Volume 1 Preparation

Part 5 Lattice gauge theory

Kiichi, Shiori and Rin · Edition of 2026-11-14 · https://computoergosum.com/en/principia/yang-mills/

> This PDF is a print of the chapters of the web edition. Interactive figures are frozen in the state in which they were printed. Links between chapters lead to the web edition, which holds the latest text.

Chapter		Page
5-01	Wilson's lattice	2
5-02	The strong-coupling expansion	13
5-03	Reflection positivity and the transfer matrix	25
5-04	The infinite-volume limit and the cluster expansion	37
5-05	Weak coupling and the continuum limit	49
5-06	How U(1) differs from the non-abelian case	60
5-07	The Monte Carlo method	74
5-08	Measuring glueballs and the string tension	86
5-09	What to hold fixed when taking the limit	99
5-10	Lattice gauge theory with finite groups	114
5-11	Detour: Bakry–Émery and one-link integration	125

Wilson's lattice — put group elements on the edges, and gauge invariance survives without approximation

This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — Group elements (link variables) are placed on the edges of the lattice, and the action is built from products around small squares (plaquettes). That action is exactly gauge invariant whatever the lattice spacing, approaches the continuum action at the rate a^2 for smooth fields, and this fixes $\beta = 2N/g^2$. The measure is defined without gauge fixing.

Prerequisite chapters — 3-03 (the action) and 1-04 (SU(2), SU(3)). Conventions follow 3-02, parallel transport 1-07, Haar measure 1-08 § 03.

Said first — This is a chapter of definitions and classical properties. The continuum limit of § 05 is how the action approaches its limit for smooth fields, a different thing from the continuum limit of the quantum theory (5-05). This chapter shows nothing about the mass gap.

Order of this chapter

1. Why a lattice — to a finite-dimensional integral. Detour: the summer of 1973
2. Link variables — group elements on the edges (Table 1)
3. Plaquettes and Wilson loops — only closed paths survive (Figure 1, Table 2)
4. The Wilson action — five properties
5. The classical continuum limit — where β comes from (Figure 2, Table 3), converting conventions (Table 4)
6. The measure on the lattice — no gauge fixing needed
7. What survives, what breaks — choosing a regularisation
8. The map of Part 5 — two limits

01

Why a lattice — making the integral finite-dimensional

Euclideanised quantum Yang–Mills theory (4-02) formally integrates $e^{-S[A]}$ over all connections A. But there is no uniform measure on an infinite-dimensional space (1-09 § 02), and the integral counts multiple times the A related by gauge transformations (3-04, 4-08). Wilson replaced spacetime by a lattice. This is the abstract of the 1974 paper.

> "It is shown how to quantize a gauge field theory on a discrete lattice in Euclidean space-time, preserving exact gauge invariance and treating the gauge fields as angular variables (which makes a gauge-fixing term unnecessary)."

"Angular variables" means holding the field as group elements rather than as elements of the Lie algebra. The problem statement (Jaffe–Witten § 6.5) writes that, for compact groups and compactified spacetime, "the lattice approximation reduces the functional integration to a finite-dimensional integral", and continues that one must check that the expectation values of gauge-invariant observables exist in the limit of lattice spacing 0 and infinite volume Known . We look at the limits in § 08.

Conventions. Euclidean $\mathbb{R}^4$, indices 1 to 4, $\hbar = c = 1$. As in 3-02, $A_\mu = A_\mu^a T^a$ (Hermitian, $\mathrm{tr}\, T^a T^b = \frac{1}{2}\delta^{ab}$), $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu]$, and the continuum action is $S = \frac{1}{2}\int \mathrm{tr}\, F_{\mu\nu} F_{\mu\nu}\, d^4x$ (summed over all indices). β is the inverse lattice coupling (a different quantity from the β function $\beta(g)$ of 4-06), and θ is just an angle (not the θ term).

Detour: the summer of 1973

A subsection you may skip.

In Wilson's recollection (published 2005), after the asymptotic freedom papers (published June 1973) he came to QCD from the lattice familiar from statistical mechanics, and in the summer of 1973 he gave a talk at Orsay (he does not remember the talk himself, and writes that he knows of it from a reference in a paper of Balian–Drouffe–Itzykson). He writes that by then he knew that the gauge variables "must be unitary matrices from the gauge group itself, rather than being variables from the infinitesimal Lie algebra of the group", that there is a separate gauge invariance at each point, and that finite results come out without gauge fixing. He also mentions that, since the coefficient is $1/g^2$, the strong-coupling expansion (5-02) was immediately visible, that J. Smit was developing a formulation of lattice gauge theory, that Polyakov had obtained the same thing independently, and the model of Wegner (1971) (2-09) Known .

02

Link variables — group elements on the edges

On the lattice $a\mathbb{Z}^4$ of spacing a (in finite volume, a periodic lattice with L points per side), the edge joining x and $x + a\hat{\mu}$ is called a link. In the continuum theory the gauge field was a tool for comparing frames at neighbouring points (3-01). On the lattice the tool itself

becomes the variable: on each link we put $U_{x,\mu} \in G$ and read it as "carrying the value at $x + a\hat{\mu}$ into the frame of x". A gauge transformation, by $\Omega(x) \in G$ at each point, acts as follows.

$$U_{x+a\hat{\mu},\,-\mu} = U_{x,\mu}^{-1}, \qquad U'_{x,\mu} = \Omega(x)\, U_{x,\mu}\, \Omega(x + a\hat{\mu})^{-1} \quad (1)$$

If a matter field transforms as $\psi \mapsto \Omega\psi$, the difference $U_{x,\mu}\psi(x + a\hat{\mu}) - \psi(x)$ changes only by $\Omega(x)$ (the four-direction version of the chain of 3-02). From a smooth A, the link is the inverse of parallel transport (1-07) along the edge. Divide the edge into n equal parts and multiply the small rotations from the left, starting from the initial end.

$$U_{x,\mu} = \lim_{n\to\infty} \prod_{k=0}^{n-1} e^{-igh A_\mu(x+(k+\frac{1}{2})h\hat{\mu})} = \exp\bigl(-iga\, A_\mu(x + \tfrac{a}{2}\hat{\mu}) + O(a^3)\bigr), \qquad h = \tfrac{a}{n} \quad (2)$$

(1) is the transformation law $A' = \Omega A \Omega^{-1} - \frac{i}{g}(\partial\Omega)\Omega^{-1}$ of 3-02 accumulated along the edge Known . Example: for U(1), $U = e^{-i\theta}$, and if $\Omega(x) = e^{i\alpha(x)}$, (1) is $\theta \mapsto \theta + \alpha(x + a\hat{\mu}) - \alpha(x)$ (2-03 § 03). For SU(2), $U = u_0 + i\,\mathbf{u}\cdot\boldsymbol{\sigma}$ ($u_0^2 + |\mathbf{u}|^2 = 1$), each link is a point of S^3, and $\frac{1}{2}\mathrm{tr}\, U = u_0$.

A periodic lattice has $4L^4$ links, and the configuration space G^{4L^4} is a compact manifold of real dimension $4(N^2-1)L^4$ (1-05). The gauge freedom G^{L^4} is a quarter of it, the same proportion as removing 1 gauge component from the 4 components in the continuum.

Table 1 Sizes of periodic lattices Computed

L	Links	Plaquettes	Real variables for SU(3)
4	1,024	1,536	8,192
16	262,144	393,216	2,097,152
32	4,194,304	6,291,456	33,554,432

03

Plaquettes and Wilson loops — only closed paths survive

A single link changes sandwiched between the Ω at its two ends, but when two links are joined end to end, the $\Omega^{-1}\Omega$ at the joining point cancels. Going round a closed path, only the Ω at the starting point remains, on both sides. The smallest closed path is a square of side a: the plaquette.

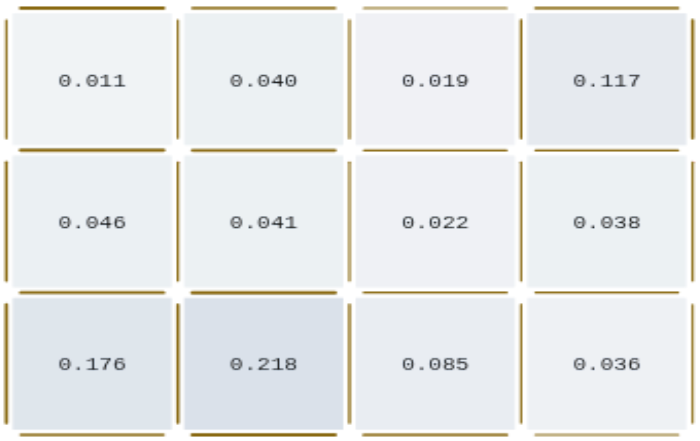
$$U_P = U_{x,\mu}\, U_{x+a\hat{\mu},\nu}\, U_{x+a\hat{\nu},\mu}^{-1}\, U_{x,\nu}^{-1}, \qquad U'_P = \Omega(x)\, U_P\, \Omega(x)^{-1} \quad (3)$$

So $\mathrm{tr}\, U_P$ is invariant without approximation, whatever the lattice spacing and however rough the configuration. For a general closed path C, the trace of the product taken in the order of the path,

$$W_C = \frac{1}{N}\,\mathrm{tr} \prod_{\ell\in C} U_\ell \quad (4)$$

is called a Wilson loop. Made from a smooth field via (2), it approaches the continuum Wilson loop (the trace of the holonomy of 1-07) as the lattice path approaches a closed curve (Chatterjee 2018 § 4) Known . Rectangular loops are used to test for confinement (5-02, 17-02). The order of multiplication matters. Multiplying the same four links in the order $U_{x,\mu}U_{x+a\hat{\nu},\mu}U_{x+a\hat{\mu},\nu}^{-1}U_{x,\nu}^{-1}$, factors such as $\Omega(x+a\hat{\nu})^{-1}\Omega(x+a\hat{\mu})$ get stuck between them and do not cancel. For U(1) the order does not matter (2-03).

Figure 1 A 5×4-point patch of SU(2). Moving the slider s (the strength of the gauge transformation) changes the thickness of the edges ($1-\frac{1}{2}\operatorname{tr}U_\ell$) but not the numbers on the faces ($1-\frac{1}{2}\operatorname{tr}UP$). "Break the order" makes the face numbers move too.




```
s = 0.00, Σ_P(1−½trU_P) = 0.849977, Σ_ℓ ½trU_ℓ = 27.0024
```

Computed The numbers in this figure were computed on this machine. The browser computes the same configuration with the same formulas as Table 2 (if JavaScript is disabled, Table 2 gives the content).

Table 2 Sums over the configuration of Figure 1 (31 edges, 12 faces)

s	$\sum_\ell \frac{1}{2}\operatorname{tr}U_\ell$	$\sum_P(1-\frac{1}{2}\operatorname{tr}UP)$	$\sum_P(1-\frac{1}{2}\operatorname{tr})$ **in broken order**
0	27.0024	0.849977	0.5719
0.5	17.5284	0.849977	3.0662
1	0.6552	0.849977	5.1220

Computed Applying a random gauge transformation to a random configuration on a four-dimensional periodic lattice with 4 points per side, the change in the action of § 04 was 2×10^{-13} for SU(2) and 0 for SU(3). Building the sum of squares of the differences $\partial_\mu A_\nu-\partial_\nu A_\mu$ from the logarithms of the links (elements of the Lie algebra), for SU(2) 34,707 changes to 35,659. Even with the commutator added, for finite gauge transformations the transformation law of A is inhomogeneous (the term $-\frac{i}{g}(\partial\Omega)\Omega^{-1}$ below (1)), so the differences are not exactly invariant. This is the meaning of "group elements rather than Lie algebra variables" in the recollection in the detour.

> 04

The Wilson action — five properties

$$S_W(U) = \beta \sum_P \Big(1 - \frac{1}{N} \operatorname{Re}\operatorname{tr} U_P\Big), \qquad \beta = \frac{2N}{g^2} \quad (G = SU(N)) \tag{5}$$

The sum takes every plaquette once ($6L^4$ of them on a periodic lattice). There are five properties Known .

1. **Gauge invariant**: each term is unchanged by (3).
2. **Real**: reversing the orientation leaves $\operatorname{Re}\operatorname{tr} U_P^\dagger$ the same.
3. **Bounded**: the eigenvalues have absolute value 1, so each term is between 0 and 2. The maximum is 2 for SU(2) ($U_P = -1$) and $3/2$ for SU(3) (all three eigenvalues $e^{2\pi i/3}$, or all three $e^{-2\pi i/3}$) Computed .
4. **Minimum**: $S_W = 0$ only when all $U_P = 1$. A pure gauge $\Omega(x)\Omega(x+a\hat{\mu})^{-1}$ is an example. On a periodic lattice, so is a configuration with commuting elements V_μ placed uniformly in each direction. If its product once round, V_μ^L, is not 1, it is not a pure gauge, because for a pure-gauge configuration the product once round is $\Omega(x)\Omega(x)^{-1} = 1$.
5. **Local**: each term depends on four links only.

Chatterjee (2018) writes $\beta' \sum_P \operatorname{Re}\operatorname{tr}(I - U_P)$, so $\beta' = \beta/N$ (the same holds for the β of SU(2) in 7-02).

05

The classical continuum limit — where $\beta = 2N/g^2$ comes from

Map a smooth A to links via (2) and look at $a \to 0$. Expanding (3) with the Baker–Campbell–Hausdorff formula $e^X e^Y = e^{X+Y+\frac{1}{2}[X,Y]+\cdots}$, which combines a product of exponentials into one, gives at the centre c of the plaquette

$$U_P = \exp\big(-iga^2 F_{\mu\nu}(c) + O(a^3)\big) \tag{6}$$

The difference of A on opposite edges gives $-iga^2(\partial_\mu A_\nu - \partial_\nu A_\mu)$, and the commutator gives $(-iga)^2[A_\mu, A_\nu]$; together they make the field strength (curvature = the shift after going once round a small loop; 1-07). Correspondence with the notation of 1-07: the anti-Hermitian connection and curvature of 1-07 correspond to $-igA$ and $-igF$ of this chapter, and parallel transport in 1-07 carries from the start to the end, so the links and plaquettes of this chapter are its inverses. This is why the sign in the exponent of (6) is opposite to 1-07's $U_p \approx \exp(-a^2 F)$; the value of the trace does not change. For Hermitian X, $\operatorname{Re}\operatorname{tr} e^{-iX} = \operatorname{tr}\cos X = N - \frac{1}{2}\operatorname{tr} X^2 + \cdots$, so

$$1 - \frac{1}{N}\operatorname{Re}\operatorname{tr} U_P = \frac{g^2 a^4}{2N} \operatorname{tr} F_{\mu\nu}(c)^2 + O(a^5) \tag{7}$$

Taking the centre of the expansion at the centre of the plaquette, the a^5 terms cancel by the symmetry of orientations, and the remaining error is a^6 per plaquette Known . Summing a^{-4} plaquettes, the relative error is a^2, and this is the term that Symanzik improvement (5-05) removes. Counting plaquettes by a point x and a pair $\mu < \nu$ and replacing $\sum_x a^4$ by an integral,

$$SW \approx \frac{\beta g^2}{2N} \sum_{\mu<\nu} \int \operatorname{tr} F_{\mu\nu}^2 \, d^4x = \frac{\beta g^2}{2N} \, S \quad (8)$$

It agrees with the continuum action S when $\beta = 2N/g^2$, and this is where the β in (5) comes from Known . Chatterjee (2018) § 3 does the same calculation and writes that in four dimensions β is classically independent of a, that in the quantum theory physicists believe β should grow like $\log(1/a)$ though there are also doubts, and that "the question remains an open mathematical problem" Physics (5-05).

Converting conventions — five ways of writing one coupling

The same coupling g^2 is written with different letters and different normalisations in different references. This is why the same condition looks like different numbers in different chapters. The standard of this book is Wilson's $\beta = 2N/g^2$ in (5) (the table of notation A-01 § 06, where Table 7 lists six ways of writing it, adding g^2, with the same numerical examples). The other ways of writing are fixed by comparing term by term with the action (5).

$$\beta = \frac{2N}{g^2}, \qquad \beta_{\mathrm{Ch}} = \frac{\beta}{N} = \frac{2}{g^2}, \qquad \beta_{\mathrm{SZZ}} = \frac{\beta}{N^2}, \qquad \lambda = g^2 N = \frac{2N^2}{\beta} = \frac{2}{\beta_{\mathrm{SZZ}}}, \qquad \alpha = \frac{g^2}{4\pi} \quad (8a)$$

β_{Ch} is the form writing the action as $\beta_{\mathrm{Ch}} \sum_P \operatorname{Re}\operatorname{tr}(I - U_P)$ (without $1/N$; (3) of 7-10), β_{SZZ} is 't Hooft's convention writing the inverse coupling in front of the action as $N\beta_{\mathrm{SZZ}}$ (Shen–Zhu–Zhu; 5-04), λ is the 't Hooft coupling of large N (17-04), and α is the coupling strength used in physics. Differences in the constant term of the action disappear in the normalisation Z and do not change the measure.

Table 4 Converting conventions, with numerical examples (the Shen–Zhu–Zhu condition is in four dimensions, $|\beta$ $1/(16(d-1)) = 1/48$) Computed

Form	Conversion from β	SU(2), $\beta = 2.5$	SU(3), $\beta = 6.0$	SZZ condition: SU(2)	SZZ condition
Wilson's β (this book)	$2N/g^2$	2.5	6.0	$\beta < 0.0833$	β
β_{Ch} (the form without $1/N$)	β/N	1.25	2.0	< 0.0417	
β_{SZZ} ('t Hooft's convention)	β/N^2	0.625	0.6667	$< 1/48 = 0.0208$	

Form	Conversion from β	SU(2), $\beta = 2.5$	SU(3), $\beta = 6.0$	SZZ condition: SU(2)	SZZ condition
't Hooft coupling λ	$2N^2/\beta$	3.2	3.0	$\lambda > 96$	
g^2	$2N/\beta$	1.6	1.0	$g^2 > 48$	
$\alpha = g^2/4\pi$	$N/(2\pi\beta)$	0.1273	0.0796	$\alpha > 3.820$	

Example: the Shen–Zhu–Zhu condition is $1/48$ independent of N when written in β_{SZZ}, $\lambda > 96$ when written in λ, and $N^2/48$ when converted to Wilson's β. That N disappears in β_{SZZ} and λ is the advantage of 't Hooft's convention. The SU(2) value $\beta = 2.5$ often used in numerical work is $\alpha \approx 0.13$, and the edge of the condition is $\alpha \approx 3.8$; the two are about 30 times apart. The conversion is this book's arithmetic, and the conditions themselves are to be read in the normalisation of each reference (5-04, 5-09, 7-10).

Looking at a single plaquette

For the constant SU(2) field $A_1 = cT^1$, $A_2 = cT^2$ there are no derivative terms, and $F_{12} = gc^2T^3$ comes only from the commutator (Example 2 in 3-02 § 04). The links are $e^{-i\theta\sigma 1/2}$ and $e^{-i\theta\sigma 2/2}$ ($\theta = gac$), and the plaquette is the group commutator $U_1U_2U_1^{-1}U_2^{-1}$. We multiply in the form $u_0 + i\,\mathbf{u}\cdot\boldsymbol{\sigma}$ of § 02. From the product of Pauli matrices $\sigma_j\sigma_k = \delta_{jk} + i\epsilon_{jkl}\sigma_l$,

$$(a_0 + i\,\mathbf{a}\cdot\boldsymbol{\sigma})(b_0 + i\,\mathbf{b}\cdot\boldsymbol{\sigma}) = a_0b_0 - \mathbf{a}\cdot\mathbf{b} + i\,(a_0\mathbf{b} + b_0\mathbf{a} - \mathbf{a}\times\mathbf{b})\cdot\boldsymbol{\sigma}$$

Putting $c = \cos\frac{\theta}{2}$ and $s = \sin\frac{\theta}{2}$, $U_1 = (c;\,-s,0,0)$ and $U_2 = (c;\,0,-s,0)$ (writing u_0 and $\mathbf{u}$ side by side), $U_1U_2 = (c^2;\,-cs,-cs,-s^2)$ and $U_1^{-1}U_2^{-1} = (U_2U_1)^{-1} = (c^2;\,cs,cs,-s^2)$. The u_0 of the product of the two is $\frac{1}{2}\mathrm{tr}\,U_P$, so

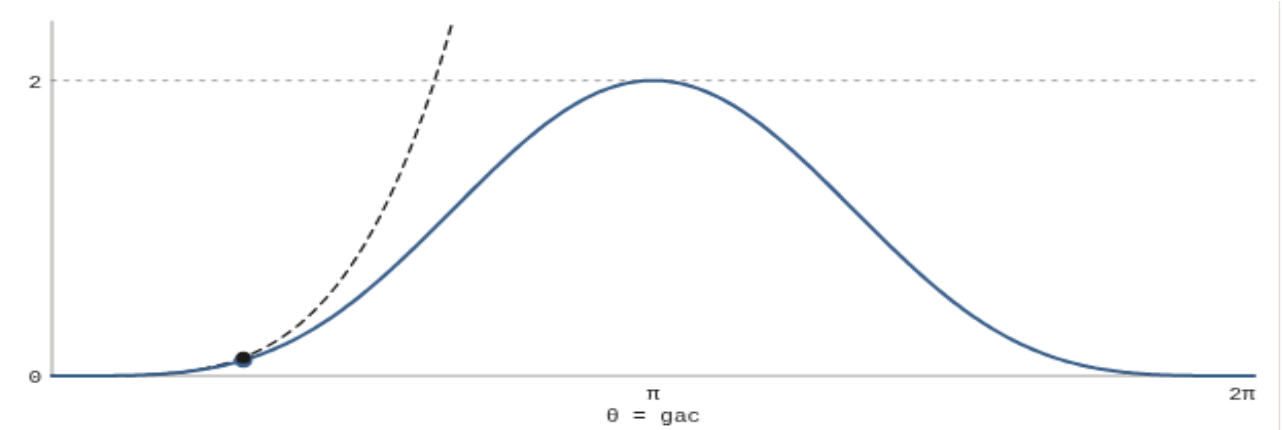
$$\tfrac{1}{2}\mathrm{tr}\,U_P = c^4 - \left(-c^2s^2 - c^2s^2 + s^4\right) = (c^2+s^2)^2 - 2s^4 = 1 - 2\sin^4\frac{\theta}{2}$$

and the following closed form comes out.

$$1 - \tfrac{1}{2}\mathrm{tr}\,U_P = 2\sin^4\frac{\theta}{2}, \qquad \frac{g^2a^4}{2N}\,\mathrm{tr}\,F_{12}^2 = \frac{\theta^4}{8} \quad (9)$$

For small θ the ratio is $1 - \theta^2/6 + \cdots$, and the relative error is proportional to a^2. For large θ the lattice value does not exceed 2, and returns to 0 at $\theta = 2\pi$. The link sees the field only in the form e^{-igaA}, so it does not distinguish a field that is too strong from a weak one. This boundedness is one reason why the integral in § 06 is finite.

