Title

The threshold conjecture for the energy critical hyperbolic Yang--Mills equation

Permalink

https://escholarship.org/uc/item/2mh8x71c

Authors

Oh, Sung-Jin
Tataru, Daniel

Publication Date

2017-09-25

Peer reviewed

THE THRESHOLD CONJECTURE FOR THE ENERGY CRITICAL HYPERBOLIC YANG–MILLS EQUATION

SUNG-JIN OH AND DANIEL TATARU

ABSTRACT. This article represents the fourth and final part of a four-paper sequence whose aim is to prove the Threshold Conjecture as well as the more general Dichotomy Theorem for the energy critical $4 + 1$ dimensional hyperbolic Yang–Mills equation. The Threshold Theorem asserts that topologically trivial solutions with energy below twice the ground state energy are global and scatter. The Dichotomy Theorem applies to solutions in arbitrary topological class with large energy, and provides two exclusive alternatives: Either the solution is global and scatters, or it bubbles off a soliton in either finite time or infinite time.

Using the caloric gauge developed in the first paper [37], the continuation/scattering criteria established in the second paper [38], and the large data analysis in an arbitrary topological class at optimal regularity in the third paper [39], here we perform a blow-up analysis which shows that the failure of global well-posedness and scattering implies either the existence of a soliton with at most the same energy bubbling off, or the existence existence of a nontrivial self-similar solution. The proof is completed by showing that the latter solutions do not exist.

CONTENTS

1. Introduction

This article represents the fourth and last of a four-paper sequence devoted to the study of finite energy solutions to the energy critical $4 + 1$ dimensional hyperbolic Yang–Mills equation. The four installments of the series are concerned with

(a) the *caloric gauge* for the hyperbolic Yang–Mills equation, [37];
(b) large data *energy dispersed* caloric gauge solutions, [38];
(c) *topological classes* of connections and large data local well-posedness, [39];
(d) the Threshold Conjecture and soliton bubbling/scattering dichotomy for large data solutions, present article.

A short overview of the four papers is provided in the survey paper [40].

Our first goal in this paper is to prove the Threshold Conjecture for the hyperbolic Yang–Mills equation; this asserts that the solution is global and scatters for all topologically trivial data with energy below $2E_{\mathrm{GS}}$, where E_{GS} represents energy of the ground state (i.e., lowest energy steady state) for this problem.

Secondly, we consider solutions with energy above this threshold, and prove the following Dichotomy Theorem: either (i) the solution is topologically trivial, global and scatters, or (ii) it "bubbles off" a soliton in either finite time (which corresponds to blow-up) or in infinite time. Here "soliton bubbling off" means that a sequence of symmetry- and gauge-equivalent solutions must converge to a soliton, namely a Lorentz transform of a steady state.

As a main common component of both theorems, we separately state and prove a Bubbling Theorem, which provides a necessary and sufficient condition for soliton bubbling off purely in terms of the the energy distribution of the solution.

The paper is organized as follows. In the first section we provide some background material on the hyperbolic Yang–Mills equation, and then we give the statements of the main results in Theorem 1.7 (the Bubbling Theorem), Theorem 1.8 (the Threshold Theorem), and finally Theorem 1.11 (the Dichotomy Theorem). In the second section we provide a brief overview of the results in the first three papers of the sequence [37],[38],[39]. The remainder of the paper is devoted to the proof of the main results.

1.1. **The energy critical hyperbolic Yang–Mills equation.**

1.1.1. *Lie groups and algebras.* Let $\mathbf{G}$ be a compact noncommutative Lie group and $\mathfrak{g}$ its associated Lie algebra. We denote by $Ad(O)X = OXO^{-1}$ the action of $\mathbf{G}$ on $\mathfrak{g}$ by conjugation (i.e., the adjoint action), and by $ad(X)Y = [X, Y]$ the associated action of $\mathfrak{g}$, which is given by the Lie bracket. We introduce the notation $\langle X, Y \rangle$ for a bi-invariant inner product on $\mathfrak{g}$,

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

or equivalently

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

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

1.1.2. *The hyperbolic Yang–Mills equation.* Let $\mathbb{R}^{1+4}$ be the $(4+1)$ dimensional Minkowski space with the standard Lorentzian metric $\mathbf{m} = \mathrm{diag}(-1, 1, 1, 1, 1)$. Denote by $A_\alpha : \mathbb{R}^{1+4} \to \mathfrak{g}$, $\alpha = 0, 1, \ldots, 4$, a connection 1-form[1] taking values in the Lie algebra $\mathfrak{g}$, and by $\mathbf{D}_\alpha$ the associated covariant differentiation,

$$\mathbf{D}_\alpha B := \partial_\alpha B + [A_\alpha, B],$$

acting on $\mathfrak{g}$-valued functions B. The commutator of two covariant derivatives takes the form $(\mathbf{D}_\alpha \mathbf{D}_\beta - \mathbf{D}_\beta \mathbf{D}_\alpha) B = [F_{\alpha\beta}, B]$, where F is the curvature tensor

$$F_{\alpha\beta} := \partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta].$$

The curvature tensor obeys the *Bianchi identity*, namely

$$\mathbf{D}_\alpha F_{\beta\gamma} + \mathbf{D}_\beta F_{\gamma\alpha} + \mathbf{D}_\gamma F_{\alpha\beta} = 0.$$

The (hyperbolic) *Yang–Mills* equation for A is the Euler–Lagrange equation associated with the formal Lagrangian action functional

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

Here we are using the standard convention of raising or lowering indices using the metric $\mathbf{m}$, as well as summing up repeated upper and lower indices. Thus, the Yang–Mills equation takes the form

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

There is a natural energy-momentum tensor associated to the Yang–Mills equation, namely

$$T_{\alpha\beta}(A) = 2\langle F_\alpha{}^\gamma, F_{\beta\gamma} \rangle - \frac{1}{2} \mathbf{m}_{\alpha\beta} \langle F_{\gamma\delta}, F^{\gamma\delta} \rangle.$$

If A solves the Yang–Mills equation (1.1) then $T_{\alpha\beta}$ is divergence free,

$$\partial^\alpha T_{\alpha\beta} = 0. \tag{1.2}$$

Integrating this for $\beta = 0$ yields the *conserved energy*

$$\mathcal{E}(A) = \mathcal{E}_{\{t\}\times\mathbb{R}^4}(A) = \int_{\{t\}\times\mathbb{R}^4} T_{00} \, dx = \int_{\{t\}\times\mathbb{R}^4} \frac{1}{2} \langle F_{jk}, F^{jk} \rangle + \langle F_{0j}, F_0{}^j \rangle \, dx, \tag{1.3}$$

which is constant in time. Here we are using the convention of using roman indices $j, k, \ldots$ for the spatial indices $\{1, \ldots, 4\}$. For a general subset $U \subset \mathbb{R}^4$, we define the local energy in U to be

$$\mathcal{E}_{\{t\}\times U}(A) = \int_{\{t\}\times U} T_{00} \, dx.$$

---

[1] The geometric setting for the hyperbolic Yang–Mills equation is the space of connections on a vector bundle on a Lorentzian manifold; here, for simplicity, we give a concrete formulation on $\mathbb{R}^{1+4}$. For a more geometric description, we refer the reader to [39].

1.1.3. *Symmetries.* The group of symmetries for the the Yang–Mills equation play a key role in our analysis. Its components are as follows:
   (1) Translations, both in space and in time;
   (2) the Lorentz group of linear coordinate changes;
   (3) the scaling group,
$$A(t, x) \to \lambda A(\lambda t, \lambda x).$$
The conserved energy functional $\mathcal{E}$ is invariant with respect to scaling precisely in dimension $4 + 1$. For this reason we call the $4 + 1$ problem *energy critical*; this is one of the motivations for our interest in this problem.

1.1.4. *Gauge invariance and Yang–Mills solutions.* In order to study the Yang–Mills equation as a well-defined evolution in time, we first need to address its gauge invariance. Given a **G**-valued function O on $\mathbb{R}^{1+4}$, we introduce the notation
$$O_{;\alpha} = \partial_\alpha O O^{-1}.$$
Such a function O induces the gauge transformation
$$A_\alpha \longrightarrow \mathcal{G}(O)A := Ad(O)A_\alpha - O_{;\alpha},$$
under which equation (1.1) is invariant. In order to uniquely determine the solutions to the Yang–Mills equation, one needs to add an additional set of constraint equations which uniquely determine a gauge. This procedure is known as *gauge fixing*.

The choice of a gauge plays a central role in the study of the Yang–Mills equation. There are multiple interesting classical gauge choices, e.g. the Lorenz gauge, the temporal gauge and the Coulomb gauge. Neither of these is well-suited for the global (in spacetime) large data problem, and a main goal of our first paper [37] is to introduce a better alternate gauge choice, namely the *caloric gauge*. We briefly return to the issue of gauge choice in Section 1.2, and then give a more detailed discussion in Section 2.

1.1.5. *Initial data sets.* In order to consider the Yang–Mills problem as an evolution equation we need to consider its initial data sets. An *initial data set* for (1.1) is a pair of $\mathfrak{g}$-valued 1-forms (a_j, e_j) on $\mathbb{R}^4$. We say that (a_j, e_j) is the initial data set for a Yang–Mills solution A if
$$(A_j, F_{0j})|_{\{t=0\}} = (a_j, e_j).$$
Note that (1.1) imposes the condition that the following equation be true for any initial data for (1.1):
$$\mathbf{D}^j e_j = 0. \tag{1.4}$$
Here, $\mathbf{D}^j$ denotes the covariant derivative with respect to the a_j connection. This equation is the *Gauss* (or the *constraint*) *equation* for (1.1). In what follows, we denote by $f = f_{ij}$ the curvature of a. We refer to Section 1.4 for the notation $H^1(\mathcal{O})$, $H^1_{loc}(\mathcal{O})$ etc. concerning function spaces.

Definition 1.1. a) A *regular initial data set* for the Yang–Mills equation is a pair $(a_j, e_j) \in H^N_{loc} \times H^{N-1}$, $N \geq 2$, also with $f \in H^{N-1}$, which has finite energy and satisfies the constraint equation (1.4).

b) A *finite energy initial data set* for the Yang–Mills equation is a pair $(a_j, e_j) \in H^1_{loc} \times L^2$, with $f \in L^2$, and which satisfies the constraint equation (1.4).

4

We remark that the family of regular initial data sets is dense in the class of finite energy data. This is not entirely trivial due to the nonlinear constraint equation.

1.1.6. *Yang–Mills solutions.* Due to the gauge invariance properties, we need to be more careful than usual about what we call a solution to the hyperbolic Yang–Mills equation:

Definition 1.2. a) Let $N \geq 2$. A *regular solution* to the Yang–Mills equation in an open set $\mathcal{O} \subset \mathbb{R}^{1+4}$ is a connection A in $\mathcal{O}$ obeying $(A, \partial_t A) \in C_t H^N_{loc} \times C_t H^{N-1}_{loc}(\mathcal{O})$, whose curvature satisfies $F \in C_t H^{N-1}_{loc}(\mathcal{O})$ and which solves the equation (1.1).
b) A *finite energy solution* to the Yang–Mills equation in the open set $\mathcal{O}$ is a connection A obeying $(A, \partial_t A) \in C_t H^1_{loc} \times C_t L^2_{loc}(\mathcal{O})$, whose curvature satisfies $F \in C_t L^2(\mathcal{O})$ and which is the limit of regular solutions in this topology.

We carefully remark that this definition does not require a gauge choice. Hence, at this point solutions are still given by equivalence classes. Corresponding to the above classes of solutions, we have the classes of gauge transformations which preserve them:

Definition 1.3. a) Let $N \geq 2$. A *regular gauge transformation* in an open set $\mathcal{O} \subset \mathbb{R}^{1+4}$ is is a map

$$O : \mathcal{O} \to \mathbf{G}$$

with the following regularity properties:

$$(O_{;t,x}, \partial_t O_{;t,x}) \in C_t H^{N+1}_{loc} \times C_t H^N_{loc}(\mathcal{O}).$$

b) An *admissible gauge transformation* in an open set $\mathcal{O} \subset \mathbb{R}^{1+4}$ is a similar map with the following regularity properties:

$$(O_{;t,x}, \partial_t O_{;t,x}) \in C_t H^1_{loc} \times C_t L^2_{loc}(\mathcal{O}).$$

Using this notion we can now talk about gauge-equivalent connections:

Definition 1.4. Two finite energy connections $A^{(1)}$ and $A^{(2)}$ in an open set $\mathcal{O} \subset \mathbb{R}^{1+4}$ are gauge equivalent if there exists an admissible gauge transformation O so that

$$A^{(2)}_\alpha = \mathcal{G}(O)A^{(1)}(= OA^{(1)}_\alpha O^{-1} - \partial_\alpha O O^{-1}).$$

We list some simple properties of finite energy connections and admissible gauge transformations in an open set $\mathcal{O}$ (see [39]):
- If $A^{(1)}$ and $A^{(2)}$ are finite energy gauge-equivalent connections then the bounds for the corresponding gauge transformation O depend only on the corresponding bounds for $A^{(1)}$ and $A^{(2)}$.
- If $A^{(1)}$ and $A^{(2)}$ are regular gauge-equivalent connections then the corresponding gauge transformation O is also regular, with uniform bounds in terms of $A^{(1)}$, $A^{(2)}$.
- The family of regular admissible gauge transformations is dense in the family of admissible gauge transformations.
- If $A^{(1)}$ and $A^{(2)}$ are gauge-equivalent finite energy connections, then $A^{(1)}$ is a finite energy solution to the Yang–Mills equation (1.1) if and only if $A^{(2)}$ is.
- If A is a finite energy connection then its equivalence class $[A]$ is closed in the corresponding topology.

In terms of local well-posedness, it is easier to work in a gauge. At this point we know that (see the more detailed discussion in Section 2):

(i) Small data global well-posedness holds in the Coulomb gauge [30], caloric gauge [38] and temporal gauge [38].

(ii) Large data local well-posedness holds for large caloric data in the caloric gauge [38] and for arbitrary large data in the temporal gauge [39].

(iii) Uniqueness of finite energy solutions (up to gauge transformations) [39].

1.1.7. *Topological classes.* The family of finite energy Yang–Mills data sets (a, e) is not a connected topological space in the above topologies. Instead, they are classified according to their *topological class*, see [39] and also the discussion in Section 2. The topological class is easily seen to be preserved dynamically for both regular and finite energy solutions to the hyperbolic Yang–Mills equation.

A special role in the present paper is played by the class $[0]$ of 0, whose elements we call *topologically trivial connections.* These have the equivalent characterization that they can be described using $\dot{H}^1$ connections [39], see Theorem 2.15 below. The topologically trivial connections are the subject of both the first two papers [37] and [38] in our four-paper series, as well as of the Threshold Theorem below.

For our purposes here, we will use a specific topological invariant, namely the *characteristic number* $\boldsymbol{\chi}$ defined by

$$\boldsymbol{\chi}(a) = \int_{\mathbb{R}^4} -\langle f \wedge f \rangle = \frac{1}{4} \int_{\mathbb{R}^4} -\langle f_{ij}, f_{k\ell} \rangle \, \mathrm{d}x^i \wedge \mathrm{d}x^j \wedge \mathrm{d}x^k \wedge \mathrm{d}x^\ell,$$

which depends only on the topological class $[a]$ of a. Two key properties of $\boldsymbol{\chi}$ are that $\boldsymbol{\chi}([0]) = 0$ and the pointwise bound

$$|\langle f \wedge f \rangle| \leq \frac{1}{2} \langle f_{jk}, f^{jk} \rangle \leq T_{00}(a), \tag{1.5}$$

which is referred to as the *Bogomoln'yi bound*; see [39] for their proofs.

In the case $\mathbf{G} = SU(2)$ the topological class of a is fully described by the characteristic number $\boldsymbol{\chi}$, which is in fact a multiple of the second Chern number c_2 computed from a. The Chern number c_2 turns out to be an integer, and each such integer defines a connected component in the space of finite energy connections in $\mathbb{R}^4$. For a general Lie group $\mathbf{G}$ the characteristic number $\boldsymbol{\chi}$ of a connection a provides only a partial description of the topological class of a.

1.1.8. *Harmonic Yang–Mills connections and the ground state.* A *harmonic Yang–Mills connection* in $\mathbb{R}^4$ is a $\dot{H}^1_{loc}$ connection a which is a critical point for the (static) energy functional

$$\mathcal{E}_e(a) = \int_{\mathbb{R}^4} \frac{1}{2} \langle f, f \rangle.$$

On the one hand they are the steady states for the hyperbolic Yang–Mills flow, and on the other hand they are celebrated objects in geometric analysis with spectacular applications to four-dimensional topology; see [7].

The Euler–Lagrange equation satisfied by a takes the form

$$\mathbf{D}^\ell f_{\ell j} = 0,$$

which becomes an elliptic system for a in a suitable gauge (e.g. Coulomb).

The key elliptic regularity result is as follows:

Theorem 1.5 (Uhlenbeck [55, 56]). *Harmonic Yang–Mills connections $a \in H^1_{loc}(\mathbb{R}^4)$ are smooth in a suitable gauge. More generally, $H^{\frac{d}{2}}_{loc}$ harmonic Yang–Mills connections in any d-dimensional Riemannian manifold (M, g) are smooth.*

As far as the energy of harmonic Yang–Mills connections and the energy in different topological classes is concerned, the key properties are as follows, see [39]:

Theorem 1.6. *Let $\mathbf{G}$ be a noncommutative compact Lie group. Let*

$$E_{GS} = \inf\{\mathcal{E}_e(a) : a \text{ is a nontrivial harmonic Yang–Mills connection on a } \mathbf{G}\text{-bundle on } \mathbb{R}^4\}.$$

Then the following statements hold.

(1) *There exists a nontrivial harmonic Yang–Mills connection a so that $\mathcal{E}_e(a) = E_{GS} < \infty$.*
(2) *Let a be any nontrivial harmonic Yang–Mills connection. Then either $\mathcal{E}_e(a) \geq 2E_{GS}$, or*

$$|\boldsymbol{\chi}| = \mathcal{E}_e(a) \geq E_{GS}.$$

This result is a combination of classical results [2, 7] concerning energy minimizing solutions within a topological class (called *instantons*), as well as a recent energy lower bound for the non-minimizing solutions proved by Gursky–Kelleher–Streets [16]. For a derivation, see [39, Section 6].

As a corollary, Theorem 1.6 shows that in the class of topologically trivial connections, harmonic Yang–Mills connections must have energy at least $2E_{GS}$. Based on this, we will call *subthreshold data/solution* any topologically trivial hyperbolic Yang–Mills data/solution with energy below $2E_{GS}$.

1.2. **The main results.** We consider the Cauchy problem for the hyperbolic Yang–Mills equation (1.1) with finite energy data (a, e). As discussed earlier, this problem is known to be locally well-posed [38, 39] for large data and globally well-posed for small data [30]. Here we are interested in the global large data problem, and we seek to address the following two questions:

- *Global well-posedness*;
- *Scattering of the solution.*

Preliminary remarks on each point in relation to the notion of the topological class of solutions are in order.

Global well-posedness. Because of the finite speed of propagation and the small data result, a classical argument shows that at the blow-up time T we must have energy concentration in a backward light cone centered at a point (T, X),

$$\underline{C}^{(T,X)} = \{(t, x) \in \mathbb{R}^{1+4} : |x - X| < T - t\}.$$

In particular, the question of global well-posedness is of local nature, i.e., has nothing to do with the topological class of the initial data.

Scattering. In a classical sense, a solution A for the Yang–Mills equation would be scattering if as t approaches infinity, $A(t)$ approaches a free wave. Such a definition is unrealistic in our situation. In the first place, it is gauge-dependent. Secondly, the small data result in [30] shows that, even in a favorable gauge, classical scattering cannot occur, and instead one needs to consider some form of *modified scattering*. Even so, there is no chance of scattering

7

unless the solution A is topologically trivial; this is due to the fact that any solution which decays in a scale invariant L^p norm for $p > 2$ must be topologically trivial. We refer to Remark 1.10 for a description of our notion of scattering.

We now present our main results, which are divided into two classes. The first consists of a gauge-independent bubbling off result. In a nutshell, it asserts that time-like energy concentration implies soliton bubbling off. To state it, we need some notation. Given a backward (resp. forward) light cone

$$^{(T,X)}\underline{C} = \{(t,x) \in \mathbb{R}^{1+4} : |x - X| < T - t\}$$

$$\left(\text{resp. } {}^{(T,X)}C = \{(t,x) \in \mathbb{R}^{1+4} : |x - X| < t - T\}\right),$$

we introduce the time slices

$$^{(T,X)}\underline{S}_t = {}^{(T,X)}\underline{C} \cap (\{t\} \times \mathbb{R}^4)$$

$$\left(\text{resp. } {}^{(T,X)}S_t = {}^{(T,X)}\underline{C} \cap (\{t\} \times \mathbb{R}^4)\right),$$

and for $0 < \gamma < 1$, the time-like cone

$$^{(T,X)}\underline{C}_\gamma = \{(t,x) \in \mathbb{R}^{1+4} : |x - X| < \gamma(T - t)\}$$

$$\left(\text{resp. } {}^{(T,X)}C_\gamma = \{(t,x) \in \mathbb{R}^{1+4} : |x - X| < \gamma(t - T)\right).$$

When the tip (T, X) coincides with the spacetime origin, we omit the superscript (T, X) and write $C = {}^{(0,0)}C$, $S_t = {}^{(0,0)}S$, $C_\gamma = {}^{(0,0)}C_\gamma$ etc.

For any future time-like vector, which in general takes the form $(1, v)$ with $|v| < 1$, we denote by L_v the Lorentz transformation[2] with velocity v.

Then we have:

Theorem 1.7 (Bubbling Theorem). *a) Let A be a finite energy Yang–Mills connection which blows up in finite time at (T, X). Assume in addition that for some $0 < \gamma < 1$ we have*

$$\limsup_{t \nearrow T} \mathcal{E}_{(T,X)\underline{C}_\gamma \cap (T,X)\underline{S}_t}(A) > 0. \tag{1.6}$$

Then there exists a sequence of points $^{(T,X)}\underline{C} \ni (t_n, x_n) \to (T, X)$ and scales $r_n > 0$ with the following properties:

(1) Time-like concentration,

$$\limsup_{n \to \infty} \frac{x_n - X}{|t_n - T|} = v, \qquad \text{for some } |v| < 1.$$

(2) Below self-similar scale,

$$\limsup_{n \to \infty} \frac{r_n}{|t_n - T|} = 0.$$

(3) Convergence to soliton:

$$\lim_{n \to \infty} r_n \mathcal{G}(O_n) A(t_n + r_n t, x_n + r_n x) = L_v Q(t, x) \quad \text{in } H^1_{loc}([-1/2, 1/2] \times \mathbb{R}^4)$$

for some sequence of admissible gauge transformations O_n and finite energy harmonic Yang–Mills connection Q. Here, L_v is the Lorentz transformation with velocity v.

---

[2]More concretely, when $v \neq 0$, L_v is the linear transformation on $\mathbb{R}^{1+4}$ that preserves $\mathbf{m}$, maps $(1, v)$ to $(\sqrt{1 - |v|^2}, 0)$ and equals the identity in $\text{span}\{(1, 0), (1, v)\}^\perp$. When $v = 0$, L_0 is simply the identity.

b) Let A be a finite energy Yang–Mills connection which is global forward in time. Assume in addition that for some $0 < \gamma < 1$ we have

$$\limsup_{t \nearrow \infty} \mathcal{E}_{C_\gamma \cap S_t}(A) > 0, \tag{1.7}$$

where we recall that $C = {}^{(0,0)}C$, $S_t = {}^{(0,0)}S_t$, $C_\gamma = {}^{(0,0)}C_\gamma$ etc[3]. Then there exists a sequence of points $C \ni (t_n, x_n) \to \infty$ and scales $r_n > 0$ with the following properties:

(1) Time-like concentration,

$$\limsup_{n \to \infty} \frac{x_n}{t_n} = v, \qquad \text{for some } |v| < 1.$$

(2) Below self-similar scale,

$$\limsup_{n \to \infty} \frac{r_n}{t_n} = 0.$$

(3) Convergence to soliton:

$$\lim_{n \to \infty} r_n \mathcal{G}(O_n) A(t_n + r_n t, x_n + r_n x) = L_v Q(t, x) \quad \text{in } H^1_{loc}([-1/2, 1/2] \times \mathbb{R}^4)$$

for some sequence of admissible gauge transformations O_n and finite energy harmonic Yang–Mills connection Q. Here, L_v is the Lorentz transformation with velocity v.

Next, we turn to the second class of main results, which concern global well-posedness and scattering properties of (1.1). For this, we need to briefly introduce our gauge choices:

Caloric gauge. This is our main choice of gauge, in which we have the strongest gauge-dependent control of solutions. We say that a connection a on $\mathbb{R}^4$ is in *caloric gauge* if its Yang–Mills heat flow

$$\partial_s A_j(x, s) = \mathbf{D}^\ell F_{\ell j}(x, s), \qquad A_j(x, s = 0) = a_j(x)$$

exists globally in heat-time s and $\lim_{s \to \infty} A(s) = 0$. Denoting by $\mathcal{C}$ the manifold of finite energy caloric connections, and by $T^{L^2}\mathcal{C}$ the completion of its tangent space in L^2, a solution to the Yang–Mills equation in the caloric gauge can be interpreted as a continuous curve $(A_x, \partial_t A_x)(t)$ in $T^{L^2}\mathcal{C}$ (see [37] and Section 2.1).

The Yang–Mills equation written in this gauge has a favorable structure, akin to the classical Coulomb gauge. But in contrast to the Coulomb gauge, the caloric gauge may be imposed for all subthreshold data (to be discussed below), making it a natural setting for the Threshold Theorem.

Temporal gauge. This is a classical gauge defined by the condition

$$A_0 = 0,$$

which plays an auxiliary role in our work. The structure of (1.1) in this gauge is less favorable, but nevertheless it has the advantage of respecting causality (i.e., finite speed of propagation) of (1.1).

Direct analysis of (1.1) in the temporal gauge at energy regularity is fraught with difficulties; however, we observe a suitable structure in the caloric gauge, which allows us to transfer some (but not all) bounds to the temporal gauge. These bounds are enough to establish small energy global well-posedness, which can then be turned into large data local

---

[3]Since this part concerns the limit $t \to \infty$, the precise choice of the tip $(0, 0)$ is irrelevant; any choice leads to an equivalent statement.

well-posedness in temporal gauge by causality (see [39] and Section 2.3). This result provides a suitable setting for considering evolution of arbitrary finite energy data, albeit with more indirect control.

We refer to the beginning of Section 2 for a further discussion of various gauges that arise in our work.

We now present the Threshold Theorem, which asserts global well-posedness and scattering for initial data with energy below a sharp threshold. In view of existence of solitons, which are counterexamples for scattering, the threshold may first appear to be the ground state energy E_{GS}. However, as we aim for scattering, we would need to limit ourselves to the class of topologically trivial connections, in which the ground state energy is $2E_{\mathrm{GS}}$ by Theorem 1.6. Thus our result is as follows:

Theorem 1.8 (Threshold Theorem). *The Yang–Mills equation* (1.1) *is globally well-posed in the caloric gauge for all topologically trivial initial data below the energy threshold $2E_{GS}$ and the corresponding solutions scatter in the following sense:*

a) (Regular data) For regular data $(a_j, b_{0j}) \in T^{L^2}\mathcal{C} \cap \dot{\mathcal{H}}^N$, then there exists a unique global regular caloric solution $(A_j, \partial_0 A_j) \in C(\mathbb{R}, T^{L^2}\mathcal{C} \cap \dot{\mathcal{H}}^N)$, also with $(A_0, \partial_0 A_0) \in C(\mathbb{R}, \dot{\mathcal{H}}^1 \cap \dot{\mathcal{H}}^N)$, which has a Lipschitz dependence on the initial data locally in time in the $\dot{\mathcal{H}} \cap \dot{\mathcal{H}}^N$ topology.

b) (Rough data) The flow map admits an extension

$$T^{L^2}\mathcal{C} \ni (a_j, b_j) \to (A_\alpha, \partial_t A_\alpha) \in C(\mathbb{R}, T^{L^2}\mathcal{C})$$

and which is continuous in the $\mathcal{H} \cap \dot{\mathcal{H}}^\sigma$ topology for $\sigma < 1$ and close to 1.

c) (Weak Lipschitz dependence) The flow map is globally Lipschitz in the $\dot{\mathcal{H}}^\sigma$ topology for $\sigma < 1$, close to 1.

d) (Scattering) The S^1 norm of A is finite. More precisely,

$$\|A_x\|_{S^1} + \|\nabla A_0\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}} < \infty. \tag{1.8}$$

Here, $\dot{\mathcal{H}}^\sigma = \dot{H}^\sigma \times \dot{H}^{\sigma-1}$. For the norm S^1, see Remark 1.10. The norm $\ell^1 L^2 \dot{H}^{\frac{1}{2}}$ is defined as $\|u\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}} = \sum_k \|P_k u\|_{L^2_t \dot{H}^{\frac{1}{2}}_x}$; see Section 1.4 below for our notation and conventions.

Remark 1.9. The preceding theorem is stated for initial data $(a, b) \in T^{L^2}\mathcal{C}$ which are already in the caloric gauge. However, by our results on the Yang–Mills heat flow (in particular, the corresponding Threshold Theorem), any topologically trivial gauge covariant Yang–Mills data set $(\tilde{a}, e) \in \dot{H}^1 \times L^2$ with energy below $2E_{\mathrm{GS}}$ admits a gauge-equivalent caloric data set (a, b), with appropriate dependence properties; see Section 2.1 below.

Remark 1.10. The S^1 norm represents, with only minor changes, the same combination of Strichartz, $X^{s,b}$ and null frame norms previously used in the study of the Maxwell-Klein-Gordon equation [29, 36] and the small data problem for Yang–Mills in [30]. The S^1 bound on A_x implies a host of other dispersive bounds in the caloric gauge, including Strichartz bounds, renormalizability property, elliptic bounds for A_0 etc. In particular, finiteness of the S^1 norm of A_x can be viewed as a scattering statement, as it shows that the caloric solutions decay in Strichartz and other norms.

