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 [31], the continuation/scattering criteria established in the second paper [32], and the large data analysis in an arbitrary topological class at optimal regularity in the third paper [33], 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 of a nontrivial self-similar solution. The proof is completed by showing that the latter solutions do not exist.

CONTENTS

1. Introduction 2
   1.1. The energy critical hyperbolic Yang–Mills equation 2
   1.2. The main results 7
   1.3. A brief history and broader context 11
   1.4. Structure of the present paper 14

2. An outline of the first three papers 15
   2.1. The caloric gauge 15
   2.2. Local well-posedness and energy dispersion 18
   2.3. Topological classes and large data solutions 19

3. Monotonicity formulas 22
   3.1. Some notation 22
   3.2. The energy momentum tensor and conservation laws 22
   3.3. Monotonicity formulas 24

4. Compactness 25

5. Regularity of stationary connections 26

6. No finite energy self-similar solutions 28

7. The bubbling-off result 31

8. No null concentration 35

9. Proof of the Threshold Theorem and the Dichotomy Theorem 40
   9.1. The blow-up case 41
   9.2. The non-scattering scenario 43

References 43
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, \[31\],
(b) large data energy dispersed caloric gauge solutions, \[32\]
(c) topological classes of connections and large data local well-posedness, \[33\].
(d) soliton bubbling/scattering dichotomy for large data solutions, present article.

A shorter overview of the four papers is provided in the survey paper \[34\].

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 \(2E_{GS}\), 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 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), respectively 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 \[31\], \[32\], \[33\]. 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 \(G\) be a compact noncommutative Lie group and \(g\) its associated Lie algebra. We denote by \(Ad(O)X = OXO^{-1}\) the action of \(G\) by conjugation (i.e., the adjoint action), and by \(ad(X)Y = [X, Y]\) the associated action of \(g\), which is given by the Lie bracket. We introduce the notation \(\langle X, Y \rangle\) for a bi-invariant inner product on \(g\),

\[
\langle [X, Y], Z \rangle = \langle X, [Y, Z] \rangle, \quad X, Y, Z \in g,
\]

or equivalently

\[
\langle X, Y \rangle = \langle Ad(O)X, Ad(O)Y \rangle, \quad X, Y \in g, \quad O \in G.
\]

If \(G\) is semisimple then one can take \(\langle X, Y \rangle = -\text{tr} (ad(X)ad(Y))\) i.e. negative of the Killing form on \(g\), which is then positive definite. However, a bi-invariant inner product on \(g\) exists for any compact Lie group \(G\).
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 $m = \text{diag}(-1, 1, 1, 1, 1)$. Denote by $A_\alpha : \mathbb{R}^{1+4} \to \mathfrak{g}$, $\alpha = 1, \ldots, 4$, a connection 1-form taking values in the Lie algebra $\mathfrak{g}$, and by $D_\alpha$ the associated covariant differentiation,

$$D_\alpha B := \partial_\alpha B + [A_\alpha, B],$$

acting on $\mathfrak{g}$ valued functions $B$. The commutator of two covariant derivatives takes the form $(D_\alpha D_\beta - D_\beta D_\alpha)B = [F_{\alpha\beta}, B]$, 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

$$D_\alpha F_{\beta\gamma} + D_\beta F_{\gamma\alpha} + D_\gamma F_{\alpha\beta} = 0.$$

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

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

Here we are using the standard convention for raising indices. Thus, the Yang–Mills equation takes the form

$$D^\alpha F_{\alpha\beta} = 0.$$  \hfill (1.1)

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

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

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

$$\partial^\alpha T_{\alpha\beta} = 0.$$  \hfill (1.2)

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

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

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 general $U \subset \mathbb{R}^4$, we define the local energy in $U$ to be

$$\mathcal{E}_{\{t\} \times U}(A) = \int_{\{t\} \times U} T_{00} \, dx.$$
1.1.3. Symmetries. The group of symmetries for 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 \( E \) 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

\[ O_\alpha = \partial_\alpha OO^{-1}. \]

The equation \( (1.1) \) is invariant under the gauge transformations

\[ A_\alpha \mapsto Ad(O)A_\alpha - O_\alpha =: G(O)A, \]

with \( O \) a function taking values in the group \( G \). 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 the gauge. This procedure is known as gauge fixing.

The choice of the 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 large data problem, and a main goal of our first paper [31] 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 \( g \)-valued 1-forms \( (a_j, e_j) \) on \( \mathbb{R}^4 \). We say that \( (a_j, e_j) \) is the initial data 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 \( (1.1) \):

\[ D^je_j = 0. \quad (1.4) \]

where \( D^j \) 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 \( a \).

**Definition 1.1.**  

a) A regular initial data set for the Yang–Mills equation is a pair \( (a_j, e_j) \in H_{lo}^{N-1} \times H^{N-1}, N \geq 2 \), 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 \( (a_j, e_j) \in \dot{H}_{lo}^1 \times L^2 \), with \( f \in L^2 \), and which satisfies the constraint equation \( (1.4) \).

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 \( A \in C([0, T]; H^N_{loc}) \), whose curvature satisfies \( F \in C([0, T]; H^{N-1}_{loc}) \) 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 \( A \in C([0, T]; H^1_{loc}) \), whose curvature satisfies \( F \in C([0, T]; L^2_{loc}) \) 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 a map

\[
O : \mathcal{O} \to G
\]

with the following regularity properties:

\[
O_x, O_t \in C^1_t(H^{N+1}_{loc}).
\]

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:

\[
O_x, O_t \in C^1_t(H^1_{loc}).
\]

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 \( \mathcal{O} \subset \mathbb{R}^{4+1} \) are gauge equivalent if there exists an admissible gauge transformation \( O \) so that

\[
A^{(2)}_\alpha = OA^{(1)}_\alpha O^{-1} - \partial_\alpha OO^{-1}.
\]

We list some simple properties of finite energy connections and admissible gauge transformations in an open set \( \mathcal{O} \) (see [33]):

- 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) iff \( A^{(2)} \) 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 [24], caloric gauge [32] and temporal gauge [32].
(ii) Large data local well-posedness holds for large caloric data in the caloric gauge \[32\]
and for arbitrary large data in temporal gauge \[33\].
(iii) Uniqueness of finite energy solutions (up to gauge transformations) \[33\].

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 \[33\] 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 \(H^1\) connections \[33\], see Theorem 2.13 below. The topologically trivial
connections are the subject of both the first two papers \[31\] and \[32\] 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 \(\chi\) defined by
\[
\chi(a) = \int_{\mathbb{R}^4} -\langle f \wedge f \rangle = \frac{1}{4} \int_{\mathbb{R}^4} -\langle f_{ij}, f_{k\ell} \rangle \, dx^i \wedge dx^j \wedge dx^k \wedge dx^\ell,
\]
which depends only on the topological class \([a]\) of \(a\). Two key properties of \(\chi\) are that
\(\chi([0]) = 0\) and the pointwise bound
\[
|\langle f \wedge f \rangle| \leq \frac{1}{2} \langle f_{jk}, f^{jk} \rangle \leq T_{00}(a),
\]
which is referred to as the Bogomol’nyi bound; see \[33\] for their proofs.

In the case \(G = SU(2)\) the topological class of \(a\) is fully described by the characteristic
class, which is then a multiple of the Chern number \(c_2\) of \(a\). This can be shown 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 \(G\) the characteristic number \(\chi\) provides only a
partial description of the topological class.

1.1.8. Harmonic Yang–Mills connections and the ground state. A harmonic Yang–Mills con-
nexion in \(\mathbb{R}^4\) is a \(H^1_{\text{loc}}\) 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 \[5\].

The Euler–Lagrange equation satisfied by \(a\) takes the form
\[
D^2 f_{ij} = 0,
\]
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 \[48, 49\]). Harmonic Yang–Mills connections \(a \in H^1_{\text{loc}}(\mathbb{R}^4)\) are
smooth in a suitable gauge. More generally, \(H^2_{\text{loc}}\) 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 [33]:

**Theorem 1.6.** Let $G$ be a noncommutative compact Lie group. Let

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

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 $\mathcal{E}[a] \geq 2E_{GS}$, or $|\chi| = \mathcal{E}_e[a] \geq E_{GS}$.

This result is a combination of classical results [1, 5] 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 Gurskey–Kelleher–Street [14]. For a derivation, see [33, 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 $(a, e)$. This problem is known to be locally well-posed [32], and globally well-posed for small data [24]. Here we are interested in the large data problem, and we seek to address the following two questions:

**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 $(T, X)$,

$$C = \{|x - X| \leq T - t\}.$$

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 [24] 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 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 the following notation: Given a backward light cone

$$C = \{|x - X| \leq T - t\}$$
we introduce the time slices
\[ S_t = C \cap \{t\} \times \mathbb{R}^4, \]
and for \(0 < \gamma < 1\), the time-like cone
\[ C_\gamma = \{|x - X| \leq \gamma(T - t)|. \]

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 \(\gamma < 1\) we have
\[ \limsup_{t \to T} E_{C_\gamma \cap S_t}(A) > 0, \quad \text{where} \quad C = \{|x - X| \leq T - t\}. \tag{1.6} \]

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

1. **Time-like concentration,**
\[ \limsup_{n \to \infty} \frac{x_n - X}{|t_n - T|} = v, \quad |v| < 1. \]

2. **Below self-similar scale,**
\[ \limsup_{n \to \infty} \frac{r_n}{|t_n - T|} = 0. \]

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

for some sequence of admissible gauge transformations \(O_n\), a Lorentz transformation \(L_v\) and finite energy harmonic Yang–Mills connection \(Q\).

b) Let \(A\) be a finite energy Yang–Mills connection which is global forward in time. Assume in addition that for some \(\gamma < 1\) we have
\[ \limsup_{t \to \infty} E_{C_\gamma \cap S_t}(A) > 0, \quad C = \{|x| \leq t\}. \tag{1.7} \]

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,**
\[ \limsup_{n \to \infty} \frac{x_n}{t_n} = v, \quad |v| < 1. \]

2. **Below self-similar scale,**
\[ \limsup_{n \to \infty} \frac{r_n}{t_n} = 0. \]

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

for some sequence of admissible gauge transformations \(O_n\), a Lorentz transformation \(L_v\) and finite energy harmonic Yang–Mills connection \(Q\).
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
\[
\partial_s A_j(x, s) = D^\ell F_{\ell j}(x, s), \quad A_j(x, s = 0) = a_j(x)
\]
exists globally in heat-time \(s\) and \(\lim_{s \to \infty} A(s) = 0\). Denoting by \(C\) the manifold of finite energy caloric connections, and by \(T^{L^2}C\) 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 \(T^{L^2}C\) (see [31] 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 are enough to establish small energy global well-posedness, which can then be turned into large data local well-posedness in temporal gauge by causality (see [33] and Section 2.3). This provides a suitable setting for considering evolution of arbitrary finite energy data, albeit with more indirect control.