Figure 2 A single plaquette. The solid line is the lattice $2\sin^4(\theta/2)$, the dashed line the continuum $\theta^4/8$, the dotted line the upper bound 2. Moving the slider for $\theta = gac$ (proportional to a for a fixed field) changes the two points and the ratio.



θ = 1.00, lattice 0.10566, continuum 0.12500, ratio 0.8453

Computed The numbers in this figure were computed on this machine. The curves are (9) (differing from the matrix calculation by at most 2×10^{-16}). If JavaScript is disabled: the ratio (lattice / continuum) is 0.998, 0.959, 0.845 and 0.501 at $\theta = 0.1,\ 0.5,\ 1,\ 2$, and the lattice value is at its maximum 2 at $\theta = \pi$ and 0 at 2π.

Looking at a smooth four-dimensional field

We compare using a smooth SU(2) field on the torus $[0,1)^4$ (a sum of trigonometric functions with 8 wave numbers, $g = 1$, $\beta = 4$). The links truncate (2) at $n = 16$, and since the continuum action is a trigonometric polynomial, it was integrated exactly as a sum over finitely many points ($S = 184.2275$).

Table 3 Lattice spacing $a = 1/L$ and the Wilson action Computed

L	S_W	$(S_W - S)/S$	The same value $/a^2$
4	134.502	−0.2699	−4.319
8	170.242	−0.07592	−4.859
16	180.618	−0.01959	−5.016
32	183.318	−0.004938	−5.057

The difference is proportional to a^2, with coefficient near −5.06 for this field (the value depends on the field). Changing n to 8, 16, 32 changes S_W by 0.03 and 0.008 at $L = 8$. The $\|U_P - e^{-iga^2F(c)}\|/a^2$ of (6) is 0.190, 0.119, 0.065 and 0.034 from $a = 1/8$ to $1/64$, nearly halving each time a is halved (ratios from 1.6 to 1.9), shrinking in proportion to a. With the sign of the commutator reversed it stayed at 5 to 6.5 Computed .

06

The measure on the lattice — no gauge fixing needed

The quantum theory on a finite lattice is defined by the following probability measure (1-08 § 01).

$$d\mu_\beta(U) = \frac{1}{Z}\, e^{-S_W(U)} \prod_\ell dU_\ell, \qquad Z = \int e^{-S_W(U)} \prod_\ell dU_\ell \tag{10}$$

dU_ℓ is the Haar measure of G (the measure of total weight 1 unchanged by multiplication from either left or right; for SU(2), the uniform distribution on S^3). **It is finite**: $0 < e^{-S_W} \leq 1$ is continuous and the domain is compact, so $0 < Z \leq 1$, and the boundedness of § 04 also gives $Z \geq e^{-12\beta L^4}$ ((3) of 7-10) Known . For SU(3) with 4 points per side it is an ordinary multiple integral over 8,192 real dimensions.

No gauge fixing needed: Haar measure is unchanged by left and right multiplication, so moving a configuration by (1) changes neither $\prod dU_\ell$ nor S_W, and $\langle f \rangle = \langle f \circ \Omega \rangle$. In the continuum, gauge fixing was needed to choose representatives from orbits of infinite volume (4-08, 3-04), but on the lattice the whole of the gauge transformations G^{L^4} is compact with weight 1, so the multiple counting is merely averaged. This is what "making the gauge-fixing term unnecessary" in the abstract means Known .

There are two consequences. (i) For general f, $\langle f \rangle$ equals the expectation value of f averaged over gauge transformations. In particular, quantities that change non-trivially under gauge transformations (a single link U_ℓ or its components, $\mathrm{tr}\, U_\ell$, and so on) are cancelled by the Haar average and have expectation value 0 (the detour of 17-02; that this survives in infinite volume is Elitzur's 1975 theorem). Not every gauge-non-invariant quantity vanishes: the expectation value of $|\mathrm{tr}\, U_\ell|^2$ is positive. (ii) The links on a tree (a set of links containing no cycles) may be set to 1 (7-02 § 02). Typical observables are Wilson loops and the correlation between two distant loops. The rate at which the correlation decreases is the mass on the lattice (4-11, 5-09), and the lattice correlation-length problem in § 5 of Chatterjee (2018) (Chatterjee's Problem 5.1) asks about it.

07

What survives, what breaks — choosing a regularisation

A device for avoiding infinities (a regularisation) changes part of the theory, and what it keeps decides the later work. On Wilson's lattice, gauge invariance survives exactly at each a, and of rotations and translations only rotations by 90° and shifts by the lattice spacing survive (the recovery of the continuum symmetries is to be shown in the limit; 7-10; for the recovery in the construction published in 2026-11, 12-06). Reflection positivity holds (Osterwalder–Seiler 1978; 5-03), and topological charge (3-05) is not defined as it stands.

The problem statement (JW § 6.5) calls the fact that reflection positivity holds in Wilson's approximation “a major advantage” Known , because it is the key to building the Hamiltonian and the Hilbert space (4-10). By contrast, a continuum momentum cutoff breaks gauge invariance, which is restored with counterterms (7-07 § 06). JW § 6.5 also writes that when reflection positivity is lost with regularisations such as dimensional regularisation or Pauli–Villars, there are few ways of recovering it Known . On this basis, a regularisation that keeps all three (gauge invariance, reflection positivity, continuum rotations) has not been found as far as we have looked (this book's organisation; for the version that counts four by adding ultraviolet control, 7-10 § 03).

The abstract of Lüscher (1982) states that non-abelian gauge fields on a four-dimensional hypercubic lattice with small action density have "an integer topological charge Q, which is invariant under continuous deformations of the field" Known . This Q is said to return to the Chern number (3-05) in the classical continuum limit. (A note of this chapter: without the condition, $SU(N)^{\text{number of links}}$ is connected, so the only continuous integer-valued quantities are constants.) Nor is the action limited to Wilson's form: there are heat-kernel actions (Villain type; 7-02), and improved actions that add 1×2 rectangles to remove the a^2 difference of Table 3 (Symanzik improvement; 5-05). The view that all of them reach the same continuum limit is universality Physics .

08

The map of Part 5 — two limits

The theory on a finite lattice was defined by (10). The theory of the problem statement needs two limits, $L \to \infty$ and $a \to 0$ (in four dimensions thought to be paired with $\beta \to \infty$ Physics). The order is first to fix β (that is, a) and take $L \to \infty$, then $a \to 0$. The lattice correlation-length problem in § 5 of Chatterjee (2018) also takes the infinite-volume limit at fixed β as its starting point Known . Even for the first, Chatterjee (2018) § 3 writes that the uniqueness of the infinite-volume limit is not known in general for large β above two dimensions Known . The construction published in 2026-11 takes the opposite order: it holds the volume fixed and takes a $\to 0$, and then takes the limit along a subsequence of volumes (the committee's assessment is that it is proved in Lean; 8-05). This construction is read in Volume 2: the family of lattices and the order of limits in 8-01 to 8-05, and one step of renormalisation in 9-01 to 9-07. The rest of Part 5 shares the work. The expansion in small β is 5-02, the Hamiltonian 5-03, infinite volume on the strong-coupling side 5-04, $\beta \to \infty$ and $a(\beta)$ 5-05, numerical integration 5-07 and 5-08, and what to hold fixed in the limit 5-09. This chapter shows nothing about the existence of the two limits or their mass gap.

What this chapter can say

Item	What is said	What is not said
The definition and gauge invariance, finiteness of the measure	Exact at each lattice spacing (algebraic identities, compactness) Known . Also checked numerically Computed	Uniqueness of the infinite-volume limit (unknown in general at large β)
The classical continuum limit (6) to (8)	The formal expansion and $\beta = 2N/g^2$ Known . For one field, the difference proportional to a^2 Computed	Error estimates for general fields. The continuum limit of the quantum theory (built by the construction published in 2026-11, at the level of the committee's assessment; 6-00, 8-05)
Reflection positivity, topological charge	Presentation of claims in the literature	Proofs (5-03, 3-05)

Item	What is said	What is not said
The mass gap	—	This chapter shows nothing about it, on the lattice or in the continuum

Sources and reproduction

Item	Kind	Source / tools
The abstract of § 01 / the recollection in the detour	Primary source (1974: abstract only checked; 2005: text checked)	K. G. Wilson, “Confinement of Quarks”, Phys. Rev. D 10 (1974) 2445–2459 / “The Origins of Lattice Gauge Theory”, Nucl. Phys. B Proc. Suppl. 140 (2005) 3–19, arXiv:hep-lat/0412043
Quotations on finite-dimensional integrals, the two limits and reflection positivity (including the case where it is lost by regularisation)	Primary source (text checked)	A. Jaffe, E. Witten, “Quantum Yang–Mills Theory” (the official Clay problem description) § 6.5
β', the calculation of the expansion, $\log(1/a)$, the approximation of Wilson loops, uniqueness, the lattice correlation-length problem	Primary source (text checked)	S. Chatterjee, “Yang–Mills for probabilists”, arXiv:1803.01950 (2018) § 3– § 5
Topological charge	Primary source (abstract checked; agreement with the Chern number not read in the text)	M. Lüscher, “Topology of lattice gauge fields”, Commun. Math. Phys. 85 (1982) 39–48
Reflection positivity, Wegner's model, Elitzur's theorem	Bibliographic data only (checked with Crossref)	K. Osterwalder, E. Seiler, Ann. Phys. 110 (1978) 440–471 / F. J. Wegner, J. Math. Phys. 12 (1971) 2259–2272 / S. Elitzur, Phys. Rev. D 12 (1975) 3978–3982
Table 4, (8a)	This book's arithmetic (conditions in the normalisation of each reference)	H. Shen, R. Zhu, X. Zhu, Comm. Math. Phys. 400 (2023) 805–851 (abstract of arXiv:2204.12737; for the reading of the normalisation, 5-04)
Tables 1 to 3, Figure 1, Figure 2, numbers in the text	Computed on this machine	Python 3.12, NumPy 1.26, SciPy 1.11. `gauge.py` (Table 1, Table 2, § 03), `plaquette.py` ((9), the 3/2 of SU(3)), `continuum.py` (Table 3, (6); about 110 seconds). The figures compute the same formulas in plain JavaScript

Next chapter to read: 5-02 The strong-coupling expansion (expanding (10) at small β and seeing the area law).

The strong-coupling expansion — at small β the Wilson loop becomes a sum over surfaces, and the area law comes out

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — At small β the expectation value of a Wilson loop becomes a sum over surfaces bounded by the loop (the character expansion); from the smallest surface come the area law $\langle W \rangle \approx u^{RT}$ and the leading term of the string tension $\sigma_0 = -\ln u$, from bumps the first correction, and from tubes the leading term of the mass $m_0 = -4 \ln u$. We set them beside SU(2) numbers and see how far they work.

Prerequisite chapters — 5-01. Representations in 1-04, Haar measure in 1-08, mass in 4-11.

Said first — The strong-coupling area law is a theorem at small β with the lattice spacing fixed, and it holds for $U(1)$ too. It concerns the side opposite to the continuum limit ($\beta \to \infty$), and is not a proof of confinement or of the mass gap. The proof of convergence is left to 5-04.

Order of this chapter

1. The world of small β — the sum over surfaces seen in $\mathbb{Z}_2$
2. The character expansion — splitting the weight of one plaquette by representation (Table 1). Detour: the derivation of (4)
3. Tiling a surface — the leading term and the area law (Figure 1). Detour: sums over surfaces and strings
4. Bumps — the first correction and the number of dimensions
5. Tubes — the leading term of the mass and the correlation length
6. Comparing with numbers — how far it works (Figure 2, Table 2)
7. Convergence — finite boxes and infinite boxes. Where it stops
8. Is the area law a proof of confinement? — it comes out for abelian groups too

01

The world of small β — the sum over surfaces seen in $\mathbb{Z}_2$

A small coefficient $\beta = 2N/g^2$ in the Wilson action (5-01 § 04) means a large coupling g, so the expansion on this side is called the **strong-coupling expansion**. β plays the role of inverse temperature, so it is also called the **high-temperature expansion** (2-08). At $\beta = 0$ the links scatter independently with Haar measure, and since $\int_G U\, dU = 0$, the expectation value of a simple loop is 0. The abstract of Wilson 1974 writes "The lattice gauge theory has a computable strong-coupling limit; in this limit the binding mechanism applies and there are no free quarks. There is unfortunately no Lorentz (or Euclidean) invariance in the strong-coupling limit." Known .

Example: $\mathbb{Z}_2$. For the group of links ± 1 (2-09) with action $\beta \sum_P (1 - U_P)$, since $U_P = \pm 1$, the weight of one plaquette splits into two terms, up to a constant.

$$e^{\beta U_P} = \cosh\beta \left(1 + t\, U_P\right), \qquad t = \tanh\beta \quad (1)$$

Opening the product gives a term $t^{|X|} \prod_{P \in X} U_P$ for each set X of plaquettes. The average of a uniform ± 1 is 1 for even powers and 0 for odd ones, so in the partition function only sets in which every link is surrounded by an even number of plaquettes (closed surfaces) survive, and in the numerator of the Wilson loop $\prod_{\ell \in C} U_\ell$ only surfaces with boundary C. The smallest surface is a flat rectangle, so

$$\langle W_{R\times T}\rangle = t^{RT}(1 + \cdots), \qquad -\ln\langle W_{R\times T}\rangle \approx RT\,(-\ln\tanh\beta) \quad (2)$$

and the expectation value decreases exponentially in the area (the **area law**). That the gap for small β in the $\mathbb{Z}_2$ of 5-03 § 08 is close to $-4\log\tanh\beta$ is the same count.

02

The character expansion — splitting the weight of one plaquette by representation

For a general group too, the weight of one plaquette $e^{\frac{\beta}{N}\mathrm{Re}\,\mathrm{tr}\,U}$ is a function unchanged by conjugation VUV^{-1} (a **class function**). To split it by representation, we prepare just two tools.

Characters. For a representation r of the group (1-04), $\chi_r(U) = \mathrm{tr}\, r(U)$ is called its **character**. The trace is unchanged by conjugation, so characters are class functions. The characters of irreducible representations are orthonormal with respect to Haar

measure dU (the uniform probability measure on the group; 1-08 § 03) (Schur's orthogonality relations; proof omitted) Known .

$$\int_G \chi_r(U)\,\chi_s(U)\,dU = \delta_{rs}$$

Class functions can be expanded in these characters. It is the generalisation of Fourier series to non-commutative groups Known .

Integrals over classes of $SU(2)$. Elements of $SU(2)$ can be written as $U = V\,\mathrm{diag}(e^{i\theta}, e^{-i\theta})\,V^{-1}$ ($0 \le \theta \le \pi$), so a class function is a function of θ alone. Its integral takes the following form (the $SU(2)$ version of Weyl's integration formula) Known .

$$\int_{SU(2)} f(U)\,dU = \frac{2}{\pi}\int_0^\pi f(\theta)\,\sin^2\theta\,d\theta$$

Then $\mathrm{tr}\,U = 2\cos\theta$, the character of the spin-j representation is $\chi_j = \sin((2j+1)\theta)/\sin\theta$, and its dimension is $2j+1$.

Character coefficients

For a general group, the weight is expanded in characters as follows.

$$e^{\frac{\beta}{N}\mathrm{Re\,tr}\,U} = c_0(\beta)\Big[1 + \sum_{r\,/=0} d_r\,u_r(\beta)\,\chi_r(U)\Big], \qquad u_r = \frac{\int e^{\frac{\beta}{N}\mathrm{Re\,tr}\,U}\,\chi_r(U)\,dU}{d_r\int e^{\frac{\beta}{N}\mathrm{Re\,tr}\,U}\,dU} \quad (3)$$

d_r is the dimension. We call u_r the **character coefficients**, and write the one for the fundamental representation simply as u (for $\mathbb{Z}_2$, $u = t$). For $SU(2)$, the class integral above can be written with modified Bessel functions I_n (Detour: the derivation of (4)):

$$u_j = \frac{I_{2j+1}(\beta)}{I_1(\beta)}, \qquad u = \frac{I_2(\beta)}{I_1(\beta)} = \frac{\beta}{4} - \frac{\beta^3}{96} + \cdots \quad (4)$$

Known (agreeing with numerical integration to 2×10^{-13} Computed). It is the same number as the ratio λ_j/λ_0 of eigenvalues of the transfer matrix of a single link in 5-03 § 06.

If $\beta > 0$, then $0 < u_r < 1$. They are positive because each term of the opened exponential is a non-negative sum of characters of tensor products (5-03 § 06). They are below 1 because $\mathrm{Re}\,\chi_r(U) \le d_r$, with equality only when $r(U)$ is the identity matrix.

At small β, u approaches β for $\mathbb{Z}_2$, $\beta/2$ for $U(1)$, $\beta/4$ for $SU(2)$, and $\beta/(2N^2)$ for $SU(N)$ ($N \ge 3$) Computed . The reason is the first-order term. Opening the exponential of (3) to first order gives $u \approx \frac{\beta}{N^2}\int(\mathrm{Re\,tr}\,U)\,\chi_F(U)\,dU$ (F is the fundamental representation),

and this integral is $1/2$ for $N \geq 3$, since $\int (\operatorname{tr} U)^2 dU = 0$, and 1 for $\mathbb{Z}_2$ and $SU(2)$, since $\operatorname{tr} U$ is real Known .

For $SU(2)$, the β of 7-02 is $\beta/2$ of this chapter (compare the w of (3.3) in 7-02 with the u of (4) in this chapter; for $U(1)$ it is the same β).

Table 1 The character coefficient $u(\beta)$ of the fundamental representation. The action is $\beta \sum_P (1 - \frac{1}{N} \operatorname{Re} \operatorname{tr} U_P)$ in every case Computed

β	$\mathbb{Z}_2$	$U(1)$	$SU(2)$	$SU(3)$
0.5	0.4621	0.2425	0.1237	0.0289
1	0.7616	0.4464	0.2402	0.0601
2	0.9640	0.6978	0.4331	0.1286

Detour: the derivation of (4)

A subsection you may skip.

The weight of one plaquette for $SU(2)$ is $e^{\frac{\beta}{2} \cdot 2 \cos\theta} = e^{\beta \cos\theta}$. The formulas used are two standard properties of the modified Bessel functions I_n Known .

$$\int_0^\pi e^{\beta \cos\theta} \cos n\theta \, d\theta = \pi I_n(\beta), \qquad I_{n-1}(\beta) - I_{n+1}(\beta) = \frac{2n}{\beta} I_n(\beta)$$

Denominator. Since $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$,

$$\frac{2}{\pi} \int_0^\pi e^{\beta \cos\theta} \sin^2\theta \, d\theta = I_0 - I_2 = \frac{2}{\beta} I_1.$$

Numerator. By the product-to-sum formula $\sin^2\theta \cdot \frac{\sin((2j+1)\theta)}{\sin\theta} = \sin\theta \, \sin((2j+1)\theta) = \frac{1}{2}\left[\cos 2j\theta - \cos(2j+2)\theta\right]$, so

$$\frac{2}{\pi} \int_0^\pi e^{\beta \cos\theta} \cdot \tfrac{1}{2}\left[\cos 2j\theta - \cos(2j+2)\theta\right] d\theta = I_{2j} - I_{2j+2} = \frac{2(2j+1)}{\beta} I_{2j+1}.$$

The u_j of (3) is the numerator divided by $d_j = 2j+1$ and by the denominator, so $u_j = I_{2j+1}/I_1$. At small β, $I_n(\beta) \approx (\beta/2)^n/n!$ gives $u = I_2/I_1 \approx \beta/4$.

03 Tiling a surface — the leading term and the area law

Multiplying (3) over all plaquettes and opening the product, each term is an assignment of "which representation to put on which plaquette". There are two rules for integrating out the links one at a time (Schur's orthogonality relations Known ; (3.1) of 7-02 is a lemma of the same idea).

$$\int_G \chi_r(U)\,dU = \delta_{r0}, \qquad \int_G \chi_r(AU)\,\chi_s(U^{-1}B)\,dU = \delta_{rs}\,\frac{\chi_r(AB)}{d_r} \quad (5)$$

The left says "if an edge lies on only one plaquette, the result is 0", and the right says "gluing two plaquettes with the same representation along a common edge makes a larger loop, with a factor $1/d_r$". What survives are terms in which the plaquettes carrying representations form a closed surface, or a surface whose boundary is the observed loop.

Example: the 1×2 loop. The observable is $\frac{1}{d}\chi(U_C)$ (d is the dimension of the fundamental representation). The leading term is the one with representations on the two plaquettes inside the loop, $(du)^2\chi(U_{P_1})\chi(U_{P_2})$. For $SU(N)$, the surface carries the conjugate representation $\bar r$ so as to pair with the χ of the observable (for $SU(2)$, $\bar r = r$). We integrate in three steps.