Our final result, which both extends and complements the Threshold Theorem, allows for data which are either topologically nontrivial, or are topologically trivial but above the $2E_{\mathrm{GS}}$

threshold. It aims to establish the full dichotomy between the bubbling off property on the one hand, and the global well-posedness and scattering on the other:

Theorem 1.11 (Dichotomy Theorem). *The Yang–Mills equation* (1.1) *is locally well-posed in the temporal gauge for arbitrary finite energy data. Further, one of the following two properties must hold for the forward maximal solution:*

  a) The solution is topologically trivial, global, and scatters at infinity $(t = \infty)$.
  b) The solution bubbles off a soliton, in the sense that either

   (1) it blows up in finite time and the conclusion of Theorem 1.7(a) holds; or
   (2) it exists globally (forward in time) and the conclusion of Theorem 1.7(b) holds.

Of course, by time reversibility, the same conclusion holds backward in time as well.

One can view the first two theorems as corollaries of this last result, modulo the different gauge assumptions. However, we prefer to state them separately because each of them represent key and largely disjoint steps in the the proof of this last result. In addition, the Threshold Theorem represents a long sought after goal in this field.

Some further comments are in order concerning the scattering property in the first part of the last theorem. As discussed in Remark 1.10, in the context of subthreshold solutions scattering means that solutions are global in the caloric gauge with a bounded S^1 norm. As it turns out, here scattering carries almost exactly the same meaning. Precisely, we show that for large enough T, the solution admits a caloric representation on the time interval $[T, \infty)$, which has a finite S^1 norm. One consequence of this is that *scattering solutions must always be topologically trivial.*

Also, as far as the soliton bubbling off property is concerned, blow up solutions with this property are known to exist. The constructions in [25, 41] give such examples[4] whose energies may be arbitrarily close to the threshold $2E_{\mathrm{GS}}$, and the recent work [18] provides[5] a blow up solution at exactly the threshold energy $2E_{\mathrm{GS}}$. These solutions concentrate at the blow-up point following a rescaled soliton profile, where the soliton scale differs logarithmically from the self-similar scale. Similarly, solutions where bubbling occurs at infinity also exist. For Yang–Mills in the topologically trivial class, this was achieved in [17]; interestingly, the nonscattering solution in [17] has exactly the threshold energy $2E_{GS}$ (see also [18]). Such solutions have been obtained for related models such as the energy critical wave maps equation in the one-bubble case, see [8].

1.3. **A brief history and broader context.** A natural point of view is to place the present papers and results within the larger context of geometric wave equations, which also includes wave maps (WM), Maxwell–Klein–Gordon (MKG) and Einstein equations. Two common features of all these problems are that they admit a Lagrangian formulation, and have some natural gauge invariance properties. Following are some of the key developments that led to the present work.

1. The null condition. A crucial early observation in the study of both long range and low regularity solutions to geometric wave equations was that the nonlinearities appearing in the

---

[4]The constructions in [25, 41] are for $\mathbf{G} = SU(2)$ in the first topological class with $c_2 = 1$ and with energies close to E_{GS}, but a straightforward gluing argument at infinity produces the desired topologically trivial solutions.

[5]We note that [18] moreover gives a complete classification of possible dynamics at the threshold energy under equivariance symmetry.

equations have a favorable algebraic structure, which was called *null condition*, and which can be roughly described as a cancellation condition in the interaction of parallel waves. In the low regularity setting, this was first explored in work of Klainerman and Machedon [20], and by many others later on.

2. The $X^{s,b}$ spaces. A second advance was the introduction of the $X^{s,b}$ spaces[6], also first used by Klainerman and Machedon [20] in the context of the wave equation. Their role was to provide enough structure in order to be able to take advantage of the null condition in bilinear and multilinear estimates. Earlier methods, based on energy bounds, followed by the more robust Strichartz estimates, had proved inadequate to the task.

3. The null frame spaces. To study nonlinear problems at critical regularity one needs to work in a scale invariant setting. However, it was soon realized that the homogeneous $X^{s,b}$ spaces are not even well defined, not to mention suitable for this. The remedy, first introduced in work of the second author [52] in the context of wave maps, was to produce a better description of the fine structure of waves, combining frequency and modulation localizations with adapted frames in the physical space. This led to the *null frame spaces*, which played a key role in subsequent developments for wave maps.

4. Renormalization. A remarkable feature of all semilinear geometric wave equations is that while at high regularity (and locally in time) the nonlinearity is perturbative, this is no longer the case at critical regularity. Precisely, isolating the non-perturbative component of the nonlinearity, one can see that this is of paradifferential type; in other words, the high frequency waves evolve on a variable low frequency background. To address this difficulty, the idea of Tao [45], also in the wave maps context, was to *renormalize* the paradifferential problem, i.e., to find a suitable approximate conjugation to the corresponding constant coefficient problem. In the case of wave maps, the conjugating operator is essentially a gauge transform (i.e., a Lie group-valued function), while in the case of Maxwell–Klein–Gordon and Yang–Mills one needs a Lie group-valued pseudo-differential operator; see [42, 28, 29, 30] and the discussion below.

5. Induction of energy. The ideas discussed so far seem to suffice for small data critical problems. Attacking the large data problem generates yet another range of difficulties. One first step in this direction is Bourgain's *induction of energy* idea [6], which is a convenient mechanism to transfer information to higher and higher energies. We remark that an alternate venue here, which sometimes yields more efficient proofs, is the Kenig–Merle idea [19] of constructing *minimal blow-up solutions*. However, the implementation of this method in problems which require renormalization seems to cause considerable trouble. For a further discussion on this issue, we refer to [27], where this method was carried out in the cases of energy critical wave maps into the hyperbolic plane.

6. Caloric gauge. Another difficulty arising in the context of large data solutions is that of finding a good gauge, which at the same time applies to large data and at the same time has good analytic properties. The caloric gauge, used in our work, is a global version of a local caloric gauge previously introduced by the first author [32, 33], and is based on an idea proposed by Tao [46] in the wave maps context.

---

[6]The concept, and also the notation, is due to Bourgain, in the context of KdV and NLS type problems.

7. Energy dispersion. One fundamental goal in the study of large data problems is to establish a quantitative dichotomy between dispersion and concentration. The notion of *energy dispersion,* introduced in joint work [43, 44] of the second author and Sterbenz in the wave maps context, provides a convenient measure for pointwise concentration. Precisely, at each energy there is an energy dispersion threshold below which dispersion wins. We remark that, when it can be applied, the Kenig-Merle method [19] yields more accurate information; for instance, see [27]. However, the energy dispersion idea, which is what we follow in the present series of papers, is much easier to implement in conjunction with renormalization.

8. The frequency gap. One obstacle in the transition from small to large data in renormalizable problems is that the low frequency background may well correspond to a large solution. Is this fatal to the renormalized solution? The answer to that, also originating in [43, 44], is that may be a second hidden source of smallness, namely a large *frequency gap* between the high frequency wave and the low frequency background it evolves on.

9. Morawetz estimates (monotonicity formulas). The outcome of the ideas above is a dichotomy between dispersion and scattering on one hand, and very specific concentration patterns, e.g., solitons, self-similar solutions on the other hand. The Morawetz estimates, first appearing in this role in the work of Grillakis [14], are a convenient and relatively simple tool to identify and, if possible, eliminate such concentration scenarios. In the present work, in analogy with the elliptic and parabolic literature, such an estimate is interpreted as a monotonicity formula (Section 3).

We now narrow the scope of discussion, and review some earlier developments on the wave maps (WM), Maxwell–Klein–Gordon (MKG) and hyperbolic Yang–Mills (YM) equations related to the present paper.

(MKG) and (YM) above the scaling critical regularity. We start our discussion with a short and incomplete survey of the (YM) problem above the scaling critical regularity. We also discuss the (MKG) problem, which has been often studied as a simpler model for (YM) with a commutative gauge group.

In the two and three dimensional cases, which are energy subcritical, global regularity of sufficiently regular solutions was shown in the early works [10, 11]. These papers in fact handled the more general Yang–Mills-Higgs system, which includes both (YM) and (MKG) as special cases. In dimension $d = 3$, local well-posedness in the energy space of (MKG) and (YM) was proved in [21] and [22], respectively. In the higher dimensional case $d \geq 4$, an essentially optimal local well-posedness result for a model problem closely related to (MKG) was obtained in [23].

However, a new difficulty arises in the large data[7] problem for (YM): Namely, the gauge transformation law is *nonlinear* due to the noncommutative gauge group. In particular, gauge transformations into the Coulomb gauge obey a nonlinear elliptic equation, for which no suitable large data regularity theory is available. Note, in comparison, that such gauge transformations obey a linear Poisson equation in the case of (MKG). In [22], where finite energy global well-posedness of the 3+1 dimensional (YM) problem was proved, this issue was handled by localizing in spacetime via the finite speed of propagation to gain smallness,

---

[7]More precisely, a suitable scaling critical norm of the connection A (e.g., $\|A\|_{L_x^d}$) or the curvature F (e.g., $\|F\|_{L_x^{\frac{d}{2}}}$) is large.

and then working in local Coulomb gauges[8]. An alternative, more robust approach without spacetime localizations to the same problem has been put forth by the first author in [32, 33], inspired by [46]. The idea is to use an associated geometric flow, namely the *Yang–Mills heat flow*, to select a global-in-space Coulomb-like gauge for data of any size.

The energy critical (WM) problem. Before turning to the (MKG) and (YM) problems at critical regularity, we briefly recall some recent developments on the wave maps equation (WM), where many of the methods we implement here have their roots. We confine our discussion to the energy critical problem in $2+1$ dimensions, which is both the most difficult and the most relevant to our present paper. For the small data problem, global well-posedness was established in [52], [45], [53]. More recently, the *Threshold Theorem* for large data wave maps, which asserts that global well-posedness and scattering hold below the ground state energy, was proved in [43, 44] in general, and independently in [27] and [49, 47, 48, 50, 51] for specific targets (namely the hyperbolic space). See also [31] for a sharp refinement in the case of a two-dimensional target, taking into account an additional topological invariant (namely, the degree of the wave map), in analogy with the refined threshold $2E_{\mathrm{GS}}$ in our work. Our present strategy was strongly influenced by [43, 44].

For the energy critical (WM), in the important case of spherical targets, we also note the recent development due to Grinis [15], which says that along a well-chosen sequence of times, all time-like energy concentration must be in the form of a superposition of rescaled solitons. Our Bubbling Theorem (Theorem 1.7) is a first step for proving an analogous result for (YM). In [9], this was complemented with a decay of the energy near the cone when the total energy is sufficiently close to the ground state.

The (MKG) and (YM) problems at critical regularity. Next, we discuss the (YM) problems at critical regularity. As before, we simultaneously consider the corresponding problems for (MKG), which is a simpler commutative analogue of (YM).

Before discussing history, let us clarify a key structural difference between (WM) on the one hand and (MKG), (YM) on the other, whose understanding is crucial for making progress on the latter two problems. Roughly speaking, all three equations can be written in a form where the main 'dynamic variables', which we denote by ϕ, obey a possibly nonlinear gauge covariant wave equation $\Box_A \phi = \cdots$, and the associated curvature $F[A]$ is determined by ϕ. In the case of (WM), this dependence is simply algebraic, whereas for (MKG) and (YM) the curvature $F[A]$ obeys a wave equation with a nonlinearity depending on ϕ. This difference manifests in the renormalization procedure for each equation: For (WM) it suffices to use a physical space gauge transformation, whereas for (MKG) and (YM) it is necessary to use a microlocal (more precisely, *pseudo-differential*) gauge transformation that exploits the fact that A solves a wave equation in a suitable gauge.

The first such renormalization argument appeared in [42], in which global regularity of (MKG) for small critical Sobolev data was established in dimensions $d \geq 6$. This work was followed by a similar high dimensional result for (YM) in [28]. The small data result in the energy critical dimension $4+1$ was obtained in [29]. Finally, the large data result for (MKG) in dimension $4+1$ was proved by the authors in [35, 36, 34] and independently by [26]. Although the implementation differs in many places, the outline of the three papers

---

[8]On the other hand, a closely related spacetime localization approach, but relying on new "initial data surgery" techniques, is developed in [39], which yields an alternative proof of the main result of [22].

[35, 36, 34] is broadly followed in the present work. In particular we borrow a good deal of notation, ideas and estimates from both [29] and [35, 36, 34]. On the other hand, we remark that [26] followed the strategy of [27]. For the (YM) problem in $4 + 1$ dimensions, the small data global result was only recently proved in [30], which is another direct predecessor the present work.

We conclude with a remark on differences between (MKG) and (YM). The issue of non-commutative gauge group for the large data problem has already been discussed. Another important difference between (MKG) and (YM) in $4 + 1$ dimensions is that the latter problem admits nontrivial steady states (i.e., harmonic Yang–Mills connections). These solutions are known to lead to a finite time blow up; see [25, 41], so for (YM) one must prove the Threshold Theorem, instead of a unconditional result as in (MKG). Finally, (YM) is more 'strongly coupled' as a system compared to (MKG), in the sense that the connection A itself obeys a covariant wave equation. This feature necessitates a more involved renormalization procedure compared to (MKG).

Other related works. In related developments, one should also note the works [4, 5] on the closely related cubic Dirac equation, as well as the massive Dirac–Klein–Gordon system, as well as [13] on the Maxwell–Dirac equation and [12] on the massive Maxwell–Klein–Gordon system.

1.4. **Notation, conventions and preliminaries.** Here we collect more notation and conventions used in the remainder of this paper.

Asymptotic notation.
- $A \lesssim B$ and $A = O(B)$ both mean $A \leq CB$ for some constant $C > 0$. The dependence of C on various parameters is specified by subscripts. When $A \lesssim B$ and $B \lesssim A$, we write $A \sim B$.

Tensor calculus.
- We employ the usual index notation, the Einstein convention of summing up repeated upper and lower indices. We use greek indices, such as $\alpha, \beta, \gamma, \ldots$, for all coordinates $x^0 = t, x^1, x^2, x^3, x^4$, and latin indices, such as $i, j, k, \ell, \ldots$, for the spatial coordinates $x = (x^1, \ldots, x^4)$.
- Given a metric (which is usually the Minkowski metric $\mathbf{m}$ on $\mathbb{R}^{1+4}$, unless otherwise stated), we write ∇ for the associated Levi-Civita connection. Tensorial indices are raised and lowered using the metric.

Exterior differential calculus.
- The wedge product $\wedge$ and the differential d for differential forms are defined in the usual way. A k-form ω can be viewed as a k-covariant tensor; we have $\omega = \sum_{\beta_1 < \cdots < \beta_k} \omega_{\beta_1 \ldots \beta_k} \mathrm{d}x^{\beta_1} \wedge \cdots \wedge \mathrm{d}x^{\beta_k}$ where $\omega_{\beta_1 \ldots \beta_k}$ is the coordinate expression of ω as a tensor.
- ι_X is the interior product with a vector field X, i.e., $(\iota_X \omega)_{\beta_1 \ldots \beta_{k-1}} = X^\alpha \omega_{\alpha \beta_1 \ldots \beta_{k-1}}$, and $\mathcal{L}_X$ is the Lie derivative with respect to X.
- The pointwise inner product $(\cdot, \cdot)$ of two k-forms is defined so that $\{\theta^{\alpha_1} \wedge \cdots \wedge \theta^{\alpha_k}\}_{\{\alpha_1 < \cdots < \alpha_k\}}$ is an orthonormal basis, where $\{\theta^\alpha\}$ is any orthonormal basis with

15

respect to a given metric on $T^*\mathbb{R}^d$ (note that this differs by the usual induced metric for tensors by $k!$).

- The Hodge star operator $\star$ is defined so that $\eta \wedge \star\omega = (\eta, \omega) \, d\mathrm{Vol}$, where $d\mathrm{Vol}$ is the volume form. The codifferential δ is the adjoint of d with respect to the L^2-product $\int \langle \cdot, \cdot \rangle \, d\mathrm{Vol}$. The Hodge Laplacian is defined to be $-\Delta = \delta d + d\delta$, so that it agrees with the usual Laplacian $\sum_j \partial_j^2$ for 0-forms (or functions) on $\mathbb{R}^4$.

Subsets of $\mathbb{R}^d$.

- For a bounded open set $U \in \mathbb{R}^d$ and $\lambda > 0$, λU is defined to be rescaling of U by the factor λ centered at the barycenter of U.
- $B_R(x)$ is the 4-dimensional ball of radius r centered at x. When $x = 0$, we simply write $B_R = B_R(0)$.
- $\mathcal{A}_{(R',R)}(x)$ is the 4-dimensional annulus of inner and outer radii R' and R, respectively, centered at $x \in \mathbb{R}^4$. When $x = 0$, we simply write $\mathcal{A}_{(R',R)} = \mathcal{A}_{(R',R)}(0)$.
- Consider the forward light cone centered at $(0,0)$:

$$C = \{(t, x) : 0 < t < \infty, |x| < t\}.$$

For $t_0 \in \mathbb{R}$ and $I \subset \mathbb{R}$, we define

$$\begin{aligned} C_I &= \{(t, x) : t \in I, |x| < t\}, & \partial C_I &= \{(t, x) : t \in I, |x| = t\}, \\ S_{t_0} &= \{(t, x) : t = t_0, |x| < t\}, & \partial S_{t_0} &= \{(t, x) : t = t_0, |x| = t\}. \end{aligned}$$

For $\delta \in \mathbb{R}$, we define the translated cones

$$C^\delta = \{(t, x) : \max\{0, \delta\} < t < \infty, |x| < t - \delta\}.$$

The corresponding objects C_I^δ, ∂C_I^δ, $S_{t_0}^\delta$ and $\partial S_{t_0}^\delta$ are defined in the obvious manner.

Polar coordinate systems.

- *Hyperbolic polar coordinates on $C \subset \mathbb{R}^{1+4}$.* We parametrize the cone $C = \{(t, x) \in \mathbb{R}^{1+4} : |x| < t\}$ by $(t, x) = \rho y$, where $\rho = \sqrt{t^2 - |x|^2} > 0$ and $y \in \mathbb{H}^4 := \{(t, x) \in \mathbb{R}^{1+4} : t^2 - |x|^2 = 1, t > 0\}$. We write dy for the volume form on $\mathbb{H}^4$.
- *Polar coordinates on $\mathbb{R}^4$.* We parametrize $\mathbb{R}^4 \setminus \{0\}$ by $x = r\Theta$, where $r = |x|$ and $\Theta \in \mathbb{S}^3 = \{x \in \mathbb{R}^4 : |x| = 1\}$.

  Given a covariant tensor $\omega_{\beta_1 \ldots \beta_k}$ (e.g., a k-form) on $\mathbb{R}^4 \setminus \{0\}$, we use the schematic notation $\omega_{\Theta \ldots \Theta}(r, \cdot)$ for its pullback onto each constant-r sphere. To formulate the calculus of such objects, *we view each $\omega_{\Theta \ldots \Theta}(r, \cdot)$ as a 1-form on the unit sphere $\mathbb{S}^3$*. We write $\not{g}$ for the metric on the unit sphere $\mathbb{S}^3$, $\not\nabla$ for the associated Levi–Civita connection, $\not{D} = \not\nabla + ad(A)$ for the covariant derivative, $d\Theta$ for the volume form and L_Θ^p for the Lebesgue spaces with respect to $d\Theta$. We write ω^Θ_Θ for the trace with respect to $\not{g}$ of a covariant tensor; accordingly, $\not\nabla^\Theta \omega_\Theta$ is the divergence operator with respect to $\not{g}$ and $d\Theta$, and $\not{D}^\Theta \omega_\Theta$ is the covariant divergence for a $\mathfrak{g}$-valued 1-form.

  Given a subset $U \subset \mathbb{R}^4 \setminus \{0\}$, we *define* the $\|\omega_{\Theta \ldots \Theta}\|_{L^p(U)}$ to be the $L^p(U)$ norm of $|\omega_{\Theta \ldots \Theta}|$, where $|\omega_{\Theta \ldots \Theta}|^2 = \not{g}(\omega_{\Theta \ldots \Theta}, \omega_{\Theta \ldots \Theta})$. Note that $\|\omega_{\Theta \ldots \Theta}\|_{L^p(\mathcal{A}_{(R',R)})} \sim_{R',R} \|\omega_{\Theta \ldots \Theta}\|_{L^p_r((R',R);L^p_\Theta)}$ for any $0 < R' < R$.

16

Functions spaces.

- ∂ (without sub- or superscripts) is the spatial gradient $\partial = (\partial_1, \partial_2, \ldots, \partial_4)$, and ∇ is the spacetime gradient $\nabla = (\partial_0, \partial_1, \ldots, \partial_4)$. We write $\partial^{(n)}$ (resp. $\nabla^{(n)}$) for the collection of n-th order spatial (resp. spacetime) derivatives, and $\partial^{(\leq n)}$ (resp. $\nabla^{(\leq n)}$) for those up to order n.
- $\dot{W}^{\sigma,p}(\mathbb{R}^d; V)$ is the homogeneous L^p-Sobolev space of order σ for functions from $\mathbb{R}^d$ into a normed vector space V. In the special case $p = 2$, we write $\dot{H}^\sigma(\mathbb{R}^d; V) = \dot{W}^{\sigma,2}(\mathbb{R}^d; V)$. The inhomogeneous counterparts are denoted by $W^{n,p}(\mathbb{R}^d; V)$ and $H^n(\mathbb{R}^d; V)$, respectively. We often suppress $\mathbb{R}^d$ and V when it is clear from the context.
- The mixed spacetime norm $L_t^q \dot{W}_x^{n,r}$ of functions on $\mathbb{R}^{1+d}$ is often abbreviated as $L^q \dot{W}^{n,r}$.
- Generally, a function space on an open subset $U \subseteq \mathbb{R}^4$ is defined by restriction, i.e., $\|u\|_{X(U)} = \inf\{\|\tilde{u}\|_X : \tilde{u} \in X, \ \tilde{u} \restriction_U = u\}$. A similar convention applies for a function space on an open subset $\mathcal{O} \subseteq \mathbb{R}^{1+4}$.
- The local function space $X_{loc}(U)$ is defined as

$$X_{loc}(U) = \bigcap_{B_r(x):\overline{B}_r(x) \subseteq U} X(B_r(x)).$$

Littlewood–Paley theory, dyadic function spaces and frequency envelopes.

- $\{P_k\}_{k \in \mathbb{Z}}$ denotes the usual Littlewood–Paley projections in the variable $x \in \mathbb{R}^4$.
- A dyadic function space X is a collection $\{X_k\}_{k \in \mathbb{Z}}$ of normed spaces on either $\mathbb{R}^4$ or $\mathbb{R}^{1+4}$. Often we use the same space for each k, in which case we simply write $X = X_k$. We define $\ell^p X$ by the norm $\|u\|_{\ell^p X}^p = \sum_k \|P_k u\|_X^p$, with the usual modification for $p = \infty$. An important example is the L^2-Sobolev space $H^\sigma = \ell^2 H^\sigma$.
- An admissible frequency envelope c is a sequence $\{c_k\}_{k \in \mathbb{Z}}$ of positive numbers satisfying $\max\{\frac{c_j}{c_k}, \frac{c_k}{c_j}\} \leq C 2^{\delta_{fe}|k-j|}$ for some constant C depending on c and an absolute constant $\delta_{fe} > 0$ inherited from [37, 38].
- We define $\|u\|_{X_c} = \sup_k c_k^{-1} \|P_k u\|_{X_k}$. If $\|u\|_{X_c} \leq 1$, then we say that c is a frequency envelope for u in X.

1.5. **Structure of the present paper.** The remainder of the paper is structured as follows.

Section 2. Here we review the main results in the first three papers of the series [37], [38] and [39], emphasizing the parts which are needed here.

Section 3. This is where we state and prove all the conservation laws and monotonicity formulas that are used in this paper. We also explore a few consequences of the monotonicity formulas.

Section 4. We use a "good gauge" representation theorem (Theorem 2.19) for large energy Yang–Mills connections in order to prove a strong local compactness statement, Theorem 4.1, that we rely on on in our blow-up analysis.

Section 5. Here we study the regularity of connections that either stationary or self-similar, and show that such connections must be gauge equivalent to a smooth connection. This is akin to elliptic theory for harmonic Yang–Mills connections.

Section 6. We show that there does not exist any nontrivial self-similar solutions to Yang–Mills with finite energy, thus eliminating one of the main potential obstructions to our results.

Section 7. Here we carry out the blow-up analysis and prove the Bubbling Theorem (Theorem 1.7). This proof uses all of the previous five sections.

Section 8. In this section we prove that sharp energy concentration cannot occur near the null cone. This is critical in order to be able to separate the bubbling-off scenario from the scattering, energy dispersed case.

Section 9. Finally, here we complete both the proof of the Threshold Conjecture, see Theorem 1.8, and the dichotomy result in Theorem 1.11.

Appendix A. We collect some technical tools needed for our analysis of gauge transformations, especially in Sections 5, 6 and 8.

2. An outline of the first three papers

Our aim here is to provide a brief outline of the first three papers [37],[38] and [39], to the extent necessary in order to complete the proof of the large data results in the present paper. For a more comprehensive review of the full series of four papers we instead refer the reader to our survey article [40].

Let us take as a starting point of our discussion the following small data result proved earlier in [30]:

Theorem 2.1. *The hyperbolic Yang–Mills equation in $\mathbb{R}^{4+1}$ is globally well-posed in the Coulomb gauge for all initial data with small energy.*

Even before considering well-posedness results for the large data, the first difficulty one faces is that the Coulomb gauge does not appear to fully extend to large data in general, and not even to subthreshold data. For this reason, our first paper [37] is devoted solely to the gauge problem; precisely, inspired by earlier work of Tao [46] and of the first author [32, 33], we develop a new gauge for the hyperbolic Yang–Mills problem (1.1), namely the *caloric gauge*. Using this gauge, the most difficult gauge-dependent analysis of the Yang–Mills equation is carried out in [38]. The caloric gauge is the natural setting of our Threshold Theorem (Theorem 1.8).

On the other hand for the large data analysis in [39] causality plays a key role, so we shift the (soft part of the) analysis to the *temporal gauge*. The causality property of the temporal gauge allows us to formulate a local well-posedness theory of the hyperbolic Yang–Mills equation for arbitrary finite energy data, which is the setting for our Dichtomy Theorem (Theorem 1.11). We note, however, that the strong S-norm control on the solution is lost in the temporal gauge.

We summarize the discussion on various gauges so far in the following table:

Gauge choice	Definition	Appearances	Remarks
(global) Coulomb	$\partial^k A_k = 0$	Thm. 2.1	Requires small initial energy; expect S-norm control; no direct usage in this paper.
Caloric	Definition 2.5	Thms. 1.8, 1.11†, 2.10, 2.12	Requires trivial topological class and finite caloric size; expect S-norm control.
Temporal	$A_0 = 0$	Thms. 1.11†, 2.11, 2.17, 2.18	No restriction on the topological class or energy; no S-norm control.

TABLE 1. Gauge choices for the initial value problem. †: In Theorem 1.11, the initial value problem is posed in the temporal gauge, but the scattering statement involves the caloric gauge; see Remark 1.10.

Finally, we note that our Bubbling Theorem (Theorem 1.7) is formulated in a gauge-independent fashion[9]. Indeed, most of the work in the present paper is carried out in a gauge-covariant fashion, while using the gauge-dependent results in [38] and [39] at critical junctures.

2.1. **The caloric gauge.** The goal of the first paper [37] is to
- Introduce the caloric gauge;
- Show that the caloric gauge is well-defined for all subthreshold data; and
- Provide a comprehensive formulation of the hyperbolic Yang–Mills equation in the caloric gauge which is sufficiently accurate for the subsequent analysis.

The caloric gauge is defined using the Yang–Mills heat flow

$$\partial_s A_j = \mathbf{D}^k F_{jk}, \qquad A_j(s=0) = a_j \tag{2.1}$$

which implicitly assumes the gauge condition $A_s = 0$ (which we refer to as the *local caloric gauge*) relative to the fully covariant formulation of the same equation. This can be naively viewed as parabolic system for the curl of A (or the curvature F), coupled with a transport equation for the divergence of A. Because these two equations are strongly coupled, this evolution displays an interesting mix of semilinear and quasilinear features.

Our interest is in initial data a for which this solution is global, with the curvature F satisfying global parabolic bounds. To capture this, we use the L^3 norm of F as a control norm, and call it the *caloric size* of a,

$$\mathcal{Q}(a) = \int_0^\infty \int_{\mathbb{R}^4} |F|^3 dx ds. \tag{2.2}$$

We note that this is a gauge invariant quantity. For solutions with finite caloric size, we have the following structure theorem:

Theorem 2.2 ([37, Corollary 5.14]). *Let $a \in \dot{H}^1$ be a connection so that $\mathcal{Q}(a) < \infty$. Then this solution has the property that the limit*

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

---

[9]However, the notion of scattering is formulated with the help of the caloric gauge; see Remark 1.10.

19

exists in $\dot{H}^1$. Further, the limiting connection is flat, $f_\infty = 0$, and the map $a \to a_\infty$ is locally Lipschitz in $\dot{H}^1$, H^N $(N \geq 1)$ and $\dot{H}^1 \cap \dot{H}^N$ $(N \geq 2)$.

Next, using the monotonicity formula for the energy, we prove the Dichotomy Theorem for the Yang–Mills heat flow:

Theorem 2.3 ([37, Theorem 6.1]). *One of the following two properties must hold for the maximally extended $\dot{H}^1$ Yang–Mills heat flow:*
   a) The solution is global and its caloric size is finite;
   b) The solution bubbles off a nontrivial harmonic Yang–Mills connection, either

   (1) at a finite blow-up time $s < \infty$, or
   (2) at infinity $s = \infty$.

Combined with topological triviality of $\dot{H}^1$ connections, we are led to the Threshold Theorem, with the identical threshold as in the hyperbolic case:

Theorem 2.4 ([37, Theorem 6.6]). *The Yang–Mills heat flow is globally well-posed in $\dot{H}^1$ for all subthreshold initial data $a \in \dot{H}^1$. Precisely, there exists a nondecreasing function*

$$\mathcal{Q} : [0, 2E_{GS}) \to \mathbb{R}^+$$

so that for all subthreshold data a with energy $\mathcal{E}$ we have

$$\mathcal{Q}(a) \leq \mathcal{Q}(\mathcal{E}). \tag{2.3}$$

For connections with finite caloric size, we define the *(global) caloric gauge* as follows:

Definition 2.5. A connection $a \in \dot{H}^1$ is caloric if $\mathcal{Q}(a) < \infty$ and the global solution to its associated Yang–Mills heat flow has the property $a_\infty = 0$.