We now present the Threshold Theorem, which is a sharp criterion for global well-posedness and scattering in terms of the initial energy. 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 \(2E_{GS}\) 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 \(2E_{GS}\), in the following sense:

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

b) (Rough data) The flow map admits an extension
\[
T^{L^2}C \ni (a_j, b_j) \rightarrow (A_\alpha, \partial_t A_\alpha) \in C(\mathbb{R}, T^{L^2}C)
\]
and which is continuous in the \(\dot{H} \cap \dot{H}^\sigma\) topology for \(\sigma < 1\) and close to \(1\).
c) (Weak Lipschitz dependence) The flow map is globally Lipschitz in the $\dot{H}^\sigma$ topology for $\sigma < 1$, close to 1.

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

$$\|A_x\|_{S^1} + \|\nabla A_0\|_{\ell^1L^2H^{1/2}} < \infty.$$  \hspace{1cm} (1.8)

Here, $\dot{H}^\sigma = \dot{H}^\sigma \times \dot{H}^{\sigma-1}$. Some remarks concerning Theorem 1.8 are in order.

Remark 1.9. The preceding theorem is stated for initial data $(a, b) \in T^{L^2}C$ 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 $(\tilde{a}, e) \in H^1 \times L^2$ with energy below $2E_{GS}$ 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 [23, 30] and the small data problem for Yang–Mills in [24]. 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 $2E_{GS}$ 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.** 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 $S$ 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, whose energies may be arbitrarily close to the threshold \cite{19, 35}. 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, we expect solutions where bubbling occurs at infinity to also exist. Such solutions have been obtained for related models such as the energy critical Wave Maps equation, see \cite{6}. However, to the best of our knowledge such a construction has not yet been carried out for Yang–Mills.

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 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 \cite{16}, and by many others later on.

2. The $X^{s,b}$ spaces. A second advance was the introduction of the $X^{s,b}$ space\cite{16}, also first used by Klainerman and Machedon \cite{16} 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 \cite{46} 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 \cite{39}, also in the wave map context, was to renormalize the paradifferential problem, i.e., to find a suitable approximate conjugation to the corresponding constant coefficient problem.

\footnote{The constructions in \cite{19, 35} are for $G = SU(2)$ in the first topological class with $c_2 = 1$, but a straightforward gluing argument at infinity produces the desired topologically trivial solutions.}

\footnote{The concept, and also the notation, is due to Bourgain, in the context of KdV and NLS type problems.}
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 [4], 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 [15] 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 [21], 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 [26] [27], and is based on an idea proposed by Tao [40] in the Wave Maps context.

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 [37] [38] of the second author and Sterbenz in the wave map 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 [15] yields more accurate information; for instance, see [21]. 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 [37] [38], 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. 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 [12], are a convenient and relatively simple tool to 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 recall some earlier developments on the hyperbolic Yang–Mills equation (YM) related to the present paper. We start our discussion with the (YM) problem above the scaling critical regularity. In the two and three dimensional cases, which are energy subcritical, global regularity of sufficiently regular solutions was shown in the early works [8] [9]. These papers in fact handled the more general Yang–Mills-Higgs system. In dimension $d = 3$, local well-posedness in the energy space was proved in [17]. 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 [18].
However, a new difficulty arises in the large data problem for (YM): Namely, the gauge transformation law is *nonlinear* due to the non-abelian 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 [17], 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, and then working in local Coulomb gauge. An alternative, more robust approach without spacetime localizations to the same problem has been put forth by the first author in [20, 27], inspired by [40]. 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.

Before turning to the (YM) problems at critical regularity, we briefly recall some recent developments on the wave map 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 [46], [39], [47]. 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 [37, 38] in general, and independently in [21] and [13, 41, 42, 44, 45] for specific targets (namely the hyperbolic space). See also [25] 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 $2E_{GS}$ in our work. Our present strategy was strongly influenced by [37, 38].

For the energy critical (WM), in the important case of spherical targets, we also note the recent development due to Grinis [13], 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 [7], this was complemented with a decay of the energy near the cone when the total energy is sufficiently close to the ground state.

Next, we discuss the (YM) problems at critical regularity. It is instructive to simultaneously consider the corresponding problems for (MKG), which is a simpler abelian 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 $\phi$, obey a possibly nonlinear gauge covariant wave equation $\Box A \phi = \cdots$, and the associated curvature $F[A]$ is determined by $\phi$. 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 $\phi$. 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

---

4 More precisely, a suitable scaling critical norm of the connection $A$ (e.g., $||A||_{L^2}$) or the curvature $F$ (e.g., $||F||_{L^2}$) is large.

5 On the other hand, a closely related spacetime localization approach, but relying on new “initial data surgery” techniques, is developed in [33], which yields an alternative proof of the main result of [17].
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 [36], 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 [22]. The small data result in the energy critical dimension $4 + 1$ was obtained in [23]. Finally, the large data result for (MKG) in dimension $4 + 1$ was proved by the authors in [29, 30, 28] and independently by [20]. Although the implementation differs in many places, the outline of the three papers [29, 30, 28] is broadly followed in the present work. In particular we borrow a good deal of notation, ideas and estimates from both [23] and [29, 30, 28]. On the other hand, we remark that [20] followed the strategy of [21]. For the (YM) problem in $4 + 1$ dimensions, the small data global result was only recently proved in [24], which is another direct predecessor the present work.

We conclude with a remark on differences between (MKG) and (YM). The issue of non-abelian 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 [19, 35], so for (YM) one must prove the Threshold Theorem, instead of an 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).

In related developments, one should also note the works [2, 3] on the closely related cubic Dirac equation, as well as the massive Dirac-Klein-Gordon system, as well as [11] on the Maxwell-Dirac equation and [10] on the massive Maxwell- Klein-Gordon system.

1.4. **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 [31], [32], and [33], 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 for large energy Yang–Mills connections in order to prove a strong local compactness statement, Theorem 1.1, that we rely on on in our blow-up analysis.

- **Section 5.** Here we study the regularity of connections with a time-like Killing field, 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 (YM) 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 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.
Finally, here we complete both the proof of the Threshold Conjecture, see Theorem 1.8, and the dichotomy result in Theorem 1.11.

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

2. An outline of the first three papers

Our aim here is to provide a brief outline of the first three papers [31], [32] and [33], 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 [34].

Let us take as a starting point of our discussion the following small data result proved earlier in [24]:

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 [31] is devoted solely to the gauge problem; precisely, inspired by earlier work of Tao [40] and of the first author [26, 27], 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 [32]. On the other hand for the large data analysis in [33] causality plays a key role, so we shift the (soft part of the) analysis to the temporal gauge. Finally, most of the work in the present paper is carried out in a gauge-covariant fashion, while using the gauge-dependent results in [32] and [33] at critical junctures.

2.1. The caloric gauge. The goal of the first paper [31] 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

\[
\partial_s A_j = D^k F_{jk}, \quad A_j(s = 0) = a_j
\]

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

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

**Theorem 2.2** ([31, Corollary 5.14]). Let $a \in \dot{H}^1$ be a connection so that $Q(a) < \infty$. Then this solution has the property that the limit
\[ \lim_{s \to \infty} A(s) = a_\infty \]
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 $\dot{H}^1 \cap H^N$ ($N \geq 2$).

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

**Theorem 2.3** ([31, 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** ([31, 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
\[ Q : [0, 2E_{GS}) \to \mathbb{R}^+ \]
so that for all subthreshold data $a$ with energy $E$ we have
\[ Q(a) \leq Q(E). \tag{2.3} \]

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 $Q(a) < \infty$ 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 an unique\(^6\) equivalent caloric gauge. This is because $a_\infty$ 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
\[ \tilde{a} = G(O)a = OaO^{-1} - \partial_x OO^{-1} \]
is caloric.

A key result in [31] asserts that:

\(^6\)Up to constant conjugations
Theorem 2.6 ([31 Proposition 7.7 and Theorem 7.8]). The space \( C \) of all caloric connections is a \( C^1 \) submanifold of the space
\[
H = \{ a \in \dot{H}^1, \ \partial^j a_j \in \ell^1 L^2 \}
\]
In addition, for all subthreshold caloric connections with energy \( E \) and caloric size \( Q \) we have the bound
\[
\|a\|^2_H \lesssim E, Q 1
\]

The second part of the state space norm \( H \) reflects the fact that caloric connections satisfy a nonlinear form of the Coulomb gauge condition.

The second part of [31] is devoted to modeling the hyperbolic Yang–Mills equation as an evolution on the caloric manifold. Precisely, the state space for this evolution is \( T^{L^2}C \), which is the \( L^2 \) completion of the tangent space \( TC \). We will view the spatial components \((A_x, \partial_t A_x) \in T^{L^2}C\) as the dynamic variable, and the temporal parts \( A_0, \partial_0 A_0 \) as auxiliary. Correspondingly, we call a pair \((a, b) \in T^{L^2}C\) (i.e., \( a \in C, b \in T^{L^2}C \)) a caloric gauge initial data set for \([1,1]\). This is related to the gauge-covariant notion of initial data sets as follows:

Theorem 2.7 ([31 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 \((\tilde{a}, \tilde{e}) \in T^{L^2}C\) and \(a_0 \in \dot{H}^1\), so that \((\tilde{a}, \tilde{e})\) is gauge equivalent to \((a, e)\), where
\[
\tilde{e}_k = b_k - D_k^{(a)} a_0.
\]
b) Given any caloric gauge initial data set \((\tilde{a}, \tilde{e}) \in T^{L^2}C\), 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
\[
e_k = b_k - D_k^{(a)} a_0
\]
satisfies the constraint equation (1.4).