1. Integrate the edge shared by the two plaquettes with (5) right. The two loops join into the boundary loop $\chi(U_\partial)$ of the 1×2, with a factor $1/d$.
2. Integrate one of the six boundary edges with (5) right. The loop of the observable and the boundary loop are glued with opposite orientations, the remaining five links cancel in the form VV^{-1}, and $\chi(1)/d = d/d = 1$ remains.
3. Normalisation: the $\frac{1}{d}$ of the observable remains (the partition function is 1 in the leading term apart from powers of the constant c_0, so dividing by it changes nothing).

Altogether $(du)^2\cdot\frac{1}{d}\cdot 1\cdot\frac{1}{d} = u^2$, that is, $\langle W_{1\times 2}\rangle = u^2$. Tiling an $R\times T$ too, the powers of d cancel, and what remains is

$$\langle W_{R\times T}\rangle = u^{RT}\big(1+O(u^4)\big), \qquad \sigma_0 = -\ln u \quad (6)$$

The energy of a static pair of charges $V(R) = -\lim_{T\to\infty}\frac{1}{T}\ln\langle W_{R\times T}\rangle$ (17-02 § 01) is $\sigma_0 R$ in the leading term: the confining form that grows in proportion to the distance Known . Example: for $SU(2)$ at $\beta = 1$, $\sigma_0 = 1.426$, and the 2×2 loop is $u^4 = 0.0033$ Computed .

Figure 1 The smallest surface bounded by the loop (thick line), and bumps. The sliders R, T, β change the number of plaquettes and the weight, and "Play" changes the position and orientation of the bump (the remaining five faces of a cube) (the two along the fourth axis enter only the count).



β=1.5 u=0.3441 u^6=0.00166 bumps 24 ways ×u⁴=0.0140 with correction 0.00233

Computed The numbers in this figure were computed on this machine (u from the series of (4), computed on the spot). If JavaScript is disabled: the surface has RT plaquettes with weight u^{RT}, and in four dimensions there are $4RT$ bumps, each with u^4.

Detour: sums over surfaces and strings

A subsection you may skip.

The abstract of Wilson 1974 states that the strong-coupling expansion becomes a sum over quark paths and the lattice surfaces connecting them, and concludes "This structure is reminiscent of relativistic string models of hadrons." A sum over surfaces paying $u = e^{-\sigma 0}$ per plaquette looks like a sum over world sheets of a string with an action proportional to area Known . Speaking in metaphor, it is a membrane stretched on the lattice. Precisely, physics considers that the strong-coupling surfaces are stiff along the lattice and in a different state from a fluctuating continuum string (17-03 § 07) Physics .

04

Bumps — the first correction and the number of dimensions

The next surfaces that matter replace one plaquette of the flat surface by the remaining five faces of the unit cube with that plaquette as its base (Figure 1). Paying five plaquettes instead of one multiplies the weight by u^4, and there are $2(d-2)$ such cubes, on both sides along the $d-2$ axes perpendicular to the surface: 4 in four dimensions. Separated bumps may be counted independently at this order, so together they make an exponential (the count for $SU(2)$).

$$\langle W_{R\times T}\rangle \approx u^{RT}\, e^{2(d-2)RT\,u^4}, \qquad \sigma 1 = -\ln u - 4u^4, \qquad \left\langle \tfrac{1}{2}\mathrm{tr}\, UP \right\rangle \approx u + 4u^5 \quad (d=4)$$

The next corrections come from overlapping bumps, two-storey bumps and surfaces carrying higher representations (higher coefficients are in the review of Drouffe–Zuber 1983, and the higher series for the string tension of $SU(2)$ in four dimensions in

Münster 1981; see the table of sources). Justifying the exponential form uniformly in the volume is the cluster expansion (5-04, 7-04) Known .

Two dimensions have no bumps. For $d = 2$ the surface cannot leave the plane, and the leading term u^{RT} is exact for every β (7-02 § 03) Known . The difficulty in three or more dimensions begins with the surface being able to fluctuate out of the plane. Example: for $SU(2)$ at $\beta = 2$, $4u^4 = 0.141$ reaches 17% of $\sigma_0 = 0.837$, and the truncation becomes doubtful Computed .

05

Tubes — the leading term of the mass and the correlation length

Mass is the rate at which correlations decrease (4-11). The smallest closed surface joining a spatial plaquette $P = \frac{1}{2}\mathrm{tr}\, UP$ and the same shifted by t in the time direction, $P(t)$, is the $1 \times 1 \times t$ box with the two as lids. The lids are observables and carry no weight, and the $4t$ side plaquettes pay u^{4t}.

$$\langle P(t)P(0)\rangle - \langle P\rangle^2 \approx C\, u^{4t}, \qquad m_0 = -4\ln u = 4\sigma_0, \qquad \xi_0 = 1/m_0 \ (8)$$

Computed (lowest-order counting). The lowest-order form is said to be standard (Drouffe–Zuber 1983, bibliographic data only). That the gap in the $\mathbb{Z}_2$ of 5-03 § 08 is close to $-4\log\tanh\beta$ has the same form. The correlation length ξ_0 is 0.175 at $\beta = 1$ and still 0.38 lattice units at $\beta = 2.6$: the correlation vanishes within one lattice unit Computed . The continuum limit needs ξ much longer than the lattice spacing ($\xi(\beta) \to \infty$ in the lattice correlation-length problem of Chatterjee 2018 (Chatterjee's Problem 5.1); 5-09), and the leading term says the opposite.

Extending the ratio $m_0/\sqrt{\sigma_0} = 4\sqrt{\sigma_0}$ of (8) outside the region of convergence, the ratio of leading terms falls to 0 as β grows. The ratio from lattice numbers rises from 3.3 ($\beta = 2.25$) to 3.8 ($\beta = 2.6$) and approaches a constant value (17-12 § 08) Physics . This concerns the form of the leading terms, not a statement about the theory. The leading terms at $\beta = 2.6$, $a\sqrt{\sigma_0} = 0.808$ and $am_0 = 2.61$, are 6.0 and 5.1 times the $0.13395(62)$ and $0.514(8)$ of Table 1 of Lucini–Teper (2001) Computed (the literature values are Physics). From $\beta = 2.4 \to 2.6$ the literature $a\sqrt{\sigma}$ halves, while the leading term decreases by only 4%.

06

Comparing with numbers — how far it works

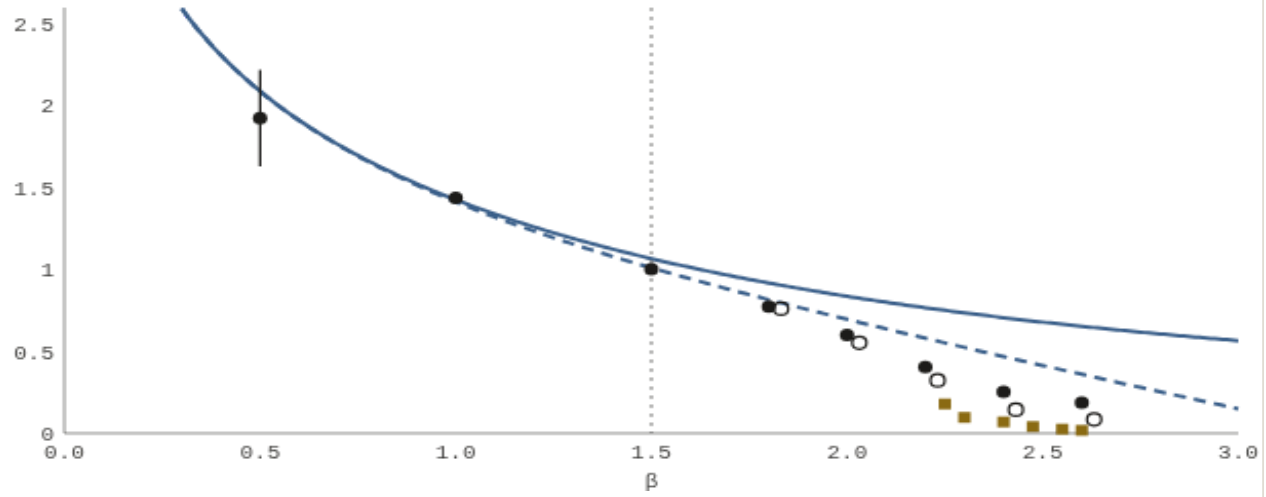
We integrated $SU(2)$ on four-dimensional periodic lattices (6 and 8 points per side) with the heat-bath method (an update that redraws the links one at a time from their

conditional distributions; 5-07), and measured the string tension with Creutz ratios ((3) of 17-02).

$$\chi(R) = -\ln \frac{W(R,R)\,W(R-1,R-1)}{W(R,R-1)^2} \tag{9}$$

The perimeter and corner terms cancel in the ratio, and with the form (7), $\chi(R) = \sigma_1$.

Figure 2 The string tension ($SU(2)$, lattice units). Solid line σ_0, dashed line σ_1, ● $\chi(2)$, ○ $\chi(3)$, ■ literature $(a\sqrt{\sigma})^2$. The slider moves the vertical line at β, and "Log scale" changes the vertical axis.



β=1.50 σ₀=1.067 σ₁=1.011 χ(2)=1.0026±0.0059

Computed The numbers in this figure were computed on this machine (curves computed on the spot, points the values of Table 2, ■ literature values). If JavaScript is disabled, Table 2 gives the content.

Table 2 Heat-bath results for four-dimensional $SU(2)$ (6 per side: $\beta \le 2$; 8: above that; statistical errors in brac and the strong-coupling formulas Computed . The last column is Lucini–Teper Physics

β	Plaquette	$u + 4u^5$	$\chi(2)$	$\chi(3)$	σ_1	Literature: plaquette / (a
0.5	0.12375(13)	0.1238	1.93(30)	–	2.089	
1.0	0.24335(14)	0.2434	1.438(32)	–	1.413	
1.5	0.36249(15)	0.3635	1.0026(59)	–	1.011	
1.8	0.44074(16)	0.4400	0.7730(21)	0.761(79)	0.816	
2.0	0.50104(22)	0.4941	0.5991(17)	0.553(33)	0.696	
2.2	0.56925(22)	0.5510	0.4054(12)	0.3231(55)	0.581	
2.4	0.63031(16)	0.6106	0.2525(11)	0.1445(23)	0.469	0.629995(17) / 0.
2.6	0.670275(68)	0.6729	0.18618(38)	0.08514(90)	0.360	0.670009(3) / 0.

There are three readings Computed . (i) For $\beta \le 1$ the plaquette agrees with $u + 4u^5$, and $\chi(2)$ with σ_1, within errors. (ii) From 1.8, $\chi(2)$ falls below σ_1, and at 2.4 it is nearly half (that the plaquette is close to $u + 4u^5$ at 2.6 is a coincidence of crossing curves). (iii) $\chi(3)$ is smaller still, yet 2 to 5 times the literature $(a\sqrt{\sigma})^2$ (small loops mix in short-

distance contributions Physics). The plaquette differs from the literature values by about 0.0003, four times the statistical error at $\beta = 2.6$. The difference may include finite-volume differences between the lattice of 8 per side and the literature's 12^4 lattice ($\beta = 2.4$) and 24^4 lattice ($\beta = 2.6$) Computed .

The abstract of Creutz (1980), who did the same kind of calculation, reports that when the lattice is made finer with the string tension held fixed, the bare coupling decreases (logarithmically) as predicted by the perturbative renormalisation group, and concludes “This supports the coexistence of confinement and asymptotic freedom for quantized non-Abelian gauge fields.” It can be read as the area law on the strong-coupling side and asymptotic freedom on the weak-coupling side appearing connected without being cut by a phase transition. That $\beta_c = \infty$ (the continuum limit only on the $\beta \to \infty$ side) for many non-abelian lattice models of interest in four dimensions is, in the words of Chatterjee (2018) § 5, at the level of “It is believed” Physics .

07

Convergence — finite boxes and infinite boxes

On a lattice with L points per side, $Z_L(\beta) = \int e^{-S_W} \prod dU_\ell$ is an integral over a finite-dimensional compact space and an entire function of β (a function expressible as a power series converging on the whole complex plane). Z_L itself is entire, but expectation values and $\ln Z_L/|\Lambda_L|$ ($|\Lambda_L|$ is the number of lattice points; the free energy) involve dividing by Z_L, so zeros of Z_L in the complex β plane shrink the radius of convergence of their expansions. The question is whether that radius of convergence survives independently of L (phase transitions in 2-08) Known . The number of connected surfaces grows with the number n of plaquettes as K^n (K fixed by the dimension), and each plaquette weighs u, so the sum converges for roughly $Ku < 1$ (a crude estimate; the precise condition is in 5-04): that is the core. The introduction of Shen–Zhu–Zhu (2023) summarises Osterwalder–Seiler (1978) as “when the coupling is sufficiently strong, the cluster expansion … for the expectation values of local observables … is convergent, uniformly in volume”, and writes that the mass gap (exponential clustering: correlations of distant observables decreasing exponentially in the distance; the mass gap of 4-11) and the existence and analyticity of the infinite-volume limit follow from it Known . SZZ themselves showed “a strictly positive mass gap” with Langevin dynamics (7-09) for $SU(N)$ with $|\beta_{\mathrm{SZZ}}| < 1/(16(d-1))$. From their formula (1.2), $\beta = N^2 \beta_{\mathrm{SZZ}}$, so for four-dimensional $SU(2)$, $\beta < 1/12$ (the conversion is this book's arithmetic). It is a sufficient condition, not the place where the expansion stops.

Where it stops

(a) Roughening: physics considers that at some β the fluctuations of the surface win, and the series for σ cannot be extended beyond it (17-03 § 07; about 2 for SU(2), with a large uncertainty in the estimate) Physics . (b) For four-dimensional $U(1)$ the area law breaks down at large β (§ 08). (c) Formulas truncated at finite order cannot express exponential smallness like $e^{-c\beta}$ as $\beta \to \infty$ (the leading term $-\ln u$ decreases only as a power, like $3/(2\beta)$). On the other hand, in the continuum limit of asymptotic freedom (4-06, 4-07), at one loop (the lowest order of the β function of 4-06) $a^2\sigma \propto e^{-(6\pi^2/11)\beta}$ should become exponentially small Physics . Taking β from 2 to 4 merely halves $-\ln u$, from 0.837 to 0.418, while $e^{-(6\pi^2/11)\beta}$ becomes about 5×10^4 times smaller Computed . The continuum-limit side cannot be reached with finite-order expansions (this book's organisation; 5-05).

The forms machine-checked on this book's definitions are in 20-01 to 20-04, and the strong-coupling area law is in 20-06. All are statements at fixed lattice spacing inside the strong-coupling window, at the opposite end from the questions on the side of the continuum limit.

08

Is the area law a proof of confinement? — it comes out for abelian groups too

The u of Table 1 is below 1 for every group, so the area law (6) comes out for $\mathbb{Z}_2$ and for $U(1)$ too. Chatterjee (2018) § 4 writes “the area law holds for any lattice gauge theory at sufficiently small β”, and goes on to say that therefore “the area law at small β cannot be evidence for quark confinement”. The proof is Osterwalder–Seiler (1978), whose abstract writes “We then prove existence and analyticity of the infinite volume limit of strongly coupled Yang-Mills theories on the lattice and we verify Wilson's confinement bound.” Known . The four-dimensional $U(1)$ lattice does not satisfy the area law at large β (Guth 1980; Guth showed it with the Villain form of the action, different from the Wilson form; a fully rigorous proof is Fröhlich–Spencer 1982; 5-06) Known . The strong-coupling area law is a property of the lattice being coarse, and does not use non-commutativity. The lattice area-law problem of the same paper (Chatterjee's Problem 4.1) asks to show, in any infinite-volume limit of a four-dimensional non-abelian compact group, for all $\beta > 0$,

$$|\langle W_{R\times T}\rangle| \le C(\beta)\, e^{-c(\beta)RT} \quad (10)$$

and writes that at large β it “remains elusive”. On the strong-coupling side, Cao–Nissim–Sheffield extended the range of the area law for $U(N)$ at large N (arXiv:2505.16585, a preprint, publication not checked; abstract: “This improves on a

classical result of Osterwalder-Seiler from 1978”), and in the 't Hooft regime (letting N grow with β proportional to N; for converting conventions, Table 4 of 5-01) they checked, with the SZZ method, a mass gap condition that the abstract cites as [DF80], and derived the area law from it (arXiv:2509.04688, a preprint, publication not checked) Known . Both are results at fixed lattice spacing. The mass gap and confinement are different statements (17-02 § 02; finite groups in 5-10), and this chapter shows neither for the continuum theory.

What this chapter can say

Item	What is said	What is not said
The character expansion, the leading terms (3) to (6)	Standard tools Known . Agreeing with numbers Computed	—
The correction (7)	Lowest-order counting for SU(2) in four dimensions. Agrees with the plaquette and $\chi(2)$ for $\beta \leq 1$ at $L = 6$ Computed	Higher coefficients, error estimates for truncation (5-04), infinite-volume values
The leading term of the mass (8)	Lowest-order counting Computed . Not compared with numbers on this machine. 5.1 times the literature value at $\beta = 2.6$	Numerical check of the mass, higher corrections
Theorems on convergence and the area law	Uniform convergence, the area law and exponential decay at small β (literature) Known	The content of the proofs. The area law at large β (the lattice area-law problem)
Confinement, the mass gap	Presentation of results on the strong-coupling side at fixed lattice spacing	Both in the continuum limit. This chapter shows nothing about them

Sources and reproduction

Item	Kind	Source / tools
Quotations in § 01 and the detour	Abstract	K. G. Wilson, Phys. Rev. D 10 (1974) 2445–2459
The results of Osterwalder–Seiler in § 07 and § 08, the quotation in § 08	Abstract (INSPIRE, Elsevier version; text not obtained)	K. Osterwalder, E. Seiler, Ann. Phys. 110 (1978) 440–471
The introduction in § 07, formula (1.2), the condition	Text § 1	H. Shen, R. Zhu, X. Zhu, Commun. Math. Phys. 400 (2023) 805–851, arXiv:2204.12737

Item	Kind	Source / tools
Quotations in § 05, § 06 and § 08, the lattice area-law problem and the lattice correlation-length problem (Chatterjee's Problems 4.1 and 5.1)	Text	S. Chatterjee, arXiv:1803.01950 (2018) § 4– § 6
Quotation in § 06	Abstract	M. Creutz, Phys. Rev. D 21 (1980) 2308–2315
Literature values in § 05, Table 2 and Figure 2	Tables 1 and 7 of the text Physics	B. Lucini, M. Teper, JHEP 0106 (2001) 050, arXiv:hep-lat/0103027
Extension of the area law in § 08	Abstracts (preprints; publication of neither checked)	S. Cao, R. Nissim, S. Sheffield, arXiv:2505.16585, arXiv:2509.04688 (2025). The [DF80] cited in the abstract of 2509.04688 is said to be Durhuus–Fröhlich 1980 (bibliographic data not checked)
The non-confining phase of $U(1)$ (Guth)	Abstract (INSPIRE)	A. H. Guth, Phys. Rev. D 21 (1980) 2291–2307
The non-confining phase of $U(1)$ (Fröhlich–Spencer)	Bibliographic data only (from Chatterjee's account)	J. Fröhlich, T. Spencer, Commun. Math. Phys. 83 (1982) 411–454
Review of higher coefficients, the lowest-order form of (8)	Bibliographic data only	J.-M. Drouffe, J.-B. Zuber, Phys. Rep. 102 (1983) 1–119
Higher series for the string tension of four-dimensional $SU(2)$	Abstract only	G. Münster, Nucl. Phys. B180 (1981) 23–60
The roughening transition in § 07	From the sources of 17-03	Itzykson, Peskin, Zuber, Phys. Lett. B 95 (1980) 259–264
Tables 1 and 2, the figures, numbers in the text	Computed on this machine	Python 3.12, NumPy, SciPy. `chars.py` (Table 1, (4)), `sc_tables.py` (§ 05, § 07), `mc_su2.py`, `run_mc.sh` (Table 2; heat bath, jackknife with 20 bins, 2 to 4 minutes per run). The figures are plain JavaScript

Next chapter to read: 5-03 Reflection positivity and the transfer matrix (extracting the Hamiltonian from the same lattice).

Reflection positivity and the transfer matrix — if the integral reflected in a mirror is positive, a Hamiltonian stands up on the lattice

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — The Wilson action is reflection positive for site reflection at every β, and for link reflection at $\beta \geq 0$. From this a positive transfer matrix and a lattice Hamiltonian stand up, and taking the time step to 0 gives the Kogut–Susskind form.

Prerequisite chapters — 5-01, 4-10. See also 2-09 (transfer matrices) and 7-02 § 03 (characters). β is the inverse lattice coupling of 5-01, a different quantity from the β function $\beta(g)$ of 4-06.

Said first — This is about finite lattices. On a finite box a positive gap is automatic, and this chapter shows nothing about whether the gap survives when the lattice spacing goes to 0 and the volume to infinity. The numbers come from calculations such as $\mathbb{Z}_2$ with space up to 4×4.

