/doktorati/2016/Boc-Thaler-14521-20.pdf.

[Brown 1961] M. Brown, "The monotone union of open $n$-cells is an open $n$-cell", *Proc. Amer. Math. Soc.* **12** (1961), 812–814. MR 0126835 Zbl 0103.39305

[Chern and Moser 1974] S. S. Chern and J. K. Moser, "Real hypersurfaces in complex manifolds", *Acta Math.* **133** (1974), 219–271. MR 0425155 Zbl 0302.32015

[Dixon and Esterle 1986] P. G. Dixon and J. Esterle, "Michael's problem and the Poincaré–Fatou–Bieberbach phenomenon", *Bull. Amer. Math. Soc.* (*N.S.*) **15**:2 (1986), 127–187. MR 854551 Zbl 0608.32008

[Docquier and Grauert 1960] F. Docquier and H. Grauert, "Levisches Problem und Rungescher Satz für Teilgebiete Steinscher Mannigfaltigkeiten", *Math. Ann.* **140** (1960), 94–123. MR 0148939 Zbl 0095.28004

[Fefferman 1974] C. Fefferman, "The Bergman kernel and biholomorphic mappings of pseudoconvex domains", *Invent. Math.* **26** (1974), 1–65. MR 0350069 Zbl 0289.32012

[Fornæss 1976] J. E. Fornæss, "An increasing sequence of Stein manifolds whose limit is not Stein", *Math. Ann.* **223**:3 (1976), 275–277. MR 0417448 Zbl 0334.32017

[Fornæss 1978] J. E. Fornæss, "A counterexample for the Levi problem for branched Riemann domains over $\mathbf{C}^n$", *Math. Ann.* **234**:3 (1978), 275–277. MR 0492397 Zbl 0427.32014

[Fornæss 2004] J. E. Fornæss, "Short $\mathbb{C}^k$", pp. 95–108 in *Complex analysis in several variables—Memorial Conference of Kiyoshi Oka's Centennial Birthday*, edited by K. Miyajima et al., Adv. Stud. Pure Math. **42**, Math. Soc. Japan, Tokyo, 2004. MR 2087041 Zbl 1071.32007

[Fornæss and Sibony 1981] J. E. Fornæss and N. Sibony, "Increasing sequences of complex manifolds", *Math. Ann.* **255**:3 (1981), 351–360. MR 615855 Zbl 0438.32012

[Fornæss and Stout 1977] J. E. Fornæss and E. L. Stout, "Polydiscs in complex manifolds", *Math. Ann.* **227**:2 (1977), 145–153. MR 0435441 Zbl 0331.32007

[Forstnerič 1986] F. Forstnerič, "Stability of polynomial convexity of totally real sets", *Proc. Amer. Math. Soc.* **96**:3 (1986), 489–494. MR 822446 Zbl 0588.32016

[Forstnerič 1992] F. Forstnerič, "An elementary proof of Fefferman's theorem", *Exposition. Math.* **10**:2 (1992), 135–149. MR 1164529 Zbl 0759.32018

[Forstnerič 1993] F. Forstnerič, "Proper holomorphic mappings: a survey", pp. 297–363 in *Several complex variables* (Stockholm, 1987/1988), edited by J. E. Fornæss, Math. Notes **38**, Princeton University Press, 1993. MR 1207867 Zbl 0778.32008

[Forstnerič 2011] F. Forstnerič, *Stein manifolds and holomorphic mappings*, Ergebnisse der Mathematik (3) **56**, Springer, Heidelberg, 2011. MR 2975791 Zbl 1247.32001

[Forstnerič 2012] F. Forstnerič, "Holomorphic families of long $\mathbb{C}^2$'s", *Proc. Amer. Math. Soc.* **140**:7 (2012), 2383–2389. MR 2898700 Zbl 1278.32012

[Forstnerič and Løw 1997] F. Forstnerič and E. Løw, "Global holomorphic equivalence of smooth submanifolds in $\mathbf{C}^n$", *Indiana Univ. Math. J.* **46**:1 (1997), 133–153. MR 1462799 Zbl 0883.32014

[Forstnerič and Rosay 1993] F. Forstnerič and J.-P. Rosay, "Approximation of biholomorphic mappings by automorphisms of $\mathbf{C}^n$", *Invent. Math.* **112**:2 (1993), 323–349. MR 1213106 Zbl 0792.32011