It is easy to see that for all connections a for which the conclusion of Theorem 2.4 holds there is a unique[10] equivalent caloric gauge. This is because a_∞ is flat and thus can be represented as

$$a_\infty = O^{-1} \partial_x O$$

for a suitable gauge transformation O. Then by gauge invariance the equivalent connection

$$\tilde{a} = \mathcal{G}(O)a = OaO^{-1} - \partial_x O O^{-1}$$

is caloric. More precisely, we have

Proposition 2.6 ([37, Proposition 7.2]). *For each $\dot{H}^1$ connection $\tilde{a}$ in $\mathbb{R}^4$ with $\mathcal{Q}(a) < \infty$, there exists an unique (up to constant gauge transformations) gauge-equivalent connection a, which is a caloric gauge connection. Further, the map $\tilde{a} \to a$ is continuous in the quotient topology defined by the distance*

$$d(a_1, a_2) = \inf_{O \in \mathbf{G}} \|O a_1 O^{-1} - a_2\|_{\dot{H}^1}.$$

A key result in [37] asserts that:

_____
[10]Up to constant conjugations

20

Theorem 2.7 ([37, Proposition 7.7 and Theorem 7.8]). *The space $\mathcal{C}$ of all caloric connections is a C^1 submanifold of the space*

$$\mathbf{H} = \{a \in \dot{H}^1 : \partial^j a_j \in \ell^1 L^2\}.$$

In addition, for all subthreshold caloric connections with energy $\mathcal{E}$ and caloric size $\mathcal{Q}$ we have the bound

$$\|a\|_{\mathbf{H}}^2 := \|a\|_{\dot{H}^1}^2 + \|\partial^j a_j\|_{\ell^1 L^2}^2 \lesssim_{\mathcal{E},\mathcal{Q}} 1.$$

The second part of the state space norm $\mathbf{H}$ reflects the fact that caloric connections satisfy a nonlinear form of the Coulomb gauge condition.

The second part of [37] is devoted to modeling the hyperbolic Yang–Mills equation as an evolution on the caloric manifold. Precisely, the state space for this evolution is $T^{L^2}\mathcal{C}$, which is the L^2 completion of the tangent space $T\mathcal{C}$. We will view the spatial components $(A_x, \partial_t A_x) \in T^{L^2}\mathcal{C}$ as the dynamic variables, and the temporal parts $A_0, \partial_0 A_0$ as auxiliary. Correspondingly, we call a pair $(a, b) \in T^{L^2}\mathcal{C}$ (i.e., $a \in \mathcal{C}$, $b \in T_a^{L^2}\mathcal{C}$) a *caloric gauge initial data set* for (1.1). This is related to the gauge-covariant notion of initial data sets as follows:

Theorem 2.8 ([37, Theorem 8.1]). *a) Given any Yang–Mills initial data pair $(a, e) \in \dot{H}^1 \times L^2$ with finite caloric size, there exists a caloric gauge data set $(\tilde{a}, b) \in T^{L^2}\mathcal{C}$ and $a_0 \in \dot{H}^1$, so that $(\tilde{a}, \tilde{e})$ is gauge equivalent to (a, e), where*

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

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

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

satisfies the constraint equation (1.4).

By this result, we may indeed fully describe Yang–Mills connections in the caloric gauge as continuous functions

$$I \ni t \to (A_x, \partial_t A_x)(t) \in T^{L^2}\mathcal{C}.$$

The equations for the dynamical variables $(A_x, \partial_t A_x)$ are proved to have the form

$$\Box_A A_k = \mathbf{P}[A_j, \partial_k A_j] + 2\Delta^{-1}\partial_k \mathbf{Q}(\partial^\alpha A_j, \partial_\alpha A_j) + R(A, \partial_t A) \tag{2.4}$$

together with a compatibility condition

$$\partial^k A_k = \mathbf{DA}(A) := \mathbf{Q}(A, A) + \mathbf{DA}^3(A), \tag{2.5}$$

where the temporal component A_0 and its time derivative $\partial_t A_0$ are uniquely determined by $(A_x, \partial_t A_x)$ and admit the expressions

$$A_0 = \mathbf{A}_0(A) := \Delta^{-1}[A, \partial_t A] + 2\Delta^{-1}\mathbf{Q}(A, \partial_t A) + \mathbf{A}_0^3(A), \tag{2.6}$$

$$\partial_t A_0 = \mathbf{DA}_0(A) := -2\Delta^{-1}\mathbf{Q}(\partial_t A, \partial_t A) + \mathbf{DA}_0^3(A). \tag{2.7}$$

Here $\mathbf{P}$ is the Leray projector, $\mathbf{Q}$ is a symmetric bilinear form with symbol[11]

$$\mathbf{Q}(\xi, \eta) = \frac{\xi^2 - \eta^2}{2(\xi^2 + \eta^2)}. \tag{2.8}$$

---

[11]Although the symbol looks anti-symmetric, it is compensated by the Lie bracket in the definition of bilinear multipliers for $\mathfrak{g}$-valued functions; see [37, Definition 3.1].

The cubic error terms are R, $\mathbf{DA}^{(3)}$, $\mathbf{A}_0^3$ and $\mathbf{DA}_0^3$ are "better behaved" in the following sense. First, we recall the following definition from [37]:

Definition 2.9 (Envelope preserving map; [37, Definition 9.1])**.** Let X, Y be dyadic norms. A map $\mathbf{F} : X \to Y$ is said to be *envelope-preserving of order* $\geq n$ ($n \in \mathbb{N}$ with $n \geq 2$) if for any admissible frequency envelope c for a in X, we have

$$\|P_k \mathbf{F}(a)\|_Y \lesssim_{\|a\|_X} c_k^n.$$

The cubic error terms R, $\mathbf{DA}^{(3)}$, $\mathbf{A}_0^3$ and $\mathbf{DA}_0^3$ are envelope preserving maps between the following spaces:

$$R : \mathrm{Str}^1 \to L^1 L^2 \cap L^2 \dot{H}^{-\frac{1}{2}}, \tag{2.9}$$

$$\mathbf{DA}^3 : \mathrm{Str}^1 \to L^1 \dot{H}^1 \cap L^2 \dot{H}^{\frac{1}{2}}, \tag{2.10}$$

$$\mathbf{A}_0^3 : \mathrm{Str}^1 \to L^1 \dot{H}^2 \cap L^2 \dot{H}^{\frac{3}{2}}, \tag{2.11}$$

$$\mathbf{DA}_0^3 : \mathrm{Str}^1 \to L^1 \dot{H}^1 \cap L^2 \dot{H}^{\frac{1}{2}}. \tag{2.12}$$

Here Str^1 collects several standard non-endpoint Strichartz norms with the appropriate scaling. Since we work with solutions with ℓ^2 dyadic summability, by the envelope preserving property, for each of the above bounds we freely gain ℓ^1 dyadic summability of the above norms. One should think of all these cubic nonlinear expressions above as playing perturbative roles in the analysis. We remark that these expressions obey nice difference bounds as well; for details, see [37, Definition 9.1 and Theorem 9.2].

2.2. Local well-posedness in the caloric gauge and energy dispersed solutions.
Our second paper [38] aims to establish both a local well-posedness result and a more refined continuation and scattering criteria for subthreshold solutions to the hyperbolic Yang–Mills equation in the caloric gauge.

In what follows, we will call hyperbolic Yang–Mills connections in the caloric gauge simply *caloric Yang–Mills waves*.

We begin with the local well-posedness result. We define the ϵ-*energy concentration scale* r_c^ϵ of a finite energy Yang–Mills initial data set (a, e) to be

$$r_c^\epsilon = r_c(E)[a, e] = \sup\{r > 0 : \mathcal{E}_{B_r}(x)[a, e] \leq \epsilon \; \forall x \in \mathbb{R}^4\}.$$

Then we have:

Theorem 2.10 (Local well-posedness in caloric gauge, [38, Theorem 1.12])**.** *There exists a non-increasing function $\epsilon_* = \epsilon_*(\mathcal{E}, \mathcal{Q}) > 0$ and a non-decreasing function $M_*(\mathcal{E}, \mathcal{Q})$ such that, the Yang–Mills equation in caloric gauge is locally well-posed on the time interval $I = [-r_c^{\epsilon_*}, r_c^{\epsilon_*}]$ for initial data with energy $\mathcal{E}$ and initial caloric size $\mathcal{Q}$. More precisely, the following statements hold.*

a) (Regular data) Let $(a, b) \in T^{L^2}\mathcal{C}$ be a smooth initial data set with energy $\mathcal{E}$ and initial caloric size $\mathcal{Q}$. Then there exists a unique smooth solution $A_{t,x}$ to the Yang–Mills equation in caloric gauge on I.

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

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

within the class of initial data with energy concentration scale $\geq r_c$.

22

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

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

d) (Weak Lipschitz dependence) Let $(a', b') \in \mathcal{C} \times L^2$ be another initial data set with similar bounds and energy concentration scale $\ge r_c$. Then for $\sigma < 1$ close to 1 we have the Lipschitz bound

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

We remark that bounds for the auxiliary variables $A_0, \partial_0 A_0$ follow a-posteriori from the S^1 bound for A_x; see [38, Theorem 5.1] for such bounds.

In particular, if the energy of the initial data set is smaller than $\epsilon_* := \min\{\epsilon_*(1), 1\}$, then the corresponding solution $A_{t,x}$ in caloric gauge exists globally and obeys the bound

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

Thus in particular this result also provides a caloric gauge version of the Coulomb gauge small data result in [30].

One downside of using either the Coulomb or caloric gauge is that causality is lost. To remedy this, in [38] we prove that the well-posedness result can also be transferred to the temporal gauge $A_0 = 0$:

Theorem 2.11 ([38, Theorem 1.17]). *The hyperbolic Yang–Mills equation in $\mathbb{R}^{4+1}$ is globally well-posed in the temporal gauge for all initial data with small energy.*

This result includes existence, uniqueness, continuous dependence on the initial data and propagation of higher regularity. In particular the finite energy solutions are identified as the unique limits of regular solutions. A downside of this theorem is that it does not provide the S^1 regularity of solutions, or any other dispersive bounds.

The second main result in [38] is the following theorem, which uses the energy dispersed norm ED defined on a time interval I by

$$\|F\|_{ED[I]} = \sup_{k \in \mathbb{Z}} 2^{-2k} \|P_k F\|_{L^\infty[I]}.$$

The result asserts that caloric solutions to Yang–Mills with sufficiently small energy dispersion are extendable and satisfy uniform bounds:

Theorem 2.12 (Regularity of energy dispersed solutions [38, Theorem 1.15]). *There exists a positive non-increasing function $\epsilon(\mathcal{E})$ and a non-decreasing function $M(\mathcal{E})$ such that if $A_{t,x}$ is a solution (in the sense of Theorem 2.10) to the Yang–Mills equation in caloric gauge on I with energy $\mathcal{E}$ and that obeys*

$$\|F\|_{ED[I]} \le \epsilon(\mathcal{E}), \qquad \mathcal{Q}(A(0)) \le 1,$$

then it satisfies the a-priori bound

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

as well as

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

Moreover, A can be continued as a solution to the Yang–Mills equation in caloric gauge past finite endpoints of I.

2.3. Topological classes and large data solutions. Unlike the first two papers, the third one [39] is concerned with large data solutions which are not necessarily topologically trivial, and thus cannot be directly studied using the global caloric gauge. The goal of [39] is two-fold:

- To describe finite energy initial data sets topologically and analytically.
- To provide a good local theory for finite energy solutions.

Here we work in two settings:

a) For initial data in $\mathbb{R}^4$ and solutions in $\mathbb{R}^{1+4}$, or time sections thereof.
b) For initial data in a ball B_R and solutions in the corresponding domain of dependence $\mathcal{D}(B_R) = \{|x| + |t| < R\}$ or time sections thereof.

In terms of the initial data, in addition to the energy, a key role is played by the ϵ-energy concentration scale localized to the ball B_R

$$r_c^\epsilon = \sup\{r > 0 : \mathcal{E}_{B_r(x) \cap B_R}[(a, e)] \leq \epsilon \; \forall x \in B_R\},$$

as well as the outer concentration radius

$$R_c^\epsilon = \inf\{r > 0 : \mathcal{E}_{B_r(x)}[(a, e)] \leq \epsilon \text{ for some } x \in \mathbb{R}^4 \}.$$

2.3.1. *Finite energy data sets.* We begin with an excision result, which provides small energy extensions outside an annulus:

Proposition 2.13 ([39, Theorem 1.16]). *Let (a, e) be a small energy data set in $B_4 \setminus B_1$. Then we can find a small energy exterior data set $(\tilde{a}, \tilde{e})$ in $\mathbb{R}^4 \setminus B_1$ which agrees with (a, e) in $B_2 \setminus B_1$. Furthermore, if (a, e) is smooth then $(\tilde{a}, \tilde{e})$ can also be chosen to be smooth.*

For initial data sets in a ball, it is useful to work with a good gauge:

Proposition 2.14 ([39, Theorem 1.4]). *Given an initial data (a, e) in B_R with finite energy and ϵ-energy concentration scale r_C, there exists a gauge-equivalent initial data $(\tilde{a}, \tilde{e})$ in B_R which satisfies the bound*

$$\|\tilde{a}\|_{\dot{H}^1 \cap L^4} \lesssim_{\epsilon, \frac{r_C}{R}} 1. \tag{2.13}$$

Consider now finite energy initial data (a, e) in $\mathbb{R}^4$. Here we need to distinguish between a in different topological classes. We begin with the topologically trivial maps:

Theorem 2.15 ([39]). *A finite energy connection a is topologically trivial if and only if it admits a representation $a \in \dot{H}^1$ in a suitable gauge.*

Finally, for topologically nontrivial initial data in $\mathbb{R}^n$ we also can find a good global gauge:

Theorem 2.16 (Good global gauge theorem [39, Theorem 1.5]). *Let $a \in H^1_{loc}$ be a finite energy connection. Then there exists a global representative a such that*

$$a = -\chi O_{(\infty);x} + b$$

where $1 - \chi$ is smooth and compactly supported, $O_{(\infty)}(x)$ is a smooth 0-homogeneous map taking values in $\mathbf{G}$ and $b \in \dot{H}^1$.

Here one can identify the topological class of a with the homotopy class $[O]$ of $O : \mathbb{S}^3 \to \mathbf{G}$. in particular O in the last theorem can be chosen arbitrarily within this homotopy class.

2.3.2. *Finite energy solutions.* A consequence of [30] and of the first two papers in the series [37, 38] is that the small data problem for the $4 + 1$ dimensional hyperbolic Yang–Mills equation is well-posed in several gauges: Coulomb, caloric, and temporal. In [39] we exploit the temporal gauge small data result, combined with causality, to obtain results for the large data problem.

For the subsequent results, set $r_c = r_c^{\epsilon_0}$ where $\epsilon_0 \ll 1$ is the energy bound for the small data result. The main local-in-time result is as follows:

Theorem 2.17 ([39, Theorem 1.22]). *a) For each finite energy data set (a, e) in $\mathbb{R}^4$ with concentration scale r_c there exists a unique finite energy solution A to (1.1) in the time interval $[-r_c, r_c]$ in the temporal gauge $A_0 = 0$, depending continuously on the initial data. Furthermore, any other finite energy solution with the same data must be gauge equivalent to A.*

b) The same result holds for data in a ball B_R and the solution in the corresponding domain of uniqueness $\mathcal{D}(B_R) \cap (I \times \mathbb{R}^4)$.

Now we consider the continuation question. The next result asserts that temporal solutions can be continued until energy concentration (i.e., a blow-up) occurs. Thus, temporal solutions are also maximal solutions for the Yang–Mills equation.

Theorem 2.18 ([39, Theorems 1.22 and 1.23]). *a) For each finite energy data set (a, e) in $\mathbb{R}^4$, let (T_{min}, T_{max}) be the maximal time interval on which the temporal gauge solution A exists. If T_{max} is finite then we have*

$$\lim_{t \to T_{max}} r_c(t) = 0.$$

Further, there exists some $X \in \mathbb{R}^4$ so that the energy concentration occurs in the backward light cone $C = \{|x - X| \leq T_{max} - t\}$ centered at (T_{max}, X), in the sense that

$$\lim_{t \nearrow T} \mathcal{E}_{C \cap S_t}(A) > \epsilon_0. \tag{2.14}$$

The similar result holds for for T_{min}.

b) The same result holds for data in a ball B_R and the solution in the corresponding domain of uniqueness $\mathcal{D}(B_R)$.

We remark that vanishing of r_c follows from [39, Theorem 1.22], and existence of an energy concentration point follows by a standard argument; see, e.g., [34, Lemma 8.1].

The temporal gauge is convenient in order to deal with causality, but not so much in terms of regularity, as it lacks good S bounds. For this reason it is convenient to borrow the caloric gauge regularity:

Theorem 2.19 ([39, Theorem 1.25]). *Let A be a finite energy Yang–Mills solution in a cone section $C_{[t_1, t_2]}$ with energy concentration scale r_c. Then in a suitable gauge A satisfies the bound*

$$\|A\|_{L^\infty(\dot{H}^1 \cap L^4)} + \|\partial_t A\|_{L^\infty L^2} + \|\partial^j A_j\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}} + \|\nabla A_0\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}} + \|\Box A_x\|_{\ell^1 L^2 \dot{H}^{-\frac{1}{2}}} \lesssim_{E, \frac{r_c}{t_2}} 1 \quad (2.15)$$

in the smaller cone $C_{[t_1, t_2]}^{4r_c}$ where the radius has been decreased by $4r_c$.

For the notation $C_{[t_1, t_2]}$ and $C_{[t_1, t_2]}^{4r_c}$, we refer to Section 1.4. The proof of this theorem requires a good gluing technique for local connections with suitable regularity; see [39] for

details. We note that the term $\|\partial^j A_j\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}}$ is missing in the statement of [39, Theorem 1.25], but is evident from the proof (see, in particular, [39, Eqs. (5.6), (5.8)]).

3. Monotonicity formulas

3.1. **The energy-momentum tensor and conservation laws.** We start by introducing the notion of a null frame and the associated null decomposition of a curvature 2-form, which provides a very useful decomposition of the energy momentum tensor. At each point $p = (t_0, x_0) \in \mathbb{R}^{1+4}$, we introduce the null pair

$$L = \partial_t + \frac{x}{|x|} \cdot \partial_x, \qquad \underline{L} = \partial_t - \frac{x}{|x|} \cdot \partial_x,$$

and also orthonormal vectors $\{e_{\mathfrak{a}}\}_{\mathfrak{a}=1,\ldots,3}$ which are orthogonal to L and $\underline{L}$. Observe that each $e_{\mathfrak{a}}$ is tangent to the sphere $\partial B_{t_0,r_0} := \{t_0\} \times \partial B_{r_0}(0)$ where $r_0 = |x_0|$. The set of vectors $\{L, \underline{L}, e_1, e_2, e_3\}$ at p is called a *null frame at p associated to $L, \underline{L}$.*

We define the *null decomposition* of the 2-form F with respect to $\{L, \underline{L}, e_{\mathfrak{a}}\}$ as

$$\alpha_{\mathfrak{a}} := F(L, e_{\mathfrak{a}}), \quad \underline{\alpha}_{\mathfrak{a}} := F(\underline{L}, e_{\mathfrak{a}}), \quad \varrho := \frac{1}{2} F(L, \underline{L}), \quad \sigma_{\mathfrak{ab}} := F(e_{\mathfrak{a}}, e_{\mathfrak{b}}).$$

Note that ϱ is a $\mathfrak{g}$-valued function, $\alpha_{\mathfrak{a}}, \underline{\alpha}_{\mathfrak{b}}$ are $\mathfrak{g}$-valued 1-forms on $\partial B_{t_0,r_0}$ and $\sigma_{\mathfrak{ab}}$ is a $\mathfrak{g}$-valued 2-form on $\partial B_{t_0,r_0}$. We define their pointwise absolute values as

$$|\alpha|^2 := \sum_{\mathfrak{a}=1,\ldots,3} \alpha_{\mathfrak{a}}^2, \quad |\underline{\alpha}|^2 := \sum_{\mathfrak{a}=1,\ldots,3} \underline{\alpha}_{\mathfrak{a}}^2, \quad |\sigma|^2 := \sum_{1 \le \mathfrak{a} < \mathfrak{b} \le 3} \sigma_{\mathfrak{ab}}^2.$$

Recall from Section 1 that the *energy-momentum tensor* associated to a connection A is

$$T_{\alpha\beta}(A) = 2\langle F_\alpha{}^\gamma, F_{\beta\gamma}\rangle - \frac{1}{2}\mathbf{m}_{\alpha\beta}\langle F_{\gamma\delta}, F^{\gamma\delta}\rangle. \tag{3.1}$$

We observe that T is a symmetric 2-tensor, which is gauge invariant at each point. Moreover for each finite energy solution solution to (1.1), the energy-momentum tensor satisfies

$$\partial^\alpha T_{\alpha\beta}(A) = 0. \tag{3.2}$$

This is verified directly for smooth connections, and it then transfers to finite energy Yang–Mills connections by approximation with smooth connections.

A simple way of obtaining energy identities for Yang–Mills equation is to contract the energy-momentum tensor with a well-chosen vector field, an then integrate over a suitable domain.

Given a vector field X on $\mathcal{O}$, we define its *deformation tensor* to be the Lie derivative of the metric with respect to X, i.e., ${}^{(X)}\pi := \mathcal{L}_X \mathbf{m}$. Using covariant derivatives, ${}^{(X)}\pi$ also takes the form

$${}^{(X)}\pi_{\alpha\beta} = \boldsymbol{\nabla}_\alpha X_\beta + \boldsymbol{\nabla}_\beta X_\alpha$$

or expressed in coordinates

$${}^{(X)}\pi_{\mu\nu} = X(\mathbf{m}_{\mu\nu}) + \partial_\mu(X^\alpha)\mathbf{m}_{\alpha\nu} + \partial_\nu(X^\alpha)\mathbf{m}_{\alpha\mu} \tag{3.3}$$

Using the deformation tensor, we define the associated 1- and 0-currents of A as

$${}^{(X)}P_\alpha(A) := T_{\alpha\beta}(A)X^\beta,$$
$${}^{(X)}K(A) := \frac{1}{2}T_{\alpha\beta}(A){}^{(X)}\pi^{\alpha\beta}. \tag{3.4}$$

26

Then by (3.2) we obtain
$$\boldsymbol{\nabla}^\alpha(^{(X)}P_\alpha(A)) = {}^{(X)}K(A). \tag{3.5}$$

Now energy identities for the Yang–Mills flow are obtained by integrating this identity over spacetime regions. Of course, this is most useful when $^{(X)}\pi$ either vanishes (i.e. X is Killing) or when it has a sign.

The simplest choice for X is $X = T = \partial_0$, the unit vector in the time direction. Then $^{(T)}\pi = 0$, so (3.5) becomes
$$\boldsymbol{\nabla}^\alpha(^{(T)}P_\alpha(A)) = 0. \tag{3.6}$$

In particular we have
$$^{(T)}P_0 = \frac{1}{2}\langle F_{jk}, F^{jk}\rangle + \langle F_{0j}, F_0{}^j\rangle = \sum_{\alpha<\beta}|F_{\alpha\beta}|^2,$$

therefore integrating (3.6) between time slices yields the well-known conservation of energy
$$\mathcal{E}_{\{t\}\times\mathbb{R}^4}(A) = \int {}^{(T)}P_0(A)\,\mathrm{d}x = \int \sum_{\alpha<\beta}|F_{\alpha\beta}|^2 \quad \text{is constant in } t.$$

In general, for $U \subset \mathbb{R}^4$ we introduce the notation
$$\mathcal{E}_{\{t\}\times U}(A) = \int_{\{t\}\times U} {}^{(T)}P_0(A)\,\mathrm{d}x.$$

We also need to use energy estimates in sections $C_{[t_1,t_2]}$ of the cone C. For this we define the *energy flux* on the lateral surface of the cone section by
$$\mathcal{F}_{\partial C_{[t_0,t_1]}}(A) = \frac{1}{2}\int_{\partial C_{[t_0,t_1]}} {}^{(T)}P_L(A)r^3\,\mathrm{d}v\mathrm{d}\sigma_{\mathbb{S}^3}$$

Then we have

Lemma 3.1. *Let A be a finite energy Yang–Mills connection on $I \times \mathbb{R}^4$ where $I \subset \mathbb{R}^+$ is an open interval. Then for every $t_0, t_1 \in I$ with $t_0 \leq t_1$, the following statements hold:*
a) The energy flux on $\mathcal{F}_{\partial C_{[t_0,t_1]}}(A)$ is non-negative and additive, i.e.,
$$\mathcal{F}_{\partial C_{[t_0,t_1]}}(A) = \mathcal{F}_{\partial C_{[t_0,t']}}(A) + \mathcal{F}_{\partial C_{[t',t_1]}}(A) \quad \text{for } t' \in [t_0, t_1]. \tag{3.7}$$

b) The energy-flux relation holds:
$$\mathcal{E}_{S_{t_1}}(A) - \mathcal{E}_{S_{t_0}}(A) = \mathcal{F}_{\partial C_{[t_0,t_1]}}(A). \tag{3.8}$$

The nonnegativity is straightforward since the flux density is expressed in terms of the curvature components in the null frame as
$$^{(T)}P_L(A) = |\varrho|^2 + |\sigma|^2 + |\alpha|^2.$$

The identities are again straightforward for smooth solutions, and obtained by approximation with smooth solutions otherwise.

A consequence of Lemma 3.1 is a simple but crucial decay result for the flux:

Corollary 3.2. *Let A be a finite energy Yang–Mills connection on $I \times \mathbb{R}^4$ where $I \subset \mathbb{R}^4$ is an open interval. Then the following statements hold.*

a) If $(0, \delta] \subseteq I$ for some $\delta > 0$, then we have

$$\lim_{t_1 \to 0} \mathcal{F}_{\partial C_{(0,t_1]}}(A) = 0, \tag{3.9}$$

b) If $[\delta, \infty) \subseteq I$ for some $\delta > 0$, then we have

$$\lim_{t_0, t_1 \to \infty} \mathcal{F}_{\partial C_{[t_0,t_1]}}(A) = 0. \tag{3.10}$$

3.2. **Monotonicity formulas.** Here we are interested in the case when the expression $^{(X)}K(A)$ is nonnegative definite. Our primary vector field here will be

$$X_0 = \frac{1}{\rho_0} \left(t \partial_t + x \cdot \partial_x \right), \qquad \rho_0 = \sqrt{t^2 - |x|^2}$$

for $(t, x) \in C$. We also introduce the null coordinates

$$u_0 = t - |x|, \qquad v_0 = t + |x|.$$

Straightforward computations (see [34]) lead to the relation

$$^{(X)}K(A) = \frac{2}{\rho_0} |\iota_{X_0} F|^2, \tag{3.11}$$

where $|\iota_{X_0} F|^2 = \mathbf{m}^{\alpha\beta} \iota_{X_0} F_\alpha \iota_{X_0} F_\beta$. Of course $\mathbf{m}$ is indefinite, but $|\iota_{X_0} F|^2$ is nonnegative due to the fact that X_0 is time-like inside the cone C. Thus the relation (3.5) becomes

$$\boldsymbol{\nabla}^\alpha \big(^{(X_0)}P_\alpha \big) = \frac{2}{\rho_0} |\iota_{X_0} F|^2 \geq 0 \tag{3.12}$$

where the interesting components of $^{(X_0)}P$ are

$$^{(X_0)}P_L = \left(\frac{v_0}{u_0} \right)^{\frac{1}{2}} |\alpha|^2 + \left(\frac{u_0}{v_0} \right)^{\frac{1}{2}} \left(|\varrho|^2 + |\sigma|^2 \right), \tag{3.13}$$

$$^{(X_0)}P_{\underline{L}} = \left(\frac{u_0}{v_0} \right)^{\frac{1}{2}} |\underline{\alpha}|^2 + \left(\frac{v_0}{u_0} \right)^{\frac{1}{2}} \left(|\varrho|^2 + |\sigma|^2 \right). \tag{3.14}$$

All expressions above are singular on the cone, so we cannot[12] directly integrate the relation (3.12) on sections of the cone C. To remedy this, we will translate the field X_0 downward by ε. Defining

$$\rho_\varepsilon = \sqrt{(t + \varepsilon)^2 + |x|^2}, \quad X_\varepsilon = \frac{1}{\rho_\varepsilon} \left((t + \varepsilon)\partial_t + x \cdot \partial_x \right)$$

$$u_\varepsilon = t + \varepsilon - |x|, \quad v_\varepsilon = t + \varepsilon + |x|,$$

we now have the shifted relations

$$\boldsymbol{\nabla}^\alpha \big(^{(X_\varepsilon)}P_\alpha \big) = \frac{2}{\rho_\varepsilon} |\iota_{X_\varepsilon} F|^2 \tag{3.15}$$

---

[12]Unless the flux is zero. This is in general not the case, instead we will work in settings where the flux is merely small.

where

$$^{(X_\varepsilon)}P_L = \left(\frac{v_\varepsilon}{u_\varepsilon}\right)^{\frac{1}{2}}|\alpha|^2 + \left(\frac{u_\varepsilon}{v_\varepsilon}\right)^{\frac{1}{2}}\left(|\varrho|^2 + |\sigma|^2\right), \tag{3.16}$$

$$^{(X_\varepsilon)}P_{\underline{L}} = \left(\frac{u_\varepsilon}{v_\varepsilon}\right)^{\frac{1}{2}}|\underline{\alpha}|^2 + \left(\frac{v_\varepsilon}{u_\varepsilon}\right)^{\frac{1}{2}}\left(|\varrho|^2 + |\sigma|^2\right). \tag{3.17}$$

We also remind the reader that since $\partial_t = \frac{1}{2}(L + \underline{L})$, we have

$$^{(X_\varepsilon)}P_0 = \frac{1}{2}\left(^{(X_\varepsilon)}P_L + {}^{(X_\varepsilon)}P_{\underline{L}}\right).$$

Integrating now the relation (3.12) over an appropriate section of the cone C yields the following:

Proposition 3.3. *Let A be a finite energy Yang–Mills connection on $[\varepsilon, 1] \times \mathbb{R}^4$, where $\varepsilon \in (0,1)$. Suppose furthermore that A satisfies*

$$\mathcal{E}_{S_1}(A) \leq E, \quad \mathcal{F}_{\partial C_{[\varepsilon,1]}}(A) \leq \varepsilon^{\frac{1}{2}}E. \tag{3.18}$$

Then

$$\int_{S_1} {}^{(X_\varepsilon)}P_0(A)\,\mathrm{d}x + \iint_{C_{[\varepsilon,1]}} \frac{2}{\rho_\varepsilon}|\iota_{X_\varepsilon}F|^2\mathrm{d}t\mathrm{d}x \lesssim E, \tag{3.19}$$

where the implicit constant is independent of ε, E.