Order of this chapter

1. Where this chapter stands — from integrals to operators
2. Time reflection on the lattice — two kinds of plane (Figure 1)
3. A one-dimensional chain — two kinds of "positive"
4. Reflection positivity of the Wilson action — two proofs
5. The transfer matrix — the lattice Hamiltonian
6. A single link — the energy of the electric field
7. The Kogut–Susskind Hamiltonian — taking the time step to 0
8. Z_2 gauge theory — a transfer matrix that is not positive (Figure 2, Table 1)
9. What is kept, and what remains — limits, improved actions, what is not said

01

Where this chapter stands — with one mirror, quantum mechanics comes back

The theory on a lattice (5-01) is written as the following probability measure.

$$d\mu_\beta(U) = \frac{1}{Z}\, e^{-S_W(U)} \prod_\ell dU_\ell, \qquad S_W = \beta \sum_P \Big(1 - \tfrac{1}{N} \operatorname{Re}\operatorname{tr} U_P\Big), \qquad \beta = \frac{2N}{g^2} \quad (1)$$

dU is the Haar measure of total weight 1, and β the coefficient of the action (for example, $\beta = 4$ for $SU(2)$ with $g^2 = 1$). The last coordinate t is called imaginary time. Here there are only integrals and expectation values, with no Hilbert space and no Hamiltonian. The mass gap is a gap in the spectrum of the Hamiltonian (4-11), so we first recover the operators. The tool is the reflection positivity of 4-10, which on the lattice can be proved from the action. § 6.5 of the problem statement writes 'Reflection positivity holds for the Wilson approximation [36], a major advantage' ([36] is Osterwalder–Seiler 1978) Known .

02

Time reflection on the lattice — two kinds of plane

For the lattice reflected in the mirror to coincide with the original, the $2t_0$ of the plane $t = t_0$ must be an integer. If t_0 is an integer, the plane passes through lattice points (**site reflection**); if it is a half-integer, it cuts the midpoints of the vertical (time-direction) links (**link reflection**).

$$\theta(\boldsymbol{x}, t) = (\boldsymbol{x}, 2t_0 - t), \qquad (\theta U)_{x \to y} = U_{\theta x \to \theta y}, \qquad (\Theta F)(U) = F(\theta U) \quad (2)$$

A link variable traversed backwards is the inverse ($U_{y\to x} = U^{-1}_{x\to y}$). Let Λ_+ be the set of links whose two ends both have $t \geq t_0$ (for link reflection, $t > t_0$), and $\mathcal{A}_+$ the set of all functions determined by those variables alone. Reflection positivity is the condition of the same form as (4) of 4-10,

$$\langle\, \Theta F \; F \,\rangle \geq 0 \qquad (F \in \mathcal{A}_+) \quad (3)$$

θ maps plaquettes to plaquettes (reversing the orientation of those containing the time direction), and $\operatorname{Re}\operatorname{tr}$ is unchanged by inversion, so $S_W(\theta U) = S_W(U)$.

Figure 1 Two kinds of mirror — Moving the plane t_0 by half a lattice unit at a time with the slider, site reflection and link reflection appear alternately. Blue is the positive half and F (a rectangular Wilson loop), brown its image.

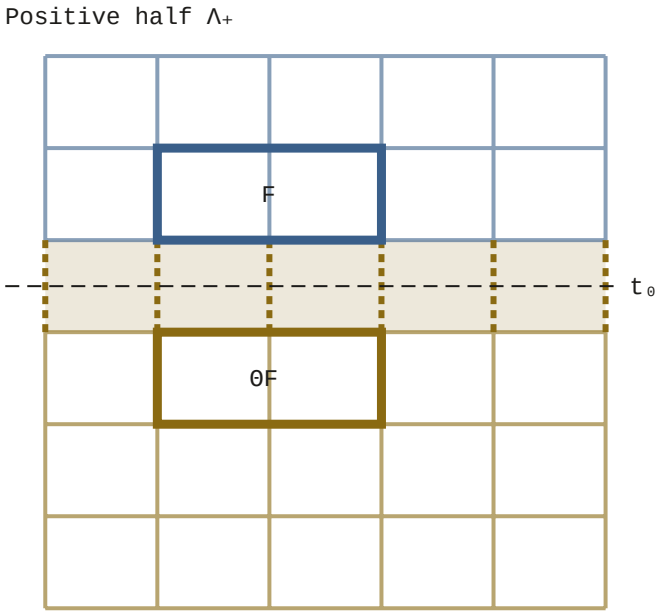


Image θΛ₊

t₀ = 1/2 (link reflection): the dashed links straddle the plane, and the shaded plaquettes join the two halves.

Example (Figure 1). At $t_0 = \frac{1}{2}$ a vertical link crossing the plane is mapped to itself with orientation reversed ($(\theta U)_b = U_b^{-1}$), belongs to neither half, and straddles the plane. The plaquettes having it as an edge join the upper and lower halves. At $t_0 = 0$ the horizontal links at $t = 0$ are mapped to themselves, and nothing straddles.

03

Seen in a one-dimensional chain — two kinds of "positive"

Put the weight $\prod_t e^{K s_t s_{t+1}}$ on a chain of spins $s_t = \pm 1$ (the one-dimensional Ising model of 2-09 with zero external field; $K < 0$ is also allowed). The transfer matrix is

$$T = \begin{pmatrix} e^{K} & e^{-K} \\ e^{-K} & e^{K} \end{pmatrix}, \qquad \lambda_0 = 2\cosh K, \quad \lambda_1 = 2\sinh K \tag{4}$$

For **site reflection** ($t_0 = 0$), fixing s_0, the sum $\varphi(s_0)$ over the upper half and the sum over the lower half are complex conjugates of each other, and whatever the sign of K, $\langle \Theta F F \rangle \propto \sum_{s_0} |\varphi(s_0)|^2 \geq 0$. For **link reflection** ($t_0 = \frac{1}{2}$) one straddling coupling remains, and

$$\langle \Theta F F \rangle \;\propto\; \sum_{s0, s1} \varphi(s_0)\; T_{s_0 s_1}\; \varphi(s_1) \tag{5}$$

is a quadratic form in T. It is non-negative for every φ only when $T \geq 0$, that is, $K \geq 0$. Example. For $F = s_1$, with link reflection $\langle s_0 s_1 \rangle = \tanh K$ becomes -0.4621 at $K = -\frac{1}{2}$ and (3) is violated, while with site reflection $\langle s_{-1} s_1 \rangle = \tanh^2 K = 0.2136 \geq 0$ Computed .

There are two kinds of "positive". One is that the entries are positive. For a matrix with all entries positive, the largest eigenvalue is not repeated and is larger than the absolute values of the other eigenvalues (the Perron–Frobenius theorem) Known . The entries of the T in (4) are positive whatever the sign of K, so $\lambda_0 > |\lambda_1|$, and the correlation $(\tanh K)^t$ decreases exponentially. The other is being positive as an operator. $T \geq 0$ only when $K \geq 0$, and if $K > 0$ one can write $T = e^{-H}$ with a real H. For $K < 0$ the correlation decreases while changing sign, and $-\log \lambda_1$ is not real. Site-reflection positivity corresponds to $T^2 \geq 0$, and link-reflection positivity to $T \geq 0$ Known .

04

Reflection positivity of the Wilson action — two proofs

Claim: on a lattice with the Wilson action, (3) holds for all real β with site reflection, and for $\beta \geq 0$ with link reflection. Reflection positivity of lattice gauge theory and the existence of a positive self-adjoint transfer matrix are due to Osterwalder–Seiler 1978. Their abstract states that physical positivity was verified, and continues 'This implies the existence of a positive self-adjoint transfer matrix' Known . The conditions on β for each kind of reflection rest on the proofs (6) to (8) below Paper . An example where link reflection actually fails at $\beta < 0$ is in section 08 Computed . Below are outlines of the proofs (with periodic time there are two planes, but it is the same).

Site reflection — fixing the variables on the plane

No plaquette straddles. Distributing the action of the plaquettes on the plane half up and half down gives $S_W = S_+ + S_-$ and $S_-(U) = S_+(\theta U)$. Fixing the variables U_0 on the plane, the lower integral becomes the complex conjugate of the upper one by the substitution $U_< = \theta U_>$ (Haar measure is invariant under inversion), so

$$\langle \Theta F F \rangle = \frac{1}{Z} \int dU_0 \left| \int dU_> \, F \, e^{-S_+} \right|^2 \geq 0 \quad (6)$$

The sign of β is used nowhere.

Link reflection — opening the straddling plaquettes into sums of squares

The straddling vertical links have distinct upper endpoints, so gauge transformations at those points can set them to 1 (7-02 § 02). For this, a general F is first replaced by $\widetilde{F}$, averaged over gauge transformations at the points of the layer just above the plane ($t = 1$). ΘF is determined by the variables with $t \leq 0$ only, and does not touch the points of that layer. So changing the integration variables $U \to U^g$ by a gauge transformation g of that layer (neither the action nor Haar measure changes), $\langle \Theta F F \rangle = \langle \Theta F F^g \rangle$, and

averaging over g gives $\langle \Theta F \widetilde{F} \rangle$. Doing the same at the mirror layer ($t = 0$), the left-hand side of (3) equals $\langle \Theta \widetilde{F}\, \widetilde{F} \rangle$. A straddling plaquette becomes $\mathrm{Re}\,\mathrm{tr}(U_{\theta\ell} U_\ell^\dagger)$, involving only the upper horizontal link ℓ and its image, and with the entries $h_{ij} = (U_\ell)_{ij} \in \mathcal{A}_+$ ($\Theta h_{ij} = (U_{\theta\ell})_{ij}$) it can be written

$$\mathrm{Re}\,\mathrm{tr}\big(U_{\theta\ell} U_\ell^\dagger\big) = \frac{1}{2} \sum_{i,j} \Big(\Theta h_{ij}\, h_{ij} + \Theta h_{ij}\, h_{ij} \Big) \tag{7}$$

Expanding the weight of the straddling plaquettes $\exp\big(\frac{\beta}{2N} \sum_k \Theta h_k h_k\big)$ (h_k ranges over all h_{ij} and h_{ij}; constant factors omitted) as a series, it takes the form $\sum_\alpha c_\alpha \Theta H_\alpha H_\alpha$ with products H_α of the h_k, and if $\beta \geq 0$ then $c_\alpha \geq 0$. The upper and lower halves no longer share variables, so

$$\langle \Theta F F \rangle \propto \sum_\alpha c_\alpha \Big| \int dU_> \, F\, H_\alpha \, e^{-S_+} \Big|^2 \geq 0 \tag{8}$$

Just as $T \geq 0$ was needed in (5) for the chain, $\beta \geq 0$ is needed here. An example where it actually fails is in section 08.

The forms machine-checked on this book's definitions are in 20-01 to 20-04.

05

The transfer matrix — the lattice Hamiltonian

Given reflection positivity, the reconstruction of 4-10 makes a Hilbert space and an operator that advances time by one lattice unit; on the lattice this operator (the transfer matrix) can be written down directly. Make time periodic (length L_t) and set as many vertical links as possible to 1 by gauge transformations (temporal gauge); then one link V_x remains on the time loop at each spatial point x. Integrating over the remaining V_x is the same as applying the gauge transformation $g_x = V_x$ to all links at one time and averaging, and that average is the projection P_0 onto gauge-invariant states. The state space is the space $\mathcal{H} = L^2(G^{E_s}, dU)$ of functions of the spatial links E_s at one time, and T is the integral operator with kernel

$$T(U, U') = e^{-\frac{1}{2} S_s(U)} \prod_{\ell \in E_s} e^{-\beta(1 - \frac{1}{N} \mathrm{Re}\,\mathrm{tr}\, U_\ell U_\ell'^{\dagger})}\, e^{-\frac{1}{2} S_s(U')} \tag{9}$$

(S_s is the action of the spatial plaquettes at one time, and the product in the middle comes from the vertical plaquettes). Then

$$Z = \mathrm{Tr}\big(P_0 T^{L_t}\big), \qquad \langle O(t) O(0) \rangle - \langle O \rangle^2 \xrightarrow{L_t \to \infty} \sum_{n \geq 1} |\langle \Omega | O | n \rangle|^2 \Big(\frac{\lambda_n}{\lambda_0} \Big)^t \tag{10}$$

(O is a gauge-invariant real observable, λ_n and $|n\rangle$ the eigenvalues and eigenvectors of $P_0 T$, $\Omega = |0\rangle$; the same calculation as (8) of 2-09).

(a) The kernel is real symmetric, so T is self-adjoint (the content of site reflection). (b) $T = VKV$ (V the operator multiplying by $e^{-S_s/2}$, K the product in the middle), and if $\beta \geq 0$ then $K \geq 0$ (section 06), so $T \geq 0$ (the content of link reflection). (c) For $\beta > 0$, V multiplies by a positive function and so has an inverse, and the λ_r of section 06 are all positive, so K does not have 0 as an eigenvalue either, and T does not have 0 as an eigenvalue Paper . Lüscher 1977 calls this property 'strictly positive' in its title. (d) On a finite box the kernel of $P_0 T$ is continuous and positive and the space of gauge orbits is compact, so by a generalisation of Jentzsch's theorem on integral operators with continuous positive kernels on compact spaces (of Krein–Rutman type; the integral-operator version of the Perron–Frobenius theorem), λ_0 is not repeated Known (the generalisation is Schaefer 1974; see the table of sources).

From here on we take $\beta > 0$. By (b)(c) all eigenvalues are positive, so since λ_0 is not repeated, $\lambda_1 < \lambda_0$. Thus on $P_0 \mathcal{H}$

$$H_a = -\frac{1}{a} \log \frac{T}{\lambda_0} \;\geq\; 0, \qquad m_L(\beta) = \log \frac{\lambda_0}{\lambda_1} \;>\; 0 \quad (11)$$

stands up (a is the lattice spacing). Example. The gap in physical units is $\Delta = m_L / a$. That $m_L > 0$ for each box ($\beta > 0$) is automatic from (d), and what is asked is whether Δ remains positive as $L \to \infty$ and $a \to 0$ (4-11 § 08, 5-09).

For how this book uses this, see 26-04 § 01 (positivity of the transfer matrix and reflection positivity applied to the bias of the twisted sectors).

06

The transfer matrix of a single link — the energy of the electric field appears

K is a product of operators k, one per link. For a single link $(kf)(U) = \int_G e^{\frac{\beta}{N} \mathrm{Re}\,\mathrm{tr}\, UU'^{-1}} f(U')\, dU'$, and the kernel is a function unchanged by conjugation $W \to hWh^{-1}$ of $W = UU'^{-1}$ (a class function). Diagonalising it needs two words from representation theory.

- We write the **character** of a representation r (of dimension d_r) as $\chi_r(W) = \mathrm{tr}\, r(W)$. It is a class function, and between irreducible representations $\int_G \chi_r(W) \chi_s(W)\, dW = \delta_{rs}$ (orthogonality) Known .
- Convolution with a class function acts as a scalar multiple on the space spanned by the matrix elements $r(U)_{ij}$ of each irreducible representation (from Schur's lemma

in 1-04). $L^2(G)$ is spanned by these matrix elements (the Peter–Weyl theorem), so k is determined by its eigenvalue on each irreducible representation Known .

For details, (3.1) in 7-02 § 03. The eigenvalue on the representation r is

$$\lambda_r = \frac{1}{d_r}\int_G e^{\frac{\beta}{N}\operatorname{Re}\operatorname{tr} W - \beta}\,\chi_r(W)\,dW \tag{12}$$

Expanding the exponential in powers of $\operatorname{tr} W$ and $\operatorname{tr} W$, each power is a character of a tensor-product representation, and the multiplicities of irreducible representations are non-negative, so if $\beta \geq 0$ then $\lambda_r \geq 0$, that is, $K \geq 0$. If $\beta > 0$ all the coefficients of the expansion are positive, and every irreducible representation appears in some tensor product of the fundamental representation and its complex conjugate Known , so $\lambda_r > 0$. For $SU(2)$, with the eigenvalues of W written $e^{\pm i\theta}$ ($0 \leq \theta \leq \pi$), $\operatorname{tr} W = 2\cos\theta$ and $\chi_j = \sin((2j+1)\theta)/\sin\theta$. The Haar measure for integrating class functions is $\frac{2}{\pi}\sin^2\theta\,d\theta$, which is the density of the uniform distribution on S^3 of 5-01 seen through the angle θ from the north pole (the simplest case of Weyl's integration formula). From these Known

$$\lambda_j = e^{-\beta}\,\frac{2}{\beta}\,I_{2j+1}(\beta), \qquad j = 0, \tfrac{1}{2}, 1, \ldots \tag{13}$$

($I_n(\beta) = \frac{1}{\pi}\int_0^\pi e^{\beta\cos\theta}\cos n\theta\,d\theta$ is the modified Bessel function; agreeing with numerical integration to relative error below 10^{-10} Computed ; the ratio λ_j/λ_0 is the u_j of (4) in 5-02). For $\beta < 0$ it is negative for half-integer j. From the asymptotic form at large β, $I_\nu(\beta) \approx \frac{e^\beta}{\sqrt{2\pi\beta}}\left(1 - \frac{4\nu^2-1}{8\beta}\right)$ (DLMF 10.40.1), the energy per step is

$$E_j = -\log\frac{\lambda_j}{\lambda_0} \;\approx\; \frac{2j(j+1)}{\beta} = \frac{g^2}{2}\,j(j+1) \tag{14}$$

$j(j+1)$ is the value of the Casimir E^aE^a (the J^2 of 1-04). Speaking in metaphor, a single link is a top spinning on $SU(2)$ (a three-dimensional sphere), and E^a is its angular momentum. Precisely, at large β, k approaches the heat kernel $e^{-\frac{g^2}{2}E^aE^a}$ on the group (7-02 § 04).

Example Computed . $E_{1/2}$ is 0.0967 at $\beta = 16$ ($g^2 = \frac{1}{4}$) ((14) gives 0.0938), and 0.0236 at $\beta = 64$ (0.0234). The deviation of the ratio $E_{1/2}/(14)$ (the relative error) shrinks from 1.032 to 1.008, roughly as $1/\beta$ (the difference itself shrinks from 0.00295 to 0.00018, as $1/\beta^2$). At strong coupling, $\beta = 1$, $E_1 = 3.24$ departs from the approximation 4. For the ratio $E_1/E_{1/2} \to 8/3$, see 7-02 § 07.

07

The Kogut–Susskind Hamiltonian — taking the time step to 0

The H_a of (11) is the logarithm of an integral operator, and its form is not visible. A local form becomes visible if only the spacing a_t in the time direction is made finer than the spatial spacing a. In four dimensions, we put separate coefficients on the plaquettes containing the time direction and those that are purely spatial,

$$\beta_t = \frac{2N}{g^2}\,\frac{a}{a_t}, \qquad \beta_s = \frac{2N}{g^2}\,\frac{a_t}{a} \tag{15}$$

From the number of plaquettes per volume $1/(a^3 a_t)$ and (7) of 5-01, the classical continuum limit is the action (8) of 5-01 in either direction. As $a_t \to 0$, $\beta_t \to \infty$, so for a step in the time direction the general form of (14), $-\log(\lambda_r/\lambda_0) \approx \frac{N}{\beta_t} C_2(r) = a_t \frac{g^2}{2a} C_2(r)$, can be used ($C_2$ is the value of the Casimir, normalised by $\operatorname{tr} T^a T^b = \frac{1}{2}\delta^{ab}$). This is because near $W = e^{iX^a T^a}$, $\frac{\beta}{N}\operatorname{Re}\operatorname{tr} W \approx \beta - \frac{\beta}{4N} X^a X^a$, and k is close to the heat kernel at time N/β (for the fundamental representation of $SU(3)$ at $\beta = 640$, the ratio is 1.003 Computed). The spatial part $V^2 = e^{-S_s}$ has $\beta_s \propto a_t$, so it has the form $e^{-a_t(\cdots)}$ from the start. Together $T = e^{-a_t H_{\mathrm{KS}} + O(a_t^2)}$, with

$$H_{\mathrm{KS}} = \frac{g^2}{2a}\sum_{\ell} E_\ell^a E_\ell^a + \frac{2N}{g^2 a}\sum_{p}\Big(1 - \frac{1}{N}\operatorname{Re}\operatorname{tr} U_p\Big) \tag{16}$$

The first term is the energy of the electric field (the spinning of the top on each link), the second that of the magnetic field, and the states are restricted to those satisfying the Gauss law ($P_0\mathcal{H}$). This form was given by Kogut–Susskind 1975, and the abstract of Creutz 1977 writes 'Taking the timelike lattice spacing to zero, we extract the Kogut-Susskind Hamiltonian from the transfer matrix in the gauge A0 = 0' Known . The coefficients of (15)(16) are calculations in the conventions of this chapter, and the limit is an approximation eigenvalue by eigenvalue. From canonical quantisation, see 4-01.

Example. When g^2 is large the first term dominates, and the first excitation is a state in which an electric field in the fundamental representation goes once round the boundary of a single plaquette, with energy $\frac{g^2}{2a}\cdot 4 \cdot C_2(\mathrm{fund})$. For $SU(N)$, $C_2(\mathrm{fund}) = \frac{N^2-1}{2N}$, and for $SU(2)$ ($j = \frac{1}{2}$, $C_2 = \frac{3}{4}$) it is $\frac{3g^2}{2a}$. The energy of a line of electric field is proportional to its length (5-02).

For how this book uses this, see 24-05 § 04 (reading the gap as the Poincaré constant of the vacuum measure on the Hamiltonian lattice).

08

Seeing everything in $\mathbb{Z}_2$ gauge theory — a transfer matrix that is not positive

Take the group to be $\mathbb{Z}_2 = \{\pm 1\}$ (the detour of 2-09, 5-10), spacetime three-dimensional (space a periodic $L \times L$ lattice), and the action $\beta \sum_P (1 - U_P)$. The factor for a single link is $e^{-\beta}$ times the matrix (4) with K replaced by β, and the ratio of its eigenvalues is $\tanh \beta$. For $\beta < 0$ it is negative.

