THE YANG–MILLS MASS GAP ON THE LATTICE

LLUIS ERIKSSON
INDEPENDENT RESEARCHER
LLUISERIKSSON@GMAIL.COM

Abstract. We prove that $SU(N_c)$ lattice Yang–Mills theory in $d = 4$ dimensions with Wilson action has a positive mass gap

$$m_{\text{gap}} \geq c(N_c) \cdot e^{-C(N_c)/g^2} > 0$$

in lattice units, uniformly in the lattice size L, for g^2 sufficiently small, assuming the compatibility of Balaban's renormalization group with the temporal transfer matrix factorization. The proof combines three ingredients: (i) Balaban's constructive renormalization group, which produces a controlled effective action at a terminal scale where $g^2 = O(1)$ and the coarsened lattice has $O(1)$ sites; (ii) the Holley–Stroock spectral gap bound from Paper III of this series, applied at the terminal scale using the orbit-space Ricci curvature $\text{Ric}_{\mathcal{B}} \geq N_c/4$ from Paper II; and (iii) a transfer-matrix trace identity showing that the physical mass gap is scale-invariant under the renormalization group. We identify the compatibility assumption as the single remaining caveat and discuss how an anisotropic variant of Balaban's RG would resolve it.

1. Introduction and main result

This is the fourth and final paper in a series [1, 2, 3] studying the Yang–Mills mass gap via the geometry of the gauge orbit space.

Theorem 1.1 (Yang–Mills lattice mass gap)**.** *For $SU(N_c)$ lattice Yang–Mills theory on $\Lambda = (\mathbb{Z}/L\mathbb{Z})^4$ with Wilson action at coupling $g^2 \leq g_0^2$ (sufficiently small), assume that Balaban's partition function identity is compatible with the temporal transfer matrix factorization (see Theorem 4.1). Then the mass gap satisfies*

$$m_{\text{gap}} \geq c(N_c) \cdot e^{-C(N_c)/g^2} > 0 \tag{1}$$

uniformly in $L_0 \leq L \leq C_0\, e^{C(N_c)/g^2}$, where $C(N_c) = 24\pi^2 \ln 2/(11 N_c)$ and $c(N_c) > 0$ is a constant depending only on N_c and the universal constants in Balaban's construction.

The proof has three steps:

1. **Balaban's RG to the terminal scale** (Section 2). The constructive renormalization group of Balaban [4] produces an effective action on a coarsened lattice with $O(1)$ sites per direction and coupling $g^2_{n_{\max}} = \gamma$.
2. **Spectral gap at the terminal scale** (Section 3). The Holley–Stroock bound from [3, Theorem 5.3], applied to the effective action at the terminal scale using $\text{Ric}_{\mathcal{B}} \geq N_c/4$ from [2, Theorem 3.1], gives $m_{n_{\max}} \geq c_0 > 0$.

Date: February 2026.

2020 *Mathematics Subject Classification.* 81T13, 81T25, 58J50, 82B28.

Key words and phrases. Yang–Mills theory, mass gap, lattice gauge theory, renormalization group, transfer matrix, Bakry–Émery theory.

(3) **RG invariance of the mass gap** (Section 4). A transfer-matrix trace identity, following from Balaban's exact preservation of the partition function and its compatibility with the temporal factorization, gives $m_{n+1} = 2m_n$ and hence $m_0 = m_{n_{\max}}/2^{n_{\max}}$.

2. Balaban's constructive RG

We use the following result of Balaban.

Theorem 2.1 (Balaban [4, Theorem 1])**.** *For $SU(N_c)$ lattice gauge theory on $\Lambda = (\mathbb{Z}/L\mathbb{Z})^4$ with $\beta = 2N_c/g^2 \geq \beta_0$: there exist effective actions S_k on coarsened lattices $\Lambda_k = (\mathbb{Z}/(L/2^k)\mathbb{Z})^4$, $0 \leq k \leq n_{\max}$, such that:*

(a) $Z(\Lambda, S_0) = Z(\Lambda_k, S_k)$ *for all* k.
(b) $S_k = \beta_k S_W + \sum_X \epsilon_k(X)$ *with* $|\epsilon_k(X)| \leq e^{-\kappa|X|/a_k}$, $\kappa > 0$ *universal.*
(c) $g_k^2 = 2N_c/\beta_k \leq \gamma$ *for* $k \leq n_{\max}$, *where* $\gamma > 0$ *is a small universal constant and* $\beta_k = \beta + 2b_0 k \ln 2 + O(1/\beta)$.