Using Proposition 3.3, we can also establish a version of (3.19) that is localized away from the boundary of the cone. This statement will be useful for propagating lower bounds in a time-like region towards $(0,0)$.

Proposition 3.4. *Let A a finite Yang–Mills connection in $[\varepsilon, 1] \times \mathbb{R}^4$, where $\varepsilon \in (0,1)$. Suppose furthermore that A satisfies (3.18). Then for $2\varepsilon \leq \delta_0 < \delta_1 \leq t_0 \leq 1$, we have*

$$\int_{S_1^{\delta_1}} {}^{(X_0)}P_0(A)\,\mathrm{d}x \leq \int_{S_{t_0}^{\delta_0}} {}^{(X_0)}P_0(A)\,\mathrm{d}x + C\left((\delta_1/t_0)^{\frac{1}{2}} + |\log(\delta_1/\delta_0)|^{-1}\right)E. \tag{3.20}$$

The proofs of Propositions 3.3 and 3.4 are similar to those of Propositions 5.4 and 5.5 in [34], respectively, and thus are omitted.

4. A COMPACTNESS RESULT

Here we establish the following compactness result for a locally uniformly bounded sequence of Yang–Mills connections which are asymptotically stationary:

Theorem 4.1. *Let $A^{(n)}$ be a sequence of finite energy Yang–Mills connections in $[-2,2]\times B_{2R}$ which is uniformly bounded, in the sense that the norms*

$$\|A^{(n)}\|_{L^\infty(\dot{H}^1 \cap L^4)}, \ \|\partial_t A^{(n)}\|_{L^\infty L^2}, \ \|\nabla A_0^{(n)}\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}}, \ \|\Box A_x^{(n)}\|_{\ell^1 L^2 \dot{H}^{-\frac{1}{2}}} \tag{4.1}$$

on $[-2,2] \times B_{2R}$ are uniformly bounded in n. Assume moreover that

$$\lim_{n\to\infty} \|\iota_V F^{(n)}\|_{L^2([-2,2]\times B_{2R})} = 0,$$

where V is a smooth time-like vector field (i.e., $\mathbf{m}(V,V) < 0$). Then on a subsequence we have

$$A^{(n)} \to A \qquad in \ H^1([-1,1] \times B_R)$$

where A solves the hyperbolic Yang–Mills equation (in the sense of distributions), satisfies $\iota_V F = 0$, and has regularity

$$A_x \in \ell^1 H^{\frac{3}{2}}([-1,1] \times B_R), \qquad \nabla A_0 \in \ell^1 L^2 H^{\frac{1}{2}}([-1,1] \times B_R).$$

Clearly, by scaling, this result is applicable to any spacetime cylinder $(t_0 - 2T, t_0 + 2T) \times B_{2R}(x_0)$. In the sequel, Theorem 4.1 will be applied to a sequence $A^{(n)}$ on nested open sets $\mathcal{O}^{(n)}(\subset \mathcal{O}^{(n+1)})$, which satisfies the hypothesis on each spacetime cylinder inside $\mathcal{O}^{(n)}$ for sufficiently large n after taking a suitable gauge transformation; see Section 7 below. Thus, on any open set $\mathcal{O}' \subset \overline{\mathcal{O}'} \subset \bigcup_n \mathcal{O}^{(n)}$, we will extract a subsequential limit via a diagonal procedure possessing the local regularity

$$A_x \in \ell^1 H^{\frac{3}{2}}_{loc}(\mathcal{O}'), \quad \nabla A_0 \in \ell^1 L^2 H^{\frac{1}{2}}_{loc}(\mathcal{O}').$$

Proof. Let χ be a smooth cutoff supported in $[-2,2] \times B_{2R}$, which is identically 1 in $[-1,1] \times B_R$. Consider the sequence $\{\chi A^{(n)}\}$, which is now globally defined in $\mathbb{R}^{1+4}$. Moreover, we claim that $\{\chi A^{(n)}\}$ is uniformly bounded with respect to n in the global-in-spacetime version of the norms listed in (4.1); the bound depends on R and the corresponding norms of $A^{(n)}$ on $[-2,2] \times B_{2R}$. Indeed, it is straightforward to reduce the claim to the following global-in-spacetime commutator bounds:

$$\|[\nabla, \chi]B\|_{L^\infty L^2} \lesssim_R \|B\|_{L^\infty L^4},$$
$$\|[\nabla, \chi]B\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}} \lesssim_R \|\nabla B\|_{L^\infty L^2},$$
$$\|[\Box, \chi]B\|_{\ell^1 L^2 \dot{H}^{-\frac{1}{2}}} \lesssim_R \|\nabla B\|_{L^\infty L^2},$$

These commutator bounds, in turn, follow from the usual Littlewood–Paley trichotomy analysis.

Next, we microlocally split the connections $\chi A^{(n)}$ into a high modulation part and a low modulation part

$$\chi(t,x)A^{(n)} = A^{(n),lo} + A^{(n),hi} := \eta(D_{t,x})\chi(t,x)A^{(n)} + (1 - \eta(D_{t,x}))\chi(t,x)A^{(n)}$$

i.e., with a classical order zero multiplier η which is supported in a small neighborhood $\{||\tau| - |\xi|| < \kappa|\xi|\}$ of the null cone $\{\tau^2 = \xi^2\}$ and which is identically 1 in the smaller neighborhood $\{||\tau| - |\xi|| < \frac{\kappa}{2}|\xi|\}$. We choose κ small enough so that ∂_0 and $V^\alpha \partial_\alpha$ are microlocally elliptic (i.e., $|\tau| \gtrsim |\tau| + |\xi|$ and $|V^0 \tau + V^k \xi_k| \gtrsim |\tau| + |\xi|$) on the support of $\eta(\tau,\xi)\chi(t,x)$, which is possible since ∂_0 and $V^\alpha \partial_\alpha$ are time-like.

Since $\Box$ is microlocally elliptic in the support of $1 - \eta(\tau, \xi)$, the uniform bound for $\Box(\chi A^{(n)})$ implies that the high modulation parts $A_x^{(n),hi}$ are uniformly bounded in $\ell^1 H^{\frac{3}{2}}$. The same happens with the $\nabla(\chi A_0^{(n)})$ component in its entirety in $\ell^1 L^2 H^{\frac{1}{2}}$. On a subsequence we get convergence in $H^1([-1,1] \times B_R)$ for $A_x^{(n),hi}$ and $A_0^{(n)}$ with the limits $A_x^{hi} \in \ell^1 H^{\frac{3}{2}}([-1,1] \times B_R)$ and $\nabla A_0 \in \ell^1 L^2 H^{\frac{1}{2}}([-1,1] \times B_R)$, respectively.

It remains to consider the low modulation part of $A_x^{(n)}$. For this we expand $\iota_V F$ as

$$(\iota_V F)_\beta = V^\alpha F_{\alpha\beta} = V^\alpha(\partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta])$$

Separating the cases $\beta = 0$ and $\beta \neq 0$, we view this as a system for A_k of the form

$$V^\alpha \partial_\alpha A_k - \partial_k(V^j A_j) = \left(-(\iota_V F)_k + V^\alpha[A_\alpha, A_k] + V^0 \partial_k A_0\right) - (\partial_k V^j)A_j, \qquad (4.2)$$

$$\partial_0(V^j A_j) = \left(-(\iota_V F)_0 + V^j(\partial_j A_0 + [A_j, A_0])\right) + (\partial_0 V^j)A_j. \qquad (4.3)$$

30

Here the LHS can be viewed as a system in A_k which is microlocally elliptic of order 1 on the support of $\eta(\tau, \xi)\chi(t, x)$. To exploit this fact, we apply $\eta(D_{t,x})\chi(t, x)$ to the both sides and rewrite the above system as

$$V^\alpha \partial_\alpha(\eta(D_{t,x})\chi A_k) - \partial_k(\eta(D_{t,x})\chi V^j A_j) = \eta(D_{t,x})\chi(\text{RHS of } (4.2))$$
$$+ [V^\alpha \partial_\alpha, \eta(D_{t,x})\chi]A_k - [\partial_k, \eta(D_{t,x})\chi]V^j A_j$$
$$\partial_0(\eta(D_{t,x})\chi V^j A_j) = \eta(D_{t,x})\chi(\text{RHS of } (4.3)) + [\partial_0, \eta(D_{t,x})\chi]V^j A_j.$$

It is straightforward to check that the resulting RHS has size

$$o_{L^2}(1) + O_{\ell^1 H^{\frac{1}{2}}}(1),$$

where we note that only the terms of the form $(\iota_V F)_\beta$ contribute $o_{L^2}(1)$. Thus, on a subsequence, we get convergence in $H^1([-1, 1] \times B_R)$ first for $\eta(D_{t,x})\chi V^j A_j^{(n)}$ and then for $A_x^{(n),lo} = \eta(D_{t,x})(\chi A_x)$, with the limits in $\ell^1 H^{\frac{3}{2}}([-1, 1] \times B_R)$. Now the convergence of $F^{(n)}$ in $L^2([-1, 1] \times B_R)$ is easy to establish. $\square$

5. Regularity of stationary connections

Here we consider the solutions provided as limiting connections in Theorem 4.1 (see the discussion following the theorem). These have the local regularity

$$A_x \in \ell^1 H_{loc}^{\frac{3}{2}}(\mathcal{O}), \qquad \nabla A_0 \in \ell^1 L^2 H_{loc}^{\frac{1}{2}}(\mathcal{O}) \tag{5.1}$$

on some open subset $\mathcal{O}$ of $\mathbb{R}^{1+4}$, and satisfy

$$\iota_V F = 0. \tag{5.2}$$

We further specialize to the following two cases:

(i) V is constant and time-like, or
(ii) $V = S = x^\alpha \partial_\alpha$ is the scaling vector field.

Moreover, in Case (ii), A is defined in the forward light cone, i.e., $\mathcal{O} \subset C$. The goal of this section is to establish the following (qualitative) regularity result.

Proposition 5.1. *Let A be a hyperbolic Yang–Mills connection on an open set $\mathcal{O} \subset \mathbb{R}^{1+4}$ that satisfies the above properties. Then in any open set $\mathcal{O}'$ such that $\overline{\mathcal{O}'} \subset \mathcal{O}$, the connection A is gauge-equivalent, via continuous local gauge transformations to a smooth connection. More precisely, there exists an open covering $\{B\}$ of $\mathcal{O}'$ and a continuous gauge transformation O on each B such that $\tilde{A} = \mathcal{G}(O)A$ is smooth.*

Passing to a smaller set $\mathcal{O}'$ makes the statement and the proof simple. This point will not be problematic for our application, thanks to the following extension result for a smooth hyperbolic Yang–Mills connection satisfying (5.2):

Proposition 5.2. *Let A be a hyperbolic Yang–Mills connection on an open set $\mathcal{O} \subset \mathbb{R}^{1+4}$ that satisfies the above properties. Assume furthermore that A is smooth. Then A is gauge-equivalent, via a smooth gauge transformation, to a smooth connection $\tilde{A} = \mathcal{G}(O)A$ obeying $\iota_V \tilde{A} = 0$ and $\mathcal{L}_V \tilde{A} = 0$. Moreover, the extension of $\tilde{A}$ to $\bigcup_{s \in \mathbb{R}} {}^{(V)}\Phi_s(\mathcal{O})$ via $\mathcal{L}_V \tilde{A} = 0$, which we still denote by $\tilde{A}$, remains a smooth hyperbolic Yang–Mills connection satisfying (5.2).*

Here, $^{(V)}\Phi_s$ is the one-parameter family of diffeomorphisms generated by the vector field V and $\mathcal{L}_V$ is the Lie derivative with respect to V, i.e., $\mathcal{L}_V \tilde{A} = \frac{\mathrm{d}}{\mathrm{d}s}{}^{(V)}\Phi_s^* \tilde{A}|_{s=0}$.

In the sequel, we will apply Proposition 5.1 so that

$$\mathcal{O}' = \begin{cases} (-\frac{1}{4}, \frac{1}{4}) \times \mathbb{R}^4 & \text{in Case (i),} \\ C^{\frac{3}{2}}_{[\frac{3}{2}, \infty)} & \text{in Case (ii).} \end{cases}$$

In both cases $\mathcal{O}'$ is contractible so that the gauge-equivalent smooth connection may be represented by a single $\mathfrak{g}$-valued 1-form $\tilde{A}$ on $\mathcal{O}'$. By continuity of the local gauge transformations, it follows that the global gauge transformation O from A to $\tilde{A}$ is continuous. Moreover, by the formula

$$O_{;\alpha} = Ad(O)A - \tilde{A},$$

regularity (5.1) and smoothness of $\tilde{A}$, O is admissible in the sense of Definition 1.3. Finally, by Proposition 5.2 the smooth connection $\tilde{A}$ extends to a smooth stationary Yang–Mills connection on $\mathbb{R}^{1+4} = \bigcup_{s \in \mathbb{R}} {}^{(V)}\Phi_s(\mathcal{O}')$ in Case (i), and to a smooth self-similar Yang–Mills connection on $C = \bigcup_{s \in \mathbb{R}} {}^{(V)}\Phi_s(\mathcal{O}')$ in Case (ii); see Section 7 below.

The remainder of the section is devoted to the proofs of Propositions 5.1 and 5.2.

Proof of Proposition 5.1. We first describe the main idea. In both cases, the basic observation is that we can use the relation $\iota_V F = 0$ to change the Yang–Mills equation to be elliptic in spacetime. More precisely, A obeys the Yang–Mills equation with the (inverse) metric changed from $\mathbf{m}^{-1}$, which is Lorentzian, to

$$\mathbf{e}^{-1} = \mathbf{m}^{-1} - \frac{2}{\mathbf{m}(V, V)} V \otimes V. \tag{5.3}$$

Indeed, by the variational formulation of the Yang–Mills equation, it suffices to show that, under the condition (5.2), the Lagrangian with respect to $\mathbf{m}$ agrees with that with respect to $\mathbf{e}$, i.e.,

$$(\mathbf{m}^{-1})^{\alpha\alpha'}(\mathbf{m}^{-1})^{\beta\beta'} F_{\alpha\beta} F_{\alpha'\beta'} \mathrm{dVol}_{\mathbf{m}} = (\mathbf{e}^{-1})^{\alpha\alpha'}(\mathbf{e}^{-1})^{\beta\beta'} F_{\alpha\beta} F_{\alpha'\beta'} \mathrm{dVol}_{\mathbf{e}},$$

where $\mathrm{dVol}_{\mathbf{m}}$ (resp. $\mathrm{dVol}_{\mathbf{e}}$) is the volume form associated with $\mathbf{m}$ (resp. $\mathbf{e}$). This property is easily verified by choosing at each point an $\mathbf{m}$-orthonormal frame that includes $\hat{V} = |\mathbf{m}(V, V)|^{-1/2} V$, which is also $\mathbf{e}$-orthonormal (and vice versa) by (5.3).

Therefore, our connection A is harmonic in $(\mathcal{O}, \mathbf{e})$. Since it has local regularity $A_x \in \ell^1 H^{\frac{3}{2}}_{loc}(\mathcal{O})$ and $\nabla A_0 \in \ell^1 L^2 H^{\frac{1}{2}}_{loc}(\mathcal{O})$, which is the critical regularity in 5 dimensions, by an argument similar to Theorem 1.5 we conclude it is locally smooth in a suitable gauge, and thus globally smooth in a suitable gauge.

In Case (i), $\mathbf{e}$ is simply the Euclidean metric. In Case (ii), $\mathbf{e}$ takes the form

$$\mathbf{e} = \mathrm{d}\rho^2 + \rho^2 \mathrm{d}y^2$$

in the hyperbolic polar coordinates (ρ, y) on $C \subset \mathbb{R}^{1+4}$ (see Section 1.4), where $\mathrm{d}y^2$ denotes the standard metric on the hyperbolic space $\mathbb{H}^4$ with sectional curvature -1. The one difference is that our elliptic problems now have variable coefficients.

For a more detailed analysis, we implement the ideas above in three steps:

Step 1: Local Coulomb gauge with respect to e. Let $x \in \mathcal{O}'$. Our aim is to place the equations in a local Coulomb gauge with respect to the metric $\mathbf{e}$,

$$\boldsymbol{\nabla}^\alpha A_\alpha = 0,$$

in a sufficiently small ball B in $\mathcal{O}$ centered at x via a gauge transformation O with regularity

$$O_{;\alpha} \in L^5(B), \quad \nabla O_{;\alpha} \in L^2 H^{\frac{1}{2}}(B). \tag{5.4}$$

Here and in the sequel, $\boldsymbol{\nabla}$ is the Levi-Civita connection associated with the Riemannian metric $\mathbf{e}$ as in (5.3), and we raise and lower greek (spacetime) indices using $\mathbf{e}$.

Let $\bar{\mathbf{e}}_{\alpha\beta} = \mathbf{e}_{\alpha\beta}(x)$. We take the ball B centered at x to be small enough so that

$$\|\nabla A\|_{L^2 H^{\frac{1}{2}}_B} + \|A\|_{L^5(B)}, \quad \|\mathbf{e}_{\alpha\beta} - \bar{\mathbf{e}}_{\alpha\beta}\|_{C^2(2B)}$$

are sufficiently small (here, $2B$ is the double enlargement of B). Then we may find extensions of A and $\mathbf{e}$ from B to $\mathbb{R}^5$ such that $\mathbf{e} - \bar{\mathbf{e}}$ is supported in $2B$ and $\|\nabla A\|_{L^2 \dot{H}^{\frac{1}{2}}} + \|\mathbf{e} - \bar{\mathbf{e}}\|_{C^2}$ is small. We look for a global gauge transformation O into the Coulomb gauge; note that $\Omega_\alpha = O^{-1}\partial_\alpha O$ must solve

$$\begin{cases} \boldsymbol{\nabla}^\alpha \Omega_\alpha = \boldsymbol{\nabla}^\alpha A_\alpha + [\Omega^\alpha, A_\alpha], \\ \boldsymbol{\nabla}_\alpha \Omega_\beta - \boldsymbol{\nabla}_\beta \Omega_\alpha = -[\Omega_\alpha, \Omega_\beta]. \end{cases} \tag{5.5}$$

By Proposition A.7 and the smallness properties of A, $\mathbf{e} - \bar{\mathbf{e}}$ (as well an extra iteration procedure to include the term $[\Omega^\alpha, A_\alpha]$), we may find a unique solution Ω to (5.5) such that $\Omega \in L^5$ and $\nabla\Omega \in L^2 \dot{H}^{\frac{1}{2}}$. Integrating the system of ODEs $O^{-1}\partial_\alpha O = \Omega$, for which the curl condition serves as the compatibility condition needed for integrability[13], we find a gauge transformation O in B with regularity

$$O^{-1}\partial_\alpha O \in L^5(B), \quad \nabla(O^{-1}\partial_\alpha O) \in L^2 H^{\frac{1}{2}}(B), \tag{5.6}$$

such that $\tilde{A} = \mathcal{G}(O)A$ obeys the Coulomb condition in B. Note that $O_{;\alpha} = Ad(O)(O^{-1}\partial_\alpha O)$ has the regularity (5.6) as well.

Step 2: Elliptic regularity. Here we examine the output connection $\tilde{A}$, which a-priori has the same regularity as $O_{;\alpha}$, i.e.,

$$\tilde{A} \in L^5(B), \quad \nabla\tilde{A} \in L^2 L^{\frac{8}{3}}(B),$$

solves the harmonic Yang–Mills equation

$$\tilde{\mathbf{D}}^\alpha \tilde{F}_{\alpha\beta} = 0,$$

where $\tilde{\mathbf{D}} = \boldsymbol{\nabla} + ad(\tilde{A})$, and satisfies the Coulomb gauge condition

$$\boldsymbol{\nabla}^\alpha \tilde{A}_\alpha = 0.$$

Together these form an elliptic system for $\tilde{A}_\alpha$ of the form

$$\Delta_{\tilde{A}} \tilde{A}_\alpha - Ric[\mathbf{e}]_{\alpha\beta} \tilde{A}^\beta = [\tilde{A}^\beta, \tilde{\mathbf{D}}_\alpha \tilde{A}_\beta],$$

where $\Delta_{\tilde{A}} = \tilde{\mathbf{D}}^\beta \tilde{\mathbf{D}}_\beta$ and $Ric[\mathbf{e}]$ is the Ricci curvature of $(B, \mathbf{e})$. Since a-priori $\tilde{A} \in L^5(B)$, which is critical regularity, by (perturbative) elliptic theory it follows that the solutions are

---

[13]To make this procedure rigorous, one first approximate A by smooth 1-forms, so that the corresponding Ω's are smooth, then take the limit by compactness using the bounds on Ω.

smooth in any smaller ball $B' \subset \overline{B'} \subset B$; for simplicity, we shrink the ball B so that $\tilde{A}$ is smooth on B.

Step 3: The regularity of the gauge. So far, we have shown that at each point $x \in \mathcal{O}'$, there exists a ball $B \ni x$ in $\mathcal{O}$ and a gauge transform O with regularity (5.6) on B such that $\tilde{A} = \mathcal{G}(O)A$ is smooth. Here the goal is to use (5.1) to boost this regularity to

$$\nabla O_{;\alpha} \in \ell^1 L^2 H^{\frac{1}{2}}(B). \tag{5.7}$$

By localization of the global-in-spacetime bound

$$\|u\|_{\ell^1 L^\infty \dot{H}^1} \lesssim \|\nabla u\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}}, \tag{5.8}$$

which follows by applying the trace theorem to each $P_k u$ and summing up in k, it would follow that $O_{;\alpha} \in \ell^1 L^\infty H^1(B)$. By Lemma A.4, we would have $O \in C^0(B)$ as well, as desired.

By (5.1) (as well as $A_\alpha \in L^\infty H^1_{loc}(\mathcal{O})$), (5.6), smoothness of $\tilde{A}$ and the formula

$$O_{;\alpha} = Ad(O)A_\alpha - \tilde{A}_\alpha, \tag{5.9}$$

it follows that

$$O_{;\alpha} \in L^\infty H^1(B), \quad \nabla O_{;\alpha} \in L^2 H^{\frac{1}{2}}(B). \tag{5.10}$$

To exploit the additional regularity $\nabla A \in \ell^1 L^2 H^{\frac{1}{2}}(B)$ from (5.1), we note that the following global-in-spacetime bounds hold:

$$\|uv\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}} \lesssim \|\nabla u\|_{L^2 \dot{H}^{\frac{1}{2}}} \|\nabla v\|_{L^2 \dot{H}^{\frac{1}{2}}}$$
$$\|Ad(O)u\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}} \lesssim_{\|O\|_{L^\infty}, \|O^{-1}\|_{L^\infty}} \|u\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}} + \|\nabla O_{;x}\|_{L^2 \dot{H}^{\frac{1}{2}}} \|u\|_{L^2 \dot{H}^{\frac{1}{2}}}.$$

Both estimates are straightforward to establish by standard Littlewood–Paley trichotomy, so we omit the proof. Localization of the above estimates, combined with (5.1) and (5.10), imply the desired estimate (5.7). $\qquad\square$

Proof of Proposition 5.2. Solving the nonlinear transport equation $V^\alpha \partial_\alpha O = O \iota_V A$ on $\mathcal{O}$ (with arbitrary smooth data), we find a smooth gauge transformation O on $\mathcal{O}$ such that $\tilde{A} = \mathcal{G}(O)A$ obeys $\iota_V \tilde{A} = 0$. By Cartan's formula, note that

$$\mathcal{L}_V \tilde{A} = \iota_V d\tilde{A} + d\iota_V \tilde{A} = \iota_V \tilde{F} - \frac{1}{2}\iota_V[\tilde{A} \wedge \tilde{A}] + d\iota_V \tilde{A} = 0,$$

where in the last equality, we used (5.2) for the first term and $\iota_V \tilde{A} = 0$ for the others. Thus the extension of $\tilde{A}$ to $\bigcup_{s \in \mathbb{R}} {}^{(V)}\Phi_s(\mathcal{O})$ by solving $\mathcal{L}_V \tilde{A} = 0$, which amounts to solving an ODE along each integral curve of V, is well-defined. That the extension satisfies $\iota_V \tilde{A} = 0$ is clear, whereas (5.2) follows by reversing the preceding computation. By Cartan's formula applied to $\mathcal{L}_V F$, it also follows that $\mathcal{L}_V F = 0$.

It remains to show that the extension still solves the Yang–Mills equation. First, note that (1.1) may be rewritten in the form

$$\tilde{\mathbf{d}}(\star F) = 0,$$

where $\tilde{\mathbf{d}}$ is the covariant exterior derivative associated with $\tilde{A}$[14] and $\star$ is the Hodge star operator associated with $\mathbf{m}$. By $[d, \mathcal{L}_V] = 0$ and $\mathcal{L}_V \tilde{A} = 0$, it follows that $[\mathcal{L}_V, \tilde{\mathbf{d}}] = 0$.

---

[14]This operator is characterized by linearity and $\tilde{\mathbf{d}}(u \otimes \omega) = \tilde{\mathbf{D}}u \wedge \omega + u d\omega$, where u is a $\mathfrak{g}$-valued function and ω is a k-form.

Moreover, since $\mathcal{L}_V \mathbf{m} = 0$ in Case (i) and $\mathcal{L}_V \mathbf{m} = 2\mathbf{m}$ in Case (ii), it follows that $[\mathcal{L}_V, \star]F = c \star F$, where $c = 0$ in Case (i) and $c = 1$ (i.e., the spacetime dimension minus 4) in Case (ii). In conclusion, $\mathcal{L}_V(\tilde{\mathbf{d}}(\star F)) = c\tilde{\mathbf{d}}(\star F)$, from which the desired conclusion follows. □

6. No finite energy self-similar solutions

One of the main enemies in proving the bubbling-off result is given by self-similar solutions. Here we prove that no nontrivial finite energy self-similar Yang–Mills connections exist:

Theorem 6.1. *There are no smooth nontrivial self-similar solutions to the hyperbolic Yang–Mills equation (i.e., $\iota_S F = 0$) defined on the whole forward light cone C which have finite energy.*

In our application, smoothness of the self-similar solution in C (the *open* forward light cone) follows from the results in Section 5, but the only information a-priori available near the boundary is the finite energy condition. The main issue in the proof of Theorem 6.1 is indeed the analysis near the boundary.

Proof. We proceed in several steps.

Step 1. We first recast the problem in hyperbolic polar coordinates, parametrizing the forward light cone C as

$$C = \{\rho y \in \mathbb{R}^{1+4} : \rho \in (0, \infty),\ y \in \mathbb{H}^4\},$$

where we remind the reader that $\rho = \sqrt{t^2 - |x|^2}$ and $\mathbb{H}^4 = \{(t, x) \in C : \rho = 1\}$ (see Section 1.4). The Minkowski metric becomes

$$\mathrm{d}s^2 = -\mathrm{d}\rho^2 + \rho^2 \mathrm{d}y^2,$$

where $\mathrm{d}y^2$ denotes the induced metric on $\mathbb{H}^4$; as is well-known, it is the standard metric on the hyperbolic space $\mathbb{H}^4$ with sectional curvature -1.

Inside the light cone C, the self-similarity condition $\iota_S F = 0$ becomes $\iota_{\partial_\rho} F = 0$. By Proposition 5.2, we may make a smooth gauge transformation to make $A_\rho = \iota_{\partial_\rho} A = 0$. Then our connection is still smooth, and also independent of ρ in the sense that $\mathcal{L}_{\partial_\rho} A = 0$ (see the proof of Proposition 5.2). We may furthermore check that the pullback of A to $\mathbb{H}^4 = \{\rho = 1\}$, which we still denote by A, is a solution to the harmonic Yang–Mills equation in $\mathbb{H}^4$.

Using the stereographic projection, we represent $\mathbb{H}^4$ as the unit disc $\mathbb{D}^4$ in $\mathbb{R}^4$ with metric

$$\mathrm{d}s^2 = \Omega^2 \mathrm{d}x^2, \qquad \Omega = \frac{2}{1 - |x|^2}.$$

By conformal invariance of the harmonic Yang–Mills equation in dimension 4, the conformal factor drops out, and we obtain the elliptic Yang–Mills system in $\mathbb{D}^4$

$$\mathbf{D}^i F_{ij} = 0.$$

with respect to the Euclidean metric $\mathrm{d}s^2 = \mathrm{d}x^2$.

We now move the finite energy condition in hyperbolic polar coordinates, then translate it to $\mathbb{D}^4$. This computation is equivalent to that in [34, Section 7.2], and yields

$$\int_{\mathbb{D}^4} \frac{1 + r^2}{1 - r^2} |F|^2 \mathrm{d}x < \infty, \tag{6.1}$$

where $r = |x|$. At this point we know that the connection is smooth inside $\mathbb{D}^4$, but nothing about its behavior at the boundary. Let $\mathcal{A}_{(\frac{1}{2},1)} = \{x \in \mathbb{D}^4 : \frac{1}{2} < |x| < 1\}$. We claim that there exists a gauge such that $A \in H^1(\mathcal{A}_{(\frac{1}{2},1)})$ and

$$\int_{\mathcal{A}_{(\frac{1}{2},1)}} \frac{1}{1-r^2}|\nabla A|^2 + \frac{1}{(1-r^2)^3}|A|^2 \, \mathrm{d}x < \infty. \tag{6.2}$$

Assuming the claim, the proof of the theorem may be completed as follows. By (6.2), it follows that $A|_{\partial \mathbb{D}^4}$ vanishes. Thus its zero extension $\bar{A}$ outside the ball is also in H^1, and its curvature $\bar{F}$ is the zero extension of F. We conclude that the zero extension of $A \in H^1(\mathbb{D}^4)$ satisfying (6.2) still solves the harmonic Yang–Mills system

$$\bar{\mathbf{D}}^i \bar{F}_{ij} = 0.$$

By the classical elliptic regularity results of Uhlenbeck (Theorem 1.5), the connection $\bar{A}$ is gauge equivalent to a smooth connection $\tilde{A}$ in $\mathbb{R}^4$. To continue we write elliptic equations for $\tilde{F}$,

$$\Delta_{\tilde{A}} \tilde{F} = [\tilde{F}, \tilde{F}]$$

Since $\tilde{F}$ has compact support and $\tilde{A}$ is smooth (thus bounded in the support of $\tilde{F}$), it follows from the classical elliptic unique continuation result[15] due to Aronszajn [1] that $\tilde{F} = 0$. Thus the connection A is trivial.