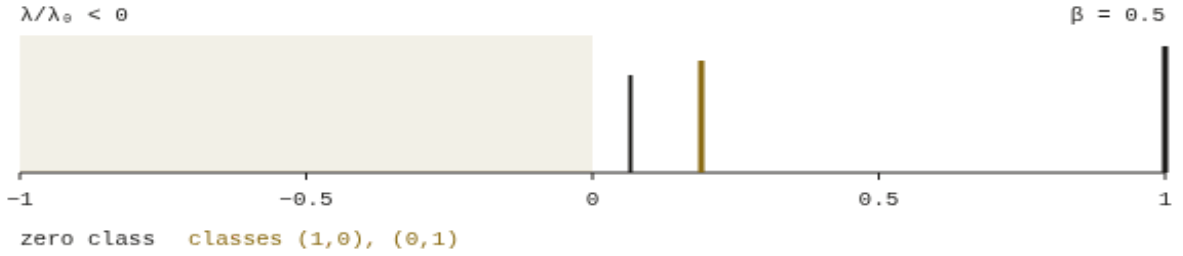
In the eigenvectors of the factor of each link (the electric-field basis), it is decided for each link whether a line of electric field passes through it. The Gauss law says that an even number of lines meet at each point, gauge-invariant states are labelled by collections e of closed lines, and K is a diagonal matrix determined only by the total length $|e|$ of the lines.

$$K(e) \;\propto\; (\tanh \beta)^{|e|} \qquad (17)$$

Closed lines fall into four classes by the parities of their winding numbers in the x and y directions, and V (which adds or removes the line round a plaquette's boundary) preserves the class. Adding a boundary does not change the parity of the length, so lines in class $(1,0)$ have length of the same parity as L. Hence for $\beta < 0$ with L odd, $K \leq 0$ in classes $(1,0)$ and $(0,1)$, and $T = VKV \leq 0$. If L is even, the ratios λ/λ_0 are the same for β and $-\beta$. Flipping the signs of the x-direction links on every other row gives $U_p \to -U_p$, and $V(-\beta)$ becomes similar to a constant multiple of $V(\beta)$. If L is even, all closed lines have even length, so the K of (17) also changes only by a constant factor Paper (agreeing for all eigenvalues at $L = 2$ and the top ones at $L = 4$ Computed).

Each class has dimension 2^{L^2-1}; $L \leq 3$ is diagonalised directly and $L = 4$ with the Lanczos method (a method that finds only the extreme eigenvalues of a large symmetric matrix by iteration) (at $L = 3$ the two agree to the 8 to 11 digits printed Computed). Let $F = W(1)$ be the Wilson line going once round in the x direction at time 1. The left-hand side of (3) for link reflection ($t_0 = \frac{1}{2}$) is $\langle W(0)W(1)\rangle = \sum_n |\langle 0|W|n\rangle|^2 \lambda_n/\lambda_0$ (n in class $(1,0)$), and at $L = 3$, $\beta = -0.5$ it is -0.1306, negative. For site reflection ($t_0 = 1$, $F = W(2)$), $\langle W(0)W(2)\rangle = +0.0183$. Monte Carlo without the transfer matrix ($3 \times 3 \times 18$) also gives $-0.1305(18)$ and $+0.0189(11)$, agreeing within errors Computed .

Figure 2 Eigenvalues of the $\mathbb{Z}_2$ transfer matrix (space 3×3) — Moving β (−1 to 2) with the slider moves the vertical lines to the positions of the eigenvalue ratios λ/λ_0 (black: λ_0 and λ_1 of the zero class; brown: the one of largest absolute value in classes $(1,0)$ and $(0,1)$).



β = 0.5: zero class λ_1/λ_0 = 0.0663 (m = 2.71), largest λ/λ_0 in class (1,0) = 0.190. All eigenvalues positive.

Computed The numbers in this figure were computed on this machine.

Computed Table 1 The gap $m = \log(\lambda_0/\lambda_1)$ of the zero class (states with no winding lines), and the lowest energy $E_w = \log(\lambda_0/\lambda_0^{(1,0)})$ of class $(1,0)$ (lattice units).

β	m (L=3)	m (L=4)	−4 log tanh β	E_w (L=3)	E_w (L=4)
0.10	9.233	9.233	9.224	6.915	9.223
0.75	1.587	1.153	1.816	0.1860	0.1443
1.00	2.991	2.840	1.089	0.0200	0.00459

How to read it. (i) For $\beta > 0$ all eigenvalues are positive and $m > 0$ ((d) of section 05). (ii) At small β, $m \approx -4 \log \tanh \beta$ (a line of length 4; the m_0 of 5-02), and $E_w \approx -L \log \tanh \beta$ (proportional to the length of the winding line: the confining form). (iii) At large β, E_w becomes rapidly smaller with L. (iv) Near $\beta = 0.75$, m moves a lot with L. Three-dimensional $\mathbb{Z}_2$ gauge theory is dual to the three-dimensional Ising model and is said to have a phase transition (Wegner 1971) Known , and the m of finite boxes alone does not determine the infinite-volume gap. (v) For $\beta < 0$ with $L = 3$, all 512 eigenvalues of classes $(1,0)$ and $(0,1)$ are at most 0 (the smallest is $-0.788\lambda_0$ at $\beta = -1$), and $\langle W(0)W(t)\rangle$ changes sign with the parity of t (-0.0027 at $\beta = -0.5$, $t = 3$) Computed .

09

What is kept, and what remains

Kept in the limit. (3) is an inequality, so if expectation values converge along a sequence of lattices and the mirror plane lies on a lattice plane at every stage, the limiting measure is also reflection positive (7-01 § 07) Known . In the four-dimensional non-abelian case, the existence of the continuum limit was shown by the construction published in 2026-11, and the reflection positivity of the limiting correlations is also stated there (the committee's assessment is that it is proved in Lean; 12-01, 12-05). The limit, however, is at the level of correlation functions; a limit as a measure has not been constructed (6-04).

Can be lost when the action is changed. For improvements adding dimension-6 terms (the Symanzik improvement of 5-05), the abstract of Lüscher–Weisz 1984 writes

'physical positivity is lost in general', and states that the transfer matrix can have complex eigenvalues (correlations with damped oscillation) and that negative weights necessarily appear in the spectral decomposition of the two-point function Known . The $\beta < 0$ of section 08 is the simplest example of a transfer matrix that is no longer positive (the eigenvalues are real and negative). For continuum regularisations, see 4-10 § 08.

What is not said. The tools of this chapter do not answer whether $mL > 0$ survives as $L \to \infty$ (the strong-coupling side is 5-04) and $a \to 0$. The recovery of rotational symmetry is also a task on the side of the limit.

What this chapter can say

Content	Status
Reflection positivity of lattice gauge theory and a positive self-adjoint transfer matrix	Known (abstract of Osterwalder–Seiler 1978)
The conditions on β in (3): all β for site reflection, $\beta \geq 0$ for link reflection	Paper ((6) to (8) of this chapter)
An example where link reflection fails at $\beta < 0$	Computed (the $\mathbb{Z}_2$ of section 08, $L = 3$)
T self-adjoint, positive for $\beta \geq 0$	Known (Osterwalder–Seiler 1978, Lüscher 1977)
For $\beta > 0$, T does not have 0 as an eigenvalue	Paper ((c) of section 05, $\lambda_r > 0$ of section 06)
Simplicity of λ_0, $mL > 0$ for each box ($\beta > 0$)	Known (a generalisation of Jentzsch's theorem; Schaefer 1974)
(12) to (16)	The formulas are Known ; the numbers of section 06 and the check of the general form for $SU(3)$ are Computed . The coefficients of (15)(16) are in this chapter's conventions, and $a_t \to 0$ is an approximation eigenvalue by eigenvalue
Table 1 and Figure 2 for $\mathbb{Z}_2$ ($L \leq 4$)	Computed (diagonalisation and independent Monte Carlo)
The mass gap, the continuum limit	Nothing shown

Sources and reproduction

Item	Kind	Source / tools
Table 1, Figure 2, numbers in the text	Computed	python3 (NumPy, SciPy): `table1.py`, `one_link.py`, `z2_transfer.py`, `check_lanczos.py`, `even_L_check.py`, `winding_corr.py`, `mc_check.py`, `fig_data.py`, `su3_casimir.py`. $\langle W(0)W(t)\rangle$ in two ways, by spectral representation and by Monte Carlo
Problem statement § 6.5	Known	A. Jaffe, E. Witten, "Quantum Yang–Mills Theory" (2000). Text checked
Reflection positivity, the transfer matrix	Known	K. Osterwalder, E. Seiler, "Gauge field theories on a lattice", Ann. Phys. 110 (1978) 440–471 (abstract checked, text not checked) / M. Lüscher, Commun. Math. Phys. 54 (1977) 283–292 (title and abstract checked)
(16), improved actions	Known	J. Kogut, L. Susskind, Phys. Rev. D 11 (1975) 395–408 / M. Creutz, Phys. Rev. D 15 (1977) 1128–1136 / M. Lüscher, P. Weisz, Nucl. Phys. B 240 (1984) 349–361. Abstracts of all checked
Perron–Frobenius, Jentzsch	Known	O. Perron, Math. Ann. 64 (1907) 248–263 / R. Jentzsch, "Über Integralgleichungen mit positivem Kern", J. reine angew. Math. 141 (1912) 235–244. Bibliographic data of both checked with Crossref; text bibliographic data only (Jentzsch's claim [for a continuous positive kernel on a square, the spectral radius is a simple eigenvalue with a positive eigenfunction] checked in the secondary source arXiv:1910.01120). Generalisation to kernels on compact spaces: the theorem on p. 337 of H. H. Schaefer, *Banach Lattices and Positive Operators*, Springer (1974) (for an integral operator with kernel $N \geq 0$, some power of which is compact, if the integral of the kernel linking any set S of positive measure with its complement is positive, then the spectral radius is a simple eigenvalue with an eigenfunction positive almost everywhere; if $N > 0$, the other eigenvalues are smaller in absolute value). Bibliographic data only (the claim checked through the citations in arXiv:1910.01120 and arXiv:1501.06011; a continuous positive kernel on a finite box is Hilbert–Schmidt and so satisfies the assumptions). M. G. Krein, M. A. Rutman, Uspekhi Mat. Nauk 3:1(23) (1948) 3–95 (English translation Amer. Math. Soc. Translation 26, 1950). Bibliographic data checked with Math-Net.Ru; text bibliographic data only
The asymptotic form of I_ν	Known	NIST DLMF 10.40.1, 10.17.1. Formulas checked
The duality of $\mathbb{Z}_2$ and Ising	Known	F. J. Wegner, J. Math. Phys. 12 (1971) 2259–2272. Bibliographic data only

Next chapter to read: 5-04 The infinite-volume limit and the cluster expansion.

The infinite-volume limit and the cluster expansion — at small β the expansion in lumps converges independently of the box, and the limit and a lattice gap follow

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — An infinite-volume limit can always be taken along a subsequence, but it need not be unique. At small β the expansion in lumps of plaquettes converges independently of the box, and the existence of the limit, analyticity and exponential decay of correlations (a gap in lattice units) follow. Uniqueness independent of the boundary condition was shown by Shen–Zhu–Zhu along the road of Langevin dynamics (§ 08). Both guarantees cover a much narrower range.

Prerequisite chapters — 5-02. See also 4-11, 5-03 and 7-04. β is the inverse lattice coupling of 5-01, a different quantity from the β function $\beta(g)$ of 4-06.

Said first — This is about the side of small β at fixed lattice spacing, the shore opposite the continuum limit $\beta \to \infty$. This chapter shows nothing about the infinite-volume limit or the mass gap of the continuum theory. Table 3 is a fit of a crude estimate, not the best value in the literature.

Order of this chapter

1. From the box to all of space — what the infinite-volume limit is
2. A limit can always be taken, but it need not be unique (Figure 1, Table 1)
3. Detour: Elitzur's theorem — quantities that change under gauge transformations are 0 in every limit
4. Splitting into lumps — only closed lumps survive
5. Counting lumps — their number grows exponentially, and their weight falls exponentially (Table 2)

6. The condition for convergence — Kotecký–Preiss (Figure 2, Table 3)
7. What comes out of the expansion — the limit, analyticity, the lattice gap
8. Another road — Langevin dynamics and the Poincaré inequality
9. Where it stops — the strong-coupling shore and the continuum-limit shore (Table 4)

01

From the box to all of space — what the infinite-volume limit is

Lattice gauge theory is first defined on a box Λ_L of side L. The weight of the Wilson action (5-01) together with the Haar measure is an ordinary probability measure on a compact space.

$$\langle F\rangle_L = \frac{1}{Z_L}\int F(U)\, e^{-S_W(U)} \prod_{\ell\in\Lambda_L} dU_\ell, \qquad S_W = \beta\sum_P \Big(1 - \tfrac{1}{N}\operatorname{Re}\operatorname{tr} U_P\Big) \quad (1)$$

The mass gap is a statement about distances stretched to infinity (4-11 § 02), so we consider the limit of larger and larger boxes. We call a continuous function determined by finitely many links a **local observable**. When a probability measure μ on the whole lattice satisfies

$$L_k \to \infty \text{ along a sequence, and for every local observable } F \quad \langle F\rangle_{L_k} \longrightarrow \int F\, d\mu \quad (2)$$

we call μ an **infinite-volume limit**. It is weak convergence of measures. Chatterjee (2018) § 3 writes "a weak limit of the above probability measures as $\Lambda \uparrow \mathbb{Z}^n$" Known .

Example: in two dimensions with free boundary, fixing the gauge along a tree makes the plaquette variables independent, so in every box $\langle W_{R\times T}\rangle = u^{RT}$ (5-02 § 04), and the limit is trivial Known .

02

A limit can always be taken, but it need not be unique

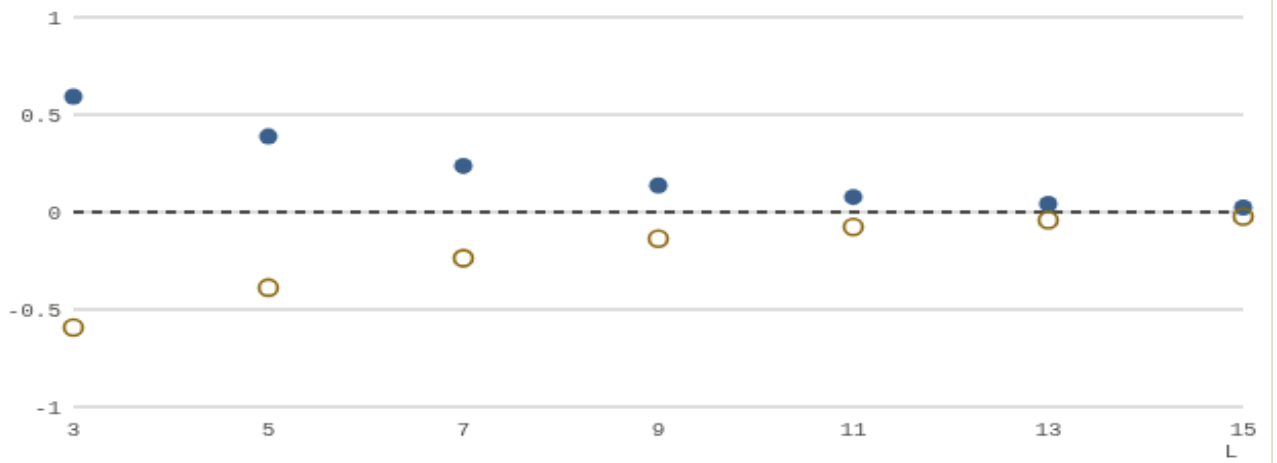
Existence comes almost free from compactness. $\langle F\rangle_L$ is a bounded sequence, so it has a convergent subsequence, and by a diagonal argument it can be made to converge simultaneously for countably many functions that approximate the local observables Known .

The hard part is **uniqueness**. Chatterjee (2018) § 3 writes “The infinite volume limit may or may not be unique. Indeed, the uniqueness (or non-uniqueness) is in general unknown for lattice gauge theories in dimensions higher than two when β is large.” Known . We look at an example where it fails in the two-dimensional Ising model (2-09). We fix the outer boundary of an $L \times L$ box to ± 1 and measure the mean of the central spin.

$$\langle \sigma 0 \rangle_L^{\pm} \xrightarrow{L\to\infty} \pm m^*, \qquad m^* = \begin{cases} 0 & \beta \le \beta_c \\ \left(1 - \sinh^{-4} 2\beta\right)^{1/8} & \beta > \beta_c \end{cases}, \qquad \beta_c = \tfrac{1}{2}\ln(1+\sqrt{2})$$

The right-hand side is the exact formula for the spontaneous magnetisation Known . At $\beta > \beta_c$ a different limit appears for each boundary condition (a phase transition; 2-08).

Figure 1 The central spin of a box whose outer boundary is + (●) or − (○) ($L = 3, \dots, 15$); the dashed lines are $\pm m^*$. The knob, "Play" and the rows of Table 1 change β.



β = 0.30, L=15: ±0.023, m* = 0.000

Computed The numbers in this figure were computed on this machine (column transfer matrix). If JavaScript is disabled, Table 1 holds the content.

Table 1 $\langle \sigma 0 \rangle_L^{\pm}$ (with the opposite sign for the − boundary). Selecting a row changes β in Figure 1 Computed

β	$L = 5$	$L = 15$	m^*	**Reading**
0.30	**0.3882**	**0.0231**	**0**	**The influence of the boundary dies away exponentially**
0.44	0.8473	0.7544	0	Near the critical point. The boxes are small
0.50	0.9241	0.9119	0.9113	Splits into $\pm m^*$

The distance between the centre and the outer boundary is $(L+1)/2$, and at $\beta = 0.2$, $\ln(\langle\sigma 0\rangle 13 / \langle\sigma 0\rangle 15) = 1.219$ is almost the same as the decay rate of correlations $2(K^* - K) = 1.223$ (2-09 § 06) Computed . **The influence of the boundary dies away as fast as correlations do.** The tool that shows decay of correlations in gauge theory is the expansion in lumps from § 04 on.

03

Detour: Elitzur's theorem — quantities that change under gauge transformations are 0 in every limit

A section you may skip.

In gauge theory the boundary cannot pick a direction. If we make a gauge transformation g at a single point x inside the box (outgoing links $gU\ell$, incoming links $U\ell g^{-1}$), neither S_W nor the Haar measure nor the boundary changes, so we can average over g, and for a link ℓ leaving x

$$\langle \operatorname{tr} U_\ell \rangle_L = \left\langle \operatorname{tr}\left(\int_G g \, dg \; U_\ell \right) \right\rangle_L = 0 \quad (4)$$

holds in every box, for every boundary condition, and hence in every limit. $\int_G g\,dg = 0$ follows from the fact that multiplying by a central element ωI ($\omega = e^{2\pi i/N}$ for $SU(N)$, -1 for $U(1)$ and $\mathbb{Z}_2$, and so on) does not change the Haar measure, so $\int g\,dg = \omega \int g\,dg$. The abstract of Elitzur (1975) is "It is argued that a spontaneous breaking of local symmetry for a symmetrical gauge theory without gauge fixing is impossible." Known . Precisely, what can be said is that "quantities that change under gauge transformations are not markers of order", not that there is no phase transition. Differences of phase can appear in gauge-invariant quantities (the area law, the Polyakov loop; 18-03) (5-06).

04

Splitting into lumps — a gas of lumps of plaquettes

At small β the weight of each plaquette is almost constant. We split the weight as "constant $\times$ (1 + small deviation)" and expand the product.

$$e^{\frac{\beta}{N}\operatorname{Re}\operatorname{tr} U} = c_0(\beta)\big(1 + f(U)\big), \qquad c_0 = \int_G e^{\frac{\beta}{N}\operatorname{Re}\operatorname{tr} U}\, dU, \qquad \int_G f\, dU = 0 \quad (5)$$

We split the set X of plaquettes that pick up the deviation into connected parts by the rule **they are connected if they share a link**. The integrals over parts that share no link are independent, so they factorise.

$$\frac{Z_L}{(e^{-\beta} c_0)^{|\mathcal{P}|}} = \sum_{\substack{\{X_1,\dots,X_k\} \\ \text{sharing no link pairwise}}} \prod_{i=1}^{k} z(X_i), \qquad z(X) = \int \prod_{P \in X} f(U_P) \prod_{\ell \in X} dU_\ell \quad (6)$$

$|\mathcal{P}|$ is the number of plaquettes in the box. This is exactly the polymer model of 7-04 § 03 Known . Two facts about the activities matter.

$$|z(X)| \le \varepsilon^{|X|}, \quad \varepsilon(\beta) = \sup_U |f(U)|; \qquad z(X) = 0 \ \text{(if some link lies in only one plaquette of}$$

The right-hand statement holds because that link appears in only one $f(AU_\ell B)$, and by the invariance of the Haar measure $\int f(AU_\ell B)\,dU_\ell = \int f = 0$. What survives are only **closed lumps**, in which every link is surrounded by two or more plaquettes. The smallest is the surface of a unit cube (6 plaquettes).

Example: for $\mathbb{Z}_2$, $f = \tanh\beta \cdot U$, and the activity of the cube is $\tanh^6\beta$. For $SU(2)$ we expand each face as $f = \sum_{j\,/=0} d_j\, u_j\, \chi_j$ by (3) of 5-02 and integrate by (5) of 5-02; only the terms in which every face carries the same representation j survive. For a closed surface the rule is known that its contribution is $d_j^{F-E+V}\, u_j^F$ (F, E, V are the numbers of faces, edges and vertices, $F - E + V$ is the Euler characteristic of the surface, and the surface of a cube is a sphere, so $6 - 12 + 8 = 2$) Known . For the cube this can be checked as follows. Gluing adjacent faces 5 times by the formula on the right of (5) ($1/d_j$ each time) turns the 6 faces into a single disc. In the product around its rim, the remaining 7 links cancel in order, starting with those adjacent as in $U_\ell\, U_{\bar\ell}^{-1}$, and the identity matrix remains. Hence $d_j^6 \cdot d_j^{-5} \cdot \chi_j(1) = d_j^2$.

$$z(\text{cube}) = \sum_{j=1/2,1,\dots} (2j+1)^2\, u_j^6 \approx 4u^6, \qquad u_j = \frac{I_{2j+1}(\beta)}{I_1(\beta)} \tag{8}$$

At $\beta = 1$, 7.682×10^{-4}, and direct Monte Carlo integration gives $7.63(8) \times 10^{-4}$ Computed .