By this result, we may indeed fully describe Yang–Mills connections in the caloric gauge as continuous functions
\[
I \ni t \to (A_x, \partial_t A_x)(t) \in T^{L^2}C.
\]
The equations for the dynamical variables \((A_x, \partial_t A_x)\) are proved to have the form
\[
\Box A_k = \mathbf{P}[A_j, \partial_t A_j] + 2\Delta^{-1} \partial_k Q(\partial^\alpha A_j, \partial_\alpha A_j) + R(A, \partial_t A) \quad (2.4)
\]
together with a compatibility condition
\[
\partial^k A_k = \mathbf{D} A(A) := Q(A, A) + \mathbf{D} A^3(A), \quad (2.5)
\]
where the temporal component \( A_0 \) and its time derivative \( \partial_0 A_0 \) are uniquely determined by \((A_x, \partial_t A_x)\) and admit the expressions
\[
A_0 = A_0(A) := \Delta^{-1}[A, \partial_t A] + 2\Delta^{-1} Q(A, \partial_t A) + A_0^3(A), \quad (2.6)
\]
\[
\partial_t A_0 = \mathbf{D} A_0(A) := -2\Delta^{-1} Q(\partial_t A, \partial_t A) + \mathbf{D} A_0^3(A). \quad (2.7)
\]
Here \( \mathbf{P} \) is the Leray projector, \( Q \) is a symmetric bilinear form with symbol\[^7\]
\[
Q(\xi, \eta) = \frac{\xi^2 - \eta^2}{2(\xi^2 + \eta^2)}.
\]
\[^7\]Although the symbol looks anti-symmetric, it is compensated by the Lie bracket in the definition of bilinear multipliers for \( g \)-valued functions; see [31 Definition 3.1].
and the cubic error terms $R$, $\mathbf{D}A_3^{(3)}$, $A_0^3$ and $\mathbf{D}A_0^3$ are Lipschitz “envelope preserving” maps as follows:

$$R : \text{Str}^1 \to L^1 L^2 \cap L^2 H^{-\frac{1}{2}}, \quad (2.9)$$
$$\mathbf{D}A_3 : \text{Str}^1 \to L^1 \dot{H}^1 \cap L^2 \dot{H}^{\frac{1}{2}}, \quad (2.10)$$
$$A_0^3 : \text{Str}^1 \to L^1 \dot{H}^2 \cap L^2 \dot{H}^{\frac{3}{2}}, \quad (2.11)$$
$$\mathbf{D}A_0^3 : \text{Str}^1 \to L^1 \dot{H}^1 \cap L^2 \dot{H}^{\frac{1}{2}}. \quad (2.12)$$

Here $\text{Str}^1$ collects several standard non-endpoint Strichartz norms with the appropriate scaling. The envelope preserving property here asserts that dyadic frequency envelope bounds are preserved in the estimates for both the nonlinear expressions $R$, $\mathbf{D}A_3$, $A_0^3$ and $\mathbf{D}A_0^3$ and for corresponding differences. In particular, for each of the above bounds we freely gain $\ell^1$ dyadic summability of the above norms. One should think of all these cubic nonlinear expressions above as playing perturbative roles in the analysis. For details, see [31, Theorem 9.2].

2.2. Local well-posedness and energy dispersion. Our second paper [32] 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 $\epsilon$-energy concentration scale $r_\epsilon$ of a finite energy Yang–Mills initial data set $(a, e)$ to be

$$r_\epsilon = r_\epsilon(a, e) = \sup \{ r : \mathcal{E}(a, e) \leq \epsilon \}.$$

Then we have:

**Theorem 2.8 (Local well-posedness in caloric gauge, [32, Theorem 1.12]).** There exists a non-increasing function $\epsilon_* = \epsilon_*(\mathcal{E}, Q) > 0$ and a non-decreasing function $M_*\epsilon(a, e)$ such that, the Yang–Mills equation in caloric gauge is locally well-posed on the time interval $I = [-r_\epsilon, r_\epsilon]$ for initial data with energy $\mathcal{E}$ and initial caloric size $Q$. More precisely, the following statements hold.

a) (Regular data) Let $(a, b) \in T^{L^2}C$ be a smooth initial data set with energy $\mathcal{E}$ and initial caloric size $Q$. 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

$$T^{L^2}C \ni (a, b) \mapsto (A_{t,x}, \partial_t A_{t,x}) \in C(I, T^{L^2}C)$$

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

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

$$\|A_{t,x}\|_{S_1[I]} \leq M_*\epsilon(\mathcal{E}, Q).$$

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

$$\|A_{t,x} - A'_{t,x}\|_{S_\sigma[I]} \lesssim M_*\epsilon(\mathcal{E}, Q, \sigma) \| (a, f) - (a', f') \|_{\dot{H}^\sigma \times \dot{H}^{\sigma-1}}.$$
We remark that bounds for the auxiliary variables $A_0, \partial_0 A_0$ follow a-posteriori from the $S^1$ bound for $A_t$; see [32, Theorem 5.1] for such bounds.

In particular, if the energy of the initial data set is smaller than $\epsilon^*: = \min\{\epsilon^2(1), 1\}$, then the corresponding solution $A_{t,x}$ in caloric gauge exists globally and obeys the bound

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

Thus in particular this result also provides a caloric gauge version of the Coulomb gauge small data result in [24].

One downside of using either the Coulomb or caloric gauge is that causality is lost. To remedy this, in [32] we prove that the well-posedness result can also be transferred to the temporal gauge $A_0 = 0$:

**Theorem 2.9** ([32, 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 [32] 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 2.10** (Regularity of energy dispersed solutions [32, 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 2.8) to the Yang–Mills equation in caloric gauge on $I$ with energy $\mathcal{E}$ and that obeys

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

then it satisfies the a-priori bound

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

as well as

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

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 [33] 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 [33] 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 $D(B_R) = \{|x| + |t| < R\}$ or time sections thereof.

In terms of the initial data, in addition of the energy a key role is played by the $\epsilon$-energy concentration scale

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

as well as the outer concentration radius

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

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

**Proposition 2.11** ([33, 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 $\mathbb{R}^4 \setminus B_2 \setminus B_1$ 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 2.12** ([33, Theorem 1.4]). Given an initial data $(a, e)$ in $B_R$ with finite energy and $\epsilon$-energy concentration scale $r_C$, there exists a gauge equivalent initial data $(\tilde{a}, \tilde{e})$ in $B_R$ which satisfies the bound

$$\|\tilde{a}\|_{H^1 \cap L^4} \lesssim \epsilon, \frac{r_C}{R}. \quad (2.13)$$

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 2.13** ([33]). A finite energy connection $a$ is topologically trivial iff it admits a representation $a \in \check{H}^1$ in a suitable gauge.

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

**Theorem 2.14** (Good global gauge theorem [33, 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 $G$ and $b \in \check{H}^1$.

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

2.3.2. Finite energy solutions. A consequence of [24] and of the first two papers in the series [31, 32] 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 [33] 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 2.15 ([33, 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. blow-up) occurs. Thus, temporal solutions are also maximal solutions for the Yang–Mills equation.

Theorem 2.16 ([33, 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

\[
\lim_{t \to T_{\max}} \mathcal{E}_{C \cap \Sigma_t}(A) > \epsilon_0 \tag{2.14}
\]

The similar result holds for \(T_{\min}\).

b) The same result holds for data in a ball \(B_R\) and the solution in the corresponding domain of uniqueness \(C_R\).

We remark that vanishing of \(r_c\) follows from [33, Theorem 1.22], and existence of an energy concentration point follows by a standard argument; see, e.g., [28, 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 2.17 ([33, 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

\[
\|A\|_{L^\infty(\dot{H}^{1/2} \cap L^4)} + \|\partial_t A\|_{L^\infty L^2} + \|\partial^j A_j\|_{\ell^1 H^{1/2}} + \|A_0\|_{\ell^2 H^1} + \|\Box A_x\|_{L^2 H^{-1/2}} \lesssim E, r_c^4 \tag{2.15}
\]

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

The proof of this theorem requires a good gluing technique for local connections with suitable regularity. A simpler application of the same ideas also yields the following result, which state in a schematic form:

Theorem 2.18 ([33]). Let \(A\) be a rough connection in \(\mathbb{R}^5\) or in a contractible subset of \(\mathbb{R}^4\), which locally admits smooth gauge equivalent representations. Then there exists a global gauge equivalent smooth representation of \(A\).

Here, by rough we mean defined on a vector bundle with local transition maps \(O\) in the critical space \(W^{2, d}_{loc}\), with local gauge potentials in \(W^{1, d}_{loc}\) \((d \geq 4)\). Consideration of critical regularity solutions to Yang–Mills necessitates handling such rough objects; we refer to [33] for details.
3. Monotonicity formulas

3.1. Some notation. Here we introduce some notation for geometric subsets of $\mathbb{R}^{1+4}$ and $\mathbb{R}^4$. The forward light cone

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

will play a central role in this paper. For $t_0 \in \mathbb{R}$ and $I \subset \mathbb{R}$, we define

$$ C_I := \{(t, x) : t \in I, |x| \leq t \}, \quad \partial C_I := \{(t, x) : t \in I, |x| = t \}, $$

$$ S_{t_0} := \{(t, x) : t = t_0, |x| \leq t \}, \quad \partial S_{t_0} := \{(t, x) : t = t_0, |x| = t \}. $$

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

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

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

At each point $p = (t_0, x_0) \in \mathbb{R}^{1+4}$, we introduce the null pair

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

and also orthonormal vectors $\{e_a\}_{a=1, \ldots, 3}$ which are orthogonal to $L$ and $\underline{L}$. Observe that each $e_a$ 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 $\{L, \underline{L}, e_a\}$ as

$$ \alpha_a := F(L, e_a), \quad \underline{\alpha}_a := F(\underline{L}, e_a), \quad \varrho := \frac{1}{2} F(L, \underline{L}), \quad \sigma_{ab} := F(e_a, e_b). $$

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

$$ |\alpha|^2 := \sum_{a=1, \ldots, 3} \alpha_a^2, \quad |\underline{\alpha}|^2 := \sum_{a=1, \ldots, 3} (\underline{\alpha}_a)^2, \quad |\sigma|^2 := \sum_{1 \leq a < b \leq 3} \sigma_{ab}^2. $$

3.2. The energy momentum tensor and conservation laws. The energy-momentum tensor associated to a connection $A$ is

$$ T_{\alpha\beta}(A) = 2\langle F_\alpha^\gamma, F_{\beta\gamma} \rangle - \frac{1}{2} m_{\alpha\beta} \langle F_{\gamma\delta}, F^\gamma_{\delta} \rangle. \quad (3.1) $$

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

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

This is verified directly for smooth connections, and it then transfers to finite energy (YM) 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., $(X)\pi := \mathcal{L}_X m$. Using covariant derivatives, $(X)\pi$ also takes the form

$$ (X)\pi_{\alpha\beta} = \nabla_\alpha X_\beta + \nabla_\beta X_\alpha. $$
or expressed in coordinates

\[(X)^{\pi}_{\mu\nu} = X(m_{\mu\nu}) + \partial_\mu(X^\alpha)m_{\alpha\nu} + \partial_\nu(X^\alpha)m_{\alpha\mu}\] (3.3)

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

\[(X)P_\alpha(A) \coloneqq T_{\alpha\beta}(A)X^\beta,\]
\[(X)K(A) \coloneqq \frac{1}{2}T_{\alpha\beta}(A)(X)^{\pi}_{\alpha\beta}.\] (3.4)

Then by (3.2) we obtain

\[\nabla^\alpha((X)P_\alpha(A)) = (X)K(A).\] (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

\[\nabla^\alpha((T)P_\alpha(A)) = 0.\] (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

\[E_{(t)}(A) = \int (T)P_0(A) \, dx = \int \sum_{\alpha<\beta}|F_{\alpha\beta}|^2 \text{ is constant in } t.\]

In general, for \(U \subset \mathbb{R}^4\) we introduce the notation

\[E_{(t)\times U}(A) = \int_{(t)\times U} (T)P_0(A) \, dx.\]