[Kaliman and Kutzschebauch 2011] S. Kaliman and F. Kutzschebauch, "On the present state of the Andersén–Lempert theory", pp. 85–122 in *Affine algebraic geometry*, edited by D. Daigle et al., CRM Proc. Lecture Notes **54**, American Mathematical Society, Providence, RI, 2011. MR 2768636 Zbl 1266.32028

[Lárusson 2010] F. Lárusson, "What is . . . an Oka manifold?", *Notices Amer. Math. Soc.* **57**:1 (2010), 50–52. MR 2590114 Zbl 1191.32004

[Pinchuk and Khasanov 1987] S. I. Pinchuk and S. V. Khasanov, "Asymptotically holomorphic functions and their applications", *Mat. Sb.* (*N.S.*) **134(176)**:4 (1987), 546–555, 576. In Russian; translated in *Math. USSR-Sb.* **62**:2 (1989), 541–550. MR 933702 Zbl 0639.32005

[Poincaré 1907] H. Poincaré, "Les fonctions analytiques de deux variables et la représentation conforme", *Rend. Circ. Mat. Palermo* **23** (1907), 185–220.

[Ritter 2013] T. Ritter, "A strong Oka principle for embeddings of some planar domains into $\mathbb{C} \times \mathbb{C}^*$", *J. Geom. Anal.* **23**:2 (2013), 571–597. MR 3023850 Zbl 1269.32012

[Rosay and Rudin 1988] J.-P. Rosay and W. Rudin, "Holomorphic maps from $\mathbf{C}^n$ to $\mathbf{C}^n$", *Trans. Amer. Math. Soc.* **310**:1 (1988), 47–86. MR 929658 Zbl 0708.58003

[Stein 1956] K. Stein, "Überlagerungen holomorph-vollständiger komplexer Räume", *Arch. Math.* (*Basel*) **7** (1956), 354–361. MR 0084836 Zbl 0072.08002

[Stolzenberg 1966] G. Stolzenberg, "The analytic part of the Runge hull", *Math. Ann.* **164** (1966), 286–290. MR 0203080 Zbl 0141.27303

[Stout 1971] E. L. Stout, *The theory of uniform algebras*, Bogden & Quigley, Tarrytown-on-Hudson, NY, 1971. MR 0423083 Zbl 0286.46049

[Stout 2007] E. L. Stout, *Polynomial convexity*, Progress in Mathematics **261**, Birkhäuser, Boston, 2007. MR 2305474 Zbl 1124.32007

[Varolin 2001] D. Varolin, "The density property for complex manifolds and geometric structures", *J. Geom. Anal.* **11**:1 (2001), 135–160. MR 1829353 Zbl 0994.32019

[Wermer 1959] J. Wermer, "An example concerning polynomial convexity", *Math. Ann.* **139** (1959), 147–150. MR 0121500 Zbl 0094.28302

[Wold 2008] E. F. Wold, "A Fatou–Bieberbach domain in $\mathbb{C}^2$ which is not Runge", *Math. Ann.* **340**:4 (2008), 775–780. MR 2372737 Zbl 1137.32002

[Wold 2010] E. F. Wold, "A long $\mathbb{C}^2$ which is not Stein", *Ark. Mat.* **48**:1 (2010), 207–210. MR 2594593 Zbl 1189.32004

Received 4 Mar 2016. Revised 19 Jul 2016. Accepted 28 Aug 2016.

LUKA BOC THALER: luka.boc@pef.uni-lj.si
*Faculty of Education, University of Ljubljana, Kardeljeva ploščad 16, SI-1000 Ljubljana, Slovenia*
and
*Institute of Mathematics, Physics and Mechanics, Jadranska 19, SI-1000 Ljubljana, Slovenia*

FRANC FORSTNERIČ: franc.forstneric@fmf.uni-lj.si
*Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, SI-1000 Ljubljana, Slovenia*
and
*Institute of Mathematics, Physics and Mechanics, Jadranska 19, SI-1000 Ljubljana, Slovenia*

## Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at msp.org/apde.

**Originality**. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

**Language**. Articles in APDE are usually in English, but articles written in other languages are welcome.

**Required items**. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

**Format**. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

**References**. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

**Figures**. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

**White space**. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

**Proofs**. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

# Analysis & PDE

Volume 9 No. 8 2016