05

Counting lumps — their number grows exponentially, and their weight falls exponentially

We fix four words (for details, 7-04 § 03 and § 04). We call the X of (6) **polymers** (lumps). Two lumps that share a link are called **incompatible**. A sequence of lumps joined into one by the incompatibility relation (the same lump may appear any number of times) is called a **cluster**. The logarithm of the left-hand side of (6) can be rewritten as a sum of terms, one for each cluster, and the question is whether that sum converges absolutely independently of the box Known .

Convergence is a contest between **how fast the number** of lumps grows and the weight ε^n. Let c_n be the number of sets of n plaquettes, connected through links, containing a root plaquette. A plaquette has 4 links and a link is contained in $2(d-1)$ plaquettes, so the plaquettes that share a link with it number $D = 4(2(d-1)-1) = 4(2d-3)$ (20 in four dimensions). There is the upper bound of Lemma 2.1(c) of Borgs–Chayes–Kahn–Lovász (2010) Known .

$$c_n \le \frac{1}{n}\binom{nD}{n-1}, \qquad \lim_{n\to\infty}\left(\frac{1}{n}\binom{nD}{n-1}\right)^{1/n} = \frac{D^D}{(D-1)^{D-1}} \approx 53.0 \quad (D=20)$$ (9)

Table 2 c_n in four dimensions (enumerated by Redelmeier's method) Computed

n	c_n	Ratio	Closed lumps	Upper bound (9)
2	20	20.0	0	20
3	458	22.9	0	590
4	11 132	24.3	0	20 540
5	280 175	25.2	0	784 245
6	7 218 768	25.8	4	31 763 004

The ratio is about 26 at $n = 6$, and the growth rate 53 of the upper bound is about twice that. There are no closed lumps for $n \le 5$, and the 4 at $n = 6$ are four cubes Computed . To guarantee convergence with this way of counting, ε must be smaller than the reciprocal of the growth rate of lumps (a value expected to be smaller than $1/26$ from the ratios in Table 2, and larger than $1/53$ from (9)).

As a metaphor, for each plaquette it is a contest between a cost of $\ln(1/\varepsilon)$ and a gain of about $\ln 26$ from the number of choices. Precisely, if the gain wins, the expansion merely becomes unusable; a phase transition need not occur (7-04 § 10).

06

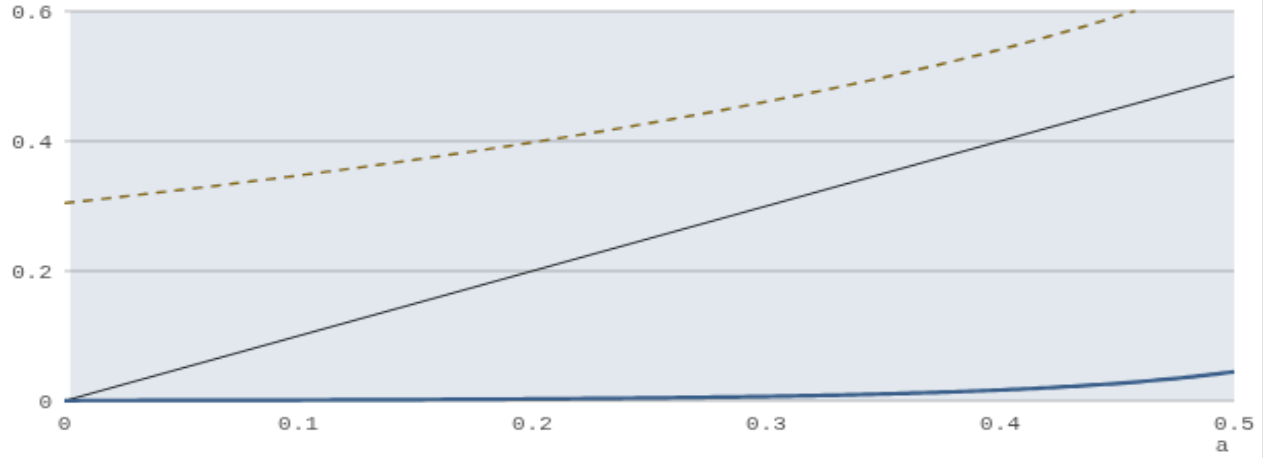
The condition for convergence — applying the Kotecký–Preiss condition

The Kotecký–Preiss condition turns the contest into a theorem (7-04 § 06). In the form of a theorem: if there is some $a > 0$ such that the left-hand inequality of (10) holds for every lump X, then the sum over clusters for the logarithm in (6) converges absolutely independently of the box Known . One reads it as giving each lump a budget $a|X|$, and requiring the total weight of its incompatible partners (multiplied by $e^{a|X'|}$) to fit within the budget.

$$\sum_{X' \text{ sharing a link with } X} |z(X')|\, e^{a|X'|} \le a|X| \quad \Longleftarrow \quad 8(d-1)\, G\big(\varepsilon e^{a}\big) \le a, \qquad G(x) = \sum_{1}$$

The right-hand side is the count that X has at most $4|X|$ links, and that a lump containing a given link contains one of its $2(d-1)$ plaquettes. c'_n is the number of closed lumps (Table 2 for $n \le 6$, (9) for $n \ge 7$). Maximising $\varepsilon = x\, e^{-8(d-1)G(x)}$ over $x = \varepsilon e^{a}$ gives the upper limit $\varepsilon_{\max}$ of the guarantee, and solving $\varepsilon(\beta) = \varepsilon_{\max}$ gives the range of β.

Figure 2 Condition (10) in four dimensions. The thin line is the right-hand side a, the thick line the left-hand side $24G(\varepsilon e^{a})$ (closed lumps), and the dashed line all connected lumps. If there is a band where the thick line lies below the thin line, convergence is guaranteed. The knob changes β and the buttons the group.



SU(2), β = 0.0100, ε = 0.0100, a satisfying the condition: 0.003–0.500

Computed The numbers in this figure were computed on this machine. If JavaScript is disabled, Table 3 holds the content.

Table 3 The upper limit of β guaranteed by (10) (0.0087 when all connected lumps are counted) Computed , and the condition of Shen–Zhu–Zhu converted to the Wilson β of this book Known (conversion of conventions in 5-01 Table 4)

Four dimensions	$\varepsilon_{\max}$	$\mathbb{Z}_2$	$U(1)$	$SU(2)$	$SU(3)$
Closed lumps	0.0161	0.0161	0.0160	0.0160	0.0160
Shen–Zhu–Zhu Known	–	–	–	0.0833	0.1875

What is guaranteed is about $\beta < 0.016$, and since $\varepsilon \approx \beta$ it hardly depends on the group. This is two orders of magnitude narrower than $\beta \leq 1$, where the strong-coupling formula agreed with the numbers in 5-02 Table 2. The sources are the supremum norm (for $SU(2)$ at $\beta = 0.1$, $\varepsilon = 0.104$, while $u = 0.025$), the crudeness of (9), and ignoring cancellations Computed . Table 3 is not "where the expansion stops".

07

What comes out of the expansion — the limit, analyticity, exponential clustering, the lattice gap

When (10) holds independently of the box, the following come out of 7-04 § 08 Known .

(a) Analyticity. The limit of $\ln Z_L$ per plaquette is an analytic function of β (it converges even when extended to complex β), and there is no phase transition in that range. **(b) Existence of the limit.** $\langle F \rangle_L$ converges without taking a subsequence (the influence of the boundary dies away exponentially, as in Table 1). What can be said is the existence of the limit along the sequence with that boundary condition; uniqueness independent of the boundary condition relies on the road of § 08. **(c) Exponential clustering.** In the connected correlation of two local observables only

clusters touching both supports survive, and joining a distance r needs about r plaquettes or more.

$$\big|\langle FG\rangle - \langle F\rangle\langle G\rangle\big| \le C_{F,G}\, e^{-m\,\mathrm{dist}(F,G)}, \qquad m > 0 \quad (11)$$

Example: the smallest lump joining two plaquettes is a tube with $4t$ side plaquettes having those two as its ends (a closed surface together with the ends), and the leading term is $m_0 = -4\ln u$ (5-02 § 05). (11) justifies this independently of the box.

(d) The lattice gap. We normalise the transfer matrix T (5-03 § 05) so that its eigenvalue on the vacuum Ω is 1. For the state $F\Omega$ made by a gauge-invariant local observable F, with its vacuum component subtracted, (11) in the time direction says that $\langle F\Omega, T^t F\Omega\rangle$ decreases at the rate e^{-mt} for every F. Such states are dense in the space of all gauge-invariant states orthogonal to the vacuum, so by the spectral theorem the spectrum of T there lies in $[0, e^{-m}]$ (4-11 § 02). That is, $\|T\| \le e^{-m}$ on the gauge-invariant states orthogonal to the vacuum, and the gap of $-\ln T$ is at least m.

In the literature, Shen–Zhu–Zhu (2023) § 1 summarises the content of Osterwalder–Seiler (1978) as "the existence of a mass gap (exponential clustering) follows from convergence of the cluster expansion, so do existence of the infinite volume limit and analyticity of Schwinger functions in the inverse coupling.", and writes of uniqueness only weakly, that it "should also follow". Chatterjee (2018) § 6 writes that "quark confinement and the existence of a positive self-adjoint transfer matrix" were shown Known . The abstract of the original paper also says "We then prove existence and analyticity of the infinite volume limit of strongly coupled Yang-Mills theories on the lattice and we verify Wilson's confinement bound." (we have not read the text).

The forms machine-checked on this book's definitions are in 20-01 to 20-04.

08

Another road — Langevin dynamics and the Poincaré inequality

There is also a road that gets there without counting lumps. We prepare three words.

Three words — Langevin dynamics, Ricci curvature, the Hessian

Langevin dynamics is a dynamics that moves the link variables randomly in the form $dQ = -\nabla S_W(Q)\,dt + \sqrt{2}\,dB$. ∇ is the gradient on the group of the link variables, and B is Brownian motion on the group (an independent fluctuation on each link). It is built so that the balance between the force down the slope and the fluctuations leaves the measure (1) unchanged (has it as invariant measure) (7-09). **Ricci curvature** is how curved the group is, seen as a curved surface. Where it is positively curved like a sphere, geodesics leaving nearby points gather together. For $SU(N)$ it is the same

positive constant everywhere ($N/2$ in the normalisation of the inner product of Shen–Zhu–Zhu). The **Hessian** is the second derivative of the action S_W taken on the group.

The Poincaré inequality and the Bakry–Émery criterion

The key is the Poincaré inequality (1-10).

$$\mathrm{Var}_\mu(F) \le C_P\, \mathbb{E}_\mu\big[\,|\nabla F|^2\,\big] \quad (12)$$

That C_P does not depend on the box is what "a uniform estimate" means. In (9) of 1-10, in one dimension $V'' \ge \kappa > 0$ gave $C_P \le 1/\kappa$. On a group the curvature adds to this κ, and the Bakry–Émery criterion takes the following form Known .

$$\text{Ricci curvature} + \text{Hessian of the action} \;\ge\; K_S > 0 \quad\Longrightarrow\quad (12)\text{ holds with a box-independe}$$

The strong-coupling condition of Shen–Zhu–Zhu (Shen–Zhu–Zhu's Assumption 1.1) is $K_S = N/2 - 8N|\beta_{\mathrm{SZZ}}|(d-1) > 0$ with $SU(N)$. The term on the right comes from the estimate that the Hessian does not fall below $-8N|\beta_{\mathrm{SZZ}}|(d-1)$. The point is that the Hessian stays of order $|\beta|N$ (an example with one link is in 5-11). As a metaphor, it is the state in which the force pulling back to the bottom of a round bowl is stronger than the bumps of the action. Precisely, what is compared is the curvature $N/2$ with the size $8N|\beta_{\mathrm{SZZ}}|(d-1)$ by which the Hessian can become negative.

The theorem of Shen–Zhu–Zhu

Shen–Zhu–Zhu (2023), under $|\beta_{\mathrm{SZZ}}| < 1/(16(d-1))$ with $SU(N)$ ($\beta = N^2\beta_{\mathrm{SZZ}}$; conversion in 5-01 Table 4), write in the abstract “The finite volume Yang–Mills measures converge to this unique invariant measure in the infinite volume limit” and “the infinite volume measure has a strictly positive mass gap”. Their uniqueness theorem for the invariant measure (Shen–Zhu–Zhu's Theorem 1.2) shows that the invariant measure of the Langevin dynamics is unique (1), and that the limit of the measures on periodic boxes agrees with it independently of the subsequence (2). Since the group is compact, the sequence of measures on boxes is always tight, and the “every tight limit” of (2) may be read as a subsequential limit in the sense of § 02. Remark 1.3 states that replacing the periodic boundary by Dirichlet or other boundary conditions gives the same invariant measure in the limit. It is one answer, in the strong-coupling range, to the question of § 02 Known . For four-dimensional $SU(2)$ this is $\beta < 1/12$.

Note Two gaps

The $1/C_P$ of (12) is the gap in the artificial Langevin time, a different quantity from the gap of the transfer matrix (18-03). The mass gap of Shen–Zhu–Zhu is the

exponential decay of correlations in distance on the lattice (not in the artificial Langevin time) derived from (12) (Shen–Zhu–Zhu's Corollary 1.6), a statement in the same sense as (11).

09

Where it stops — the strong-coupling shore and the continuum-limit shore

Table 4 The scale of β for four-dimensional $SU(2)$ (action (1))

β	What is said	Level
< 0.016	The fit in this chapter (Table 3)	Computed
< 0.083	A unique limit and exponential decay of correlations (Shen–Zhu–Zhu)	Known
$\lesssim 1$	The strong-coupling formula agrees with the numbers (5-02 Table 2)	Computed
≈ 2	Roughening of the surface (5-02 § 07)	Physics
2.3 and above	Numbers extrapolated to the continuum limit (5-08)	Physics
$\to \infty$	The continuum limit (5-05). A limit of correlation functions was built by the construction published in 2026-11 (along a subsequence of volumes; 6-00). There is no theorem on uniqueness or on a gap in physical units (6-04)	The committee's assessment

The meaning of "infinite-volume limit" changes between the top and the bottom of the table. The limit in this chapter is at fixed lattice spacing. In the continuum limit, $a \to 0$ while the box in lattice units also goes to infinity, and the gap in lattice units should close as $m(\beta) \to 0$ (5-09) Physics . What is needed is a gap uniform in physical units, and there is no small ε there. The problem statement § 6 writes “no present ideas point the direction to establish the existence of a mass gap that is uniform in the volume. Nor do present methods suggest how to obtain the existence of the infinite volume limit $T^4 \to R^4$.” Known .

Also, analyticity says nothing outside the small disc of complex β guaranteed by (10). For four-dimensional $U(1)$ the area law fails at large β (Guth 1980; according to Chatterjee (2018) § 4, the fully rigorous proof is Fröhlich–Spencer 1982; 5-06), while that non-abelian groups have no critical point at finite β (the correlation length diverges only as $\beta \to \infty$) is at the level of Chatterjee (2018) § 5, “It is believed that ... many of the non-Abelian lattice models of interest have βc = ∞.” Physics (7-10).

What this chapter can say

Matter	Said	Not said
(2)–(8)	Existence of subsequential limits, Elitzur, the gas of lumps Known . Table 1 and the activity of the cube Computed	Uniqueness at large β
Tables 2 and 3	The numbers for $n \le 6$, and the crude sufficient condition (10) Computed	The best range in the literature, the β at which the expansion stops
(a)–(d), (11), (12)	An account of known results on the strong-coupling side (Osterwalder–Seiler from the abstract and secondary sources). Uniqueness independent of the boundary condition is, from the expansion in lumps, at the level of "should also follow"; Shen–Zhu–Zhu showed it along the road of Langevin dynamics Known	The details of the proofs. Uniqueness and the gap in the continuum limit. This chapter shows nothing about them

Sources and reproduction

Item	Kind	Source / tools
Quotations in § 01, § 02, § 07 and § 09	Text § 3, § 4, § 5, § 6	S. Chatterjee, arXiv:1803.01950 (2018)
Quotations in § 07 and § 08, Assumption 1.1, Theorem 1.2, Remark 1.3, Theorem 1.4, Corollary 1.6, formulas (1.2), (4.7), (4.8)	Abstract, text § 1 and § 4	H. Shen, R. Zhu, X. Zhu, Commun. Math. Phys. 400 (2023) 805–851, arXiv:2204.12737
Osterwalder–Seiler in § 07	Abstract (positivity of the transfer matrix, existence and analyticity of the infinite-volume limit at strong coupling, Wilson's confinement bound; the INSPIRE abstract). Bibliographic data only for the text	K. Osterwalder, E. Seiler, Ann. Phys. 110 (1978) 440–471
Original source (including (10))	Bibliographic data only (content from the two above and 7-04)	R. Kotecký, D. Preiss, CMP 103 (1986) 491. E. Seiler, *Gauge theories as a problem of constructive quantum field theory and statistical mechanics*, Springer (1982)
Quotation in § 03	Abstract	S. Elitzur, Phys. Rev. D 12 (1975) 3978–3982
(9)	Text, Lemma 2.1(c)	C. Borgs, J. Chayes, J. Kahn, L. Lovász, arXiv:1002.0115 (2010)
Enumeration method of Table 2	Bibliographic data only	D. H. Redelmeier, Counting polyominoes: yet another attack, Discrete Math. 36 (1981) 191–203

Item	Kind	Source / tools
The rule d^{F-E+V} in (8)	Text § 3.4 (the gluing formulas (3.17) and (3.19), and that $(\dim R)^{2-2p}$ appears for a closed surface; in the form of the two-dimensional theory)	S. Cordes, G. Moore, S. Ramgoolam, Lectures on 2D Yang–Mills theory, equivariant cohomology and topological field theories, arXiv:hep-th/9411210 (1994)
The problem statement in § 09	Text § 6 (checked word for word in the material)	A. Jaffe, E. Witten, Quantum Yang–Mills Theory, Clay Mathematics Institute (2000)
The m^* of (3) / the phases of $U(1)$	Bibliographic data only	C. N. Yang, Phys. Rev. 85 (1952) 808–816 / A. H. Guth, Phys. Rev. D 21 (1980) 2291, J. Fröhlich, T. Spencer, Commun. Math. Phys. 83 (1982) 411
Tables 1–3, the figures, (8)	Computed on this machine	Python 3, NumPy, SciPy: ising_box.py, animals.py, kp.py, cube.py, cube_check.py

Next chapter to read: 5-05 Weak coupling and the continuum limit (the opposite shore). For the general theory, 7-04.

2026-10-01 · chapter Yang–Mills and the mass gap Part 5 Lattice gauge theory continuum limit

Weak coupling and the continuum limit — raising β makes the lattice finer, and asymptotic freedom sets how fast

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — Raising β brings the lattice spacing $a(\beta)$ towards 0, and a is an output, fixed by calibrating against a physical quantity. The difference between "scaling" and "asymptotic scaling". Lattice errors vanish as a^2 and can be reduced by improved actions and continuum extrapolation.

Prerequisite chapters — 5-01 (Wilson's lattice) and 4-06 (the β function). For the general theory of the continuum limit, 2-08 § 08; for the mass gap, 4-11. The β of this chapter is the inverse lattice coupling $2N/g_0^2$ of 5-01, a different quantity from the β function $\beta(g)$ of 4-06. The "lattice β function" of § 03 refers to how this β depends on a, (3).

Said first — The asymptotic form of $a(\beta)$ and the a^2 errors are matters at the level of perturbation theory and numerics Physics . The continuum limit of the four-dimensional non-abelian lattice was built by the construction published in 2026-11, for a family whose bare coupling is chosen to match the running of asymptotic freedom (the committee's assessment is that it is proved in Lean; 6-00). A limit in the form with the scale set by a physical quantity, the limit of the ratios (4), and the mass gap have not been shown (6-04). This chapter too shows nothing about the continuum limit or the mass gap.

Order of this chapter

1. The continuum limit is approaching a critical point
2. The lattice spacing is an output — what to calibrate with
3. The lattice β function and asymptotic scaling — the $\Delta\beta$ that halves a (Table 1, Figure 1). Detour: the price of the continuum limit
4. Scaling — dimensionless ratios stand still

5. The form of lattice errors — a^2 seen in a free field (Table 2, Figure 2)
6. Symanzik's effective theory and improved actions — adding rectangles to remove a^2 (Table 3), logarithmic corrections in the quantum theory. Detour: the price of improvement
7. Continuum extrapolation — drawing a straight line against a^2
8. The continuum limit seen from mathematics — what has been shown and what has not

---

01

The continuum limit is approaching a critical point

A lattice theory has no unit of length; what comes out are numbers of the form "so many lattice spacings". For example, the correlation length $\xi_{\text{lat}}(\beta)$ when the correlation of two plaquettes decreases as $e^{-|x|/\xi_{\text{lat}}}$ (4-11). A mass in physical units is obtained by dividing by the lattice spacing.

$$\Delta = \frac{1}{a(\beta)\,\xi_{\text{lat}}(\beta)} \quad (1)$$

To hold Δ fixed and take $a \to 0$ on the lattice with $\beta = 2N/g_0^2$ (5-01), we need $\xi_{\text{lat}} \to \infty$, that is, to approach a critical point. Chatterjee (2018) § 5 summarises the physical usage as “the model has a continuum limit if there is a critical point βc ∈ [0,∞] such that as β → βc, the correlation length tends to infinity”, and for four-dimensional non-abelian models writes “It is believed … βc = ∞” Known . $\beta = \infty$ is $g_0 \to 0$, the same direction as asymptotic freedom (4-06).

There are two limits: first hold β fixed and make the box infinite (5-04), and afterwards take $\beta \to \infty$.