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

\[\mathcal{F}_{\partial C_{[t_0,t_1]}}(A) = \frac{1}{2} \int_{\partial C_{[t_0,t_1]}}^{} (T)P_L(A)r^3 \, d\sigma_{\mathbb{S}^3}\]

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.,

\[\mathcal{F}_{\partial C_{[t_0,t_1]}}(A) = \mathcal{F}_{\partial C_{[t_0,t']}}(A) + \mathcal{F}_{\partial C_{[t',t_1]}}(A) \quad \text{for } t' \in [t_0, t_1].\] (3.7)

b) The energy-flux relation holds:

\[E_{S_{t_1}}(A) - E_{S_{t_0}}(A) = \mathcal{F}_{\partial C_{[t_0,t_1]}}(A).\] (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
\[
\lim_{t_1 \to 0} F_{\partial C(0, t_1)}(A) = 0,
\]
(3.9)
b) If \([\delta, \infty) \subseteq I\) for some \( \delta > 0 \), then we have
\[
\lim_{t_0, t_1 \to \infty} F_{\partial C(t_0, t_1)}(A) = 0.
\]
(3.10)

### 3.3. 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
\[
X_0 = \frac{1}{\rho_0} (t \partial_t + x \cdot \partial_x), \quad \rho_0 = \sqrt{t^2 - |x|^2}
\]
for \((t, x) \in C\). We also introduce the null coordinates
\[
u_0 = t - |x|, \quad v_0 = t + |x|.
\]

Straightforward computations (see [28]) lead to the relation
\[
(X) K(A) = \frac{2}{\rho_0} |\iota_{X_0} F|^2,
\]
(3.11)
where \( |\iota_{X_0} F|^2 = m^{\alpha \beta} \iota_{X_0} F_\alpha \iota_{X_0} F_\beta \). Of course \( m \) 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
\[
\nabla^\alpha (X_0 P_\alpha) = \frac{2}{\rho_0} |\iota_{X_0} F|^2 \geq 0
\]
(3.12)
where the interesting components of \((X_0) P\) are
\[
(X_0) P_L = (u_0 / v_0)^{-1/2} |\alpha|^2 + (v_0 / u_0)^{1/2} \left( |\varrho|^2 + |\sigma|^2 \right),
\]
(3.13)
\[
(X_0) P_L = (u_0 / v_0)^{1/2} |\alpha|^2 + (v_0 / u_0)^{-1/2} \left( |\varrho|^2 + |\sigma|^2 \right).
\]
(3.14)

All expressions above are singular on the cone, so we cannot directly integrate the relation (3.12) on sections of the cone \( C \). To remedy this, we will translate the field \( X_0 \) downward by \( \epsilon \). Defining
\[
\rho_\epsilon = \sqrt{(t + \epsilon)^2 + |x|^2}, \quad X_\epsilon = \frac{1}{\rho_\epsilon} ((t + \epsilon) \partial_t + x \cdot \partial_x)
\]
\[
u_\epsilon = t + \epsilon - |x|, \quad v_\epsilon = t + \epsilon + |x|,
\]

8Unless the flux is zero. This is in general not the case, instead we will work in settings where the flux is merely small.
we now have the shifted relations
\[ \nabla^a (X_F P_a) = \frac{2}{\rho_\varepsilon} |t_{X_F} F|^2 \]  
(3.15)

where
\[ (X_F) P_L = \left( \frac{v_\varepsilon}{u_\varepsilon} \right)^{\frac{1}{2}} |\alpha|^2 + \left( \frac{u_\varepsilon}{v_\varepsilon} \right)^{\frac{1}{2}} \left( |\rho|^2 + |\sigma|^2 \right), \]  
(3.16)
\[ (X_F) P_L = \left( \frac{u_\varepsilon}{v_\varepsilon} \right)^{\frac{1}{2}} |\alpha|^2 + \left( \frac{v_\varepsilon}{u_\varepsilon} \right)^{\frac{1}{2}} \left( |\rho|^2 + |\sigma|^2 \right). \]  
(3.17)

We also remind the reader since \( \partial_t = \frac{1}{2} (L + \bar{L}) \), we have
\[ (X_F) P_0 = \frac{1}{2} (X_F) P_L + (X_F) P_L \]  
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
\[ \mathcal{E}_{S_1}(A) \leq E, \quad \mathcal{F}_{\partial C_{[\varepsilon, 1]}}(A) \leq \varepsilon^{\frac{3}{2}} E. \]  
(3.18)

Then
\[ \int_{S_1} (X_F) P_0(A) \, dx + \int \int_{C_{[\varepsilon, 1]}} \frac{2}{\rho_\varepsilon} |t_{X_F} F|^2 \, dt \, dx \lesssim E, \]  
(3.19)
where the implicit constant is independent of \( \varepsilon, E \).

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
\[ \int_{S_1} (X_F) P_0(A) \, dx \leq \int_{S_{t_0}} (X_F) P_0(A) \, dx + C \left( \frac{1}{\delta_1/t_0} + \log(\delta_1/\delta_0) \right) E. \]  
(3.20)

The proofs of the last two propositions are similar to arguments in [28] and are omitted.

4. Compactness

Here we establish the following compactness result for caloric Yang–Mills connections which are nearly stationary:

**Theorem 4.1.** Let \( A^{(n)} \) be a sequence of finite energy Yang–Mills connections in \([-2, 2] \times \mathbb{R}^4\) which is locally uniformly bounded in the sense of (2.15). Assume that
\[ \lim_{n \to \infty} \| t_{X_F} F \|_{L^2([-2, 2] \times B_{2R}(0))} = 0 \]
where \( X \) is a smooth time-like vector field (i.e., \( m(X, X) < 0 \)). Then on a subsequence we have
\[ A^{(n)} \to A \quad \text{in} \quad H^1([-1, 1] \times B_R(0)) \]
where $A$ is a stationary Yang–Mills wave, $\iota_X F = 0$, with regularity

$$A_x \in \ell^1 H_{\text{loc}}^{\frac{3}{2}}, \quad \nabla A_0 \in \ell^1 L^2 H_{\text{loc}}^{\frac{3}{2}}.$$  

**Proof.** Let $\chi$ be a smooth cutoff supported in $[-2, 2] \times B_{2R}(0)$, which is identically 1 in $[-1, 1] \times B_R(0)$. We microlocally split the connections $\chi A^{(n)}$ into a high modulation part and a low modulation part

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

i.e., with a classical order zero multiplier $\eta$ which is supported in a small neighborhood of the null cone \{$\tau^2 = \xi^2$\} and which is identically 1 in a smaller neighborhood. We choose $\eta$ so that $\partial_0$ and $X^\alpha \partial_\alpha$ are microlocally elliptic on its support, which is possible since $\partial_0$ and $X^\alpha \partial_\alpha$ are time-like.

Since $\Box$ is microlocally elliptic in the support of $1 - \eta(\tau, \xi)$, the bound for the $\Box A$ component of the norms in (2.15) implies that the high modulation parts $A^{(n), hi}$ are uniformly bounded in $\ell^1 H^{\frac{3}{2}}$. The same happens with the $\nabla A_0$ component in its entirety in $\ell^1 L^2 H^{\frac{3}{2}}$. On a subsequence we get convergence in $H^1$ for $A_x$, respectively $\ell^1 L^2 H^{\frac{3}{2}}$ for $\nabla A_0$.

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

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

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

$$X^\alpha \partial_\alpha A_k - X^j \partial_j A_j = - (\iota_X F)_k + X^\alpha [A_\alpha, A_k] + X^0 \partial_k A_0$$

$$X^k \partial_k A_k = - (\iota_X F)_0 + X^j (\partial_j A_0 + [A_j, A_0])$$

Here the LHS can be viewed as a system in $A_k$ which is microlocally elliptic of order 1 on the null cone (or more precisely, on the support of $\chi(t, x)\eta(\tau, \xi)$), whereas the RHS is

$$O_{L^2}(1) + O_{\ell^1 H^{\frac{3}{2}}}(1)$$

Thus on a subsequence we get convergence in $H^1$ for $A^{(n), lo}$, with the limit in $\ell^1 H^{\frac{3}{2}}$. Now the convergence of $F$ in $L^2$ is easy to establish. \(\square\)

## 5. Regularity of stationary connections

Here we consider the local solutions provided as limiting connections in Theorem 4.1. These have the local regularity\(^9\)

$$A_x \in \ell^1 H_{\text{loc}}^{\frac{3}{2}}, \quad A_0 \in \ell^1 H_{\text{loc}}^{\frac{1}{2}}$$

and satisfy

$$\iota_V F = 0,$$  

(5.2)

where either

(i) $V$ is constant, time-like, or

(ii) $V = S$, scaling field, inside the null cone

We claim that

**Proposition 5.1.** Any hyperbolic Yang–Mills connection $A$ as above is smooth and either (i) stationary or (ii) self-similar (i.e., $\mathcal{L}_V A = 0$) after a suitable gauge transformation.

\(^9\)Here, $H^{1, \frac{3}{2}}$ is defined with the norm $\|\nabla (\cdot)\|_{L^2 H^{\frac{3}{2}}}$. 

-26
We remark that the regularity of the gauge transformation is implicit from the regularity of $A$.

**Proof.** Consider the first case. In spacetime we can use the relation $\iota_V F = 0$ to change the equations to elliptic, with the metric changed from $m$ to $e = m + CV \otimes V$ with $C \gg 1$. Then our connection $A$ is harmonic in $(\mathbb{R}^5, e)$. Since it has local regularity $H^{\frac{3}{2}}_{\text{loc}}$ which is the critical regularity in 5 dimensions, by Theorem 1.5 we conclude it is locally smooth in a suitable gauge, and thus globally smooth in a suitable gauge.

In the second case we recast the problem in hyperbolic coordinates, and use the relation $\iota_S F = 0$ to view the problem as an elliptic problem there. 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:

**Local Coulomb gauge.** We place the equations in a local Coulomb gauge with respect to the metric $e$,

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

Thus we seek a gauge transformation $O$ which solves

$$\partial^\alpha [O^{-1} A_\alpha O - O^{-1} O_\alpha] = 0,$$

or equivalently

$$D^\alpha O_{\alpha} = \partial^\alpha A_\alpha.$$ 

We combine this with the compatibility conditions

$$\partial_\alpha O_{\alpha} - \partial_\beta O_{\beta} = [O_{\alpha}, O_{\beta}]$$

to obtain a nonlinear div-curl system. To solve this system in a small ball $B$ we impose the Neumann type boundary conditions

$$O_{\alpha} \nu^\alpha = 0 \quad \text{on } \partial B.$$ 

The ball $B$ is chosen small enough with the property that

$$\|\nabla A\|_{L^2 L^2(\mathbb{R}_0^5)} \ll 1.$$ 

This is critical regularity, and the smallness guarantees solvability for the above system. We obtain a solution $O_{\alpha}$ with regularity

$$O_{\alpha} \in L^2 H^{\frac{3}{2}} \cap H^{1, \frac{3}{2}}.$$ 

(5.3)

**Elliptic regularity.** Here we examine the output connection $\tilde{A}$, which a-priori has the same regularity as $O_{\alpha}$ above. It solves the harmonic Yang–Mills equation

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

and the Coulomb gauge condition

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

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

$$\Delta_{\tilde{A}} \tilde{A}_\alpha = [\tilde{A}_\beta, \tilde{D}_\alpha \tilde{A}_\beta].$$
Since a-priori $\tilde{A}$ has critical regularity, by elliptic theory it follows that the solutions are smooth. By Theorem 2.18 we can patch these local Coulomb connections to obtain a global smooth one which is still gauge equivalent to the initial connection $A$.