It remains to prove our claim, and show that a representation satisfying (6.2) exists.

Step 2. For $0 < d < \frac{1}{6}$, $k \geq 0$ and $2 \leq p \leq \infty$, we claim that

$$\|\mathbf{D}^{(k)}F\|_{L^p(\mathcal{A}_{(1-2d,1-d)})} \lesssim d^{-\alpha+\frac{1}{2}}\|d^{-\frac{1}{2}}F\|_{L^2(\mathcal{A}_{(1-4d,1-\frac{d}{2})})}, \quad \alpha = k + 2 - \frac{4}{p}. \tag{6.3}$$

This is proved by applying Uhlenbeck's lemma (Theorem A.9) in balls of size proportional to their distance to the boundary, and using the fact that the harmonic Yang–Mills equation becomes strictly elliptic in the Coulomb gauge, which allows us to use interior elliptic regularity.

Step 3. In the remainder of this proof, we work in the polar coordinates $x = r\Theta$ on $\mathbb{D}^4 \setminus \{0\}$. As stated in Section 1.4, we write $A_\Theta(r, \cdot)$ for the pullback of a 1-form A to each constant r-sphere, which *we then view as a 1-form on the unit sphere* $(\mathbb{S}^3, \mathbf{g})$. Alternatively, one may think of the whole analysis in the rest of this proof as taking place on $(0,1)_r \times \mathbb{S}^3_\Theta$ equipped with the metric $\mathrm{d}s^2 = \mathrm{d}r^2 + \mathrm{d}\Theta^2$.

Roughly speaking, the idea is to fix the gauge by specifying the conditions

$$A_r = 0 \text{ in } \mathcal{A}_{(\frac{1}{2},1)}, \quad A(r, \cdot) \to 0 \text{ and } \boldsymbol{\nabla}^\Theta A_\Theta(r, \cdot) \to 0 \text{ as } r \to 1, \tag{6.4}$$

which is possible, at least heuristically, by the decay of F as $r \to 1$ and Uhlenbeck's lemma. For technical reasons, however, we proceed slightly differently and work with a sequence of gauges approximately satisfying (6.4).

We start with the connection $A^{(0)}$ in the exponential gauge at the origin (i.e., $A_r^{(0)} = 0$ in $\mathbb{D}^4$ and $A^{(0)}(0) = 0$). Observe that $A^{(0)}$ is smooth. By (6.1), we may find a sequence $r_n \to 1$ such that $\|(1-r_n)^{-\frac{1}{2}}F(r_n, \cdot)\|_{L^2_\Theta} \to 0$. Viewing $A^{(0)}_\Theta(r_n, \cdot)$ as a connection 1-form on the unit sphere $\Theta \in \mathbb{S}^3$ and applying Uhlenbeck's lemma on $\mathbb{S}^3$ (Proposition A.11), we find gauge

---

[15]For further references on unique continuation for second order elliptic PDEs, see, for instance, [24]

transformations $O^{(n)} = O^{(n)}(\Theta)$ on $\mathbb{S}^3$ such that the following property holds: Viewing $O^{(n)}$ as defined on $\mathbb{D}^4 \setminus \{0\}$ by $O^{(n)}(r, \Theta) = O^{(n)}(\Theta)$, the representation $A^{(n)} = \mathcal{G}(O^{(n)})A^{(0)}$ obeys

$$\boldsymbol{\nabla}^\Theta A_\Theta^{(n)}(r_n, \cdot) = 0, \quad \|\boldsymbol{\nabla}_\Theta A_\Theta^{(n)}(r_n, \cdot)\|_{L_\Theta^2} \lesssim \|F(r_n, \cdot)\|_{L_\Theta^2}, \tag{6.5}$$

as well as

$$A_r^{(n)}(r, \cdot) = 0 \quad \text{for } 0 < r < 1, \tag{6.6}$$

simply due to $A_r^{(0)} = 0$ and the r-independence of $O^{(n)}$. Thanks to smoothness of $A^{(0)}$ and (6.3) (elliptic regularity for F), note that $O^n \in C^\infty(\mathbb{S}^3)$.

Step 4. Let

$$D^2 = \int_{\mathcal{A}_{(\frac{1}{3},1)}} \frac{1}{1-r}|F|^2 \, \mathrm{d}x, \quad \epsilon_n^2 = \frac{1}{1-r_n} \int_{\mathbb{S}^3} |F|^2(r_n, \Theta) \, \mathrm{d}\Theta,$$

where $\mathcal{A}_{(\frac{1}{3},1)} = \{x \in \mathbb{D}^4 : \frac{1}{3} < |x| < 1\}$. By (6.1), $D^2 < \infty$, and by construction, $\epsilon_n \to 0$.

From now on, we work in the gauge constructed in the previous step. Without loss of generality, we may assume that $r_n > \frac{1}{2}$. Note that on the annulus $\mathcal{A}_{(\frac{1}{2},r_n)} = \{x \in \mathbb{D}^4 : \frac{1}{2} < r < r_n\}$, we have the equivalence $\|g\|_{L^p(\mathcal{A}_{(\frac{1}{2},r_n)})} \sim \|g\|_{L_r^p L_\Theta^p((\frac{1}{2},r_n)\times\mathbb{S}^3)}$ for any $1 \le p \le \infty$, where the constant is independent of n.

As $A_r^{(n)} = 0$ and $\partial_r A_\Theta^{(n)} = F_{r\Theta}^{(n)}$, we immediately have

$$\int_{1/2}^{r_n} \int \frac{1}{1-r}|\partial_r A^{(n)}|^2 \, \mathrm{d}r\mathrm{d}\Theta \lesssim D^2. \tag{6.7}$$

To proceed, recall the following elementary inequality (essentially one-dimensional Hardy's inequality): For $1 \le p \le \infty$ and $0 < r \le r_n$, we have

$$\int_r^{r_n} \left((1-r')^{\beta-1}|g|\right)^p \, \mathrm{d}r' \lesssim_{\beta,p} \int_r^{r_n} \left((1-r')^\beta |\partial_r g|\right)^p \, \mathrm{d}r' + \left((1-r_n)^{\beta-\frac{p-1}{p}}|g|\right)^p (r_n) \tag{6.8}$$

provided that $\beta < \frac{p-1}{p}$.

In our gauge, (6.8) implies

$$\int_{1/2}^{r_0} \int \frac{1}{(1-r)^3}|A^{(n)}|^2 \, \mathrm{d}r\mathrm{d}\Theta \lesssim D^2 + \frac{1}{(1-r_0)^2} \int |A^{(n)}|^2(r_0, \Theta) \, \mathrm{d}\Theta, \tag{6.9}$$

for any $1/2 < r_0 \le r_n$.

Step 5. To complete the proof, in view of (6.7) and (6.9), it remains to establish

$$\limsup_{n\to\infty} \int_{1/2}^{r_n} \int \frac{1}{1-r^2}|\boldsymbol{\nabla}_\Theta A^{(n)}|^2 \, \mathrm{d}r\mathrm{d}\Theta \lesssim_D 1. \tag{6.10}$$

Once (6.10) is proved, then it is a routine matter to extract a limit $O^{(n)} \rightharpoonup O$ in $H^2(\mathbb{D}^4)$ such that $A = Ad(O)A^{(0)} - O_{;x}$ obeys the desired vanishing condition (6.2).

As a first attempt to prove (6.10), note that we have control of $\partial_r\partial_\Theta A_\Theta^{(n)} = \mathbf{D}_\Theta^{(n)} F_{r\Theta}^{(n)} + O(A_\Theta^{(n)}, F_{r\Theta}^{(n)})$ by (6.3). However, if we naively use the L^2 bound in (6.3), we encounter a logarithmic divergence. To rectify this, we use an additional cancellation from the harmonic Yang–Mills equation.

37

The idea is to compute the div-curl system on $\mathbb{S}^3$ satisfied by $A_\Theta^n(r, \cdot)$. First, note that

$$\mathbf{D}^\Theta F_{r\Theta} = r^2 \mathbf{D}^j (\iota_{\partial_r} F)_j = rx^j \left(\mathbf{D}^i F_{ij} \right), \tag{6.11}$$

where the last term is zero if the harmonic Yang–Mills equation holds. Therefore, we have

$$\partial_r \mathbf{\nabla}^\Theta A_\Theta^{(n)} = \mathbf{\nabla}^\Theta F_{r\Theta}^{(n)} = (\mathbf{D}^{(n)})^\Theta F_{r\Theta}^{(n)} - ad(A^{(n)\Theta}) F_{r\Theta}^{(n)}) = -ad(A^{(n)\Theta}) F_{r\Theta}^{(n)}.$$

We furthermore note that

$$\int_{1/2}^{r_n} \int \frac{1}{(1-r)^2} |A^{(n)}|^4 \, drd\Theta \lesssim D^4 + \frac{1}{1-r_n} \int |A^{(n)}|^4 (r_n, \Theta) \, d\Theta \lesssim D^4 + (1-r_n)\epsilon_n^4 \tag{6.12}$$

by (6.3) with $(k,p) = (0,4)$, (6.8) with $p = 4$ and (6.5). Then thanks to (6.3) and (6.12), we have

$$\limsup_{n\to\infty} \|\partial_r \mathbf{\nabla}^\Theta A_\Theta^{(n)}\|_{L_r^2 L_\Theta^2((1/2,r_n)\times\mathbb{S}^3)} \lesssim D^2.$$

Recall from (6.5) that $\mathbf{\nabla}^\Theta A_\Theta^{(n)}(r_n, \Theta) = 0$. By (6.8), it follows that

$$\limsup_{n\to\infty} \|(1-r)^{-1} \mathbf{\nabla}^\Theta A_\Theta^{(n)}\|_{L_r^2 L_\Theta^2((1/2,r_n)\times\mathbb{S}^3)} \lesssim D^2.$$

On the other hand, by the schematic relation

$$(\mathrm{d}A)_{\Theta\Theta} = F_{\Theta\Theta} - [A_\Theta, A_\Theta]$$

and the bound (6.12), we have

$$\limsup_{n\to\infty} \|(1-r)^{-\frac{1}{2}} (\mathrm{d}A^{(n)})_{\Theta\Theta}\|_{L_r^2 L_\Theta^2((1/2,r_n)\times\mathbb{S}^3)} \lesssim D + D^2.$$

By the div-curl estimate on $\mathbb{S}^3$, we obtain

$$\limsup_{n\to\infty} \|(1-r)^{-\frac{1}{2}} \mathbf{\nabla}_\Theta A_\Theta^{(n)}\|_{L_r^2 L_\Theta^2((1/2,r_n)\times\mathbb{S}^3)} \lesssim D + D^2,$$

which implies the desired bound (6.10). □

7. The bubbling-off result

In this section we prove the bubbling off result in Theorem 1.7. Much of the argument is similar to that in [34, Sections 8.3–8.6] (see also [44, Sections 6.5–6.8]), from which we borrow many results.

Throughout this section, we assume that A is a finite energy Yang–Mills connection satisfying the hypothesis of either Theorem 1.7.a) (finite time blow-up at (T, X)) or b) (infinite time blow-up). We write $E = \mathcal{E}(A)$ and $E_1 = \limsup_{t \nearrow T} \mathcal{E}_{C_\gamma \cap S_t}(A)$, where $T = \infty$ in the infinite time blow-up case. Moreover, in the finite time blow-up case, we translate the point (T, X) to $(0, 0)$, and reverse the time direction so that the blow up occurs inside C backward in time towards $(0, 0)$.

Our first goal is to prove that from the connection A we can extract a sequence of smooth connections $A^{(n)}$ in increasing cone sections and with decreasing fluxes:

Lemma 7.1. *There exists a sequence of smooth hyperbolic Yang–Mills connections $A^{(n)}$ in cone sections $C_{[\varepsilon_n,1]}$ with $\varepsilon_n \to 0$ which satisfy the following properties:*

(1) *Closeness to A.* There exists a sequence $\tilde{A}^{(n)}$ of rescaled and translated copies of A so that

$$\lim_{n\to\infty} \sup_{t\in[\varepsilon_n,1]} \|A^{(n)} - \tilde{A}^{(n)}\|_{\dot{H}^1\cap L^4(S_t)} = 0 \tag{7.1}$$

(2) *Bounded energy in the cone.*

$$\mathcal{E}_{S_t}(A^{(n)}) \leq E + o(1) \quad \text{for every } t \in [\varepsilon_n,1], \tag{7.2}$$

(3) *Decaying flux on ∂C.*

$$\mathcal{F}_{[\varepsilon_n,1]}(A^{(n)}) \leq \varepsilon_n^{\frac{1}{2}} E, \tag{7.3}$$

(4) *Time-like energy concentration at $t=1$.*

$$\mathcal{E}_{C_\gamma\cap S_1}(A^{(n)}) \geq E_1 > 0 \tag{7.4}$$

with some $\gamma < 1$.

Proof. To clarify the ideas we assume at first that A is smooth in the closure of C. We start from the flux-energy relation (3.8), which shows that the flux decays toward the tip of the cone in the blow-up case,

$$\lim_{t\searrow 0} \mathcal{F}_{(0,t]}(A) = 0.$$

respectively toward infinity in the non-scattering case,

$$\lim_{t\nearrow\infty} \mathcal{F}_{[t,\infty)}(A) = 0.$$

Using these properties, it easily follows in both cases that we can find a sequence of connections $A^{(n)}$ which are obtained from A simply by rescaling.

Suppose now that A is a finite energy solution inside the cone. Then its energies on each time slice $\mathcal{E}_{S_t}(A)$ are still well defined and nondecreasing in t. Thus its fluxes are at least formally defined via the energy flux relation (3.8), therefore we obtain the sequence $\tilde{A}^{(n)}$ of rescaled copies of A which satisfies the properties (2)–(4). We now consider a smooth approximation $\tilde{A}^{(n)}_\delta$ of $\tilde{A}^{(n)}$ so that

$$\sup_{t\in[\varepsilon_n+\delta,1-\delta]} \|\tilde{A}^{(n)}_\delta - \tilde{A}^{(n)}\|_{\dot{H}^1\cap L^4(S_t^\delta)} \leq \frac{1}{n}.$$

Then the desired smooth connections $A^{(n)}$ are obtained by slightly (by $O(\delta)$ to be precise) translating and rescaling $\tilde{A}^{(n)}_\delta$ provided that δ is small enough, depending on n. $\qquad\square$

At this point, we may apply Proposition 3.3 to $A^{(n)}$ and obtain

$$\iint_{C_{[\varepsilon_n,1]}} \frac{2}{\rho_{\varepsilon_n}} |\iota_{X_{\varepsilon_n}} F^{(n)}|^2 \mathrm{dt}\mathrm{d}x \lesssim E. \tag{7.5}$$

This property implies a decay of $F^{(n)}$ towards the tip of the cone C for large n, which is the key ingredient of the proof (see Lemma 7.3.(4) below).

Next, we show that the energy concentration at time $t=1$ persists in time:

Lemma 7.2. *Let $A^{(n)}$ be the sequence of smooth Yang–Mills connections in the previous lemma. Then there exist $E_2 > 0$ and $\gamma_2 \in (0,1)$ such that*

$$\int_{C_{\gamma_2}\cap S_t} {}^{(X_0)}P_0(A^{(n)})\,\mathrm{d}x \geq E_2 \quad \text{for every } t \in [\varepsilon_n^{\frac{1}{2}}, \varepsilon_n^{\frac{1}{4}}]. \tag{7.6}$$

39

This is a gauge independent property, which follows from the localized monotonicity formulas as in the (MKG) case, via Proposition 3.4; see [34, Proof of Lemma 8.10] for details.

At this point, we can freely replace ${}^{(X_0)}P$ in (7.6) by ${}^{(T)}P$ (i.e., the energy density) at the expense of adjusting E_2 as we are away from the cone. Now a final rescaling leads us to

Lemma 7.3. *There exists a sequence of smooth Yang–Mills connections $A^{(n)}$ on $[1, T_n] \times \mathbb{R}^4$ with $T_n \to \infty$ satisfying the following properties:*

(1) *Closeness to A. There exists a sequence $\tilde{A}^{(n)}$ of rescaled and translated copies of A so that*
$$\lim_{n \to \infty} \sup_{t \in [1, T_n]} \|A^{(n)} - \tilde{A}^{(n)}\|_{\dot{H}^1 \cap L^4(S_t)} = 0 \tag{7.7}$$

(2) *Bounded energy in the cone,*
$$\mathcal{E}_{S_t}(A^{(n)}) \leq E + o(1), \qquad \text{for every } t \in [1, T_n], \tag{7.8}$$

(3) *Nontrivial energy in a time-like region,*
$$\mathcal{E}_{C_{\gamma_2} \cap S_t}(A^{(n)}) \geq E_2 \quad \text{for every } t \in [1, T_n], \tag{7.9}$$

(4) *Asymptotic self-similarity,*
$$\iint_K |\iota_{X_0} F^{(n)}|^2 \mathrm{d}t\mathrm{d}x \to 0 \quad \text{as } n \to \infty \tag{7.10}$$
for every compact subset K of the interior of $C_{[1,\infty)}$.

Here, E_2 is the constant from Lemma 7.2, after making an adjustment mentioned before Lemma 7.3. This lemma is essentially rescaling and pigeonhole principle; see [34, Proof of Lemma 8.11]. Note that Property (4) follows from (7.5), which in turn was a consequence of the monotonicity formula (Proposition 3.3).

To proceed, we introduce few definitions. For each $j = 1, 2, \ldots,$ let
$$C_j = \{(t, x) \in C^1_{[1,\infty)} : 2^j \leq t < 2^{j+1}\},$$
$$\tilde{C}_j = \{(t, x) \in C^{1/2}_{[1/2,\infty)} : 2^j \leq t < 2^{j+1}\}.$$

Note that C_j, respectively $\tilde{C}_j$, is simply the set of points in the truncated cone $C_{[2^j, 2^{j+1})}$ at distance ≥ 1, respectively $\geq \frac{1}{2}$, from the lateral boundary ∂C.

We have the following lemma, which is basically [34, Lemma 8.12], for understanding concentration scales:

Lemma 7.4. *Let $A^{(n)}$ be a sequence of hyperbolic Yang–Mills connections as in the previous lemma. Let E_0 be sufficiently small. Then for each $j = 1, 2, \cdots,$ after passing to a subsequence, one of the following alternatives holds:*

(1) *Concentration of energy: There exist points $(t_n, x_n) \in \tilde{C}_j$, scales $r_n \to 0$ and $0 < r = r(j) < 1/4$ such that the following bounds hold:*
$$\mathcal{E}_{\{t_n\} \times B_{r_n}(x_n)}(A^{(n)}) = \epsilon_0, \tag{7.11}$$
$$\sup_{x \in B_r(x_n)} \mathcal{E}_{\{t_n\} \times B_{r_n}(x)}(A^{(n)}) \leq \epsilon_0, \tag{7.12}$$
$$\frac{1}{4r_n} \int_{t_n - 2r_n}^{t_n + 2r_n} \int_{B_r(x_n)} |\iota_{X_0} F^{(n)}|^2 \, \mathrm{d}t\mathrm{d}x \to 0 \quad \text{as } n \to \infty. \tag{7.13}$$

40

(2) *Uniform non-concentration of energy:* There exists $0 < r = r(j) < 1/4$ such that the following bounds hold:

$$\mathcal{E}_{C_\gamma \cap S_t}(A^{(n)}) \geq E_2 \quad \text{for } t \in [2^j, 2^{j+1}), \tag{7.14}$$

$$\sup_{(t,x) \in C_j} \mathcal{E}_{\{t\} \times B_r(x)}(A^{(n)}) \leq \epsilon_0, \tag{7.15}$$

$$\iint_{\widetilde{C}_j} |\iota_{X_0} F^{(n)}|^2 \, dt dx \to 0 \quad \text{as } n \to \infty. \tag{7.16}$$

In applying this lemma there are two scenarios we need to consider. Either

(i) Property (1) holds for some j, or
(ii) Property (2) holds for all j.

(i) Concentration scenario. Now we need to run a compactness argument. On a subsequence we can assume that

$$\lim_{n \to \infty} \frac{x_n}{t} = v, \qquad |v| < 1.$$

We denote $V = (1, v)$, which is a future pointing time-like vector.

We restrict the connections $A^{(n)}$ to the regions $[t_n - r_n, t_n + r_n] \times B_r(x_n)$, with fixed $r \ll 1$. Then we rescale to unit time and translate to center them at $(0, 0)$. We obtain a sequence of Yang–Mills connections $A^{(n)}$ in the time interval $[-1, 1]$ with the properties that

$$\mathcal{E}_{\{0\} \times B_1(0)}(A^{(n)}) = E_0, \tag{7.17}$$

$$\sup_{x \in B_{R_n}(0)} \mathcal{E}_{\{0\} \times B_1(x)}(A^{(n)}) \leq E_0, \tag{7.18}$$

$$\int_{-1}^{1} \int_{B_{R_n}(0)} |\iota_V F^{(n)}|^2 \, dt dx \to 0 \quad \text{as } n \to \infty, \tag{7.19}$$

where $R_n = r_n^{-1} r \to \infty$. In (7.19), we replaced X_0 by $V = \sqrt{1 - |v|^2} X_0|_{(1,v)}$ by the convergence $\frac{x_n}{t} \to v$, the decay $r_n \to 0$ and 0-homogeneity of X_0.

Our next aim is to find a sequence of admissible gauge transformations such that, after passing to a subsequence and for some A on $[-1/2, 1/2] \times \mathbb{R}^4$ such that

$$A \in \ell^1 H_{loc}^{\frac{3}{2}}, \qquad A_0 \in \ell^1 H_{loc}^{1,\frac{1}{2}},$$

we have

$$\mathcal{G}(O^{(n)}) A^{(n)} \to A \quad \text{in } H_{loc}^1([-1/2, 1/2] \times \mathbb{R}^4),$$

and in addition

$$\iota_V F = 0 \quad \text{in } [-1/2, 1/2] \times \mathbb{R}^4.$$

So far, we have only used energy considerations, which are gauge independent. For the next step, however, we need better regularity information, so for each $R = 1, 2, \ldots$, we place the above solutions in a "good gauge" $A^{R(n)}$, as provided by Theorem 2.19 on $[-1, 1] \times B_{2R}$ for n sufficiently large $r/r_n \gg R$ (the theorem is stated on a truncated cone $C_{[t_1,t_2]}$, but the domain $[-1, 1] \times B_{2R}$ is essentially the same). We apply the compactness result in Theorem 4.1 to conclude that on a subsequence we have local convergence to a finite energy

41

Yang–Mills connection A^R in $[-1/2, 1/2] \times B_R$. Moreover, for $R' < R$, note that $A^{R'(n)}$ and $A^{R(n)}$ are connected by a gauge transformation $O^{RR'(n)}$ in $B_{2R'}$ such that

$$\nabla O_{;t,x}^{RR'(n)} \in \ell^1 L^2 H_{loc}^{\frac{1}{2}}([-1, 1] \times B_{2R'}),$$

with uniform bounds on compact subsets (indeed, for $O_{;t}$, we use the regularity $\nabla A_0 \in \ell^1 L^2 H^{\frac{1}{2}}$, whereas for $\partial O_{;x}$ we use $\partial^k A_k \in \ell^1 L^2 \dot{H}^{\frac{1}{2}}$ and Lemma A.3. Finally, for $\partial_t O_{;x}$ we use $\partial_t O_{;j} = \partial_j O_{;t} + [O_{;j}, O_{;t}]$). Then, passing to a subsequence, we obtain a gauge transformation $O_{;t,x}^{RR'} \in \ell^1 L^2 H_{loc}^{\frac{1}{2}}([-1/2, 1/2] \times B_{R'})$, which is admissible, such that $A^R = \mathcal{G}(O^{RR'})A^{R'}$. By patching together A^R for $R = 1, 2, \dots$ (see, for instance, [39, Section 3.5, Scenario (3)]), we obtain a global solution A on $[-1/2, 1/2] \times \mathbb{R}^4$, as desired.

By the regularity result in Proposition 5.1, the connection A is gauge equivalent to a smooth connection in the domain $(-1/4, 1/4) \times \mathbb{R}^4$, which we still denote be A, that moreover satisfies
  (i) Nontriviality, $\mathcal{E}(A) > 0$;
  (ii) Finite energy, $\mathcal{E}(A) \leq E$;
  (iii) Stationarity, $\iota_V F = 0$.
Applying Proposition 5.2, we may place A in the gauge $\iota_V A = 0$, and extend it to the whole spacetime $\mathbb{R}^{1+4}$. Then A is a Lorentz transform of a nontrivial harmonic Yang–Mills connection Q, namely $A = L_v Q$. The theorem is proved in this case.

(ii) Non-concentration scenario. The argument is similar here. By Theorem 2.19 we can use gauge-equivalent representations of the connections $A^{(n)}$ which are in a "good gauge" as provided by Theorem 2.19, and thus are bounded in the sense of (2.15), uniformly on compact subsets of $C_{[1,T)}^1$ for each $T = 2, 3, \dots$. There applying the compactness result in Theorem 4.1 for each T, and patching together the resulting limits as in Case (i), we obtain a global nontrivial self-similar, finite energy connection A in $C_{[\frac{3}{2}, \infty)}^{\frac{3}{2}}$. Applying the regularity result in Proposition 5.1, and then Proposition 5.2 with $\mathcal{O}' = C_{[2, \infty)}^2$, we obtain a smooth global self-similar solution with finite energy inside the light cone C. The nontrivial energy of $A^{(n)}$ inside the cone (7.14) insures that this limiting connection is nontrivial. But such a connection does not exist by Theorem 6.1.

8. No null concentration

A key step in the transition from Theorem 1.7 to Theorems 1.8 and 1.11 is to deal with the possibility that the energy stays concentrated near the boundary of the light cone. Whereas it is not implausible that the energy near the cone must necessarily decay to zero (in particular, see [9] for the small data wave maps problem), at this point we are not able to prove this. Instead, here we prove a weaker statement which asserts that if almost all energy stays near the cone, then our connection admits an energy dispersed caloric representation:

Theorem 8.1. *Let A be a finite energy Yang–Mills connection on $\{1\} \times \mathbb{R}^4$. Suppose that*

$$\mathcal{E}_{S_1^\gamma}(A) + \mathcal{E}_{\{1\} \times \mathbb{R}^4 \setminus S_1}(A) \leq \epsilon_1, \tag{8.1}$$

and

$$\mathcal{E}_{S_1}(A) + \int_{S_1} {}^{(X_\varepsilon)} P_0(A) \, \mathrm{d}x \leq E_3. \tag{8.2}$$

42

Given any $\epsilon, E_3 > 0$, for sufficiently small ϵ_1, ε and γ close enough to 1 (depending only on ϵ, E_3), there exists a caloric gauge representation of the connection $A_x(1)$ so that

$$\|A_x(1)\|_{\dot{H}^1} \lesssim_{E_3} 1, \qquad \mathcal{Q}(A_x(1)) \lesssim_{E_3} 1,$$

whereas

$$\|A_x(1)\|_{L^4} + \|F(1)\|_{\dot{W}^{-1,4}} \lesssim_{E_3} \epsilon.$$

We emphasize that the term $\|F(1)\|_{\dot{W}^{-1,4}}$ contains both spatial and temporal components F_{jk} and F_{0j}, respectively, of F. In our application, control of the second term on the LHS of (8.2) will come from the monotonicity formula (Proposition 3.3).

As an immediate consequence of the last bound, we obtain the smallness of the fixed-time energy dispersion in the caloric gauge:

Corollary 8.2. *The caloric connection $A(1)$ provided by the above theorem satisfies*

$$\sup_k 2^{-2k} \|P_k F(1)\|_{L^\infty} \lesssim \epsilon. \tag{8.3}$$

The hypothesis of the theorem involves the full connection A at time $t = 1$, which includes both information about A_x, A_0 and the corresponding curvature components F_{ij} and F_{0j}. Our first task is to peel off the nonessential parts A_0 and F_{0j} and to reduce the problem to a statement about only the spatial part of the connection. To state the result, we introduce an orthonormal frame (e_r, e_1, e_2, e_3) at every point of $\mathbb{R}^4 \setminus \{0\}$, where $e_r = \partial_r$ in the polar coordinates (r, Θ) and $\{e_\mathfrak{a}\}_{\mathfrak{a}=1,2,3}$ is an orthonormal frame at x tangent to the sphere $\partial B_r(0)$ (with $r = |x|$). We also fix a small constant $0 < \delta_0 \ll \frac{1}{100}$.

Proposition 8.3. *Let A be an $\dot{H}^1$ connection in $\mathbb{R}^4$ with energy at most E_3, which satisfy the following properties:*

i) $F_{\mathfrak{a}\mathfrak{b}} = F(e_\mathfrak{a}, e_\mathfrak{b})$ ($\mathfrak{a}, \mathfrak{b} = 1, 2, 3$) is small in L^2,

$$\|F_{\mathfrak{a}\mathfrak{b}}\|_{L^2} \leq \epsilon. \tag{8.4}$$

ii) $F_{r\mathfrak{a}} = F(e_r, e_\mathfrak{a})$ ($\mathfrak{a} = 1, 2, 3$) is small outside an annulus,

$$\|F_{r\mathfrak{a}}\|_{L^2(\{\frac{7}{8} \leq |x| \leq 1\}^c)} \leq \epsilon.$$

iii) $\mathbf{D}^\mathfrak{a} F_{r\mathfrak{a}}$ (i.e., the covariant angular divergence) is small in $\dot{H}_A^{-1}$,

$$\|\mathbf{D}^\mathfrak{a} F_{r\mathfrak{a}}\|_{\dot{H}_A^{-1}} \leq \epsilon. \tag{8.5}$$

Assume that ϵ is sufficiently small,

$$\epsilon \ll_{E_3} 1.$$

Then there exists a caloric gauge representation of the connection A so that

$$\|A\|_{\dot{H}^1} \lesssim_{E_3} 1, \qquad \mathcal{Q}(A) \lesssim_{E_3} 1, \tag{8.6}$$

whereas[16]

$$\|A\|_{L^4} \lesssim_{E_3} \epsilon^{\frac{3}{8}(1-\delta_0)}. \tag{8.7}$$

---

[16]The factor $\frac{3}{8}$ can be improved to $\frac{3}{4}$ by using further techniques in [37], but for our purposes it is unnecessary.