02

The lattice spacing is an output — what to calibrate with

How many fm a is, is decided afterwards by matching the lattice value $\widehat{Q} = a^k Q$ of a quantity Q of mass dimension k against its physical value.

$$aQ(\beta) = \left(\widehat{Q}(\beta)/Q_{\text{phys}}\right)^{1/k} \quad (2)$$

As standards one uses the string tension σ (Creutz 1980 fixed a by “hold fixed the string tension”), the length r_0 defined from the force, and t_0 of the gradient flow (Lüscher

2010) Known . r_0 is the length fixed by $r_0^2 F(r_0) = 1.65$ from the force $F(r)$ acting between a pair of static charges (Sommer 1994; 4-06 § 07). t_0 is a length fixed by smoothing the field with a diffusion equation (the gradient flow) (18-04 § 08 calls the same thing the Wilson flow).

Example: for pure SU(3), r_0/a is 2.922 at $\beta = 5.7$ and 9.74 at 6.4 (Table 1 of Necco–Sommer 2002; the values for $\beta \le 6.4$ are reproduced from earlier calculations). Setting $r_0 \approx 0.5$ fm (a reference value used in QCD; the introduction of Necco–Sommer) gives $a \approx 0.171$ fm and 0.051 fm (1 fm = 10^{-15} m) Computed . The physical value of r_0 is an input from outside the lattice.

Changing the standard changes $a(\beta)$ slightly at finite β, and the ratio of the two a is fixed by the ratio $\widehat{Q}_1/\widehat{Q}_2$ measured on the lattice. That this ratio settles down is the scaling of § 04. In the mathematical formulation (§ 08) too the choice of calibration is an issue, and the existence of $\varepsilon(\beta)$ in either convention is unproved (5-09).

03

The lattice β function and asymptotic scaling

To keep the same physics as the lattice is made finer, the bare coupling g_0 is weakened. How fast begins with the scheme-independent b_0, b_1 (4-06 § 05) Known . Reading $\mu = 1/a$ and rewriting in terms of $\beta = 2N/g_0^2$ gives

$$\frac{d\beta}{d\ln(1/a)} = 4Nb_0 + \frac{8N^2 b_1}{\beta} + O(\beta^{-2}),$$
$$a\Lambda_{\mathrm{L}} = (b_0 g_0^2)^{-b_1/(2b_0^2)} e^{-1/(2b_0 g_0^2)} (1 + O(g_0^2)) \quad (3)$$

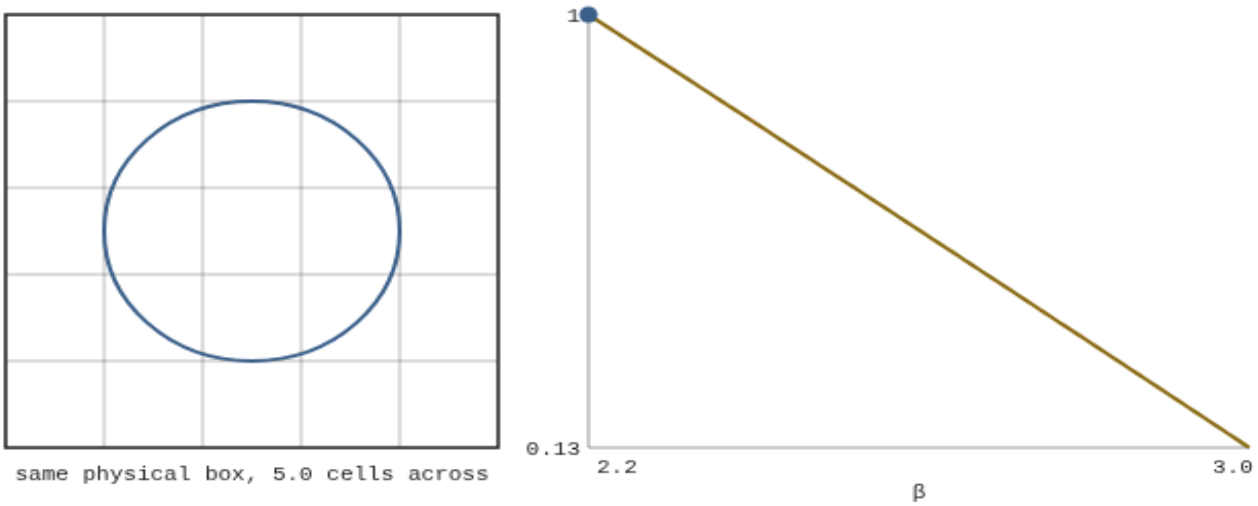
and the lower formula is formula (8) of 4-06, obtained by integrating the upper one. The constant of integration Λ_{L} (4-07) does not come out, but the ratio $a(\beta)/a(\beta')$ does. At one loop $a \propto e^{-\beta/(4Nb_0)}$, and the step that halves a is $\Delta\beta = 4Nb_0 \ln 2$: 0.2575 for SU(2) from $4Nb_0 = 11/(3\pi^2)$, and 0.5794 for SU(3) Computed .

Table 1 $a(\beta)$ from formula (3) (two loops) Computed

Group	β	a/a_{ref}	Lattice points in the same box ×	$\Delta\beta$ to halve
SU(2)	2.2	1	1	0.2760
SU(2)	2.6	0.3656	56.0	0.2732
SU(2)	3.0	0.1323	3263.5	0.2711
SU(3)	5.7	1	1	0.6155
SU(3)	6.3	0.5088	14.9	0.6121

Group	β	a/a_{ref}	Lattice points in the same box ×	$\Delta\beta$ to halve
SU(3)	6.9	0.2579	225.9	0.6092

Figure 1 Raising β makes the lattice in the same physical box (left) finer according to formula (3), and more lattice cells span the correlation length (circle). The right shows the logarithm of a/a_{ref}.



SU(2) β = 2.20, a/a(2.2) = 1.0000, ξ in lattice units ×1.00, lattice points ×1.0, Δβ to halve a = 0.2760

Computed The numbers in this figure were computed on this machine (formula (3) computed on the spot; agrees with Table 1 to 4 digits. If JavaScript is disabled, Table 1).

That lattice numbers follow (3) is called **asymptotic scaling**. Creutz (1980) saw for SU(2) "a logarithmic decrease of the bare coupling constant in a manner consistent with the perturbative renormalization-group prediction" Known . The step in pure SU(3) data is 0.354 at $\beta = 5.70$ and 0.524 at 6.20, short of the two-loop 0.61 (Table 4 of 4-06) Physics . The abstract of Lepage–Mackenzie (1993) says that expanding in a coupling defined by a physical quantity makes perturbation theory and numerics agree well. The abstract also says that if that coupling is used instead of the bare coupling, scaling already follows perturbation theory from about $\beta = 5.7$. Much of the deviation seen with the bare coupling comes from the choice of coupling.

Detour: the price of the continuum limit

A subsection you may skip.

Holding the box fixed and halving a multiplies the lattice points by 16 in four dimensions. For SU(2), raising β from 2.2 to 3.0 multiplies them by about 3,300 (Table 1). As a metaphor, it is "a picture where each doubling of resolution means 16 times the pixels, and each stroke gets slower too". Precisely, the cost is the number of lattice points times the autocorrelation time, and the latter is also expected to grow as a power of ξ_{lat} (5-07) Physics .

04

Scaling — dimensionless ratios stand still

What the continuum limit needs is that **ratios** of physical quantities stop depending on β.

$$R(\beta) = \widehat{Q}_1(\beta)/\widehat{Q}_2(\beta) = R_0\bigl(1 + s\,a^2 Q_2^2 + \cdots\bigr) \quad (4)$$

being constant within errors is called **scaling**, and that range of β the scaling window. Example: Necco–Sommer (2002) write, for the ratio r_c/r_0 of two lengths defined by the force, “No dependence on the resolution is seen within the errors of below 1%”, and obtain 0.5133(24) by extrapolation Known . $m/\sqrt{\sigma}$ is looked at in 5-08.

Scaling is ratios standing still, and is necessary for the continuum limit. **Asymptotic scaling** is $a(\beta)$ itself following (3), an additional claim tying the lattice to the Λ of perturbation theory Physics . If two quantities deviate from (3) by the same amount, the deviation cancels in the ratio. Lattice results such as $m/\sqrt{\sigma}$ rely on scaling (18-04 § 05).

05

The form of lattice errors — a² seen in a free field

We see why the correction in (4) starts at a^2 in a free field, which can be solved exactly. In the symmetric difference $f(x+a) - 2f(x) + f(x-a)$, odd powers of a cancel. For a free scalar field of mass m, the transfer matrix (5-03) fixes the mass E seen on the lattice Known . Concretely, looking for the pole of the lattice propagator $1/(\hat{p}^2 + m^2)$ ($\hat{p}^2$ is formula (6) below) at zero spatial momentum and imaginary momentum $p = iE$ gives $\frac{4}{a^2}\sinh^2\frac{aE}{2} = m^2$, and rewriting this gives the upper formula below.

$$\begin{aligned} \cosh(aE) &= 1 + \tfrac{1}{2}(am)^2, \\ \frac{E}{m} &= 1 - \frac{(am)^2}{24} + \frac{3(am)^4}{640} - \cdots \end{aligned} \quad (5)$$

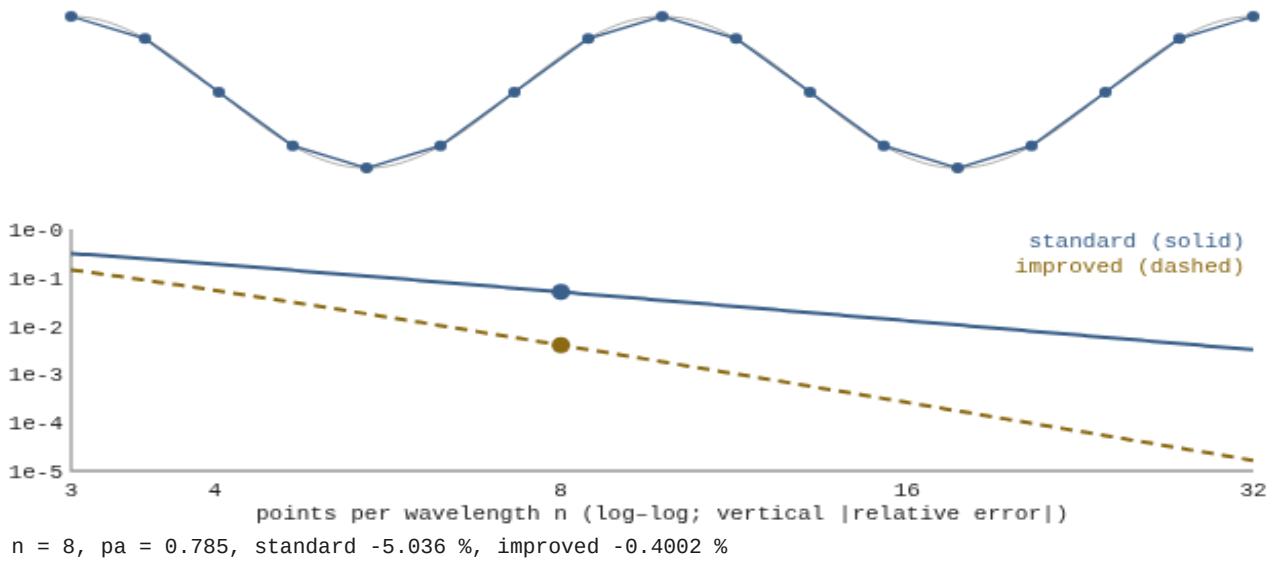
Table 2 The lattice mass (5) of the free field and the two-point extrapolation (8) Computed

am	$(E/m-1)/(am)^2$	Extrapolation error
1.0	−0.03758	-9.9×10^{-4}
0.4	−0.04093	-2.9×10^{-5}
0.2	−0.04148	-1.9×10^{-6}
0.1	−0.04162	-1.2×10^{-7}

The second column approaches $-1/24 = -0.04167$. The same holds in the spatial directions: applying the lattice Laplacian to a wave of wave number p gives $\hat{p}^2$ instead of p^2. Building a difference that uses points two steps away, with weights $\frac{4}{3}$, $-\frac{1}{3}$, removes the a^2 term. This is the simplest example of improvement.

$$\hat{p}^2 = \frac{4}{a^2}\sin^2\frac{pa}{2} = p^2\left(1 - \frac{(pa)^2}{12} + \cdots\right),$$
$$\hat{p}^2_{\text{imp}} = \frac{4}{3}\hat{p}^2 - \frac{1}{3a^2}\sin^2(pa) = p^2\left(1 - \frac{(pa)^4}{90} + \cdots\right) \quad (6)$$

Figure 2 Changing the number n of lattice points per wavelength moves the sample points above and the positions of the errors of the standard (solid) and improved (dashed) differences below (log–log).



n = 8, pa = 0.785, standard -5.036 %, improved -0.4002 %

Computed The numbers in this figure were computed on this machine ((6) computed on the spot). If JavaScript is disabled: $\hat{p}^2/p^2 - 1$ at $n = 4, 8, 16, 32$ is −0.189, −0.0504, −0.0128, −0.00321 for the standard difference and −0.0543, −0.00400, −0.000261, −0.0000165 for the improved one. Doubling n makes the standard error about 1/4 and the improved one about 1/16.

06

Symanzik's effective theory and improved actions

Symanzik (1983) read a lattice theory at small a as "the continuum theory plus terms with powers of a", and says in the abstract “Corrections to continuum theory results stemming from finite lattice spacing can be diminished systematically by use of lattice actions that also include suitable irrelevant terms” Known . For pure gauge theory it is thought to take the following form Physics .

$$\mathcal{L}_{\text{eff}} = \frac{1}{2}\operatorname{tr} F_{\mu\nu}F_{\mu\nu} + a^2\sum_{i=1}^{3} c_i(g_0)\,\mathcal{O}_i + O(a^4) \quad (7)$$

The first term has the same normalisation as 5-01 (g_0 put into the covariant derivative). $\mathcal{O}_i$ are gauge-invariant terms of mass dimension 6 ($\mathrm{tr} F^2$ has dimension 4, so a^2 multiplies them; 4-04). According to the abstract of Lüscher–Weisz (1985), if only spectral quantities (masses of stable particles, the static potential and so on) are to be improved, one of the three coefficients can be chosen freely Known . **Why there is no a^1**: a term of dimension 5 has two field strengths and one derivative, so 5 indices. The lattice is invariant under reversing each axis, so the indices of each direction must appear an even number of times, and an odd number of indices cannot be placed Physics . This is an argument about the action and spectral quantities (masses). For quantities that involve operators, such as the force between static charges, it has to be shown separately that the correction starts at a^2, and Necco–Sommer (2002) § 4 gives that argument.

Adding rectangles

We add 1×2 rectangles to the plaquette with weights c_0, c_1 ($c_0 + 8c_1 = 1$). A rectangle has two orientations in each plane, and has twice the area so acts 4 times as strongly on the leading term; hence 8. The lowest-order (classical) coefficients are attributed to Lüscher–Weisz (1985), and for the action that adds only rectangles they are $c_1 = -\frac{1}{12}$ ($c_0 = \frac{5}{3}$) Known . Table 3 reproduces this number. We check it with a smooth U(1) field on a two-dimensional torus (a sum of six wave numbers; the flux computed exactly with an antiderivative).

Table 3 Relative error of the action for a smooth U(1) field in two dimensions (classical) Computed

$L = 1/a$	**Wilson**	$/a^2$	**Improved** $c_1 = -\frac{1}{12}$	$/a^4$
4	−0.2718	−4.35	−0.1229	−31.5
16	−0.02121	−5.43	-1.06×10^{-3}	−69.2
32	−0.005370	−5.50	-6.94×10^{-5}	−72.8
64	−0.001347	−5.52	-4.39×10^{-6}	−73.7

The ratio of errors on halving a is 3.95→3.99 (→4) for Wilson and 15.2→15.8 (→16) for the improved action. The Wilson coefficient approaches the expected value $-\frac{1}{12}\int |\nabla F|^2 / \int F^2 = -5.522$. At $L = 16$, taking $c_1 = 0, -\frac{1}{24}, -\frac{1}{12}, -\frac{1}{8}$ makes the error divided by a^2 move linearly as $-5.43, -2.85, -0.27, +2.31$ Computed . For a four-dimensional non-abelian field, see Table 3 of 5-01.

Coefficients and logarithmic corrections in the quantum theory

In the quantum theory c_i depend on g_0 (one loop: Lüscher–Weisz 1985). Because of asymptotic freedom, the leading error comes with a logarithmic correction of the form $a^2[\bar{g}^2(a^{-1})]^{\hat{\gamma}_1}$, and Husung–Marquard–Sommer (2020) conclude for Yang–Mills that

“the log-corrections are a rather weak modification” Physics . Here $\bar{g}^2(a^{-1})$ is the running coupling at the scale $1/a$ (4-06), which approaches 0 like $1/\ln(1/(a\Lambda))$. $\hat{\gamma}_1$ is an exponent fixed for each correction term (a number that plays the role of an anomalous dimension). The error therefore takes the form a^2 times a power of $\ln(1/a)$.

Detour: the price of improvement

A subsection you may skip.

The Wilson action has reflection positivity and gives a positive transfer matrix, that is, a lattice Hamiltonian (5-03). The abstract of Lüscher–Weisz (1984) showed that adding operators of dimension 6 means “physical positivity is lost in general”: a transfer matrix can be defined but complex eigenvalues can appear, and the spectral decomposition of two-point functions always has contributions with negative weight Known . It is a choice between taking precision and taking a lattice Hamiltonian (4-10) (18-04).

07

Continuum extrapolation — drawing a straight line against a²

A computer cannot reach $a = 0$, so one measures a ratio R at several β, draws a straight line against a^2 on the horizontal axis, and takes the intercept.

$$R(a) = R_0 + c\,a^2 + O(a^4) \quad \Longrightarrow \quad R_0 \approx \tfrac{1}{3}\big(4R(a/2) - R(a)\big) \ (8)$$

The third column of Table 2 is the error of this two-point extrapolation, and the error drops to a^4. Even if the slope c differs from action to action, the intercept R_0 should be common (universality; § 08). For the Wilson case of Table 3 too, building it from $L = 16, 32$ takes the relative error from 0.54 % to 0.009 % Computed . In actual numerics, Necco–Sommer (2002) fit the force in the form $(1 + s(a/r_c)^2)$ and report that the slope s is not significant Known .

The pitfalls are points that are too coarse (at $L = 4$ in Table 3, $/a^2$ is off by 20 %), the logarithmic corrections of § 06, and boxes that are too small. All are procedures resting on the premise that $R(a)$ has the form (8), and the premise is not a theorem Physics .

08

The continuum limit seen from mathematics — what has been shown and what has not

Chatterjee (2018) § 3 writes “the general belief in the physics community is that β should scale like log(1/ε) in dimension four, although there are doubts about this belief and the question remains an open mathematical problem” (written in 2018) Known . Integrating the upper formula of (3) gives $\beta \approx 4Nb_0 \log(1/a)$, and this is that $\log(1/\varepsilon)$. The continuum string-tension problem in § 5 of the same paper (Chatterjee's Problem 5.2) asks whether, in the infinite-volume theory as $\beta \to \infty$, there exist sequences $\varepsilon(\beta) \to 0$, $c(\beta) \to \infty$ and a nonzero d such that, for a rectangular loop of height R and width T,

$$\log\left\langle W_{\gamma_{R/\varepsilon,\,T/\varepsilon}}\right\rangle = -c\,(R+T) - d\,RT + o(1) \quad (9)$$

holds. d is the continuum string tension, and about the perimeter term it says “it is not clear to me whether the renormalization term c(R+T) is indeed necessary” Known .

In two dimensions it has been shown that the Wilson, Villain and Manton lattice actions go to the same Yang–Mills measure (Chevyrev–Shen, the trivial bundle on the torus; 7-02) Known . Four-dimensional U(1) behaves differently (there is a massless photon phase on the weak-coupling side, and the critical point is at finite β; 5-06), and the renormalisation steps of the non-abelian case are treated in 7-07. In the four-dimensional non-abelian case, the construction published in 2026-11 showed, for a family of Wilson actions with a chosen sequence of spacings, that the correlations of renormalised local observables converge (the committee's assessment is that it is proved in Lean; the volume limit is along a subsequence; 8-02, 8-05). The bare coupling of that family is chosen so that the ratio of β to log(1/a) approaches a fixed value (10-06). Independence of the action and of the sequence of spacings, the limit of the ratios (4), and the continuum string-tension problem are not theorems (6-04). The tables and figures of this chapter are procedures that assume the limit, not evidence of its existence. Volume 2 reads this construction in the order: the objects (8-01 to 8-05), one step of renormalisation (9-01 to 9-07), and matching the coupling (10-01 onward).