The regularity of the gauge. Here we consider the gauge transformation $O$ from the initial connection $A$ to the smooth connection $\tilde{A}$. By construction it has the regularity as in (5.3). Here we use (5.1) to boost this regularity to

$$O;_{x,t} \in H^{1+\frac{1}{2}}.$$  (5.4)

The second part is immediate from the regularity of $A_0$ in (5.1). For the first part we use the divergence property to obtain

$$\partial^k O;_{i,k} \in H^2.$$  

We combine this with the curl system for $O;_{i,k}$ to obtain an elliptic system which we solve in $H^2$. Here we can freely use the bound

$$O;_{i,k} \in L^\infty H^1$$

which follows by the trace theorem from (5.3). The time derivative of $O;_{x}$ is then estimated in $H^\frac{1}{2}$ from the spacetime curl equations. \hfill \Box

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 nontrivial self-similar solutions inside the light cone $C$ which have finite energy.

*Proof.* We proceed in several steps.

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

$$C = \{ e^\rho y : \rho \in \mathbb{R}, y \in \mathbb{H}^4 \}.$$  

The Minkowski metric becomes

$$ds^2 = e^{2\rho}(-(d\rho)^2 + (dy)^2),$$

where $(dy)^2$ denotes the standard metric on the hyperbolic space $\mathbb{H}^4$ with sectional curvature $-1$. The curvature tensor $F$ expressed in this frame has components

$$(F_{yy}, F_{y\rho}).$$

The self-similarity condition $\iota_S F = 0$ becomes $\iota_{\partial_\rho} F = 0$, i.e.

$$F_{y\rho} = 0.$$  

By Theorem 5.1 our connection is smooth in a suitable gauge. We can further switch this smooth representation into the $\rho$-temporal gauge $A_\rho = 0$. Then our connection is still smooth, and also independent of $\rho$.

Thus we have reduced the problem to a problem on the hyperbolic space, where the four remaining curvature conditions read

$$D^i F_{ij} = 0 \quad \text{in} \quad \mathbb{H}^4.$$  

28
at fixed $\rho$, where the conformal factor $e^{2\rho}$ drops out. In other words, $A$ is now a harmonic connection in $\mathbb{H}^4$.

We represent $\mathbb{H}^4$ as the unit disc $\mathbb{D}^4$ in $\mathbb{R}^4$ with metric

$$(ds)^2 = \Omega^2(dy)^2, \quad \Omega = \frac{2}{1-y^2}. $$

Again the conformal factor drops out, and we obtain the elliptic Yang–Mills system in $\mathbb{D}^4$

$$D^i F_{ij} = 0.$$ with respect to the Euclidean metric.

We now move to the finite energy condition in hyperbolic coordinates. This computation is equivalent to that in [28, Section 7.2], and yields

$$\int_{\mathbb{D}^4} \frac{1 + r^2}{1 - r^2} |F|^2 dy < \infty. \quad (6.1)$$

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

$$\int_{\{ \frac{1}{2} < r < 1 \}} \frac{1}{1 - r^2} |\nabla A|^2 + \frac{1}{(1 - r^2)^3} |A|^2 dy < \infty. \quad (6.2)$$

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 $\tilde{A}$ outside the ball is also in $\tilde{H}^1$, and its curvature $\tilde{F}$ 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

$$\tilde{D}^k \tilde{F}_{kj} = 0.$$ By the elliptic regularity results in the previous section, the connection $\tilde{A}$ is gauge equivalent to a smooth connection $A^1$ in $\mathbb{R}^4$. To continue we write elliptic equations for $F^1$,

$$\Delta_{A^1} F^1 = [F^1, F^1]$$

Since $A^1$ is smooth and $F^1$ has compact support, by standard elliptic unique continuation results it follows that $F^1 = 0$. Thus the connection $A$ is trivial.

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

Step 2. We introduce $R_d = \{ x \in \mathbb{D} : 1 - 2d < |x| < 1 - d \}$ and $\tilde{R}_d = \{ x \in \mathbb{D} : 1 - 4d < |x| < 1 - \frac{d}{2} \}$. For $k \geq 0$ and $2 \leq p \leq \infty$, we claim that

$$\| D^{(k)} F \|_{L^p(R_d)} \lesssim d^{-\alpha + \frac{1}{2}} \| d^{-\frac{1}{2}} F \|_{L^2(\tilde{R}_d)}, \quad \alpha = k + 2 - \frac{4}{p}. \quad (6.3)$$

This is proved by applying Uhlenbeck’s lemma 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. Roughly speaking, the idea is to fix the gauge by specifying the conditions

$$A_r = 0 \text{ in } \mathbb{D}^4, \quad A(r, \Theta) \to 0 \text{ and } \text{div}_\Theta A(r, \Theta) \text{ as } r \to 1, \quad (6.4)$$
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., $A^0 = 0$ on $\mathbb{D}^4$ 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 $\|(1-r_n)^{-\frac{1}{2}}F^0(r_n, \cdot)|L^2_\Theta \to 0$. Viewing $A_\Theta(r_n, \cdot)$ as a connection 1-form on the unit sphere $\Theta \in S^3$ and applying Uhlenbeck’s lemma, we find gauge transformations $O^n = O^n(\Theta)$ on $S^3$ such that

$$A^n_r(r, \Theta) = 0, \ \div_\Theta A^n_\Theta(r_n, \Theta) = 0, \ |A^n_\Theta(r_n, \Theta)|_{L^2_\Theta} \lesssim |F(r_n, \cdot)|_{L^2_\Theta} \quad (6.5)$$

where

$$A^n = \mathrm{Ad}(O^n)A^0 - O^n_{;x},$$

viewing $O^n$ as defined on $\mathbb{D}^4$ by $O^n(r, \Theta) = O^n(\Theta)$. Thanks to smoothness of $A^0$ and (6.3) (elliptic regularity for $F$), note that $O^n \in C^\infty(S^3)$.

**Step 4.** Let

$$D^2 = \int_{\mathbb{D}^4} \frac{1}{1-r}|F|^2 \, dx, \ \epsilon^2_n = \frac{1}{1-r_n} \int_{S^3} |F|^2(r_n, \Theta) \, d\Theta.$$ 

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 $r_n > \frac{1}{2}$. For each fixed $\frac{1}{2} < r < r_n$, we regard $A^n_\Theta(r, \cdot)$ as a 1-form on $S^3$. For $p, q \in [1, \infty]$, we use the notation

$$\|g\|_{L^p(r \in I; L^q_\Theta)} = \|g(r, \Theta)\|_{L^q_\Theta} \|L^q_\Theta(r \in I).$$

Note that on the annulus $R_n = \{ x \in \mathbb{D}^4 : 1/2 < r < r_n \}$, we have the equivalence $\|g\|_{L^p(R_n)} \approx \|g\|_{L^p(1/2 < r < r_n; L^q_\Theta)}$ for any $1 \leq p \leq \infty$, where the constant is independent of $n$.

As $A^n_r = 0$ and $\partial_r A^n_\Theta = F^n_\Theta$, we immediately have

$$\int_{1/2}^{r_n} \int \frac{1}{1-r} |\partial_r A^n|^2 \, dr \, d\Theta \lesssim D^2. \quad (6.6)$$

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

$$\int_r^{r_n} \left( (1-r')^{\beta-1}|g| \right)^p dr' \lesssim_{\beta, p} \int_r^{r_n} \left( (1-r')^{\beta} |\partial_r g| \right)^p dr' + \left( (1-r_n)^{\beta-\frac{p-1}{p}} |g| \right)^p (r_n) \quad (6.7)$$

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

In our gauge, (6.7) implies

$$\int_{1/2}^{r_n} \int \frac{1}{(1-r)^3} |A^n|^2 \, dr \, d\Theta \lesssim D^2 + \frac{1}{(1-r_0)^2} \int |A^n|^2(r_0, \Theta) \, d\Theta, \quad (6.8)$$

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

**Step 5.** To complete the proof, in view of (6.6) and (6.8), it remains to establish

$$\limsup_{n \to \infty} \int_{1/2}^{r_n} \int \frac{1}{1-r^2} |\partial_r A^n|^2 dr \, d\Theta \lesssim D. \quad (6.9)$$
Once (6.9) is proved, then it is a routine matter to extract a limit \( O^n \to O \) in \( H^2(\mathbb{D}^4) \) such that \( A = \text{Ad}(O)A^0 - O \), obeys the desired vanishing condition (6.2).

As a first attempt to prove (6.9), note that we have control of \( \partial_\theta \partial_\phi A^n_\Theta = D^\phi F^n_{\Theta \phi} + O(A^n_\Theta, F^n_{\Theta \phi}) \) by (6.3). However, if we naively use the \( L^2 \) bound in (6.3), we encounter logarithmic divergence. To rectify this, we use an additional cancellation from the harmonic Yang–Mills equation.