43

We remark that the assumptions in the proposition are all formulated in a gauge invariant fashion. Most notably, assumption (iii) involves the space $\dot{H}_A^{-1}$, which is the dual of the space $\dot{H}_A^1$ with norm

$$\|B\|_{\dot{H}_A^1}^2 = \|\mathbf{D}_A B\|_{L^2}^2.$$

In particular, nothing is assumed about the $\dot{H}^1$ size of A and its various components. This turns out to be a problem in the proof, where it would be very convenient to have as a starting point a connection A with some good bounds. To address this difficulty, the key ingredient of the proof of the proposition is the following lemma, which we now state in the polar coordinates $x = r\Theta$:

Lemma 8.4. *Let A be a connection which satisfies the hypotheses of Proposition 8.3. Then there exists a gauge-equivalent connection B which has the following properties:*

(1) *B is bounded in $\dot{H}^1$,*
$$\|B\|_{\dot{H}^1} \lesssim_{E_3} 1.$$

(2) *B is small away from the unit sphere,*
$$\|B\|_{\dot{H}^1(\{\frac{3}{4} \le |x| \le 1\}^c)} \lesssim_{E_3} \epsilon.$$

(3) *B_r is small in $\mathcal{A}_{(\frac{2}{3}, \frac{4}{3})}$,*
$$\|B_r\|_{\dot{H}^1(\mathcal{A}_{(\frac{2}{3}, \frac{4}{3})})} \lesssim_{E_3} \epsilon^{1-\delta_0}.$$

(4) *B_Θ has small angular derivatives in $\mathcal{A}_{(\frac{2}{3}, \frac{4}{3})}$,*
$$\|\boldsymbol{\nabla}_\Theta B_\Theta\|_{L^2(\mathcal{A}_{(\frac{2}{3}, \frac{4}{3})})} + \|r^{-1} B_\Theta\|_{L^2(\mathcal{A}_{(\frac{2}{3}, \frac{4}{3})})} \lesssim_{E_3} \epsilon^{1-\delta_0}.$$

We remark that the connection B provided by the above lemma has the property that it is small in L^4. This is explained in what follows.

From properties (2)–(4) and Hardy's inequality, it follows that each component B_j in the rectangular coordinates obeys

$$\|r^{-1} B_j\|_{L^2} \lesssim_{E_3} \epsilon^{1-\delta_0}.$$

Thus we may localize B_j via a smooth cutoff outside the annulus $\{\frac{3}{4} \le |x| \le 1\}$, and show that the L^4 norm of this portion is small using property (2). To bound the L^4 norm of the localized remainder, the following variant of the Sobolev (or Bernstein) inequality applies:

Lemma 8.5. *Let u be supported in an annulus $\mathcal{A}_{(r_0, r_1)}$. Then*

$$\|u\|_{L^4} \lesssim_{r_0, r_1} \|u\|_{\dot{H}^1}^{\frac{1}{4}} \left(\|\boldsymbol{\nabla}_\Theta u\|_{L^2} + \|r^{-1} u\|_{L^2} \right)^{\frac{3}{4}}.$$

Proof. In what follows, we suppress the dependence of constants on r_0, r_1. Using a smooth partition of unity in the angular variables, we may assume that u is supported in an angular sector $\Gamma = \{r\Theta \in \mathcal{A}_{(r_0, r_1)} : \Theta \in \kappa\}$, where κ is a spherical cap in $\mathbb{S}^3$. Then we may use a diffeomorphism from κ to a ball $B \subset \mathbb{R}^3$ to map Γ to $[-1, 1] \times B \subset \mathbb{R}^4$.

We are left to prove

$$\|u\|_{L^4} \lesssim \|u\|_{H^1}^{\frac{1}{4}} \|u\|_{L^2 H^1}^{\frac{3}{4}} \tag{8.8}$$

for a function u supported on the cylinder $[-1, 1] \times B \subset \mathbb{R}^4$; here, the mixed norms are defined with respect to x^1 and $x' = x^2, x^3, x^4$. By the Littlewood–Paley inequality, it suffices

to verify this inequality for a single piece $P_k u$. We also introduce the Littlewood–Paley projections P'_j associated with x'. Then by Bernstein's inequality,

$$\|P_k P'_j u\|_{L^4} \lesssim 2^{\frac{1}{4}k} 2^{\frac{3}{4}j} \min\{2^{-j}\|u\|_{L^2 H^1}, 2^{-k}\|u\|_{H^1}\},$$

and the LHS vanishes for $j \geq k + O(1)$. Now summing up in j, the desired bound follows. $\square$

It follows that

$$\|B\|_{L^4} \lesssim_{E_3} \epsilon^{\frac{3}{4}(1-\delta_0)},$$

which in turn shows that in this gauge, B is energy dispersed (i.e., $2^{-k}\|P_k B\|_{L^\infty} \lesssim_{E_3} \epsilon^{\frac{3}{4}(1-\delta_0)}$).

One minor downside of Lemma 8.4 is that the polar coordinates are not so convenient to use near zero and near infinity. However, both near zero and near infinity we have small L^2 curvature, so we may directly apply Uhlenbeck's lemmas; see Theorems A.9 and A.10 in the appendix. Thus, after standard partitioning and regluing operations, the problem reduces to the simpler case when we work in an annulus:

Lemma 8.6. *Let $A \in \dot{H}^1$ be a connection in the annulus $\mathcal{A}_{(1,2)}$ with energy at most E_3, which has the following properties (all norms are implicitly defined on $\mathcal{A}_{(1,2)}$ by restriction):*
  i) Small tangential curvature,

$$\|F_{\Theta\Theta}\|_{L^2} \leq \epsilon.$$

  ii) Small angular covariant divergence of the transversal curvature,

$$\|\mathbf{D}^\Theta F_{\Theta r}\|_{\dot{H}_A^{-1}} \leq \epsilon. \tag{8.9}$$

If $\epsilon \ll_{E_3} 1$ then there is a gauge equivalent connection B in the Coulomb gauge[17] with the following properties:
  a) Bounded size,

$$\|B\|_{\dot{H}^1} \lesssim_{E_3} 1.$$

  b) Small components:

$$\|B_r\|_{\dot{H}^1} + \|\mathbf{\nabla}_\Theta B_\Theta\|_{L^2} + \|r^{-1} B_\Theta\|_{L^2} \lesssim_{E_3} \epsilon^{1-\delta_0}.$$

Remark 8.7. As a corollary of Lemma 8.5 and Properties (a)–(b) in Lemma 8.6, we have

$$\|B_r\|_{L^4} \lesssim_{E_3} \epsilon^{1-\delta_0}, \qquad \|B_\Theta\|_{L^4} \lesssim_{E_3} \epsilon^{\frac{3}{4}(1-\delta_0)}. \tag{8.10}$$

Indeed, the bound for B_r is simply the Sobolev embedding. To apply Lemma 8.5 to B_Θ, we need to find an $\dot{H}^1$-extension $\bar{B}_\Theta$ of B_Θ outside $\mathcal{A}_{(1,2)}$ that is supported in (say) $\mathcal{A}_{(\frac{1}{2}, \frac{5}{2})}$ and

$$\|\partial_r \bar{B}_\Theta\|_{L^2} \lesssim_{E_3} 1, \quad \|\mathbf{\nabla}_\Theta \bar{B}_\Theta\|_{L^2} + \|r^{-1} \bar{B}_\Theta\|_{L^2} \lesssim_{E_3} \epsilon^{1-\delta_0}.$$

For this purpose, we take an even reflection of B_Θ across the boundaries of $\partial \mathcal{A}_{(1,2)}$ (see Lemma A.2) and apply a radial cutoff that equals one on $\mathcal{A}_{(1,2)}$ and supported in $\mathcal{A}_{(\frac{1}{2}, \frac{5}{2})}$.

Conversely, if such a bound holds then the $\dot{H}^1$ type bounds follow by solving linear elliptic systems; see the proof of Lemma 8.6 below.

We now successively prove the above results in reverse order:

---

[17]Here we reserve the right to choose the metric favorably.

Proof of Lemma 8.6. We seek the connection B, which is gauge-equivalent to A, so that it satisfies the Coulomb gauge condition $\boldsymbol{\nabla}^k B_k = 0$ with respect to the metric

$$\mathbf{e} = \mathrm{d}r^2 + \mathrm{d}\Theta^2,$$

and with the boundary condition

$$B_r = 0 \qquad \text{on } \partial\mathcal{A}_{(1,2)}.$$

We claim that such a connection exists, and satisfies the conclusion of the lemma.

We use an elliptic bootstrap (i.e., a continuity) argument. Suppose that we have a continuous one parameter family of connections

$$A^{(h)} \in (\dot{H}^1 \cap L^4)(\mathcal{A}_{(1,2)}), \qquad h \in [0,1], \qquad A^{(0)} = 0, \quad A^{(1)} = A$$

so that the hypotheses of the lemma hold uniformly in $h \in [0,1]$. Then the Coulomb connection $B^{(0)} = 0$ is the obvious solution when $h = 0$, and we seek to extend this property by continuity up to $h = 1$. For this we consider the following bootstrap assumption:

The connection $A^{(h)}$ admits a Coulomb gauge representation $B^{(h)}$ as above, and which satisfies the additional property

$$\|B^{(h)}\|_{H^1} \leq C_0, \quad \|B^{(h)}\|_{L^4} \leq \epsilon^{\frac{3}{4}(1-\delta_0)}. \tag{8.11}$$

We will establish that, if C_0 is large enough and ϵ is sufficiently small (depending on E_3 and chosen in this order), the set

$$H = \{h \in [0,1]: \quad (8.11) \text{ holds}\},$$

which trivially contains 0, is both open and closed, and thus contain $h = 1$.

Fractional Sobolev spaces and elliptic operators on $\mathbb{S}^3$ and $\mathcal{A}_{(1,2)}$. In what follows, we will employ fractional Sobolev spaces on $\mathbb{S}^3$ and $\mathcal{A}_{(1,2)}$, where we will distinguish between tangential and tranversal regularities in the latter case (for more details, see below). We start with the case of the unit sphere $\mathbb{S}^3$. Denote by $\mathcal{X}$ a finite set of smooth vector fields X on $\mathbb{S}^3$ that spans the tangent space at each point (e.g., the set of rotations Ω_{jk} in the $x^j x^k$-plane for $j, k = 1, \ldots, 4$ would do). The L^2-Sobolev space of k-forms on $\mathbb{S}^3$ of order $m \in \mathbb{N}$ is defined by the norm

$$\|\omega\|^2_{H^m(\mathbb{S}^3;\Lambda^k)} = \sum_{m'=0}^{m} \sup_{\substack{X_1,\ldots X_{m'} \in \mathcal{X} \\ X'_1,\ldots,X'_k \in \mathcal{X}}} \|X_1 \cdots X_{m'}(\omega(X'_1, \ldots, X'_k))\|^2_{L^2(\mathbb{S}^3)}.$$

Clearly, any different choice of $\mathcal{X}$ gives rise to an equivalent norm. As usual, these spaces are extended to negative orders by duality, and to fractional orders by complex interpolation.

A basic operator in this setting is the Hodge Laplacian on k-forms, which we denote by Δ_k. It is a second order elliptic operator that is nonpositive on L^2. Thus, for any $\alpha > 0$ and $\gamma \in \mathbb{R}$, $(-\Delta_k + 1)^\alpha : H^{\gamma+\alpha}(\mathbb{S}^3;\Lambda^k) \to H^\gamma(\mathbb{S}^3;\Lambda^k)$ has a well-defined inverse, which we denote by $(-\Delta_k + 1)^{-\alpha}$. Moreover, the first and second de Rham cohomology groups of $\mathbb{S}^3$ are trivial, so Δ_1 and Δ_2 have trivial kernel by the Hodge theorem. Thus, for $k = 1, 2$, the preceding discussion holds with $-\Delta_k + 1$ replaced by $-\Delta_k$.

Next, we consider the domain $\mathcal{A}_{(1,2)}$. For the moment, we view it as an open submanifold $(1,2) \times \mathbb{S}^3$ of the compact manifold $\mathcal{M} = (\mathbb{R}/4\mathbb{Z})_r \times \mathbb{S}^3_\Theta$, equipped with the product metric

$\mathbf{e} = \mathrm{d}r^2 + \mathrm{d}\Theta^2$. We say that a vector field X (resp. a k-form) on $\mathcal{M}$ is *tangential* (to the constant-r spheres) if $\mathrm{d}r(X) = 0$ (resp. $\iota_{\partial_r}\omega = 0$). A general k-form ω on $\mathcal{M}$ may be decomposed into its tangential part, which may be identified with the pullback $\omega_{\Theta\ldots\Theta}$, and its transversal part $\iota_{\partial_r}\omega$, which is a tangential $(k-1)$-form.

Let $\mathcal{X}_{tan}$ be a finite set of smooth tangential vector fields on $\mathcal{A}_{(1,2)}$ that spans the tangent space of $\{r = const\}$ at every point. For $\sigma \in \mathbb{R}$ and $m \in \mathbb{N}$, we define the anisotropic L^2-Sobolev norm of order (σ, m) for a tangential k-form ω by

$$\|\omega\|^2_{H^{\sigma,m}(\mathcal{M};\Lambda^k_{tan})} = \sum_{m'=0}^{m} \sup_{\substack{X_1,\ldots X_{m'} \in \mathcal{X}_{tan} \\ X_1',\ldots,X_k' \in \mathcal{X}_{tan}}} \|X_1 \cdots X_{m'}(\omega(X_1',\ldots,X_k'))\|^2_{H^\sigma(\mathcal{M})}.$$

Again, any other choice of $\mathcal{X}_{tan}$ gives rise to an equivalent norm. This definition is extended to negative m by duality, and to fractional orders by complex interpolation (while keeping σ fixed). Then the norm $H^{\sigma,\gamma}((1,2) \times \mathbb{S}^3, \Lambda^k_{tan})$ is defined by restriction (cf. Section 1.4).

The Hodge Laplacian $\not{\Delta}_k$ acts on a tangential k-form ω by viewing each $\omega(r,\Theta)$ as a k-form on the unit sphere $\mathbb{S}^3_\Theta$. It is not difficult to verify (via induction on $\gamma \in \mathbb{N}$, duality for $\gamma < 0$ and interpolation for $\gamma \in \mathbb{R} \setminus \mathbb{Z}$) that $(-\not{\Delta}_k + 1)^\alpha : H^{\sigma,\gamma+\alpha}(\mathcal{M}, \Lambda^k_{tan}) \to H^{\sigma,\gamma}(\mathcal{M}, \Lambda^k_{tan})$ is invertible for any $\alpha > 0$, $\sigma, \gamma \in \mathbb{R}$; the same property holds on $(1,2) \times \mathbb{S}^3$, too. Moreover, it is clear that $H^{\sigma,\gamma}(\mathcal{M}; \Lambda^k_{tan})$ admits the following useful spectral characterization in terms of the commuting self-adjoint operators $-\mathcal{L}^2_{\partial_r} + 1$ and $-\not{\Delta}_k + 1$ on $L^2(\mathcal{M}; \Lambda^k_{tan})$:

$$\|\omega\|_{H^{\sigma,\gamma}(\mathcal{M};\Lambda^k_{tan})} \sim \|(-\mathcal{L}^2_{\partial_r} - \not{\Delta}_k + 2)^\sigma (-\not{\Delta}_k + 1)^\gamma \omega\|_{L^2(\mathcal{M};\Lambda^k_{tan})}. \tag{8.12}$$

The same conclusions hold with $-\not{\Delta}_k + 1$ replaced by $-\not{\Delta}_k$ if $k = 1, 2$.

Finally, to connect back to the original setting, we note that for each $\sigma \in \mathbb{R}$ and $m \in \mathbb{N}$, we have the equivalence

$$\|\omega\|_{H^{\sigma,m}((1,2)\times\mathbb{S}^3,\Lambda^k_{tan})} \sim \sum_{m'=0}^{m} \sup_{\substack{X_1,\ldots X_{m'} \in \mathcal{X}_{tan} \\ X_1',\ldots,X_k' \in \mathcal{X}_{tan}}} \|X_1 \cdots X_{m'}(\omega(X_1',\ldots,X_k'))\|^2_{H^\sigma(\mathcal{A}_{(1,2)})}$$

where the norm $H^\sigma(\mathcal{A}_{(1,2)})$ on the RHS is defined with respect to the Euclidean metric on $\mathbb{R}^4$. Note also the equivalences $\|\cdot\|_{H^\sigma((1,2)\times\mathbb{S}^3)} = \|\cdot\|_{H^{\sigma,0}((1,2)\times\mathbb{S}^3)} \sim \|\cdot\|_{H^\sigma(\mathcal{A}_{(1,2)})}$ and $\|\cdot\|_{L^p_r L^p_\Theta((1,2)\times\mathbb{S}^3)} \sim \|\cdot\|_{L^p(\mathcal{A}_{(1,2)})}$. Accordingly, in what follows we will refer to these norms simply by H^σ and L^p, respectively, without any possibility of confusion.

A-priori estimates for $B^{(h)}$. We now begin our proof in earnest. First, we establish some a-priori bounds for $B^{(h)}$, which improve the bootstrap assumption (8.11) and also imply the desired bounds stated in the lemma.

In what follows, we suppress h and just write $B = B^{(h)}$. We also abuse the notation a bit and write F for the curvature 2-form associated to B. We omit the dependence of constants on E_3 and write c for a small positive constant that may vary from line to line. We use the roman indices $a, b, \ldots$ for coordinates on $\mathbb{S}^3$ and use the metric $\mathrm{d}\Theta^2$ to raise and lower these indices. We also suppress Λ^k_{tan} in the norms when the degree of the differential form is clear from the context.

Note that $(\iota_{\partial_r}F) = F_{r\Theta}$, viewed as a tangential 1-form, satisfies the following div-curl system on each sphere:

$$\begin{cases} \slashed{\nabla}^a F_{ar} = \slashed{D}^a F_{ar} - [B^a, F_{ar}], \\ \partial_a F_{br} - \partial_b F_{ar} = -\partial_r F_{ab} - [B_r, F_{ab}] - [B_a, F_{rb}] + [B_b, F_{ra}]. \end{cases}$$

Thus

$$\slashed{\Delta}_1 F_{ar} + \partial_a[B^b, F_{br}] + \slashed{\nabla}^b([B_b, F_{ra}] - [B_a, F_{rb}]) = G, \tag{8.13}$$

where

$$G = \partial_a(\slashed{D}^b F_{br} - [B^b, F_{br}]) - \slashed{\nabla}^b(\partial_r F_{ba} + [B_r, F_{ba}]). \tag{8.14}$$

We claim that

$$\|[B^a, F_{ar}]\|_{H^{-1+\delta_0, -\delta_0}((1,2)\times\mathbb{S}^3)} \lesssim_{C_0} \epsilon^c \|F_{\Theta r}\|_{H^{-1+\delta_0, 1-\delta_0}((1,2)\times\mathbb{S}^3)}. \tag{8.15}$$

We defer the proof until later, but remark that this estimate (barely) fails when $\delta_0 = 0$; this failure is the reason why we introduce $\delta_0 > 0$ and the fractional anisotropic Sobolev spaces. Combined with the bound $\|F_{\Theta r}\|_{H^{-1+\delta_0, 1-\delta_0}((1,2)\times\mathbb{S}^3)} \lesssim \|F\|_{L^2(\mathcal{A})} \lesssim 1$, which is obvious from (8.12), it follows that the terms involving B_a on the LHS may be absorbed into the main term. Therefore,

$$\|F_{\Theta r}\|_{H^{-1+\delta_0, 1-\delta_0}((1,2)\times\mathbb{S}^3)} \lesssim \|G\|_{H^{-1+\delta_0, -1-\delta_0}((1,2)\times\mathbb{S}^3)}.$$

On the one hand, by (8.14) and the assumptions, G is $O(\epsilon)$ in $H^{-1,-1}((1,2)\times\mathbb{S}^3)$. On the other hand, by (8.13) and the estimate

$$\|\partial_a[B^b, F_{br}] + \slashed{\nabla}^b([B_b, F_{ra}] - [B_a, F_{rb}]))\|_{H^{0,-2}((1,2)\times\mathbb{S}^3)} \lesssim_{C_0} \epsilon^c \|F\|_{L^2((1,2)\times\mathbb{S}^3)}, \tag{8.16}$$

whose proof we also defer until later, it follows that G is $O(1) + O_{C_0}(\epsilon^c)$ in $H^{0,2}((1,2)\times\mathbb{S}^3)$. By interpolation, which is obvious from (8.12), we then have

$$\|F_{\Theta r}\|_{H^{-1+\delta_0, 1-\delta_0}((1,2)\times\mathbb{S}^3)} \lesssim \epsilon^{1-\delta_0},$$

provided that ϵ is small enough depending on C_0.

Next, we can write an elliptic equation for B_r in $(1,2)\times\mathbb{S}^3$,

$$(\partial_r^2 + \slashed{\Delta}_0)B_r + \slashed{\nabla}^a[B_a, B_r] = \slashed{D}^a F_{ar} - [B^a, F_{ar}],$$

with the Dirichlet boundary condition $B_r = 0$ on $\partial((1,2)\times\mathbb{S}^3)$. Here the RHS has size $O(\epsilon) + O_{C_0}(\epsilon^c\epsilon^{1-\delta_0})$ in $H^{-1}((1,2)\times\mathbb{S}^3)$ because of the hypothesis, (8.15) and the simple embedding (see (8.12))

$$H^{-1+\delta_0, -\delta_0}((1,2)\times\mathbb{S}^3) \hookrightarrow H^{-1}((1,2)\times\mathbb{S}^3).$$

Also the coefficient B_a on the left is small in L^4. Hence the elliptic problem is uniquely solvable, and the solution B_r satisfies

$$\|B_r\|_{H^1} \lesssim \epsilon + C_{C_0}\epsilon^c\epsilon^{1-\delta_0}.$$

Finally, for B_Θ we have the following div-curl system on each sphere:

$$\begin{cases} \slashed{\nabla}^a B_a = -\partial_r B_r, \\ \partial_a B_b - \partial_b B_a = F_{ab} - [B_a, B_b]. \end{cases}$$

The first RHS is $O(\epsilon) + O_{C_0}(\epsilon^c\epsilon^{1-\delta_0})$ in L^2, whereas the second RHS is $O(\epsilon)$. It follows that

$$\|\slashed{\nabla}_\Theta B_\Theta\|_{L^2} + \|B_\Theta\|_{L^2} \lesssim \epsilon + C_{C_0}\epsilon^c\epsilon^{1-\delta_0}.$$

48

On the other hand we can use $F_{r\Theta}$ to bound

$$\|\boldsymbol{\nabla}_r B_\Theta\|_{L^2} \leq \|F_{r\Theta}\|_{L^2} + \epsilon + C_{C_0}\epsilon^c\epsilon^{1-\delta_0} \lesssim 1,$$

if ϵ is sufficiently small depending on C_0. Thus, we have proved that the conclusion of the lemma holds for $h \in H$. Furthermore, we have

$$\|B\|_{H^1} \lesssim 1, \quad \|B\|_{L^4} \lesssim (\epsilon + C_{C_0}\epsilon^c\epsilon^{1-\delta_0})^{\frac{3}{4}}, \tag{8.17}$$

where we used Remark 8.7 for the second estimate. Once we choose C_0 large enough and ϵ sufficiently small, the bootstrap assumption for B is improved.

Proof of (8.15) **and** (8.16). To conclude the proof of the a-priori estimates, it remains to establish (8.15) and (8.16). In both cases, the idea is to reduce the problem to global-in-spacetime estimates via localization and change of variables (cf. proof of Lemma 8.5). The key ingredient in the reduction are the invariance of the anisotropic Sobolev spaces on $(1,2) \times \mathbb{S}^3$ under multiplication by smooth functions and pullback by diffeomorphisms, both of which are straightforward to verify.

First, we extend B_Θ and $F_{\Theta r}$ by an even reflection across the boundaries of $(1,2) \times \mathbb{S}^3$ and apply a smooth cutoff that equals 1 on $(1,2) \times \mathbb{S}^3$ and is supported in $(\frac{1}{2}, \frac{5}{2}) \times \mathbb{S}^3$. Using a partition of unity on $\mathbb{S}^3$, it suffices to consider B_Θ and $F_{\Theta r}$ that are supported in $(\frac{1}{2}, \frac{5}{2}) \times \kappa$, where κ is a spherical cap in $\mathbb{S}^3$. Finally, we use the invariance under pullback by diffeomorphisms to straighten κ to a unit ball B' in $\mathbb{R}^3$, and also the invariance under multiplication by smooth functions to strip away the (variable coefficient) metric $\mathbf{e}$ and the volume form.

As a result, estimates (8.15) and (8.16) are reduced, respectively, to the following estimates for functions u, v on $\mathbb{R}^4$:

$$\|(-\Delta+1)^{-\frac{1-\delta_0}{2}}(uv)\|_{L^2 H^{-\delta_0}} \lesssim \epsilon^c \|(-\Delta+1)^{-\frac{1-\delta_0}{2}}v\|_{L^2 H^{1-\delta_0}}, \tag{8.18}$$

$$\|uv\|_{L^2 H^{-1}} \lesssim \epsilon^c \|v\|_{L^2}, \tag{8.19}$$

where the mixed norms are defined with respect to x^1 and $x' = (x^2, x^3, x^4)$, and u obeys

$$\|u\|_{H^1} \lesssim 1, \quad \|u\|_{L^4} \lesssim \epsilon^{\frac{3}{4}(1-\delta_0)}. \tag{8.20}$$

Before we turn to the proof of (8.18) and (8.19), we first deduce from (8.20)

$$\|u\|_{L^\infty L^3} \lesssim \epsilon^c. \tag{8.21}$$

We introduce the inhomogeneous Littlewood–Paley projections $\{\tilde{P}_j\}_{j \geq 0}$ on $\mathbb{R}^4$ (i.e., $\tilde{P}_0 = P_{\leq 0}$ and $\tilde{P}_j = P_j$ for $j \geq 1$), as well as their analogues $\{\tilde{P}'_j\}_{j \geq 0}$ defined with respect to $x' = (x^2, x^3, x^4)$. In view of the refined Sobolev embedding [3, Theorem 1.43][18] on $\mathbb{R}^3$,

$$\|u\|^3_{L^\infty L^3} \lesssim \|u\|^2_{L^\infty H^{\frac{1}{2}}} \sup_j \|\tilde{P}'_j u\|_{L^\infty},$$

it suffices to show that $\|u\|_{L^\infty H^{\frac{1}{2}}} \lesssim 1$ and $\sup_j \|\tilde{P}'_j u\|_{L^\infty} \lesssim \epsilon^c$. The former assertion follows from (8.20) and the trace theorem. For the latter assertion, we introduce a parameter $m > 1$

---

[18]To be precise, [3, Theorem 1.43] is formulated in terms of homogeneous spaces, but the inhomogeneous version stated here follows immediately.

and estimate

$$2^{-j}\|\tilde{P}_j' u\|_{L^\infty} \lesssim 2^{-j}\|\tilde{P}_j'\tilde{P}_{\leq j+m}u\|_{L^\infty} + 2^{-j}\|\tilde{P}_j'\tilde{P}_{>j+m}u\|_{L^\infty}$$

$$\lesssim \sum_{k\leq j+m} 2^k\|\tilde{P}_k u\|_{L^4} + \sum_{k>j-m} 2^{\frac{1}{2}j}\|\tilde{P}_k u\|_{L^\infty L^2}$$

$$\lesssim 2^m\|u\|_{L^4} + 2^{-\frac{1}{2}m}\|u\|_{H^1}.$$

Then using (8.20) and optimizing the choice of m, the desired estimate follows.

Next, we establish (8.18). We normalize v so that $\|(-\Delta+1)^{-\frac{1-\delta_0}{2}}v\|_{L^2 H^{1-\delta_0}} \leq 1$. We decompose $uv = \sum_{j,k,\ell \geq 0} \tilde{P}_j(\tilde{P}_k u \tilde{P}_\ell v)$ and divide the proof into the following (overlapping) cases:

(1) *Low-High interaction,* $|j-\ell| < 3$, $k < j+5$. We introduce the exponents $2-$ and $6-$ defined by the relations $\frac{1}{2-} = \frac{1}{2} + \frac{\delta_0}{3}$ and $\frac{1}{6-} = \frac{1}{6} + \frac{\delta_0}{3}$. We estimate

$$2^{(-1+\delta_0)j}\|\tilde{P}_j(\tilde{P}_{<j+5}u\tilde{P}_\ell v)\|_{L^2 H^{-\delta_0}} \lesssim 2^{(-1+\delta_0)j}\|\tilde{P}_j(\tilde{P}_{<j+5}u\tilde{P}_\ell v)\|_{L^2 L^{2-}}$$

$$\lesssim 2^{(-1+\delta_0)j}\|\tilde{P}_{<j+5}u\|_{L^\infty L^3}\|\tilde{P}_\ell v\|_{L^2 L^{6-}}$$

$$\lesssim \|u\|_{L^\infty L^3} 2^{(-1+\delta_0)\ell}\|\tilde{P}_\ell v\|_{L^2 H^{1-\delta_0}},$$

which is acceptable thanks to (8.21).

(2) *High-Low interaction,* $|j-k| < 3$, $\ell < j+5$. This case can be handled similarly as in the Low-High interaction case; we even get an additional gain of $2^{(-1+\delta_0)(j-\ell)}$.

(3) *High-High interaction,* $|k-\ell| < 3$, $j < \min\{k,\ell\} - 3$. Let $m > 1$ be a parameter to be fixed later, and let $(\frac{3}{2}-)^{-1} = \frac{2}{3} + \frac{\delta_0}{3}$. For $\ell > j+m$, we estimate

$$2^{(-1+\delta_0)j}\|\tilde{P}_j(\tilde{P}_k u\tilde{P}_\ell v)\|_{L^2 H^{-\delta_0}} \lesssim 2^{(-1+\delta_0)j}\|\tilde{P}_j(\tilde{P}_k u\tilde{P}_\ell v)\|_{L^2 L^{2-}}$$

$$\lesssim 2^{\delta_0 j}\|\tilde{P}_j(\tilde{P}_k u\tilde{P}_\ell v)\|_{L^1 L^{\frac{3}{2}-}}$$

$$\lesssim 2^{-\delta_0(\ell-j)}2^k\|\tilde{P}_k u\|_{L^2} 2^{(-1+\delta_0)\ell}\|\tilde{P}_\ell v\|_{L^2 L^{6-}}.$$

After summation, the contribution of these terms is $O(2^{-\delta_0 m})$. On the other hand, for $\ell \leq j+m$, we estimate

$$2^{(-1+\delta_0)j}\|\tilde{P}_j(\tilde{P}_k u\tilde{P}_\ell v)\|_{L^2 H^{-\delta_0}} \lesssim 2^{(1-\delta_0)(\ell-j)}\|\tilde{P}_k u\|_{L^\infty L^3} 2^{(-1+\delta_0)\ell}\|\tilde{P}_\ell v\|_{L^2 L^{6-}},$$

which contributes $O(2^{(1-\delta_0)m}\|u\|_{L^\infty L^3})$ after summation. Using (8.21) and optimizing the choice of m, the desired estimate follows.