Remark 2.2. At the terminal scale $k = n_{\max}$: the coarsened lattice $\Lambda_{n_{\max}}$ has $L_{n_{\max}} = L/2^{n_{\max}}$ sites per direction. For L a sufficiently large multiple of $2^{n_{\max}}$: $L_{n_{\max}} \geq 2$. By choosing L to be a small multiple of $2^{n_{\max}}$, we may arrange $L_{n_{\max}} = L/2^{n_{\max}} \leq C_0$ for a fixed constant C_0 independent of L.

3. Spectral gap at the terminal scale

Theorem 3.1 (Gap at the terminal scale)**.** *The spectral gap of the transfer matrix $T_{n_{\max}}$ associated with the effective action $S_{n_{\max}}$ satisfies*

$$m_{n_{\max}} \geq c_0(N_c, \gamma, \kappa) > 0. \tag{2}$$

Proof. We apply the argument of [3, Theorem 5.3] to the effective theory at scale $n_{\max}$.

Step 1: Orbit space geometry. The orbit space $\mathcal{B}_{n_{\max}} = \mathcal{A}_{n_{\max}}/\mathcal{G}_{n_{\max}}$ of the coarsened lattice has:

- Dimension $n = 3L_{n_{\max}}^3(N_c^2 - 1) = O(1)$ (since the spatial slice is three-dimensional and $L_{n_{\max}} = O(1)$).
- Ricci curvature $\mathrm{Ric}_{\mathcal{B}} \geq N_c/4$ by [2, Theorem 3.1]. This bound is intrinsic to the orbit space and independent of the effective action.
- Compact (quotient of a product of compact groups by a compact group).

Step 2: The Gibbs measure. The effective Hamiltonian at scale $n_{\max}$ defines a Gibbs measure on $\mathcal{B}_{n_{\max}}$:

$$d\mu = e^{-f}\, d\,\mathrm{Vol}_{\mathcal{B}}/Z, \qquad f = \frac{2V_{\text{eff}}}{g_{n_{\max}}^2},$$

where $V_{\text{eff}} = \beta_{n_{\max}} V_W + \sum_X \epsilon_{n_{\max}}(X)/(2/g_{n_{\max}}^2)$ is the effective potential.

The oscillation of f is bounded:

$$\mathrm{osc}(f) = \sup f - \inf f \leq \frac{2}{\gamma}\Big(\beta_{n_{\max}} \max V_W + \sum_X |\epsilon_{n_{\max}}(X)|\Big).$$

Since $V_W \leq$ (number of plaquettes) $\cdot\, 2/N_c = O(L_{n_{\max}}^4) = O(1)$ and $\sum_X |\epsilon_{n_{\max}}(X)| \leq C_\epsilon(\kappa, d) < \infty$ (finite sum of exponentially decaying terms on a lattice of $O(1)$ sites):

$$\mathrm{osc}(f) \leq \frac{K}{\gamma}$$

where K depends only on N_c, κ, and d.

Step 3: Holley–Stroock bound. By the Holley–Stroock bounded perturbation lemma [5]: if $d\mu = e^{-f} d\nu/Z$ with the uniform measure ν, then the log-Sobolev constant satisfies $\alpha(\mu) \geq \alpha(\nu) \cdot e^{-\mathrm{osc}(f)}$.

Since $\mathrm{Ric}_{\mathcal{B}} \geq N_c/4$: the Bakry–Émery criterion [6] gives $\alpha(\nu) \geq N_c/2$ for the uniform measure. Therefore:

$$\alpha(\mu) \geq \frac{N_c}{2}\, e^{-K/\gamma}.$$

The spectral gap of the Laplacian with respect to μ satisfies $\lambda_1 \geq \alpha(\mu)/2 \geq (N_c/4)\, e^{-K/\gamma}$.