The idea is to compute the div-curl system on \( S^3 \) satisfied by \( A^n_\Theta(r, \cdot) \). We compute

\[
\partial_r \text{div}_\Theta A^n_\Theta = \text{div}_\Theta F^n_{\Theta r} = \text{div}_{A^n_\Theta} F^n_{\Theta r} + O(A^n_\Theta, F^n_{\Theta r}) = O(A^n_\Theta, F^n_{\Theta r}).
\]

where \( \text{div}_{A^n_\Theta} F^n_{\Theta r} = 0 \) by the harmonic Yang–Mills equation. We furthermore note that

\[
\int_{1/2}^{r_n} \int \frac{1}{(1-r)^2} |A^n|^4 \, dr \, d\Theta \lesssim D^4 + \int \frac{1}{1-r_n} \int |A^n|^4 (r_n, \Theta) \, d\Theta \lesssim D^4 + (1-r_n) r_n^4. \tag{6.10}
\]

by (6.3) with \((k, p) = (0, 4)\), (6.7) with \( p = 4 \) and (6.5). Then thanks to (6.3) and (6.10), we have

\[
\limsup_{n \to \infty} \| \partial_r \text{div}_\Theta A^n_\Theta \|^2_{L^2_t(1/2 < r < r_n; L^2_\Theta)} \lesssim D^2.
\]

Recall from (6.5) that \( \text{div}_\Theta A^n_\Theta(r_n, \Theta) = 0 \). By (6.7), it follows that

\[
\limsup_{n \to \infty} \|(1-r)^{-1} \text{div}_\Theta A^n_\Theta \|^2_{L^2_t(1/2 < r < r_n; L^2_\Theta)} \lesssim D^2.
\]

On the other hand, by the schematic relation

\[
\text{curl}_\Theta A = F_{\Theta r} + O(A_\Theta, A_\Theta)
\]

and the bound (6.10), we have

\[
\limsup_{n \to \infty} \|(1-r)^{-\frac{1}{2}} \text{curl}_\Theta A^n_\Theta \|^2_{L^2_t(1/2 < r < r_n; L^2_\Theta)} \lesssim D + D^2.
\]

By the div-curl estimate on the 3-spheres, we obtain

\[
\limsup_{n \to \infty} \|(1-r)^{-\frac{1}{2}} \partial_\Theta A^n_\Theta \|^2_{L^2_t(1/2 < r < r_n; L^2_\Theta)} \lesssim D + D^2,
\]

which implies the desired bound (6.9).

\[\square\]

7. The Bubbling-Off Result

In this section we prove the bubbling off result in Theorem [7.1]. Much of the argument is similar to that in [28, Sections 8.3–8.6] (see also [38, Sections 6.5–6.8]), from which we borrow much of the results.

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

\[
\limsup_{n \to \infty} \| A^{(n)} - \tilde{A}^{(n)} \|_{\dot{H}^1 \cap L^4(S_t)} = 0 \tag{7.1}
\]

(2) Bounded energy in the cone.

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

31
(3) Decaying flux on \( \partial C \).
\[
\mathcal{F}_{[\varepsilon, 1]}(A^{(n)}) \leq \varepsilon \frac{1}{n},
\]  
(7.3)

(4) Time-like energy concentration at \( t = 1 \).
\[
\mathcal{E}_{C, \gamma \cap S_1}(A^{(n)}) \geq E_1 > 0
\]  
(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,

\[
\lim_{t \uparrow T} \mathcal{F}_{[t, T]}(A) = 0.
\]

respectively toward infinity in the non-scattering case,

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

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

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

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 \( \delta \) is small enough, depending on \( n \).

\[ \square \]

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

\[
\int_{C_{\gamma_2 \cap S_t}} (X_0) P^0(A^{(n)}) \, dx \geq E_2 \quad \text{for every } t \in [\varepsilon_n, 1].
\]  
(7.5)

This is a gauge independent property, which follows from the localized monotonicity formulas as in the (MKG) case, via Proposition 3.4; see [28, Proof of Lemma 8.10] for details.

At this point, we can freely replace \((X_0) P\) 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

\[
\lim_{n \to \infty} \sup_{t \in [1, T_n]} \| A^{(n)} - \tilde{A}^{(n)} \|_{H^1 \cap L^4(S_t)} = 0
\]  
(7.6)
(2) Bounded energy in the cone
\[ \mathcal{E}_{S_t}(A^{(n)}) \leq E(A), \quad \text{for every } t \in [1, T_n], \] (7.7)

(3) Nontrivial energy in a time-like region
\[ \mathcal{E}_{C_{\gamma} \cap S_t}(A^{(n)}) \geq E_2 \quad \text{for every } t \in [1, T_n], \] (7.8)

(4) Asymptotic self-similarity
\[ \int \int_K |\iota_{X_0} F^{(n)}|^2 \, dt \, dx \to 0 \quad \text{as } n \to \infty \] (7.9)

for every compact subset \( K \) of the interior of \( C_{[1, \infty)} \).

This is essentially rescaling and pigeonhole principle; see [28, Proof of Lemma 8.11].

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

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 \( \geq 1 \), respectively \( \geq 1/2 \), from the lateral boundary \( \partial C \).

We have the following lemma, which is basically [28, 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:
   \[ \mathcal{E}_{(t_n) \times B_{r_n}(x_n)}(A^{(n)}) = E_0, \] (7.10)
   \[ \sup_{x \in B_r(x_n)} \mathcal{E}_{(t_n) \times B_{r_n}(x)}(A^{(n)}) \leq E_0, \] (7.11)
   \[ \frac{1}{4r_n} \int_{t_n-2r_n}^{t_n+2r_n} \int_{B_r(x_n)} |\iota_{X_0} F^{(n)}|^2 \, dt \, dx \to 0 \quad \text{as } n \to \infty. \] (7.12)

2. **Uniform non-concentration of energy.** There exists \( 0 < r = r(j) < 1/4 \) such that the following bounds hold:
   \[ \mathcal{E}_{C_{\gamma} \cap S_t}(A^{(n)}) \geq E_2 \quad \text{for } t \in [2^j, 2^{j+1}), \] (7.13)
   \[ \sup_{(t, x) \in C_j} \mathcal{E}_{(t) \times B_r(x)}(A^{(n)}) \leq E_0, \] (7.14)
   \[ \int \int_{\tilde{C}_j} |\iota_{X_0} F^{(n)}|^2 \, dt \, dx \to 0 \quad \text{as } n \to \infty. \] (7.15)

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 \( j \).
The concentration scenario (i). Now we need to run a compactness argument. On a subsequence we can assume that
\[ \lim_{n \to \infty} \frac{x_n}{t} = v, \quad |v| < 1. \]

We denote \( V = (1, v) \), which is a forward 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
\[ \mathcal{E}_{[0]}(A^{(n)}) = E_0, \]
\[ \sup_{x \in B_{R_n}(0)} \mathcal{E}_{[0]}(A^{(n)}) \leq E_0, \]
\[ \int_{-1}^{1} \int_{B_{R_n}(0)} |\iota_V F^{(n)}|^2 \, dt \, dx \to 0 \quad \text{as } n \to \infty. \]

where \( R_n = r_n^{-1} r \to \infty \).

So far, we have only used energy considerations, which are gauge independent. For the next step, however, we need better regularity information, so we place the above solutions in a “good gauge”, as provided by Theorem 2.17. This allows us to assume that the connections \( A^{(n)} \) satisfy the bounds in (2.15), uniformly on bounded sets.

Now we apply the compactness result in Theorem 4.1 to conclude that on a subsequence we have local convergence to a finite energy Yang–Mills connection \( A \) with regularity \( A \in H_{lo}^{\frac{3}{2}}, \quad \partial^j A_j \in \ell^1 H_{lo}^{\frac{1}{2}}, \quad A_0 \in \ell^1 H_{lo}^{\frac{1}{2}} \)

so that
\[ A^{(n)} \to A \quad \text{in } H_{lo}^{1}([-1/2, 1/2] \times \mathbb{R}^4) = 0 \]
and in addition
\[ \iota_V F = 0 \quad \text{in } [-1/2, 1/2] \times \mathbb{R}^4 \]

In addition, by the regularity result in Proposition 5.1, the connection \( A \) is gauge equivalent to a smooth connection in the domain \([-1/2, 1/2] \times \mathbb{R}^4\). Thus the limiting connection \( A \) is smooth in \([-\frac{1}{2}, \frac{1}{2}] \times \mathbb{R}^4\) with the following properties:
(i) Nontriviality, \( E(A) > 0 \).
(ii) Stationarity \( \iota_V F = 0 \).

Placing \( A \) in the \( V \)-temporal gauge we can also make it stationary, \( \partial_V A = 0 \). Then \( A \) is a Lorentz transform of a nontrivial harmonic Yang–Mills connection \( Q \), namely \( A = L_v Q \).

Theorem is proved in this case.

The non-concentration scenario (ii). The argument is similar here. By Theorem 2.17 we can use gauge equivalent representations of the connections \( A^{(n)} \) which are in a “good gauge” as provided by Theorem 2.17 and thus are bounded in the sense of (2.15), uniformly on compact subsets of \( C_{1,\infty} \). There applying successively the compactness Theorem 4.1 and the regularity result in Proposition 5.1 in the limit we get a smooth nontrivial finite energy connection \( A \) in \( C_{1,\infty} \). Placing this in the \( X_0 \)-temporal gauge we obtain a smooth global self-similar solution with finite energy inside the light cone. The nontrivial energy of
A inside the cone insures that this limiting connection is nontrivial. But such connections do not exist by Theorem [6.1].

8. No null concentration

A key step in the transition from Theorem [1.7] to Theorems [1.8,1.11] 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 [7] 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

$$\mathcal{E}_{S_1^0}(A) + \mathcal{E}_{\{1\} \times \mathbb{R}^4 \setminus S_1}(A) \leq \epsilon_1,$$

and

$$\mathcal{E}_{S_1}(A) + \int_{S_1} (X_\epsilon) P_0(A) \, dx \leq E_3.$$

Given any $\epsilon, E_3 > 0$, for sufficiently small $\epsilon_1, \epsilon$ and $\gamma$ close enough to $1$ (depending only on $\epsilon, E$), there exists a caloric gauge representation of the connection $A$ so that

$$\|A(1)\|_{\dot{H}^1} \lesssim_{E_3} 1, \quad Q(A) \lesssim_{E_3} 1,$$

whereas

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

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

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

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

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. We state this in polar coordinates $(r, \Theta)$:

**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) $F_{\Theta \Theta}$ is small in $L^2$,

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

ii) $F_{r \Theta}$ is small outside an annulus,

$$\|F_{r \Theta}\|_{L^2(\{r_2 \leq |x| \leq r_1\})} \leq \epsilon.$$  

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

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

Assume that $\epsilon$ is sufficiently small,

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

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

$$\|A(1)\|_{\dot{H}^1} \lesssim E_3, \quad Q(A) \lesssim E_3,$$

whereas

$$\|A(1)\|_{L^4} \lesssim E^{\frac{3}{4}}.$$

(8.5)

(8.6)

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

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

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 8.4.** Let $A$ be a connection which satisfies the hypotheses of the proposition. Then there exists a gauge equivalent connection $B$ which has the following properties:

1. $B$ is bounded in $\dot{H}^1$,

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

2. $B$ is small away from the unit sphere,

$$\|B\|_{\dot{H}^1(\{\frac{3}{4} \leq |x| \leq 1\})} \lesssim E_3 \epsilon.$$

3. $B_r$ is small,

$$\|B_r\|_{\dot{H}^1} \lesssim E_3 \epsilon.$$

4. $B_\Theta$ has small angular derivatives,

$$\|\nabla_\Theta B_\Theta\|_{L^2} \lesssim E_3 \epsilon.$$

**Remark 8.5.** It follows from here by Bernstein’s inequality that

$$\|B\|_{L^4} \lesssim E^{\frac{3}{4}} \epsilon,$$

which in turn shows that in this gauge $B$ is energy dispersed.