Finally, we prove (8.19). By the Sobolev embeddings on $\mathbb{R}^3$, we have

$$\|uv\|_{L^2 H^{-1}} \lesssim \|uv\|_{L^2 L^{\frac{6}{5}}} \lesssim \|u\|_{L^\infty L^3}\|v\|_{L^2},$$

which implies (8.19) in view of (8.21).

Completion of the continuity argument. Next, we consider a perturbative problem, and prove that if B is Coulomb, $\dot{H}^1$ and small in L^4 as above, then all connections $\tilde{A}$ which are sufficiently close to A in $\dot{H}^1$ admit a similar Coulomb representation.

Abusing the notation a bit, we write A instead of B and redefine $\tilde{A}$ by applying the same gauge transformation that takes A to B. Hence, $\partial^k \tilde{A}_k$ is small. Applying a further gauge transformation (see Lemma A.1), we may assume that $\tilde{A}_r = 0$ on $\partial\mathcal{A}_{(1,2)}$ as well. Then to

find a gauge transformation O which takes $\tilde{A}$ into the Coulomb gauge, we end up having to solve for $\Omega_k = O_{;k}$ the system[19]

$$\begin{cases} \partial^k \Omega_k = \partial^k (Ad(O)\tilde{A}_k) = Ad(O)\partial^k \tilde{A}_k + [O_{;k}, Ad(O)\tilde{A}^k], \\ \partial_j \Omega_k - \partial_k \Omega_j = -[\Omega_j, \Omega_k], \end{cases} \tag{8.22}$$

with the boundary condition $\Omega_r = 0$ on $\partial \mathcal{A}_{(1,2)}$. To solve this system, we start with $O^{(0)} = Id$ and construct $\Omega^{(n)}$ by applying Proposition A.8 with $B = Ad(O^{(n-1)})A$. Then $O^{(n)}$ is constructed by integrating the system of ODEs $O^{(n)}_{;j} = \partial_j O^{(n)}(O^{(n)})^{-1} = \Omega^{(n)}$, which is possible thanks to the curl condition for $\Omega^{(n)}$. By smallness of $\|\partial^k \tilde{A}_k\|_{L^2}$ and $\|\tilde{A}\|_{L^4}$, this iteration procedures goes through and we obtain a uniform bound $\|O^{(n)}_{;x}\|_{H^1} \lesssim \|\partial^k \tilde{A}_k\|_{L^2}$. Taking the limit (along a suitable subsequence), we obtain a desired gauge transformation O that also satisfies $\|O_{;x}\|_{H^1} \ll 1$.

The a-priori bound shows that if $h \in H$ then the stronger bound (8.17) holds. Then the perturbative argument shows that for $h \in H$ there exists a fixed size neighborhood $[h - c, h + c]$ which is in H. We conclude that $H = [0, 1]$, which completes the continuity argument.

Existence of a continuous path $A^{(h)}$. The remaining issue is that of constructing a continuous path from A to 0. In effect it suffices to show that there exists an extension of A inside the full unit ball which still satisfies the assumptions of the lemma and so that A vanishes near $x = 0$. Then we can obtain the desired family by scaling[20],

$$A^{(h)}(x) = hA(hx), \qquad h \in [0, 1].$$

This can be done as follows:

(1) In a suitable gauge set $A_r = 0$ on the boundary; see Lemma A.1.
(2) Double the annulus inside, and extend the connection as odd for A_r and even for A_θ. This extension is still H^1, and the smallness hypothesis still holds in the double annulus; see Lemma A.2.
(3) Choose a sphere S within the extended part on which $F_{\Theta\Theta}$ is small in L^2. Using Uhlenbeck's lemma on the 3-sphere (Proposition A.11), we may set $\|A_\Theta\|_{H^1(S)}$ small in a suitable gauge. In addition, again using Lemma A.1, we may set $A_r = 0$ on S.
(4) Choose an extension of A inside S which is small in H^1. More precisely, since the trace of A_r vanishes on S, it follows that the extension of A_r by zero inside of S is in H^1. Similar considerations apply to A_Θ after subtracting an extension of the boundary values, which can be made to have a small H^1 norm. Overall, the H^1 norm of the extension is small inside S, so that the assumptions of the lemma are kept. Finally, by smallness we may harmlessly cutoff A_Θ near 0 as well, as desired. $\qquad\square$

---

[19]An alternative idea would have been to work with $O^{-1}\partial_k O$ as in Section 5, which has the advantage that no O appears in the div-curl system; see (5.5). However, for the boundary value problem on the annulus, the cokernel (and also the kernel) of the associated Neumann problem is nontrivial. The system (8.22) has the virtue of having a cokernel independent of O, while it depends on O for (5.5).

[20]As in the original proof of Uhlenbeck's lemma in [55].

Lemma 8.6 $\implies$ Lemma 8.4. In accordance with the choice of metric in Lemma 8.6, we endow $\mathbb{R}^4$ with a smooth Riemannian metric that coincides with $\mathrm{d}r^2 + \mathrm{d}\Theta^2$ in $\mathcal{A}_{(1,2)}$ and with the Euclidean metric $\mathrm{d}r^2 + r^2\mathrm{d}\Theta^2$ outside $\mathcal{A}_{(\frac{1}{2},\frac{5}{2})}$. We formulate the Coulomb gauge conditions in the proof with respect to this metric. As already noted in [55], Uhlenbeck's lemmas work just as well on Riemannian manifolds if we take ϵ_0 small enough; so does Lemma A.3, which is an interior elliptic regularity result .

By Uhlenbeck's lemma (Theorem A.9) we obtain a gauge-equivalent connection A_{in} in $B_{\frac{3}{4}}$ which is ϵ- small in H^1. Next, by Uhlenbeck's lemma in the exterior of a ball (Theorem A.10, we obtain a gauge-equivalent connection A_{out} in $\mathbb{R}^4 \setminus B_1$, which is ϵ- small in $\dot{H}^1 \cap L^4(\mathbb{R}^4 \setminus B_1)$. Now note that (8.5) is equivalent to (8.9), since

$$\mathbf{D}^{\mathfrak{a}} F_{\mathfrak{a}r} = r^{-2} \, \slashed{\mathbf{D}}^{\Theta} F_{\Theta r}.$$

Thus by Lemma 8.6 we obtain a gauge-equivalent connection A_{mid} in the annulus $\mathcal{A}_{(\frac{2}{3},\frac{4}{3})}$. The L^4 smallness allows us to patch the three connections cleanly (without any topological obstructions). More precisely, the Coulomb gauge conditions imply that the transition maps O in the intersections obey a favorable div-curl system. The L^4 bounds on A_{in}, A_{mid} and A_{out} imply that $O_{;x}$ (defined in each intersection) is small in L^4. Then, by the div-curl system we may upgrade this bound to smallness in $\dot{H}^1$, and then via Lemma A.3 to smallness in $\ell^1 \dot{H}^1$ (where we shrink the domain at each step). Thus each O is uniformly closed to a constant (Lemma A.4), and a standard patching argument (see, e.g., [55, Proposition 3.2]) now works. $\qquad\square$

Lemma 8.4 $\implies$ Proposition 8.3. We start with a continuity argument. Using the equivalent connection A given by the lemma, we produce a continuous family of connections $A^h = hA$ with $h \in [0,1]$ so that $A^0 = 0$ and $A^1 = A$, and which satisfies uniformly the hypotheses of the proposition.

We consider the subset H of $h \in [0,1]$ for which the following property holds:

The Yang–Mills heat flow of A is global and satisfies the bound

$$\|F\|_{L^3_{x,s}} \leq 1.$$

Clearly $0 \in H$. Also by the continuity properties of the Yang–Mills heat flow, H is closed. It remains to show that H is open, which would imply that $H = [0,1]$. For this it suffices to take the above bound as a bootstrap assumption, and show that we can improve it.

Under this assumption, it follows immediately from Proposition 2.6 that we have a gauge transformation O with

$$\|O_{;x}\|_{\dot{H}^1} \lesssim_{E_3} 1,$$

which transforms A into its caloric representation $\tilde{A}$. In turn $\tilde{A}$ must also satisfy

$$\|\tilde{A}\|_{\dot{H}^1} \lesssim_{E_3} 1.$$

Further, since A was small in L^4, it curvature is small in $\dot{W}^{-1,4}$,

$$\|F\|_{\dot{W}^{-1,4}} \lesssim_{E_3} \epsilon^{\frac{3}{4}(1-\delta_0)}.$$

By this and the bound for $O_{;x}$, the curvature of $\tilde{A}$, namely $\tilde{F} = OFO^{-1}$, must also be small,

$$\|\tilde{F}\|_{\dot{W}^{-1,4}} \lesssim_{E_3} \epsilon^{\frac{3}{4}(1-\delta_0)}.$$

Propagating this bound along the caloric flow [37, Proposition 8.9; Eq. (8.44)], we obtain

$$\|P_k\tilde{F}(s)\|_{\dot{W}^{-1,4}} \lesssim_{E_3} \epsilon^{\frac{3}{4}(1-\delta_0)}(1+2^{2k}s)^{-100},$$

On the other hand, by the bootstrap assumption and [37, Proposition 7.13; Eq. (7.20)], we also have

$$\|P_k\tilde{F}(s)\|_{L^2} \lesssim_{E_3} c_k(1+2^{2k}s)^{-100},$$

where $\|c_k\|_{\ell^2} \lesssim_{E_3} 1$. By Bernstein (for the second bound) and interpolation, we have

$$\|P_k\tilde{F}(s)\|_{L^3} \lesssim_{E_3} c_k^{\frac{2}{3}}\epsilon^{\frac{1}{4}(1-\delta_0)}2^{\frac{2}{3}k}(1+2^{2k}s)^{-100}.$$

Then by Schur's test, we obtain

$$\|\tilde{F}\|_{L^3_{s,x}} \lesssim_{E_3} \epsilon^{\frac{1}{4}(1-\delta_0)}\|c_k^{\frac{2}{3}}\|_{\ell^3} \lesssim_{E_3} \epsilon^{\frac{1}{4}(1-\delta_0)},$$

which improves the bootstrap assumption. Moreover, by Bernstein and [37, Proposition 8.9; Eq. (8.45)], it follows that $\|\tilde{A}\|_{\ell^\infty \dot{W}^{-1,\infty}} \lesssim \|\tilde{A}\|_{\ell^\infty L^4} \lesssim \epsilon^{\frac{3}{4}(1-\delta_0)}$. Then by the bound $\|\tilde{A}\|_{\dot{H}^1} \lesssim_{E_3}$ 1 and the improved Sobolev inequality [3, Theorem 1.43], (8.7) follows. This completes the proof of the proposition. $\qquad\square$

Proposition 8.3 $\implies$ *Theorem 8.1.* We express the curvature components in the null frame. By (8.1) and (8.2), as well as (3.16) and (3.17) for the expression of $^{(X_\varepsilon)}P_0(A) = \frac{1}{2}(^{(X_\varepsilon)}P_L(A) + {}^{(X_\varepsilon)}P_{\underline{L}}(A))$, the null components α, ϱ and σ are already small in L^2 provided that ϵ_1, ε and $1 - \gamma$ are sufficiently small. We now use the constraint equation to express

$$\mathbf{D}^{\mathfrak{a}}F_{\mathfrak{a}r} = \mathbf{D}^{\mathfrak{a}}(F_{\mathfrak{a}0} + F_{\mathfrak{a}L}) = -r^{-3}\mathbf{D}_r(r^3 F_{r0}) + \mathbf{D}^{\mathfrak{a}}\alpha_{\mathfrak{a}} = -r^{-3}\mathbf{D}_r(r^3\varrho) + \mathbf{D}^{\mathfrak{a}}\alpha_{\mathfrak{a}}, \qquad (8.23)$$

which implies the desired smallness of $\mathbf{D}^{\mathfrak{a}}F_{\mathfrak{a}r}$ in the gauge-invariant space H_A^{-1}. Thus, we have established that the spatial part of the connection A_x satisfies the hypotheses of Proposition 8.3.

Suppose now that A_x is in the caloric gauge and satisfies the bounds in Proposition 8.3. It remains to consider the temporal components of F. For $F_{0\mathfrak{a}}$ we write

$$F_{0\mathfrak{a}} = F_{L\mathfrak{a}} + F_{r\mathfrak{a}} = \alpha_{\mathfrak{a}} + F_{r\mathfrak{a}},$$

and the $\dot{W}^{-1,4}$ smallness follows. For F_{0r} we simply have

$$F_{0r} = \frac{1}{2}F_{L\underline{L}} = \varrho,$$

which is small even in L^2. $\qquad\square$

53

9. Proof of the Threshold Theorem and the Dichotomy Theorem

In this section, we finally prove the Threshold and Dichotomy Theorems (i.e., Theorems 1.8 and 1.11, respectively).

For both theorems, we argue by contradiction. Suppose that the conclusion of the Dichotomy Theorem (Theorem 1.11) is false, i.e., there exists a solution A for which both alternatives a) and b) are false. Then we are in one of the following two scenarios:

(i) The solution blows up in finite time, and the hypothesis of Theorem 1.7 is false near the tip of the cone C.

(ii) The solution is global but the hypothesis of Theorem 1.7 is false near the infinite end of the cone C.

On the other hand, assume now that the conclusion of the Threshold Theorem (Theorem 1.8) is false. We seek to show that the conclusion of Theorem 1.7 is false, and therefore we are again in one of the two scenarios above. To achieve this, we need to use the energy assumption $\mathcal{E}(A) < 2E_{\mathrm{GS}}$ along with vanishing of the characteristic number $\chi = 0$ (as a consequence of topological triviality $A \in \dot{H}^1$). Our argument is similar to [31] (see also [37, Section 6.2]).

If the conclusion of Theorem 1.7 were true, this would imply that a sequence of translated, rescaled and gauge transformed copies $A^{(n)}$ of A converges (modulo gauge transformations) in H^1_{loc} to a Lorentz transform of a nontrivial harmonic Yang–Mills connection $L_v Q$. This implies (spacetime) L^2_{loc} convergence of curvature tensors $F^{(n)}$, and thus for almost every $t \in (-1/2, 1/2)$ (and possibly passing to a subsequence)

$$\mathcal{E}_{\{t\} \times B_R}(A^{(n)}) = \frac{1}{2} \int_{B_R} \langle F^{(n)}, F^{(n)} \rangle(t) \to \mathcal{E}_{\{t\} \times B_R}(L_v Q) \qquad \text{for any } R > 0,$$

which in turn implies

$$\mathcal{E}(Q) \leq \mathcal{E}(A) < 2E_{\mathrm{GS}}.$$

By Theorem 1.6, the only possibility for Q is that $|\chi(Q)| = \mathcal{E}_e(Q)$. Moreover, since Lorentz transform preserves the topological class, we have $\chi(L_v(Q)) = \chi(Q)$. From here on, we assume that $\chi(Q) > 0$; the alternative case is similar.

Fix a large number $R \gg 1$ and $t \in (-1/2, 1/2)$. By topological triviality of $A^{(n)}(t)$, we have

$$0 = \chi(A^{(n)}(t)) = \int_{B_R} -\langle F^{(n)} \wedge F^{(n)} \rangle(t) + \int_{\mathbb{R}^4 \backslash B_R} -\langle F^{(n)} \wedge F^{(n)} \rangle(t).$$

Again by the (spacetime) L^2_{loc} convergence of $F^{(n)}$, there exists a subsequence such that

$$\int_{\mathbb{R}^4 \backslash B_R} \langle F^{(n)} \wedge F^{(n)} \rangle(t) = \int_{B_R} -\langle F^{(n)} \wedge F^{(n)} \rangle(t) \to \int_{B_R} -\langle F[L_v Q] \wedge F[L_v Q] \rangle.$$

By (1.5), we have

$$\mathcal{E}(A) \geq \limsup_{n \to \infty} \left(\frac{1}{2} \int_{B_R} \langle F^{(n)}, F^{(n)} \rangle(t) + | \int_{\mathbb{R}^4 \backslash B_R} \langle F^{(n)} \wedge F^{(n)} \rangle(t)| \right)$$

$$\geq \mathcal{E}_{\{t\} \times B_R}(L_v Q) + | \int_{B_R} -\langle F[L_v Q] \wedge F[L_v Q] \rangle|.$$

Sending $R \to \infty$, the RHS tends to $\mathcal{E}(L_v Q) + \chi(L_v Q) \geq 2\mathcal{E}(Q) \geq 2E_{GS}$, which is a contradiction.

It follows that the conclusion of Theorem 1.7 is false, and thus its hypothesis is false. Hence we have reduced the problem again to the above alternative (i)–(ii). From here on, the proofs of the two theorems are identical. The analysis is largely similar in the two cases (i) and (ii), but there are still some differences so we consider them separately.

Remark 9.1. A difference in the properties of the solutions is that in the subthreshold case we can work globally in the caloric gauge, whereas otherwise we need the local solutions given by Theorem 2.10. However this makes no essential differences in the proofs below.

(i) The blow-up scenario. Let $E = \mathcal{E}(A)$. If $[0, T)$ is a maximal existence time, then the temporal gauge local well-posedness result (Theorem 2.18) implies that there exists a point $X \in \mathbb{R}^4$ so that the energy does not decay to zero in the backward cone of (X, T). By translation invariance we will set $(X, T) = (0, 0)$ and, reversing time, denote its forward cone by C. Thus we now have a Yang–Mills connection A with the property that

$$\lim_{t \searrow 0} \mathcal{E}_{S_t}(A) > \epsilon_0, \tag{9.1}$$

where ϵ_0 is a universal positive constant corresponding to the small data result.

We also know that the hypothesis of Theorem 1.7 is false, which gives

$$\lim_{t \searrow 0} \mathcal{E}_{S_t^\gamma}(A) = 0, \qquad 0 < \gamma < 1. \tag{9.2}$$

We would like to use these two properties in order to show that the connection A admits a caloric representation near the tip of the cone, which is also energy dispersed. Then we could directly apply the energy dispersed result in Theorem 2.12 to conclude that the solution can be extended beyond the blow-up time $T = 0$, which is a contradiction. However, this strategy cannot work unless the energy of A is very small also outside the cone, which is not at all guaranteed a-priori. To resolve this difficulty, we first truncate the solution outside the cone in order to insure that the outer energy stays small:

Lemma 9.2. *For each $\varepsilon > 0$ there exists a $t_\varepsilon > 0$ and a finite energy Yang–Mills solution $\tilde{A}$ in $(0, t_\varepsilon]$ with the following properties:*
  (1) Gauge equivalence: $\tilde{A}$ is gauge equivalent to A in $C_{(0,t_\varepsilon]}$.
  (2) Small energy outside the cone

$$\mathcal{E}_{(\{t\} \times \mathbb{R}^4) \setminus S_t}(\tilde{A}) \leq \varepsilon^8 E \quad \text{for every } t \in (0, t_\varepsilon], \tag{9.3}$$

  (3) Small flux on ∂C

$$\mathcal{F}_{(0,t_\varepsilon]}(\tilde{A}) \leq \varepsilon^9 E. \tag{9.4}$$

Proof. The flux energy relation (3.8) implies that the flux decays to zero near the tip of the cone,

$$\lim_{t \to 0} \mathcal{F}_{(0,t]}(A) = 0.$$

so we first choose t_ε small enough so that the last condition is satisfied for A. We then choose $\delta > 0$ so that the energy of A in a δ-annulus around the cone is small,

$$\mathcal{E}_{\{t_\varepsilon\} \times \{t_\varepsilon < |x| < t_\varepsilon + 3\delta\}}(A) \leq \varepsilon^9 E.$$

55

Again by the flux energy relation, this propagates to all smaller times,

$$\mathcal{E}_{\{t\}\times\{t<|x|<t+3\delta\}}(A) \leq 2\varepsilon^9 E, \qquad 0 < t < t_\varepsilon.$$

We now reset t_ε to a smaller value,

$$t_\varepsilon \to \min\{t_\varepsilon, \delta\}.$$

For this new choice we have

$$\mathcal{E}_{\{t_\varepsilon\}\times\{t_\varepsilon<|x|<4t_\varepsilon\}}(A) \leq 2\varepsilon^9 E.$$

By (a rescaled form of) Proposition 2.13, we can truncate the data $(a, e)(t_\varepsilon)$ for A at time t_ε. We obtain a new data set $(\tilde{a}, \tilde{e})(t_\varepsilon)$ which agrees with $(a, e)(t_\varepsilon)$ inside the cone, but is small outside,

$$\mathcal{E}_{\{t_\varepsilon\}\times\{t_\varepsilon<|x|<4t_\varepsilon\}}(\tilde{A}) \lesssim \varepsilon^9 E.$$

Now we consider the solution $\tilde{A}$ generated by the truncated data $(\tilde{a}, \tilde{e})(t_\varepsilon)$ below time t_ε. For as long as it exists it is gauge equivalent with A inside the cone, since they are gauge equivalent initially, see Theorem 2.18. This shows that it cannot blow-up inside the cone. On the other hand outside the cone it satisfies the second condition in the lemma by the flux-energy relation, so it does not have enough energy blow up there either. It follows that $\tilde{A}$ persists as a finite energy solution and satisfies the conditions in the lemma up to time $t = 0$. $\qquad\square$

From here on, we work with the connection $\tilde{A}$ which satisfies the properties (9.1), (9.2) and (9.3) where for the latter we choose

$$\varepsilon \ll_E 1.$$

Since the flux decays to zero at the tip of the cone, Proposition 3.3 applied in the interval $[\varepsilon t, t]$ also implies that for small enough t we have

$$\int_{S_t} {}^{(X_{\varepsilon t})} P_0(\tilde{A}) \, \mathrm{d}x \lesssim E. \tag{9.5}$$

The bounds (9.1), (9.2) and (9.3) together with (9.5) with small enough ε guarantee that the hypothesis of Theorem 8.1 is satisfied for $\tilde{A}$ all small enough t. This shows that the connection $\tilde{A}$ admits a caloric representation in $[0, t]$, which is also energy dispersed. Thus we can apply the energy dispersed result in Theorem 2.12 to conclude that the solution can be extended beyond the blow-up time $T = 0$, which is a contradiction.

(ii) The non-scattering scenario. This is similar but simpler, as we no longer need to truncate the data for A and instead we may work directly with A. For this we choose R large enough so that the outer energy of A is small,

$$\mathcal{E}_{\{0\}\times\{|x|>R\}}(A) \leq \varepsilon^9 E,$$

and then work with the translated connection $A(t - R, x)$. This satisfies the conditions (2),(3) in Lemma 9.2 for $t \in [R, \infty)$. From this point on, the bound (9.5) must hold for all large enough t. Hence the hypothesis of Theorem 8.1 is satisfied for $\tilde{A}$ all large enough t. This shows that the connection A admits a caloric representation in $[t, \infty]$, which is also energy dispersed (in the case of Theorem 1.8, we have smallness of the energy dispersion in the original gauge, thanks to the uniqueness of the caloric gauge). Thus we can apply the

energy dispersed result in Theorem 2.12 in $[t, \infty)$ and conclude that $A \in S^1([t, \infty))$, i.e., the desired scattering result.

APPENDIX A. TOOLS FOR ANALYSIS OF GAUGE TRANSFORMATIONS

In this appendix, we collect various technical results, mostly concerning gauge transformations, that are used in the main text.

A.1. **Results from** [39]. We recall some useful results that were proved in [39]. The first result is essentially an extension to the critical regularity of a well-known result (see, for instance, [55, Lemma 2.6]).

Lemma A.1 ([39, Lemma 3.5]). *Let B be a $\mathfrak{g}$-valued function in $H^{\frac{1}{2}}(\mathbb{S}^3)$. There exists $O \in L^\infty \cap H^2(B_1)$, which depends continuously on B, such that*

$$(O, O_{;r}) \upharpoonright_{\{r=1\}} = (Id, B),$$

where $O_{;r} = \frac{x^j}{|x|} O_{;j}$. A similar construction can be done in the exterior region $\mathbb{R}^4 \setminus B_1$.

The next result works in tandem with Lemma A.1, and provides an simple way to extend a H^1 connection 1-form through a sphere.

Lemma A.2 ([39, Lemma 3.18]). *Let $A \in H^1(B_R)$ with $A_r = 0$ on ∂B_R. Extend A outside B_R by*

$$\bar{A}_r\left(\frac{R^2}{r}, \Theta\right) = -A_r(r, \Theta), \quad \bar{A}_\Theta\left(\frac{R^2}{r}, \Theta\right) = A_\Theta(r, \Theta).$$

Then the curvature 2-form $\bar{F}$ of the extension obeys

$$\bar{F}\left(\frac{R^2}{r}, \Theta\right) = F(r, \Theta) \quad \text{for } r < R.$$

A similar construction can be done starting from the exterior region $\mathbb{R}^4 \setminus B_R$.

The following results are useful tools for proving continuity of a gauge transformation in the critical regularity setting.

Lemma A.3 ([39, Lemma 3.16]). *Let $O_{;j} \in \dot{W}^{1,\frac{d}{2}}(B_{2R})$ be a solution to the div-curl system*

$$\begin{aligned}\partial_j O_{;k} - \partial_k O_{;j} &= [O_{;j}, O_{;k}] \\ \partial^\ell O_{;\ell} &= H.\end{aligned} \tag{A.1}$$

If $H \in \ell^1 L^{\frac{d}{2}}(B_{2R})$, then $O_{;x} \in \ell^1 \dot{W}^{1,\frac{d}{2}}(B_R)$ with the bound

$$\|O_{;x}\|_{\ell^1 \dot{W}^{1,\frac{d}{2}}(B_R)} \lesssim \|H\|_{\ell^1 L^{\frac{d}{2}}(B_{2R})} + \|O_{;x}\|^2_{\dot{W}^{1,\frac{d}{2}}(B_{2R})}.$$

We note that, even though the divergence $\partial^\ell O_{;\ell}$ is formulated in terms of the Euclidean metric, the lemma works with $\mathbf{e}^{k\ell} \boldsymbol{\nabla}_k O_{;\ell}$ defined with respect to any smooth metric on B_R with suitable adjustment of the constants.

Lemma A.4 ([39, Lemma 3.17]). *If $O_{;x} \in \ell^1 W^{1,\frac{d}{2}}(B)$, then O is continuous on B.*

A.2. **Solvability of div-curl systems.** Our aim in this subsection is to provide solvability results for the nonlinear div-curl system

$$
\begin{cases}
\boldsymbol{\nabla}^\ell \Omega_\ell = \boldsymbol{\nabla}^\ell B_\ell \\
\boldsymbol{\nabla}_j \Omega_k - \boldsymbol{\nabla}_k \Omega_j = -[\Omega_j, \Omega_k]
\end{cases}
\quad \text{in } \mathcal{O}, \tag{A.2}
$$

either with $\mathcal{O} = \mathbb{R}^5$ but with respect to a variable metric $\mathbf{e}$, or with $\mathcal{O} = \mathcal{A}_{(R',R)} \subset \mathbb{R}^4$ (with respect to the Euclidean metric) and with suitable boundary conditions on $\partial\mathcal{A}_{(R',R)}$; these problems arise in Sections 5 and 8, respectively. To begin with, we solve in each case the easier linear system (see Section 1.4 for the notation)

$$
\begin{cases}
\delta\omega = f \\
\mathrm{d}\omega = g
\end{cases}
\quad \text{in } \mathcal{O}, \tag{A.3}
$$

where we remind the reader that $\delta\omega = -\boldsymbol{\nabla}^\ell \omega_\ell$ for a 1-form ω.

Lemma A.5 (Linear div-curl system in $\mathbb{R}^5$). *Consider the div-curl system* (A.3) *in* $(\mathbb{R}^5, \mathbf{e})$, *where $\mathbf{e}_{\alpha\beta}$ is a smooth metric such that $\|\mathbf{e}_{\alpha\beta} - \bar{\mathbf{e}}_{\alpha\beta}\|_{L^\infty \cap \dot{H}^{\frac{5}{2}}} < \epsilon_{\mathbf{e}}$ for some constant positive definite matrix $\bar{\mathbf{e}}_{\alpha\beta}$. Suppose that f, g are in $L^2 \dot{H}^{\frac{1}{2}}$ and obey the compatibility condition*

$$
\mathrm{d}g = 0.
$$

Then for $\epsilon_{\mathbf{e}}$ sufficiently small, there exists a unique solution ω to this problem such that $\omega \in L^5$ and $\nabla\omega \in L^2 \dot{H}^{\frac{1}{2}}$, which obeys

$$
\|\nabla\omega\|_{L^2 \dot{H}^{\frac{1}{2}}} \lesssim \|f\|_{L^2 \dot{H}^{\frac{1}{2}}} + \|g\|_{L^2 \dot{H}^{\frac{1}{2}}}.
$$

Note that if $\omega \in L^5$ and $\nabla\omega \in L^2 \dot{H}^{\frac{1}{2}}$, then

$$
\|\omega\|_{L^5} \lesssim \|\omega\|_{L^2 \dot{H}^{\frac{3}{2}}} + \|\omega\|_{L^\infty \dot{H}^1} \lesssim \|\nabla\omega\|_{L^2 \dot{H}^{\frac{1}{2}}}. \tag{A.4}
$$

Indeed, the first inequality is simply the 4-dimensional Sobolev inequality and interpolation, and the second inequality follows from applying the trace theorem to each $P_k\omega$ and square summing in k.

Proof. First, we treat the case $\mathbf{e}_{\alpha\beta} = \bar{\mathbf{e}}_{\alpha\beta}$, in which case (A.3) has constant coefficients. Let

$$
\tilde{f} = (-\Delta)^{-1} f, \quad \tilde{g}_{\alpha\beta} = (-\Delta)^{-1} g_{\alpha\beta},
$$

and define

$$
\omega = \mathrm{d}\tilde{f} + \delta\tilde{g}.
$$

By the compatibility condition $\mathrm{d}g = 0$, as well as the relations $\mathrm{d}^2 = 0$ and $\delta^2 = 0$, it is easy to check that ω solves (A.3). Any other solution differs from ω by a harmonic 1-form in L^5, which must be zero; thus the uniqueness assertion follows. The desired estimate is then clear by Fourier transform.