The transfer matrix gap is $m_{n_{\max}} = (g^2_{n_{\max}}/2)\lambda_1 \geq (\gamma N_c/8)\, e^{-K/\gamma} =: c_0(N_c, \gamma, \kappa) > 0$. □

Remark 3.2. The constant c_0 is positive but potentially extremely small (exponentially small in $1/\gamma$). This is a mathematical lower bound, not a physical prediction. The actual mass gap is expected to be $O(\Lambda_{\mathrm{QCD}})$, vastly larger than this bound.

4. RG invariance of the mass gap

Theorem 4.1 (Spectral gap scaling)**.** *Assume that Balaban's partition function identity $Z(\Lambda, S_0) = Z(\Lambda_k, S_k)$ (Theorem 2.1(a)) holds for all lattice sizes L that are multiples of $2^{n_{\max}}$, and that it is compatible with the temporal transfer matrix factorization $Z_n = \mathrm{Tr}(T_n^{L_n})$. Then the mass gap scales as*

$$m_{n+1} = 2m_n \tag{3}$$

for $0 \leq n < n_{\max}$. Equivalently, the physical mass gap $m_{\mathrm{phys}} = m_n/a_n$ is scale-independent.

Proof. Step 1: Transfer matrix representation. On the lattice $\Lambda_n = (\mathbb{Z}/L_n\mathbb{Z})^4$ with temporal extent $L_n^{(t)} = L_n$ (we use the convention of equal spatial and temporal extent): the partition function admits the transfer matrix decomposition

$$Z_n = \mathrm{Tr}_{\mathcal{H}_n}\big(T_n^{L_n}\big),$$

where T_n is the transfer matrix acting on the Hilbert space $\mathcal{H}_n = L^2(\mathcal{B}_n^{\mathrm{spat}})$ of gauge-invariant functions on the spatial orbit space $\mathcal{B}_n^{\mathrm{spat}}$.

T_n is a positive self-adjoint trace-class operator on the finite-dimensional space $\mathcal{H}_n$, with eigenvalues $\lambda_0^{(n)} \geq \lambda_1^{(n)} \geq \cdots \geq 0$. The mass gap at scale n (in lattice units a_n) is

$$m_n = -\ln \frac{\lambda_1^{(n)}}{\lambda_0^{(n)}}.$$

Step 2: Trace identity. By Theorem 2.1(a): $Z_n = Z_{n+1}$ for the same physical volume. Since Λ_n has $L_n = L/2^n$ sites per direction and Λ_{n+1} has $L_{n+1} = L/2^{n+1} = L_n/2$:

$$\mathrm{Tr}_{\mathcal{H}_{n+1}}\big(T_{n+1}^{L_n/2}\big) = \mathrm{Tr}_{\mathcal{H}_n}\big(T_n^{L_n}\big). \tag{4}$$

This identity holds for all L that are multiples of $2^{n_{\max}}$. By varying L, the integer $M := L_n = L/2^n$ ranges over all multiples of $2^{n_{\max}-n}$. In particular, M takes infinitely many values tending to $+\infty$.

Step 3: Extraction of the leading eigenvalue. For $M \to \infty$:

$$\begin{aligned}\mathrm{Tr}(T_n^M) &= (\lambda_0^{(n)})^M \Big[1 + \Big(\frac{\lambda_1^{(n)}}{\lambda_0^{(n)}}\Big)^M + \cdots\Big] \\ &= (\lambda_0^{(n)})^M \big[1 + e^{-m_n M} + O(e^{-m_n' M})\big],\end{aligned} \tag{5}$$

where $m_n' = -\ln(\lambda_2^{(n)}/\lambda_0^{(n)}) > m_n$ is the gap to the third eigenvalue (assuming $\lambda_1^{(n)} > \lambda_2^{(n)}$; if degenerate, the coefficient of $e^{-m_n M}$ is the multiplicity).