One minor downside of the last lemma 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 the following Uhlenbeck lemma directly applies:

**Theorem 8.6 (Uhlenbeck [18]).** Let $A \in \dot{H}^1$ be a connection in a ball $B_R \subset \mathbb{R}^4$ that satisfies

$$\|F\|_{L^2(B_R)} < \epsilon_0.$$

(8.7)

Then there exists an admissible gauge transformation $O$, unique up to multiplication by a constant element of $G$, such that $B = \text{Ad}(O)A - O_{x^*}$ obeys

$$\partial^\ell B_\ell = 0 \text{ in } B_R, \quad x^\ell B_\ell = 0 \text{ on } \partial B_R$$

and

$$\|B\|_{\dot{H}^1(B_R)} \lesssim \|F\|_{L^2(B_R)}.$$

Thus, after standard partitioning and regluing operations, the problem reduces to the simpler case when we work in an annulus:
Lemma 8.7. Let $A \in \dot{H}^1$ be a connection in the annulus $A_{[1,2]}$ with energy at most $E_3$, which has the following properties:

i) Small tangential curvature,
\[ \|F_{\Theta\Theta}\|_{L^2} \leq \epsilon. \]

ii) Divergence property,
\[ \|D^\Theta F_{\Theta r}\|_{\dot{H}^{-1}} \leq \epsilon. \]

If $\epsilon \ll E_3$ then there is a gauge equivalent connection $B$ in the Coulomb gauge with the following properties:

a) Bounded size,
\[ \|B\|_{\dot{H}^1} \lesssim E_3. \]

b) Small components:
\[ \|B_r\|_{\dot{H}^1} + \|\partial_\Theta B_{\Theta}\|_{L^2} \lesssim E_3 \epsilon. \]

Remark 8.8. As a corollary of the lemma, we have the bound
\[ \|B_r\|_{L^1} \lesssim E_3 \epsilon, \quad \|B'_{\Theta}\|_{L^4} \lesssim E_3 \epsilon^{\frac{3}{4}}. \] (8.8)

Conversely, if such a bound holds then the $\dot{H}^1$ type bounds follow easily by solving a linear div-curl system.

We now successively prove the above results in reverse order:

Proof of Lemma 8.7. We use the metric
\[ ds^2 = dr^2 + d\Theta^2 \]
in the annulus, and seek the connection $B$ so that it satisfies the Coulomb gauge condition with respect to this metric,
\[ \nabla^k B_k = 0, \]
and with boundary condition
\[ B_r = 0 \quad \text{on } \partial A_{[1,2]}. \]
We claim that such a connection exists, and satisfies the conclusion of the Lemma.

We use an elliptic bootstrap argument. Suppose that we have a continuous one parameter family of connections
\[ A^{(h)} \in (\dot{H}^1 \cap L^4)(A_{[1,2]}), \quad h \in [0,1], \quad A^{(0)} = 0, \quad A^{(1)} = A \]
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
\[ \|B^{(h)}\|_{L^1} \leq C\epsilon^{\frac{3}{4}}. \] (8.9)

We will establish that the set
\[ H = \{ h \in [0,1]; \text{ [8.9] holds} \}, \]

\[ ^{10} \text{Here we reserve the right to choose the metric favorably.} \]
which trivially contains 0, is both open and closed, and thus contain $h = 1$.

We begin by exploring further the properties of the connections $B^{(h)}$ for $h \in H$, in order to prove some a-priori bounds. First we can write an elliptic equation for $B_r$ in $A_{[1,2]}$,

$$\Delta B_r + [B^\Theta, D_\Theta B_r] = G := D^\Theta F_\Theta + D^\Theta [B_r, B_\Theta]$$

with Dirichlet boundary condition. Here the RHS $G$ has size $\epsilon$ in $H^{-1}$ because of the hypothesis and the bootstrap assumption. Also the coefficient $B^\Theta$ on the left is small in $L^4$. Hence the elliptic problem is uniquely solvable, and the solution $B_r$ satisfies

$$\|B_r\|_{H^1} \lesssim \epsilon.$$

For $B_\Theta$ on the other hand we have a linear (angular) div-curl system

$$\nabla^a B_a = -\nabla_r B_r, \quad \nabla_a B_b - \nabla_b B_a = F_{ab} - [B_a, B_b],$$

where both RHS terms are $O(\epsilon)$ in $L^2(A_{[1,2]})$. Solving this div-curl system on the spheres, we obtain the bound

$$\|B_\Theta\|_{L^2 H_b^2} \lesssim \epsilon.$$

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

$$\|\nabla_r B_\Theta\|_{L^2} \lesssim 1.$$

Thus, we have proved that the conclusion of the lemma holds for $h \in H$.

Furthermore, by Bernstein’s inequality we can now improve the bootstrap assumption to

$$\|B\|_{L^4} \lesssim \epsilon^3. \quad (8.10)$$

Secondly, 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, $\nabla^k \tilde{A}_k$ is small. Then to find a gauge transformation $O$ which takes $\tilde{A}$ into the Coulomb gauge, we end up having to solve for $C_k = O^{-1} \partial_k O$ the elliptic system

$$D^k_A C_k = \partial^k \tilde{A}_k, \quad \partial_j C_k - \partial_k C_j = -[C_j, C_k].$$

Here we need a boundary condition $C_r = -\tilde{A}_r$ on the boundary. Everything is small here including the $A_k$ in $D_k$, so this elliptic problem can be solved. Then we obtain a solution $O$ with $O^{-1} \partial_x O = C_k$, that also satisfies

$$\|O_{,x}\|_{H^1} \ll 1.$$

The a-priori bound shows that if $h \in H$ then the stronger bound $(8.10)$ holds. Then the perturbative argument shows that for $h \in H$ there exists a fixed size neighborhood $[h - \delta, h + \delta]$ which is in $H$. We conclude that $H = [0, 1]$, which completes the continuity argument.

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\textsuperscript{11}

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

This can be done as follows:

1. In a suitable gauge set $A_r = 0$ on the boundary; see [33, Lemma 3.5].
2. Double the annulus inside, and extend the connection as odd for $A_r$ and even for $A_\theta$. This extension is still $H^1$, and the smallness hypothesis still holds in the double annulus; see [33, Lemma 3.18].
3. Choose a sphere $S$ within the extended part on which $F_\Theta$ is small in $L^2$ (this is Uhlenbeck’s lemma [48] on the sphere). In a suitable gauge (say Coulomb) we can set $A_\Theta \in H^1(S)$, small, and $A_r = 0$ on $S$ (again use [33, Lemma 3.5]).
4. Choose an extension of $A$ inside $S$ which is small in $H^1$. Note that this task is made straightforward by the smallness of the subcritical norm $\|A_\Theta\|_{H^1(S)}$.

□

Lemma 8.7 $\implies$ Lemma 8.4. By Uhlenbeck’s lemma (Theorem 8.6) we obtain a gauge equivalent connection $A_{\text{in}}$ in $B_{\frac{1}{4}}(0)$ which is $\epsilon$-small in $H^1$. Then again by Uhlenbeck’s lemma and conformal inversion, we obtain a gauge equivalent connection $A_{\text{out}}$ in $B_1(0)^c$ which is $\epsilon$-small in $H^1$. By Lemma 8.7 we obtain a gauge equivalent connection $A_{\text{mid}}$ in the annulus $A_{\frac{2}{3}, \frac{4}{3}}$. 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, and hence $O_{\neq}x$ is small in $\ell^1 H^1$. Thus each $O$ is uniformly closed to a constant, and a standard patching argument (see, e.g., [48, Proposition 3.2]) now works.

□

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

$$\|F\|_{L^{3,4}_{2,4}} \leq 1.$$*

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 that we have a gauge transformation $O$ with

$$\|O_{\neq}x\|_{H^1} \lesssim_{E_3} 1.$$

\textsuperscript{11}As in the original proof of Uhlenbeck’s lemma in [48].
which transforms $A$ into its caloric representation $\tilde{A}$. In turn $\tilde{A}$ must also satisfy
\[ \|A\|_{H^1} \lesssim E_3. \]
Further, since $A$ was small in $L^4$, it curvature is small in $\dot{W}^{-1,4}$,
\[ \|F\|_{\dot{W}^{-1,4}} \lesssim E_3 \epsilon^3. \]
By this and the bound for $O_{ix}$, the curvature of $\tilde{A}$, namely $\tilde{F} = OFO^{-1}$, must also be small,
\[ \|\tilde{F}\|_{\dot{W}^{-1,4}} \lesssim E_3 \epsilon^3. \]
Propagating this bound along the caloric flow [31, Proposition 8.9], we obtain
\[ \|F\|_{L^2_{x,t}} \lesssim E_3 \epsilon^4, \]
which improves the bootstrap assumption. This completes the proof of the proposition. $\square$

**Proposition 8.3 $\implies$ Theorem 8.1.** We express the curvature components in the null frame. Thanks to (8.1) and (8.2), for sufficiently small $\epsilon_1, \epsilon$ and $1 - \gamma$ the null components $\alpha, \varrho$ and $\sigma$ are already small in $L^2$. We now use the constraint equation to express
\[ D^\theta F_{\theta r} = D^\theta F_{\theta 0} + D^\theta F_{\theta L} = -D^r F_{r 0} + D^\theta F_{\theta L} = -D^r \rho + D^\theta \alpha_{\theta} \] (8.11)
which implies the smallness in the gauge-invariant space $H^{-1}_A$. Thus, we have established
that the spatial part of the connection $A$ satisfies the hypotheses of Proposition 8.3.

Suppose now that $A$ is in the caloric gauge and satisfies the bounds in Proposition 8.3.