In the general case, we use a simple perturbation argument. Note that (A.3) may be written as

$$
\begin{cases}
\delta\omega = f + \mathrm{Err}_{\mathbf{e}}\omega, \\
\mathrm{d}\omega = g,
\end{cases}
$$

where δ denotes the constant coefficient divergence with respect to $\bar{\mathbf{e}}$ and

$$
\mathrm{Err}_{\mathbf{e}}\omega = (\bar{\mathbf{e}}^{\alpha\beta} - \mathbf{e}^{\alpha\beta})\partial_\alpha\omega_\beta - \frac{1}{\sqrt{\det\mathbf{e}}}\partial_\alpha(\mathbf{e}^{\alpha\beta}\sqrt{\det\mathbf{e}})\omega_\beta.
$$

By the standard Moser estimates in $L^\infty \cap \dot{H}^{\frac{5}{2}}$, which is an algebra, it is straightforward to establish (for $\epsilon_{\mathbf{e}} < 1$)

$$\|\bar{\mathbf{e}}^{\alpha\beta} - \mathbf{e}^{\alpha\beta}\|_{L^\infty \cap \dot{H}^{\frac{5}{2}}} \lesssim \epsilon_{\mathbf{e}}, \quad \|\frac{1}{\sqrt{\det \mathbf{e}}} \partial_\alpha (\mathbf{e}^{\alpha\beta} \sqrt{\det \mathbf{e}})\|_{\dot{H}^{\frac{3}{2}}} \lesssim \epsilon_{\mathbf{e}}.$$

Combined with the embeddings $\dot{H}^{\frac{3}{2}} \cdot L^2 \dot{H}^{\frac{3}{2}} \hookrightarrow L^2 \dot{H}^{\frac{1}{2}}$ and $(L^\infty \cap \dot{H}^{\frac{5}{2}}) \cdot L^2 \dot{H}^{\frac{1}{2}} \hookrightarrow L^2 \dot{H}^{\frac{1}{2}}$, both of which follow from the standard Littlewood–Paley trichotomy with respect to $x^1, \ldots, x^4$, we have

$$\|\mathrm{Err}_{\mathbf{e}}\omega\|_{L^2 \dot{H}^{\frac{1}{2}}} \lesssim \epsilon_{\mathbf{e}} \|\nabla\omega\|_{L^2 \dot{H}^{\frac{1}{2}}}.$$

Thus, for $\epsilon_{\mathbf{e}}$ sufficiently small, (A.3) is solvable with $\nabla\omega \in L^2 \dot{H}^{\frac{1}{2}}$, as desired. $\qquad\square$

Next we consider the case $\mathcal{A}_{(R', R)}$. We denote by ν the outward unit normal of $\mathcal{A}_{(R', R)}$ on $\partial\mathcal{A}_{(R', R)}$.

Lemma A.6 (Linear div-curl system in a 4-dimensional annulus). *Consider the div-curl system (A.3) in $\mathcal{A}_{(R', R)} \subset \mathbb{R}^4$ with the boundary condition*

$$\iota_\nu \omega = 0 \qquad on\ \partial\mathcal{A}_{(R', R)}.$$

Suppose that f, g are in $L^2(\mathcal{A}_{(R', R)})$ and obey the compatibility conditions

$$\int_{\mathcal{A}_{(R', R)}} f = 0, \quad \mathrm{d}g = 0.$$

Then there exists a unique solution $\omega \in H^1(\mathcal{A}_{(R', R)})$ to this boundary value problem, which obeys

$$\|\omega\|_{H^1(\mathcal{A}_{(R', R)})} \lesssim_{R', R} \|f\|_{L^2(\mathcal{A}_{(R', R)})} + \|g\|_{L^2(\mathcal{A}_{(R', R)})}.$$

Proof. For simplicity, we write $\mathcal{A} = \mathcal{A}_{(R', R)}$ and omit the dependence of constants on R' and R. As in Lemma A.5, we start by solving the following boundary value problem for $\tilde{g}$:

$$\begin{cases} -\Delta\tilde{g} = g & \text{in } \mathcal{A}, \\ (\iota_\nu \mathrm{d}\tilde{g}, \iota_\nu \tilde{g}) = 0 & \text{on } \partial\mathcal{A}. \end{cases}$$

Here $-\Delta = \delta\mathrm{d} + \mathrm{d}\delta$ denotes the Hodge Laplacian. By the solvability of the absolute boundary value problem for 2-forms, we may find a unique solution $\tilde{g} \in H^2(\mathcal{A})$ to these problems; see, for instance, [54, Proposition 9.8]. We remark that for uniqueness, we use the Hodge theorem and the fact that the second de Rham cohomology group of $\mathcal{A}$ is trivial. We also note that

$$\mathrm{d}\delta\mathrm{d}\tilde{g} = \mathrm{d}(-\Delta)\tilde{g} = \mathrm{d}g = 0.$$

Testing $\mathrm{d}\tilde{g}$ against the above equation, integrating d by parts and using $\iota_\nu \mathrm{d}\tilde{g} = 0$ to make the boundary terms vanish, it follows that $\delta\mathrm{d}\tilde{g} = 0$.

Next, we solve the following Neumann boundary value problem for $\tilde{f}$:

$$\begin{cases} -\Delta\tilde{f} = f & \text{in } \mathcal{A}, \\ \iota_\nu \mathrm{d}\tilde{f} = -\iota_\nu \delta\tilde{g} & \text{on } \partial\mathcal{A}, \\ \displaystyle\int_{\mathcal{A}} \tilde{f} = 0. \end{cases}$$

To solve this problem with $\tilde{f} \in H^2(\mathcal{A})$, we need to verify the following the compatibility condition (which arises from integrating $-\Delta\tilde{f} = f$ over $\mathcal{A}$, integrating the LHS by parts and using the conditions on f and $\iota_\nu d\tilde{f}|_{\partial\mathcal{A}}$):

$$0 = \int_{\partial\mathcal{A}} \iota_\nu \delta\tilde{g} \, d\mathrm{Vol}_{\partial\mathcal{A}}.$$

Since $\iota_\nu \star 1$ is precisely the induced volume form $d\mathrm{Vol}_{\partial\mathcal{A}}$, for any sufficiently regular 3-form η defined in a neighborhood of $\partial\mathcal{A}$ we have $\iota_\nu \star \eta \, d\mathrm{Vol}_{\partial\mathcal{A}} = i^*_{\partial\mathcal{A}}\eta$, where $i_{\partial\mathcal{A}}$ is the embedding $\partial\mathcal{A} \hookrightarrow \overline{\mathcal{A}}$. It follows that

$$\iota_\nu \delta\tilde{g} \, d\mathrm{Vol}_{\partial\mathcal{A}} = d(i^*_{\partial\mathcal{A}} \star \tilde{g}),$$

so that the compatibility condition holds by the Stokes theorem.

In conclusion, $\omega = d\tilde{f} + \delta\tilde{g}$ gives a desired $H^1(\mathcal{A})$ solution to the div-curl system (A.3) with $\iota_\nu\omega|_{\partial\mathcal{A}} = 0$; uniqueness follows from the Hodge theorem and the fact that the first de Rham cohomology group of $\mathcal{A}$ is trivial. $\qquad\square$

We are now ready to state and prove the perturbative solvability results for the nonlinear div-curl system (A.2).

Proposition A.7 (Nonlinear div-curl system in $\mathbb{R}^5$). *Consider the nonlinear div-curl system (A.2) in $(\mathbb{R}^5, \mathbf{e})$, where $\mathbf{e}_{\alpha\beta}$ is a smooth metric such that $\|\mathbf{e}_{\alpha\beta} - \bar{\mathbf{e}}_{\alpha\beta}\|_{L^\infty \cap \dot{H}^{\frac{5}{2}}} < \epsilon_{\mathbf{e}}$ for some constant positive definite matrix $\bar{\mathbf{e}}_{\alpha\beta}$ and A obeys $A \in L^5$, $\nabla A \in L^2\dot{H}^{\frac{1}{2}}$ and $\|\boldsymbol{\nabla}^\alpha A_\alpha\|_{L^2\dot{H}^{\frac{1}{2}}} < \epsilon_A$. Then for $\epsilon_{\mathbf{e}}, \epsilon_A$ sufficiently small, there exists a unique solution Ω to this problem such that $\Omega \in L^5$ and $\nabla\Omega \in L^2\dot{H}^{\frac{1}{2}}$, which obeys*

$$\|\nabla\Omega\|_{L^2\dot{H}^{\frac{1}{2}}} \lesssim \|\boldsymbol{\nabla}^\alpha A_\alpha\|_{L^2\dot{H}^{\frac{1}{2}}}.$$

Proof. We aim to solve (A.2) by iteration; however, the RHS of the $d\Omega$ equation may not satisfy the compatibility condition during the iteration procedure. To rectify this issue, we use a Leray-type projection operator. For a 2-form g defined in $\mathbb{R}^5$, we introduce the operators

$$\mathbb{P}^{df} g = \delta(-\Delta)^{-1}dg, \quad \mathbb{P}^{cf} = g - \mathbb{P}^{df} g.$$

where δ and $(-\Delta)^{-1}$ are defined with respect to the constant metric $\bar{\mathbf{e}}$; such a simple choice is allowed since the condition we need to ensure $(d(\cdot) = 0)$ is independent of the metric.

Now we set up an iteration scheme by starting with $\Omega^{(0)} = 0$, and defining $\Omega^{(n)}$ by solving the following system:

$$\begin{cases} \boldsymbol{\nabla}^\alpha \Omega^{(n)}_\alpha = \boldsymbol{\nabla}^\alpha B_\alpha, \\ d\Omega^{(n)} = \dfrac{1}{2}\mathbb{P}^{cf}[\Omega^{(n-1)} \wedge \Omega^{(n-1)}]. \end{cases}$$

Using Lemma A.5, it is straightforward to show that $\Omega^{(n)}$ has a limit Ω such that $\|\nabla\Omega\|_{L^2\dot{H}^{\frac{1}{2}}} \lesssim \|\boldsymbol{\nabla}^\alpha B_\alpha\|_{L^2\dot{H}^{\frac{1}{2}}}$ and solves

$$\begin{cases} \boldsymbol{\nabla}^\alpha \Omega_\alpha = \boldsymbol{\nabla}^\alpha B_\alpha, \\ d\Omega = \dfrac{1}{2}\mathbb{P}^{cf}[\Omega \wedge \Omega] = \dfrac{1}{2}[\Omega \wedge \Omega] - Z, \end{cases}$$

where

$$Z = \frac{1}{2}\mathbb{P}^{df}[\Omega \wedge \Omega].$$

60

It remains to show that $Z = 0$. As a preparation, note that for any 1-form ω,

$$[[\omega \wedge \omega] \wedge \omega] = 0,$$

which follows from the Jacobi identity for the Lie bracket. Thus, Z obeys the identity

$$Z = \frac{1}{2}\delta(-\Delta)^{-1}\mathrm{d}[\Omega \wedge \Omega]$$

$$= \frac{1}{2}\delta(-\Delta)^{-1}[[\Omega \wedge \Omega] \wedge \Omega] - \frac{1}{2}\delta(-\Delta)^{-1}[\mathbb{P}^{df}[\Omega \wedge \Omega] \wedge \Omega]$$

$$= -\delta(-\Delta)^{-1}[Z \wedge \Omega].$$

But then since $Z \in L^2\dot{H}^{\frac{1}{2}}$ and

$$\|Z\|_{L^2\dot{H}^{\frac{1}{2}}} \lesssim \|\Omega\|_{L^5}\|Z\|_{L^2\dot{H}^{\frac{1}{2}}},$$

we have $Z = 0$ provided that ϵ_A is small enough, as desired. $\qquad\square$

Proposition A.8 (Nonlinear div-curl system in a 4-dimensional annulus)**.** *Consider the nonlinear div-curl system* (A.2) *in* $\mathcal{A}_{(R',R)} \subset \mathbb{R}^4$ *with the boundary condition*

$$\iota_\nu\omega = 0 \qquad on \; \partial\mathcal{A}_{(R',R)}.$$

Suppose that A is in $H^1(\mathcal{A}_{(R',R)})$ and obeys $\|\partial^\ell A_\ell\|_{L^2(\mathcal{A}_{(R',R)})} < \epsilon_A$. Then for ϵ_A sufficiently small (depending on R', R), there exists a unique solution $\Omega \in H^1(\mathcal{A}_{(R',R)})$ to this boundary value problem, which obeys

$$\|\Omega\|_{H^1(\mathcal{A}_{(R',R)})} \lesssim_{R',R} \|\partial^\ell A_\ell\|_{L^2(\mathcal{A}_{(R',R)})}.$$

Proof. Again, for simplicity, we write $\mathcal{A} = \mathcal{A}_{(R',R)}$ and omit the dependence of constants on R', R. As in the proof of Proposition A.7, the crucial step is to construct a suitable projection that enforces the compatibility condition. For sufficiently smooth g ($g \in H^2(\mathcal{A})$ is enough), solve the boundary value problem

$$\begin{cases} -\Delta u = \mathrm{d}g & \text{in } \mathcal{A}, \\ (\iota_\nu\mathrm{d}u, \iota_\nu u) = 0 & \text{on } \partial\mathcal{A}, \\ \qquad\qquad u \perp \mathcal{H}^3_A(\mathcal{A}), \end{cases}$$

where $\mathcal{H}^3_A(\mathcal{A})$ is the space of harmonic 3-forms η satisfying $\iota_\nu\eta = 0$; see, for instance, [54, Proposition 9.8]. Note that the solvability condition $\mathrm{d}g \perp \mathcal{H}^3_A$ is clearly satisfied. Furthermore, observe that $\mathrm{d}\delta\mathrm{d}u = \mathrm{d}^2g = 0$; thus testing by $\mathrm{d}u$ and using the boundary condition $\iota_\nu\mathrm{d}u = 0$, it follows that $\delta\mathrm{d}u = 0$. Thus, if we define

$$\mathbb{P}^{df}g = \delta u, \quad \mathbb{P}^{cf} = g - \mathbb{P}^{df}g,$$

then $\mathrm{d}\mathbb{P}^{df}g = \mathrm{d}g$ and $\mathrm{d}\mathbb{P}^{cf}g = 0$. Moreover, by the H^1 estimate for the Hodge Laplacian,

$$\|\mathbb{P}^{df}g\|_{L^2(\mathcal{A})} \lesssim \|u\|_{H^1(\mathcal{A})} \lesssim \|g\|_{L^2(\mathcal{A})},$$

by which we may extend $\mathbb{P}^{df}$ to any $g \in L^2(\mathcal{A})$.

As in Proposition A.7, we now solve the system

$$\begin{cases} \delta\Omega = \delta B, \\ \mathrm{d}\Omega = \dfrac{1}{2}\mathbb{P}^{cf}[\Omega \wedge \Omega], \end{cases}$$

with the boundary condition $\iota_\nu \Omega = 0$ by iteration. Note that $\iota_\nu B = 0$ on $\partial \mathcal{A}$ ensures the compatibility condition $\int_{\mathcal{A}} \delta B = 0$. Then to finish the proof, it suffices to show that

$$Z = \frac{1}{2} \mathbb{P}^{df}[\Omega \wedge \Omega],$$

which is a-priori in L^2, must vanish. Since $\mathbb{P}^{df} g$ is defined from dg, we may perform a similar computation as in Proposition A.7 and conclude that Z obeys a self-improving relation if ϵ_A is sufficiently small; thus $Z = 0$ as desired. $\qquad\square$

A.3. **Uhlenbeck's lemmas.** Here, we record various lemmas that allows us to pass to the Coulomb gauge under a suitable gauge-independent smallness condition. We begin with the classical results proved by Uhlenbeck. In what follows, ν denotes the outward unit normal to B_R on ∂B_R.

Theorem A.9. *Let A be a connection in a ball $B_R \subset \mathbb{R}^4$ that satisfies $A \in H^1(B_R)$ and $\|F\|_{L^2(B_R)} < \epsilon_0$. If ϵ_0 is sufficiently small, then there exists an admissible gauge transformation O, unique up to multiplication by a constant element of $\mathbf{G}$, such that $\tilde{A} = \mathcal{G}(O)A$ obeys*

$$\partial^\ell \tilde{A}_\ell = 0 \text{ in } B_R, \qquad \nu^\ell \tilde{A}_\ell = 0 \text{ on } \partial B_R$$

and

$$\|\tilde{A}\|_{\dot{H}^1(B_R)} \lesssim \|F\|_{L^2(B_R)}.$$

Theorem A.10. *Let A be a connection in $\mathbb{R}^4 \setminus B_R$ that satisfies $A \in \dot{H}^1 \cap L^4(\mathbb{R}^4 \setminus B_R)$ and $\|F\|_{L^2(\mathbb{R}^4 \setminus B_R)} < \epsilon_0$. If ϵ_0 is sufficiently small, then there exists an admissible gauge transformation O, unique up to multiplication by a constant element of $\mathbf{G}$, such that $\tilde{A} = \mathcal{G}(O)A$ obeys*

$$\partial^\ell \tilde{A}_\ell = 0 \text{ in } \mathbb{R}^4 \setminus B_R, \qquad \nu^\ell \tilde{A}_\ell = 0 \text{ on } \partial B_R$$

and

$$\|\tilde{A}\|_{\dot{H}^1(\mathbb{R}^4 \setminus B_R)} \lesssim \|F\|_{L^2(\mathbb{R}^4 \setminus B_R)}.$$

The first result is essentially [55, Theorem 1.3]; see [39, Theorem 3.11] for the uniqueness assertion at the critical regularity. The second result is the combination of [57, Theorem 4.5], which is formulated on a punctured disk, and a conformal inversion procedures, which is also in [57]. See, also, Theorem 3.12 and the proof of Theorem 1.5 in [39].

We end with a result that concerns a connection on the unit 3-sphere $\mathbb{S}^3$ whose curvature is small in L^2; note that this is a subcritical assumption.

Proposition A.11 (Subcritical Uhlenbeck on $\mathbb{S}^3$). *Let A be a connection in $\mathbb{S}^3$ that satisfies $A \in H^1_\Theta(\mathbb{S}^3)$ and $\|F\|_{L^2_\Theta(\mathbb{S}^3)} < \epsilon_0$. If ϵ_0 is sufficiently small, then there exists a H^2_Θ gauge transformation O, unique up to multiplication by a constant element of $\mathbf{G}$, such that $\tilde{A} = \mathcal{G}(O)A$ obeys $\nabla^\Theta A_\Theta = 0$ and*

$$\|\tilde{A}\|_{H^1_\Theta(\mathbb{S}^3)} \lesssim \|F\|_{L^2_\Theta(\mathbb{S}^3)}$$

This proposition is a slight strengthening of [56, Theorem 2.5]; we include a sketch of the proof for completeness.

62

Proof. We cover $\mathbb{S}^3$ by two caps $\mathcal{O}_N$ and $\mathcal{O}_S$ centered at the north and the south poles, respectively, and apply the usual Uhlenbeck's lemma (Theorem A.9) to each; we denote the resulting representations by $A^{(N)}$ and $A^{(S)}$, respectively. In the intersection, $A^{(N)} = \mathcal{G}(O^{(NS)})A^{(S)}$ for some gauge transformation $O^{(NS)} \in H^2_\Theta(\mathcal{O}_N \cap \mathcal{O}_S)$. By the Sobolev embedding, taking ϵ_0 small enough, the image of $O^{(NS)}$ is contained in a small ball near a constant element. Hence we may patch together $A^{(N)}$ and $A^{(S)}$ to obtain a global representation $\bar{A}$ such that $\|\bar{A}\|_{H^1_\Theta} \lesssim \epsilon_0$. Finally, applying a subcritical perturbative argument (see, for instance, [56, Proof of Theorem 2.5]), we find a gauge transformation from $\bar{A}$ into the Coulomb gauge. The uniqueness assertion follows also from the same perturbative argument. □

References

[1] N. Aronszajn, *A unique continuation theorem for solutions of elliptic partial differential equations or inequalities of second order*, J. Math. Pures Appl. (9) **36** (1957), 235–249.

[2] M. F. Atiyah, N. J. Hitchin, V. G. Drinfeld, and Y. I. Manin, *Construction of instantons*, Phys. Lett. A **65** (1978), no. 3, 185–187, `doi`.

[3] H. Bahouri, J.-Y. Chemin, and R. Danchin, *Fourier analysis and nonlinear partial differential equations*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 343, Springer, Heidelberg, 2011, `doi`.

[4] I. Bejenaru and S. Herr, *The cubic Dirac equation: small initial data in $H^{\frac{1}{2}}(\mathbb{R}^2)$*, Comm. Math. Phys. **343** (2016), no. 2, 515–562, `doi`.

[5] _____, *On global well-posedness and scattering for the massive Dirac–Klein–Gordon system*, J. Eur. Math. Soc. (JEMS) **19** (2017), no. 8, 2445–2467, `doi`.

[6] J. Bourgain, *Global solutions of nonlinear Schrödinger equations*, American Mathematical Society Colloquium Publications, vol. 46, American Mathematical Society, Providence, RI, 1999, `doi`.

[7] S. K. Donaldson and P. B. Kronheimer, *The geometry of four-manifolds*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 1990, Oxford Science Publications.

[8] R. Donninger and J. Krieger, *Nonscattering solutions and blowup at infinity for the critical wave equation*, Math. Ann. **357** (2013), no. 1, 89–163, `doi`.

[9] T. Duyckaerts, H. Jia, C. Kenig, and F. Merle, *Universality of blow up profile for small blow up solutions to the energy critical wave map equation*, preprint (2016), `arXiv:1612.04927`.

[10] 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** (1982), no. 2, 171–191, `link`.

[11] _____, *The global existence of Yang-Mills-Higgs fields in 4-dimensional Minkowski space. II. Completion of proof*, Comm. Math. Phys. **83** (1982), no. 2, 193–212, `link`.

[12] C. Gavrus, *Global well-posedness for the massive Maxwell-Klein-Gordon equation with small critical Sobolev data*, preprint (2016), `arXiv:1610.03581`.

[13] C. Gavrus and S.-J. Oh, *Global well-posedness of high dimensional Maxwell-Dirac for small critical data*, Mem. Amer. Math. Soc. (to appear), `arXiv:1604.07900`.

[14] M. G. Grillakis, *On the wave map problem*, Nonlinear wave equations (Providence, RI, 1998), Contemp. Math., vol. 263, Amer. Math. Soc., Providence, RI, 2000, pp. 71–84, `doi`.

[15] R. Grinis, *Quantization of time-like energy for wave maps into spheres*, Comm. Math. Phys. **352** (2017), no. 2, 641–702, `doi`.

[16] M. Gursky, C. Kelleher, and J. Streets, *A conformally invariant gap theorem in Yang-Mills theory*, preprint (2017), `arXiv:1708.01157`.

[17] J. Jendrej, *Construction of two-bubble solutions for energy-critical wave equations*, Amer. J. Math. **141** (2019), no. 1, 55–118, `doi`.

[18] J. Jendrej and A. Lawrie, *Two-bubble dynamics for threshold solutions to the wave maps equation*, preprint (2017), `arXiv:1706.00089`.

[19] 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** (2008), no. 2, 147–212, `doi`.

[20] S. Klainerman and M. Machedon, *Space-time estimates for null forms and the local existence theorem*, Comm. Pure Appl. Math. **46** (1993), no. 9, 1221–1268, `doi`.

[21] _____, *On the Maxwell-Klein-Gordon equation with finite energy*, Duke Math. J. **74** (1994), no. 1, 19–44, `doi`.

[22] _____, *Finite energy solutions of the Yang-Mills equations in* $\mathbb{R}^{3+1}$, Ann. of Math. (2) **142** (1995), no. 1, 39–119, `doi`.

[23] S. Klainerman and D. Tataru, *On the optimal local regularity for Yang-Mills equations in* $\mathbf{R}^{4+1}$, J. Amer. Math. Soc. **12** (1999), no. 1, 93–116, `doi`.

[24] H. Koch and D. Tataru, *Carleman estimates and unique continuation for second-order elliptic equations with nonsmooth coefficients*, Comm. Pure Appl. Math. **54** (2001), no. 3, 339–360, `doi`.

[25] J. Krieger, W. Schlag, and D. Tataru, *Renormalization and blow up for the critical Yang-Mills problem*, Adv. Math. **221** (2009), no. 5, 1445–1521, `doi`.

[26] J. Krieger and J. Lührmann, *Concentration compactness for the critical Maxwell-Klein-Gordon equation*, Ann. PDE **1** (2015), no. 1, Art. 5, 208.

[27] J. Krieger and W. Schlag, *Concentration compactness for critical wave maps*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2012, `doi`.

[28] J. Krieger and J. Sterbenz, *Global regularity for the Yang-Mills equations on high dimensional Minkowski space*, Mem. Amer. Math. Soc. **223** (2013), no. 1047, vi+99, `doi`.

[29] J. Krieger, J. Sterbenz, and D. Tataru, *Global well-posedness for the Maxwell-Klein-Gordon equation in* $4 + 1$ *dimensions: small energy*, Duke Math. J. **164** (2015), no. 6, 973–1040, `doi`.

[30] J. Krieger and D. Tataru, *Global well-posedness for the Yang-Mills equation in* $4 + 1$ *dimensions. Small energy*, Ann. of Math. (2) **185** (2017), no. 3, 831–893, `doi`.

[31] A. Lawrie and S.-J. Oh, *A refined threshold theorem for* $(1 + 2)$*-dimensional wave maps into surfaces*, Comm. Math. Phys. **342** (2016), no. 3, 989–999, `doi`.

[32] S.-J. Oh, *Gauge choice for the Yang-Mills equations using the Yang-Mills heat flow and local well-posedness in* H^1, J. Hyperbolic Differ. Equ. **11** (2014), no. 1, 1–108, `doi`.

[33] _____, *Finite energy global well-posedness of the Yang-Mills equations on* $\mathbb{R}^{1+3}$*: an approach using the Yang-Mills heat flow*, Duke Math. J. **164** (2015), no. 9, 1669–1732, `doi`.

[34] S.-J. Oh and D. Tataru, *Global well-posedness and scattering of the* $(4+1)$*-dimensional Maxwell-Klein-Gordon equation*, Invent. Math. **205** (2016), no. 3, 781–877, `arXiv:1503.01562`, `doi`.

[35] _____, *Local well-posedness of the* $(4 + 1)$*-dimensional Maxwell-Klein-Gordon equation at energy regularity*, Ann. PDE **2** (2016), no. 1, Art. 2, 70, `arXiv:1503.01560`, `doi`.

[36] _____, *Energy dispersed solutions for the (4+1)-dimensional Maxwell-Klein-Gordon equation*, Amer. J. Math. (2017), `arXiv:1503.01561`.

[37] _____, *The Yang-Mills heat flow and the caloric gauge*, preprint (2017), `arXiv:1709.08599`.

[38] _____, *The hyperbolic Yang-Mills equation in the caloric gauge. Local well-posedness and control of energy dispersed solutions*, preprint (2017), `arXiv:1709.09332`.

[39] _____, *The hyperbolic Yang-Mills equation for connections in an arbitrary topological class*, preprint (2017), `arXiv:1709.08604`.

[40] _____, *The Threshold Theorem for the* $(4 + 1)$*-dimensional Yang-Mills equation: An overview of the proof*, preprint (2017), `arXiv:1709.09088`.

[41] P. Raphaël and I. Rodnianski, *Stable blow up dynamics for the critical co-rotational wave maps and equivariant Yang-Mills problems*, Publ. Math. Inst. Hautes Études Sci. **115** (2012), 1–122, `doi`.

[42] I. Rodnianski and T. Tao, *Global regularity for the Maxwell-Klein-Gordon equation with small critical Sobolev norm in high dimensions*, Comm. Math. Phys. **251** (2004), no. 2, 377–426, `doi`.

[43] J. Sterbenz and D. Tataru, *Energy dispersed large data wave maps in* $2 + 1$ *dimensions*, Comm. Math. Phys. **298** (2010), no. 1, 139–230, `doi`.

[44] _____, *Regularity of wave-maps in dimension* $2 + 1$, Comm. Math. Phys. **298** (2010), no. 1, 231–264, `doi`.

[45] T. Tao, *Global regularity of wave maps. II. Small energy in two dimensions*, Comm. Math. Phys. **224** (2001), no. 2, 443–544, `doi`.

[46] _____, *Geometric renormalization of large energy wave maps*, Journées "Équations aux Dérivées Partielles", École Polytech., Palaiseau, 2004, pp. Exp. No. XI, 32.

[47] _____, *Global regularity of wave maps III. Large energy from βR^{1+2} to hyperbolic spaces*, preprint (2008), `arXiv:0805.4666`.

[48] _____, *Global regularity of wave maps IV. Absence of stationary or self-similar solutions in the energy class*, preprint (2008), `arXiv:0806.3592`.

[49] _____, *Global regularity of wave maps V. Large data local wellposedness and perturbation theory in the energy class*, preprint (2008), `arXiv:0808.0368`.

[50] _____, *Global regularity of wave maps VI. Abstract theory of minimal-energy blowup solutions*, preprint (2009), `arXiv:0906.2833`.

[51] _____, *Global regularity of wave maps VII. Control of delocalised or dispersed solutions*, preprint (2009), `arXiv:0908.0776`.

[52] D. Tataru, *On global existence and scattering for the wave maps equation*, Amer. J. Math. **123** (2001), no. 1, 37–77, `link`.

[53] _____, *Rough solutions for the wave maps equation*, Amer. J. Math. **127** (2005), no. 2, 293–377, `link`.

[54] M. E. Taylor, *Partial differential equations I. Basic theory*, second ed., Applied Mathematical Sciences, vol. 115, Springer, New York, 2011, `doi`.

[55] K. K. Uhlenbeck, *Connections with L^p bounds on curvature*, Comm. Math. Phys. **83** (1982), no. 1, 31–42, `link`.

[56] _____, *Removable singularities in Yang-Mills fields*, Comm. Math. Phys. **83** (1982), no. 1, 11–29, `link`.

[57] _____, *The Chern classes of Sobolev connections*, Comm. Math. Phys. **101** (1985), no. 4, 449–457, `link`.

KIAS, Seoul, Korea 02455
E-mail address: `sjoh@kias.re.kr`

Department of Mathematics, UC Berkeley, Berkeley, CA 94720
E-mail address: `tataru@math.berkeley.edu`