Step 4: Matching. Substituting (5) into (4) and using $M/2$ in place of M for T_{n+1}:

$$(\lambda_0^{(n+1)})^{M/2}\big[1 + e^{-m_{n+1}M/2} + \cdots\big] = (\lambda_0^{(n)})^M\big[1 + e^{-m_n M} + \cdots\big]. \quad (6)$$

Leading order. Taking logarithms and dividing by M:

$$\tfrac{1}{2}\ln\lambda_0^{(n+1)} + O(e^{-m_{n+1}M/2}/M) = \ln\lambda_0^{(n)} + O(e^{-m_n M}/M).$$

As $M \to \infty$ (through the allowed sequence):

$$\lambda_0^{(n+1)} = \big(\lambda_0^{(n)}\big)^2. \quad (7)$$

Subleading order. Using (7), divide both sides of (6) by $(\lambda_0^{(n)})^M = (\lambda_0^{(n+1)})^{M/2}$:

$$1 + e^{-m_{n+1}M/2} + O(e^{-m_{n+1}'M/2}) = 1 + e^{-m_n M} + O(e^{-m_n' M}).$$

Subtracting 1 and comparing the exponentially decaying terms for large M:

$$e^{-m_{n+1}M/2} = e^{-m_n M}\big(1 + O(e^{-\delta M})\big)$$

where $\delta = \min(m_n' - m_n,\, m_{n+1}' - m_{n+1}) > 0$. Taking logarithms:

$$-\frac{m_{n+1}}{2}\cdot M = -m_n \cdot M + O(e^{-\delta M}).$$

Dividing by M and sending $M \to \infty$:

$$m_{n+1} = 2m_n. \quad (8)$$

□

Remark 4.2. Equation (8) says that the mass gap doubles at each RG step (in lattice units). Since the lattice spacing also doubles ($a_{n+1} = 2a_n$), the physical mass gap $m_{\text{phys}} = m_n/a_n$ is invariant:

$$\frac{m_{n+1}}{a_{n+1}} = \frac{2m_n}{2a_n} = \frac{m_n}{a_n}.$$

5. Proof of the main theorem

Proof of Theorem 1.1. By Theorem 3.1: $m_{n_{\max}} \geq c_0 > 0$.

By Theorem 4.1 applied iteratively: $m_0 = m_{n_{\max}}/2^{n_{\max}}$.

By Balaban's coupling evolution (Theorem 2.1(c)): $n_{\max} = \lfloor(\beta - \beta_{n_{\max}})/(2b_0\ln 2)\rfloor = \lfloor 1/(2b_0 g^2\ln 2)\rfloor \cdot (1 + O(g^2))$, so

$$2^{n_{\max}} = \exp\Big(\frac{\ln 2}{2b_0 g^2} + O(1)\Big) = \exp\Big(\frac{C}{g^2} + O(1)\Big)$$

with $C = \ln 2/(2b_0) = 24\pi^2\ln 2/(11N_c)$.

Therefore:

$$m_{\text{gap}} = m_0 = \frac{m_{n_{\max}}}{2^{n_{\max}}} \geq \frac{c_0}{e^{C/g^2 + O(1)}} = c_0' \cdot e^{-C/g^2}$$

where $c_0' = c_0 \cdot e^{-O(1)} > 0$.

This bound is uniform in L satisfying $2^{n_{\max}+1} \leq L \leq C_0 \cdot 2^{n_{\max}}$ for a fixed constant C_0, ensuring that the coarsened lattice has $L_{n_{\max}} = L/2^{n_{\max}} \leq C_0$ sites per direction. For

$L > C_0 \cdot 2^{n_{\max}} = C_0 \cdot e^{C/g^2+O(1)}$: the lattice exceeds the correlation length $\xi \sim e^{C/g^2}$, and the mass gap is expected to equal the infinite-volume mass gap up to corrections $O(e^{-mL})$. A rigorous proof for $L > \xi$ requires a spatial cluster expansion, which we do not carry out here. □

6. Discussion

6.1. **What is proved.** Theorem 1.1 establishes a positive mass gap for $SU(N_c)$ lattice Yang–Mills theory in four dimensions at weak coupling, with exponentially small lower bound $c \cdot e^{-C/g^2}$ in lattice units. The proof is conditional on a single assumption: the compatibility of Balaban's isotropic RG with the temporal transfer matrix factorization (stated in Theorem 4.1). All other ingredients are either published results (Balaban's RG, the orbit-space Ricci curvature from Paper II, the Holley–Stroock bound from Paper III) or proved in this paper (Theorems 3.1 and 4.1).