It remains to consider the temporal components of $F$. For $F_{0\theta}$ we write
\[ F_{0\theta} = F_{L\theta} + F_{r\theta} \]
and the $\dot{W}^{-1,4}$ smallness follows. For $F_{0r}$ we simply have
\[ F_{0r} = F_{Lr} \]
which is small even in $L^2$. $\square$

**9. PROOF OF THE THRESHOLD THEOREM AND THE DICHOTOMY THEOREM**

For both theorems, we argue by contradiction. Suppose that the conclusion of the Dichotomy 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 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 $E(A) < 2E_{GS}$ 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 [25] (see also [31, 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 $L_0Q$. This
implies (spacetime) $L^2_{\text{loc}}$ convergence of curvature tensors $F^{(n)}$, and thus for almost every $t$
(and possibly passing to a subsequence)
\[ \mathcal{E}_{(t)\times B_R}(A^{(n)}) = \frac{1}{2} \int_{B_R} \langle F^{(n)}, F^{(n)} \rangle(t) \to \mathcal{E}_{(t)\times B_R}(L_0Q) \quad \text{for any } R > 0, \]
which in turn implies
\[ \mathcal{E}(Q) \leq \mathcal{E}(A) < 2E_{GS}. \]
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, $\chi(L_v(Q)) = \chi(Q)$. From here on, we assume that
$\chi(Q) > 0$; the alternative case is similar.
Fix a large number $R \gg 1$ and $t \in [-\frac{1}{2}, \frac{1}{2}]$. By topological triviality of $A^{(n)}(t)$, we have
\[ 0 = \chi(A^{(n)}(t)) = \int_{B_R(0)} -\langle F^{(n)} \wedge F^{(n)} \rangle(t) + \int_{\mathbb{R}^4 \setminus B_R(0)} -\langle F^{(n)} \wedge F^{(n)} \rangle(t). \]
Again by the (spacetime) $L^2_{\text{loc}}$ convergence of $F^{(n)}$, there exists a subsequence such that
\[ \int_{\mathbb{R}^4 \setminus B_R(0)} \langle F^{(n)} \wedge F^{(n)} \rangle(t) = \int_{B_R(0)} -\langle F^{(n)} \wedge F^{(n)} \rangle(t) \to \int_{B_R} -\langle F[L_0Q] \wedge F[L_0Q] \rangle. \]
By (1.5), we have
\[ \mathcal{E}(A) \geq \limsup_{n \to \infty} \left( \frac{1}{2} \int_{B_R} \langle F^{(n)}, F^{(n)} \rangle(t) + \int_{\mathbb{R}^4 \setminus B_R} \langle F^{(n)} \wedge F^{(n)} \rangle(t) \right) \]
\[ \geq \mathcal{E}_{(t)\times B_R}(L_0Q) + \int_{B_R} -\langle F[L_0Q] \wedge F[L_0Q] \rangle. \]
Sending $R \to \infty$, the RHS tends to $\mathcal{E}(L_0Q) + \chi(L_0Q) \geq 2\mathcal{E}(Q) \geq 2E_{GS}$, which is a contradic-
tion.
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 2.8. However this makes no difference in the proofs below.

9.1. The blow-up case. Let $E = \mathcal{E}(A)$. If $[0, T]$ is a maximal existence time, then the
temporal gauge local well-posedness result (Theorem 2.16) 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
\[ \lim_{t \searrow 0} \mathcal{E}_S(A) > E_0, \quad (9.1) \]
where $E_0$ is a universal constant corresponding to the small data result.
We also know that the hypothesis of Theorem 1.7 is false, which gives
\[ \lim_{t \searrow 0} \mathcal{E}_S^\gamma(A) = 0, \quad 0 < \gamma < 1. \quad (9.2) \]
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 2.10 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 $\tilde{\mathbf{A}}$ in $(0, t_\varepsilon]$ with the following properties:

1. **Gauge equivalence:** $\tilde{\mathbf{A}}$ is gauge equivalent to $\mathbf{A}$ in $C(0, t_\varepsilon]$.
2. **Small energy outside the cone**
   \[
   \mathcal{E}_{\{(t) \times \mathbb{R}^4 \setminus S_t\}}(\tilde{\mathbf{A}}) \leq \varepsilon^8 E \quad \text{for every } t \in (0, t_\varepsilon],
   \] (9.3)
3. **Small flux on $\partial C$**
   \[
   \mathcal{F}_{\{(0, t)\}}(\tilde{\mathbf{A}}) \leq \varepsilon^9 E.
   \] (9.4)

**Proof.** The flux energy relation (3.8) implies that the flux decays to zero near the tip of the cone,
\[
\lim_{t \to 0} \mathcal{F}_{\{(0, t)\}}(\mathbf{A}) = 0.
\]
so we first choose $t_\varepsilon$ small enough so that the last condition is satisfied for $\mathbf{A}$. We then choose $\delta > 0$ so that the energy of $A$ in a $\delta$-annulus around the cone is small,
\[
\mathcal{E}_{\{t_\varepsilon \times \{t < |x| < t_\varepsilon + 3\delta\}\}}(\mathbf{A}) \leq \varepsilon^9 E.
\]
Again by the flux energy relation, this propagates to all smaller times,
\[
\mathcal{E}_{\{t \times \{t < |x| < t + 3\delta\}\}}(\mathbf{A}) \leq 2\varepsilon^9 E, \quad 0 < t < t_\varepsilon.
\]
We now reset $t_\varepsilon$ to a smaller value,
\[
t_\varepsilon \to \min\{t_\varepsilon, \delta\}.
\]
For this new choice we have
\[
\mathcal{E}_{\{t_\varepsilon \times \{t < |x| < 4t_\varepsilon\}\}}(\mathbf{A}) \leq 2\varepsilon^9 E, \quad 0 < t < t_\varepsilon.
\]
By (a rescaled form of) Proposition 2.11 we can truncate the data $(a, e)(t_\varepsilon)$ for $\mathbf{A}$ at time $t_\varepsilon$. 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,
\[
\mathcal{E}_{\{t_\varepsilon \times \{t < |x| < 4t_\varepsilon\}\}}(\tilde{\mathbf{A}}) \lesssim \varepsilon^9 E.
\]
Now we consider the solution $\tilde{\mathbf{A}}$ generated by the truncated data $(\tilde{a}, \tilde{e})(t_\varepsilon)$ below time $t_\varepsilon$. For as long as it exists it is gauge equivalent with $\mathbf{A}$ inside the cone, since they are gauge equivalent initially, see Theorem 2.16. 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 $\tilde{\mathbf{A}}$ persists as a finite energy solution and satisfies the conditions in the lemma up to time $t = 0$. \qed
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 $\varepsilon \ll E$.

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

$$
\int_{S_t} (x_{vt}) P_0(\tilde{A}) \, dx \lesssim E. \tag{9.5}
$$

The bounds (9.1), (9.2) and (9.3) together with (9.5) with small enough $\varepsilon$ 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 2.10 to conclude that the solution can be extended beyond the blow-up time $T = 0$, which is a contradiction.

9.2. The non-scattering scenario. This is similar but simpler, as we no longer need to truncate the data for $A$ and instead is to 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 $A$ all large enough $t$. This shows that the connection $A$ admits a caloric representation in $[t, \infty]$, which is also energy dispersed. Thus we can apply the energy dispersed result in Theorem 2.10 in $[t, \infty)$ and conclude that $A \in S^1([t, \infty))$, i.e., the desired scattering result.

References

[1] 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.

[2] I. Bejenaru and S. Herr, The cubic Dirac equation: small initial data in $H^{1/2}(\mathbb{R}^2)$, Comm. Math. Phys. 343 (2016), no. 2, 515–562, doi.

[3] , 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.

[4] J. Bourgain, Global solutions of nonlinear Schrödinger equations, American Mathematical Society Colloquium Publications, vol. 46, American Mathematical Society, Providence, RI, 1999, doi.

[5] 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.

[6] 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.

[7] 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:1512.04927.

[8] 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.

[9] , 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.

[10] C. Gavrus, Global well-posedness for the massive Maxwell-Klein-Gordon equation with small critical Sobolev data, preprint (2016), arXiv:1510.03581.
[11] 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
[12] 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
[13] R. Grinis, Quantization of time-like energy for wave maps into spheres, Comm. Math. Phys. 352 (2017), no. 2, 641–702. doi
[14] M. Gursky, C. Kelleher, and J. Streets, A conformally invariant gap theorem in Yang-Mills theory, preprint (2017), arXiv:1708.01157
[15] 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
[16] 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
[17] ———, Finite energy solutions of the Yang-Mills equations in $\mathbb{R}^{3+1}$, Ann. of Math. (2) 142 (1995), no. 1, 39–119. doi
[18] S. Klainerman and D. Tataru, On the optimal local regularity for Yang-Mills equations in $\mathbb{R}^{4+1}$, J. Amer. Math. Soc. 12 (1999), no. 1, 93–116. doi
[19] 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
[20] J. Krieger and J. Lihrmann, Concentration compactness for the critical Maxwell-Klein-Gordon equation, Ann. PDE 1 (2015), no. 1, Art. 5, 208.
[21] J. Krieger and W. Schlag, Concentration compactness for critical wave maps, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2012. doi
[22] 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
[23] 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
[24] 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.
[25] 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
[26] 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
[27] ———, 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
[28] 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
[29] ———, 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
[30] ———, Energy dispersed solutions for the $(4 + 1)$-dimensional Maxwell-Klein-Gordon equation, Amer. J. Math. (2017), arXiv:1503.01561
[31] ———, The Yang-Mills heat flow and the caloric gauge, preprint (2017).
[32] ———, The hyperbolic Yang-Mills equation in the caloric gauge. Local well-posedness and control of energy dispersed solutions, preprint (2017).
[33] ———, The hyperbolic Yang-Mills equation for connections in an arbitrary topological class, preprint (2017).
[34] ———, The Threshold Theorem for the $(4 + 1)$-dimensional Yang-Mills equation: An overview of the proof, preprint (2017).
[35] 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
[36] 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
[37] 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
[38] ______, Regularity of wave-maps in dimension $2 + 1$, Comm. Math. Phys. 298 (2010), no. 1, 231–264, doi
[39] T. Tao, Global regularity of wave maps. II. Small energy in two dimensions, Comm. Math. Phys. 224 (2001), no. 2, 443–544, doi
[40] ______, Geometric renormalization of large energy wave maps, Journées “Équations aux Dérivées Partielles”, École Polytech., Palaiseau, 2004, pp. Exp. No. XI, 32.
[41] ______, Global regularity of wave maps III. Large energy from $\beta R^{1+2}$ to hyperbolic spaces, preprint (2008), arXiv:0805.4668
[42] ______, Global regularity of wave maps IV. Absence of stationary or self-similar solutions in the energy class, preprint (2008), arXiv:0806.3592
[43] ______, Global regularity of wave maps V. Large data local wellposedness and perturbation theory in the energy class, preprint (2008), arXiv:0808.0368
[44] ______, Global regularity of wave maps VI. Abstract theory of minimal-energy blowup solutions, preprint (2009), arXiv:0906.2833
[45] ______, Global regularity of wave maps VII. Control of delocalised or dispersed solutions, preprint (2009), arXiv:0908.0776
[46] D. Tataru, On global existence and scattering for the wave maps equation, Amer. J. Math. 123 (2001), no. 1, 37–77, link.
[47] ______, Rough solutions for the wave maps equation, Amer. J. Math. 127 (2005), no. 2, 293–377, link.
[48] K. K. Uhlenbeck, Connections with $L^p$ bounds on curvature, Comm. Math. Phys. 83 (1982), no. 1, 31–42, link.
[49] ______, Removable singularities in Yang-Mills fields, Comm. Math. Phys. 83 (1982), no. 1, 11–29, link.

KIAS, SEOUL, KOREA 02455
E-mail address: sjob@kias.re.kr

DEPARTMENT OF MATHEMATICS, UC BERKELEY, BERKELEY, CA 94720
E-mail address: tataru@math.berkeley.edu