6.2. **Comparison with the Millennium Problem.** The Clay Millennium Problem [7] asks for:

(1) Construction of a continuum $d = 4$ Yang–Mills QFT satisfying the Wightman (or Osterwalder–Schrader) axioms.
(2) Proof that this QFT has a mass gap $m > 0$.

Our result addresses item (2) on the *lattice*, not in the continuum. The remaining steps are:

- **Continuum limit:** Prove that Balaban's effective actions converge (in an appropriate sense) to a continuum QFT as $a \to 0$. This is within reach of Balaban's program but requires additional work on the construction of continuum operator algebras and the verification of the OS axioms.
- **Mass gap persistence:** Prove that the lattice mass gap $m(a) = c \cdot e^{-C/g^2(a)} = O(\Lambda_{\mathrm{QCD}})$ has a well-defined positive limit as $a \to 0$. Our bound gives $m(a) \geq c \cdot \Lambda_{\mathrm{QCD}} \cdot a \cdot e^{O(1)} \to 0$, which does NOT directly give a continuum mass gap. A stronger bound (e.g., $m(a) \geq c' \cdot \Lambda_{\mathrm{QCD}}$ independent of a) would require controlling the $O(1)$ constants in Balaban's construction.

Paper	Key result	Used in proof
I	Brascamp–Lieb mass gap	(not directly)
II	$\mathrm{Ric}_{\mathcal{B}} \geq N_c/4$	Theorem 3.1
III	Holley–Stroock bound; Morse–Bott structure	Theorem 3.1
IV	Trace identity; main theorem	Theorems 1.1 and 4.1

Table 1. Contributions of each paper.

6.3. **The role of each paper.**

6.4. **Strengths, limitations, and caveats.** *Strengths:* The proof is short (the new content is Theorem 4.1), modular (each ingredient is proved independently), and conditional only on the compatibility of Balaban's isotropic RG with the temporal transfer matrix factorization (see Theorem 4.1 and §6).

Limitations:

(1) The bound $c \cdot e^{-C/g^2}$ is exponentially small in $1/g^2$ and does not directly give a continuum mass gap.
(2) The constant $c_0 = (\gamma N_c/8)e^{-K/\gamma}$ from the Holley–Stroock bound is potentially extremely small.
(3) The proof applies to $d = 4$ (where Balaban's full results are available). Extension to $d = 3$ requires the completion of Balaban's program in three dimensions (cf. [8]).
(4) The transfer matrix trace identity (Theorem 4.1) assumes that Balaban's partition function identity $Z_0 = Z_n$ holds for *all* lattice sizes L that are multiples of $2^{n_{\max}}$, and that this identity is compatible with the temporal transfer matrix factorization $Z = \mathrm{Tr}(T^M)$. Balaban's isotropic RG coarsens all directions simultaneously, so the compatibility with a distinguished temporal direction requires verification. An anisotropic variant of Balaban's RG (coarsening only spatial directions) would resolve this point but has not been published. We regard this as the principal caveat of the present proof.

References

[1] L. Eriksson, *Mass gap in the weakly coupled Gribov–Zwanziger lattice gauge theory via Brascamp–Lieb log-concavity*, Preprint, 2026.
[2] L. Eriksson, *Geodesic convexity and structural limits of curvature methods for the Yang–Mills mass gap on the lattice*, Preprint, 2026.
[3] L. Eriksson, *Morse–Bott spectral reduction and the Yang–Mills mass gap on the lattice*, Preprint, 2026.
[4] T. Balaban, *Large field renormalization. II*, Comm. Math. Phys. **122** (1989), 355–392.
[5] R. Holley and D. Stroock, *Logarithmic Sobolev inequalities and stochastic Ising models*, J. Stat. Phys. **46** (1987), 1159–1194.
[6] D. Bakry and M. Émery, *Diffusions hypercontractives*, Sém. Prob. XIX, LNM **1123**, Springer, 1985, pp. 177–206.
[7] A. Jaffe and E. Witten, *Quantum Yang–Mills theory*, Clay Mathematics Institute, 2000.
[8] J. Dimock, *The renormalization group according to Balaban. I*, Rev. Math. Phys. **25** (2013), 1330010.