What this chapter can say

Matter	Said	Not said
Continuum limit = critical point (1)	An account of the physical usage Known	That ξ_{lat} exists and diverges on a four-dimensional non-abelian lattice

Matter	Said	Not said
Two-loop $a(\beta)$ (3), Table 1, Figure 1	Computed from the universal b_0, b_1 Computed	That the actual $a(\beta)$ follows (3) (in numerics a deviation remains Physics)
Scaling (4)	An account of numerical support Physics	That the ratios have a limit (no theorem)
Errors start at a^2; improvement $c_1 = -\frac{1}{12}$	Free field and two-dimensional U(1) (classical) Computed . Account of the effective theory Physics	Proofs, coefficients and logarithmic corrections in the four-dimensional quantum theory
Continuum extrapolation (8)	For free and classical fields Computed	Validity for actual numerics
Existence of the continuum limit, the mass gap	—	This chapter shows nothing about them

Sources and reproduction

Item	Kind	Source / tools
Quotations in § 01 and § 08	Primary source (text checked)	S. Chatterjee, “Yang–Mills for probabilists”, arXiv:1803.01950 (2018) § 3, § 5
Calibration by the string tension in § 02 and § 03	Primary source (abstract checked)	M. Creutz, “Monte Carlo Study of Quantized SU(2) Gauge Theory”, Phys. Rev. D 21 (1980) 2308
r_0 and t_0 in § 02	Primary source (abstract checked)	R. Sommer, Nucl. Phys. B 411 (1994) 839, arXiv:hep-lat/9310022 / M. Lüscher, JHEP 1008 (2010) 071, arXiv:1006.4518
r_0/a in § 02 (the values for $\beta \le 6.4$ in Table 1 are reproduced from earlier calculations) and $r_0 \approx 0.5$ fm, r_c/r_0 in § 04, the a^2 of the force in § 06, the fit in § 07	Primary source (text checked; bibliographic data only for § 4)	S. Necco, R. Sommer, Nucl. Phys. B 622 (2002) 328, arXiv:hep-lat/0108008
Lepage–Mackenzie in § 03	Primary source (abstract checked)	G. P. Lepage, P. B. Mackenzie, Phys. Rev. D 48 (1993) 2250, arXiv:hep-lat/9209022
Symanzik in § 06	Primary source (abstract checked)	K. Symanzik, Nucl. Phys. B 226 (1983) 187, 205

Item	Kind	Source / tools
Coefficients in § 06, the transfer matrix in the detour	Primary source (abstract checked. The number $-\frac{1}{12}$ is not checked in the text; the abstract goes as far as saying the coefficients were fixed at lowest order)	M. Lüscher, P. Weisz, Commun. Math. Phys. 97 (1985) 59; Erratum 98 (1985) 433 / Phys. Lett. B 158 (1985) 250 / Nucl. Phys. B 240 (1984) 349
Logarithmic corrections in § 06	Primary source (abstract checked)	N. Husung, P. Marquard, R. Sommer, Eur. Phys. J. C 80 (2020) 200, arXiv:1912.08498
Two-dimensional universality in § 08	Primary source (abstract checked)	I. Chevyrev, H. Shen, arXiv:2302.12160; Comm. Pure Appl. Math. (2026) (volume and pages not checked)
Tables 1–3, Figures 1 and 2, numbers in the text	Computed on this machine	Python 3.12, NumPy: `a_of_beta.py` (including the conversion of a to fm in § 02), `freefield.py`, `improve2d.py`. The formulas of the figures are checked against `jscheck.js`. For b_0, b_1 and the SU(3) step, 4-06

Next chapter to read: 5-06 How U(1) differs from the non-abelian case.

How U(1) differs from the non-abelian case — on the same lattice, an abelian group changes phase on the way

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — If the group on a four-dimensional lattice is $U(1)$, at large β one enters a phase of massless photons, and that there is a phase transition on the way is a theorem (at least for actions of Villain type) (Guth, Fröhlich–Spencer). For $SU(2)$ no jump is seen in the numbers, and it is believed that the correlation length does not diverge at finite β ($\beta_c = \infty$). The precise form of "because it is non-abelian".

Prerequisite chapters — 5-02 The strong-coupling expansion. We also borrow the phase transitions of 2-08 and the one-loop coefficient of 4-06.

Said first — The theorems are about $U(1)$, finite groups and three dimensions. For four-dimensional $SU(N)$, neither $\beta_c = \infty$ nor the gap has been proved. The numbers are observations on a small 6^4 lattice.

Order of this chapter

1. Two groups on the same lattice — sums of angles and products of matrices
2. On the strong-coupling side they cannot be told apart — both have the area law
3. Weak-coupling U(1) is a gas of photons — the Coulomb force (Figure 1)
4. Between the two shores the line breaks — Guth, Fröhlich–Spencer (Figure 2)
5. Detour: monopole loops — four dimensions and three
6. No jump is seen for SU(2) — only a crossover
7. The source of the difference is the running of the coupling — C_A and $\beta_c = \infty$
8. The precise form of "because it is non-abelian" — a table of counterexamples
9. What remains as mathematics

01

Two groups on the same lattice — sums of angles and products of matrices

Wilson's lattice (5-01) has the same form for every compact group. Group elements are placed on the edges, and the product around the four edges of a plaquette is weighted. For $U(1)$ (2-03 § 08) each edge carries an angle θ_ℓ, and the product becomes a sum of angles θ_P. For $SU(2)$ each edge carries a 2×2 matrix U_ℓ, and the product is the matrix product U_P.

$$S_{U(1)} = \beta \sum_P (1-\cos\theta_P), \quad \theta_P = \theta_{x,\mu} + \theta_{x+\hat\mu,\nu} - \theta_{x+\hat\nu,\mu} - \theta_{x,\nu}, \quad \beta = \frac{1}{g^2}; \qquad S_{SU(2)}$$

The convention for β is chosen so that the action returns to the Maxwell or Yang–Mills action in the classical continuum limit Known . The $\beta = 1/g^2$ of $U(1)$ is not the $2/g^2$ obtained by putting $N=1$ into the $\beta = 2N/g^2$ of $SU(N)$. This is because the generators are normalised differently (for $U(1)$ the edge carries e^{-iagA_μ} (the sign is the same as in (2) of 5-01), for $SU(N)$ $\operatorname{tr} T^a T^b = \delta^{ab}/2$). For $U(1)$, $\theta_P \approx -a^2 g F_{\mu\nu}$ and $1-\cos\theta_P \approx \theta_P^2/2$ make the action approximately $\frac{\beta g^2}{4}\sum_x a^4 \sum_{\mu,\nu} F_{\mu\nu}^2$, and matching it to Maxwell's $\frac14\int F^2$ gives $\beta = 1/g^2$. The difference shows in the form of the formulas. θ_P is a **linear expression** in the edge angles (the lattice curl), and the only non-linearity of the $U(1)$ action is the periodicity of $\cos$, that is, the compactness of the group. For $SU(2)$, writing $U_\ell = e^{i\vartheta_\ell\cdot\sigma}$, the exponent of U_P contains terms with the commutator $[\vartheta,\vartheta]$ (1-03). This is the self-interaction on the lattice (3-02). Example: if the angles on the four edges are $0.1, 0.2, 0.3, 0.4$, then $\theta_P = -0.4$, and the order of addition does not matter. For $SU(2)$ with $A = e^{0.3i\sigma_1}$ and $B = e^{0.3i\sigma_3}$, $\frac12\operatorname{tr}(ABA^{-1}B^{-1}) = 0.9847$, while the reordered $\frac12\operatorname{tr}(ABB^{-1}A^{-1}) = 1$ Computed . For $U(1)$ the sum of angles is 0 in both cases. Cyclic permutations (such as $BA^{-1}B^{-1}A$), however, do not change the value, because $\operatorname{tr} XY = \operatorname{tr} YX$.

02

On the strong-coupling side they cannot be told apart — both have the area law

On the side where β is small the two groups behave alike. The variable u of the character expansion of 5-02 is

$$u_{U(1)} = \frac{I_1(\beta)}{I_0(\beta)}, \qquad u_{SU(2)} = \frac{I_2(\beta)}{I_1(\beta)}, \qquad \langle W_{R\times T}\rangle \approx u^{RT}, \quad \sigma \approx -\log u \ (2)$$

(I_n are modified Bessel functions), and since $u < 1$ both have the area law. Example: for $U(1)$ at $\beta = 0.5$, $u = 0.2425$ and $\sigma \approx 1.417$; at $\beta = 1$, $u = 0.4464$ and $\sigma \approx 0.807$ Computed . The area law and exponential decay at small β are theorems for every compact group (Osterwalder–Seiler 1978) Known . They are a property of the lattice being coarse and do not use non-commutativity (5-02 § 08).

The difference shows only in numerical coefficients. As $\beta \to 0$, $u_{U(1)} \approx \beta/2$ and $u_{SU(2)} \approx \beta/4$, so $\sigma \approx -\log u$ grows in both cases in the form $\log(1/\beta)$, and the difference is the constant $\log 2$. At $\beta = 1$, $u_{SU(2)} = 0.2402$ and $\sigma \approx 1.426$, larger than for $U(1)$ at the same β ($\sigma \approx 0.807$) Computed . The convention (1) differs from group to group, so comparing sizes at the same β means little. What matters is that both have an area law of the same form.

03

Weak-coupling U(1) is a gas of photons — the Coulomb force and the perimeter law

When β is large the weight demands $\theta_P \approx 0$, and with $1 - \cos\theta_P \approx \theta_P^2/2$ the action is quadratic. Since θ_P is linear, this is a lattice Maxwell field without interaction (photons).

$$S_{U(1)} \approx \frac{\beta}{2} \sum_P \theta_P^2, \qquad \langle \theta_\mu(k)\theta_\nu(-k)\rangle = \frac{\delta_{\mu\nu}}{\beta\,\hat{k}^2}, \qquad \hat{k}^2 = \sum_\mu 4\sin^2\frac{k_\mu}{2} \quad (3)$$

The right-hand formula of (3) is the form for one fixed gauge (Feynman gauge, 4-08) Known . Gauge-invariant quantities (Wilson loops) do not depend on this choice.

That $1/\hat{k}^2$ diverges as $\hat{k} \to 0$ is what "massless" means, and correlations of the field strength decrease as the power $|x|^{-4}$ Known .

Per point there are 3 physical degrees of freedom (4 edges minus 1 for the gauge), and 6 plaquettes. From the Gaussian integral $\int x^2 e^{-x^2/2}dx / \int e^{-x^2/2}dx = 1$, the mean action of one mode with quadratic action $\frac{\lambda}{2}x^2$ is $1/2$ (equipartition). Hence $6 \cdot \frac{\beta}{2}\langle\theta_P^2\rangle = \frac{3}{2}$, that is,

$$\langle \cos\theta_P \rangle \approx 1 - \frac{1}{4\beta} \;\; (U(1)), \qquad \left\langle \tfrac{1}{2}\operatorname{tr} U_P \right\rangle \approx 1 - \frac{3}{4\beta} \;\; (SU(2)) \quad (4)$$

($SU(2)$ has 3 generators). Example: for $U(1)$ at $\beta = 2$, 0.875, and Monte Carlo gives 0.865 (Table 2) Computed .

The force between static charges

An $R \times T$ Wilson loop is the amplitude of a pair of static charges held for time T. For a Gaussian field it can be written in closed form, and as $T \to \infty$

$$V(R) = -\lim_{T\to\infty} \frac{1}{T} \log\langle W_{R\times T}\rangle = \frac{G_3(0) - G_3(R)}{\beta}, \qquad G_3(r) = \int_{[-\pi,\pi]^3} \frac{d^3k}{(2\pi)^3} \frac{e^{ik\cdot r}}{\sum_i 4\sin^2($$

remains Known . G_3 is the Green function of the three-dimensional lattice, $1/(4\pi R)$ at large distance. So V has the Coulomb form and stays at the finite value $G_3(0)/\beta$. It does not confine. Computing with $G_3(r,0,0) = \int_0^\infty e^{-6t} I_r(2t) I_0(2t)^2 dt$ gives $G_3(0) = 0.252731$ Computed . The $1/6$ at $R = 1$ is a check from $6G_3(0) - 6G_3(1) = 1$.

Computed Table 1 $\beta V(R)$ of weak-coupling $U(1)$ (formula (5)) and the Coulomb approximation

R	$\beta V(R)$	$G_3(0) - \frac{1}{4\pi R}$
1	0.16667	0.17315
4	0.23244	0.23284
16	0.24775	0.24776
∞	0.25273	0.25273

Figure 1 The static potential of $U(1)$. The knob β changes the strong-coupling σR (dashed) and the weak-coupling Coulomb form (solid)



Computed The numbers in this figure were computed on this machine (the solid line is the calculation of Table 1; the dashed line computes the series for I_1/I_0 on the spot). Near $\beta \approx 1$ neither approximation can be used (§ 04).

For finite squares $R \times R$ too, we obtained $-\beta \log\langle W\rangle$ from a Fourier sum on a periodic 64^4 lattice. At $R = 1, 4, 8, 16$ it is $0.250,\ 1.631,\ 3.584,\ 7.557$: divided by the area R^2 it goes towards 0 as $0.250 \to 0.030$, and divided by the perimeter $4R$ it settles to almost a constant, $0.063,\ 0.102,\ 0.112,\ 0.118$ (the perimeter law). The value per perimeter approaches $G_3(0)/2 \approx 0.126$ as $R \to \infty$ (the same as the limit of long rectangles). It approaches with corrections of $1/R$ and logarithms from the corners Computed .

04

Between the two shores the line breaks — Guth and Fröhlich–Spencer

That the same picture remains when the approximations are dropped is a theorem for $U(1)$. The theorem, however, is about actions of Villain type (Guth's abstract). The β $\cos$

Villain type is the action whose weight is the periodic Gaussian $\sum_n e^{-\frac{\nu}{2}(\theta+2\pi n)^2}$ in place of $e^{\beta\cos\theta}$. The abstract of Fröhlich–Spencer does not state the type of action (the text was not obtained). (2), (4), Table 2, Figure 2 and the β_T below are numbers for the Wilson action (1).

The logic of the theorem is short: if the string tension is positive at small β and 0 at large β, then $\sigma(\beta)$ cannot be an analytic function. An analytic function that is 0 on an interval is 0 everywhere. A point where a thermodynamic quantity is not analytic is a phase transition (2-08 § 04).

$$\sigma(\beta) > 0 \ (\beta < \beta_1), \quad \sigma(\beta) = 0 \ (\beta > \beta_2) \implies \sigma \text{ is not analytic somewhere in } [\beta_1, \beta_2] \quad (6)$$

The second half was shown by Guth (1980): for an action of Villain type with $g^2 < 0.168$, a lower bound on Wilson loops "has the form of a Coulomb interaction and guarantees that electric charges are not confined", and together with the strong-coupling result of Osterwalder–Seiler he concludes "this model has at least one phase transition" (abstract) Known .

The abstract of Fröhlich–Spencer (1982) is "We give a new, elementary proof for the existence of a deconfining transition to a massless (QED) phase in the four-dimensional U(1) lattice gauge theory" Known . Chatterjee (2018) § 4 introduces it as "A fully rigorous proof of Guth's theorem".

Convergence to the continuum limit — Driver

The abstract of Driver (1987) deals with the continuum limit. "compact U(1) lattice gauge theory with general energy function converges to a renormalized free electromagnetic field on the current sector as the lattice spacing approaches zero" comes with the condition "provided the coupling constant is sufficiently large". For the Wilson action it says "it is possible, by judicious choice of the Gibbs state, to get convergence for arbitrary coupling strengths" Known . The "current sector" is the range of observables that smear the field with smooth currents. Whether the "coupling constant" of the abstract is on the side of β or of g has not been checked, because we have not been able to read the text.

The jump seen in numbers

For the Wilson action the transition point is reported numerically as $\beta_T = 1.011128(11)$, with evidence of a first-order transition (Arnold et al. 2001, conference proceedings; the abstract of the follow-up proceedings hep-lat/0210010 also writes "strong evidence for a first-order phase transition") Physics . A first-order transition is one in which the mean plaquette jumps at β_T, and there the two phases can coexist.

Because they can coexist, a run started in one phase stays in that phase for a while, and runs that raise β and runs that lower it disagree. This is hysteresis.

On this machine we compared, with the Metropolis method on a 6^4 lattice, a run lowering β from above (cold start) with a run raising it from below (hot start) (at each β, 100 sweeps discarded and 100 sweeps measured).

Figure 2 The mean plaquette. The buttons choose the group, and choosing a point of β with the knob changes the vertical line and the reading (cold ●, hot ○, leading strong-coupling term dashed, leading weak-coupling term dotted)



Computed The numbers in this figure were computed on this machine (the points are Monte Carlo, of which Table 2 is an excerpt; the curves compute (2) and (4) on the spot).

Computed Table 2 The mean plaquette on a 6^4 lattice (in brackets, the error from bin averages;

underestimated near the transition)

Group	β	Cold	Hot	Strong coupling	Weak coupling
U(1)	0.98	0.5742(55)	0.5541(27)	0.4393	0.7449
U(1)	1.00	0.6425(39)	0.5684(35)	0.4464	0.7500
U(1)	1.02	0.6756(18)	0.6150(38)	0.4534	0.7549
U(1)	2.00	0.8652(4)	0.8641(5)	0.6978	0.8750
SU(2)	1.8	0.4440(10)	0.4371(18)	0.3994	0.5833
SU(2)	2.2	0.5685(20)	0.5638(18)	0.4645	0.6591
SU(2)	4.0	0.7997(8)	0.7994(6)	0.6580	0.8125

For $U(1)$ the two runs disagree for $\beta = 0.98$–1.02, and at $\beta = 1$ the difference is 0.074. It can be read as each run staying in its original phase (the hysteresis above). The $U(1)$ transition was also seen in the numerics of Creutz–Jacobs–Rebbi (1979) Physics . This table comes from short runs on a small lattice and has no power to decide the order (5-07).

05

Detour: monopole loops — four dimensions and three

A section you may skip.

θ_P does not distinguish integer multiples of 2π. Bringing the θ_P of the six faces of a small cube back into $(-\pi, \pi]$ and adding them can leave an integer multiple of 2π, and that is a lattice magnetic monopole. In four dimensions monopoles are closed loops (world lines); a loop of length ℓ pays about $\beta\varepsilon\ell$ in action and can be drawn in about μ^ℓ ways.

$$\sum_{\ell} \mu^{\ell} e^{-\beta\varepsilon\ell} = \sum_{\ell} e^{-(\beta\varepsilon - \log\mu)\ell} \tag{7}$$

If $\beta\varepsilon < \log\mu$, long loops fill space, disorder the lines of force and produce the area law. When β is large only short loops remain, and photons survive. Estimates of this kind are in Banks–Myerson–Kogut (1977) Physics . Precisely, (7) leaves out the interaction between loops, and the theorem is the one in § 04.

In three dimensions monopoles become points, and a gas of points causes Debye screening (the weakening of the far field by a gas of point charges) at every coupling. Polyakov (1977) showed confinement for three-dimensional compact $U(1)$ by this mechanism (a semiclassical argument) Physics . Göpfert–Mack (1982) proved for the Villain action that the "string tension α is nonzero for all values of the coupling constant g^2" Known . The same abstract writes that in the continuum limit the string tension in physical units goes to infinity. It is an example that confines even though abelian, and where the scale does not become a single one (17-02 § 03).

06

No jump is seen for SU(2) — only a crossover

For $SU(2)$ in Table 2, the difference between the two runs is at most about 0.007, about a tenth of that for $U(1)$. The difference 0.0069 at $\beta = 1.8$ is about 3 times the error in the table, but it is a difference between runs thermalised for only 100 sweeps at each β, and as β varies the cold and hot runs do not split into two systematic branches (how to tell insufficient thermalisation from a jump is in 5-07). The plaquette follows the leading strong-coupling term for $\beta \lesssim 1.8$, approaches the leading weak-coupling term for $\beta \gtrsim 3$, and joins the two smoothly around $\beta \approx 2.2$ Computed .

The abstract of Creutz (1980) fixes the string tension, makes the lattice finer, sees the bare coupling decrease logarithmically as the renormalisation group predicts, and writes “This supports the coexistence of confinement and asymptotic freedom” Physics .

The peak of the specific heat around the crossover is said to be the “shadow” of a nearby critical end point in a family of actions with an added term in the adjoint representation (Bhanot–Creutz 1981) Physics . Whether there is a phase transition can depend on the choice of action, so we consider it separately from the continuum limit. That for four-dimensional $SU(2)$ there is no point at finite β where the correlation length diverges is a believed statement, not a theorem. Chatterjee (2018) § 5 writes “It is believed ... βc = ∞”, that is, $\beta_c = \infty$ (5-05 § 01).

07

The source of the difference is the running of the coupling — C_A and $\beta_c = \infty$

What separates $U(1)$ and $SU(2)$ is whether the effective coupling changes with distance (4-06). At one loop

$$\mu \frac{dg}{d\mu} = -b_0 g^3, \qquad b_0 = \frac{11\, C_A}{48\pi^2}, \qquad \frac{1}{g^2(L)} \approx \frac{1}{g_0^2} - 2b_0 \log \frac{L}{a} \tag{8}$$

C_A is the Casimir number of the adjoint representation, N for $SU(N)$ and 0 for $U(1)$ Known . The right-hand formula is the value of the coupling g_0 at spacing a seen at length L.

With no matter fields, b_0 is 0 for $U(1)$, 0.04644 for $SU(2)$ and 0.06966 for $SU(3)$ Computed .

For $U(1)$, $b_0 = 0$, so if β is large the Gaussian approximation is self-consistent at all distances, and the Coulomb force of § 03 continues to large distances. For $SU(2)$,

however large β is, taking L large enough brings the right-hand side of (8) towards 0, and the coupling becomes strong. The rough guide is $L \sim a\, e^{\beta/(4Nb_0)}$, the same form as the $a(\beta)$ of 5-05 § 03. That there is no point at finite β where the correlation length becomes infinite is what "$\beta_c = \infty$" means Physics . Example: for $SU(2)$ at one loop, raising β by 1 multiplies the correlation length by about $e^{1/0.3715} \approx 15$ ($4Nb_0 = 11/(3\pi^2)$) Computed . This is a one-loop guide only; corrections from two loops and beyond, and the breaking of asymptotic scaling in the range of β actually used, are treated in 5-05.

08

The precise form of "because it is non-abelian" — four phrasings and counterexamples

Using known examples with groups and dimensions exchanged, we list which phrasings fail (see also 16-08 § 04).

Known Table 3 Phrasings, and examples that break them (on the side of large β)

Phrasing	Example that breaks it	Source
Abelian means no confinement	Three-dimensional $U(1)$ has the area law at every coupling (Villain action)	Göpfert–Mack
Non-abelian means confinement	Four-dimensional non-abelian finite groups are thought not to confine Physics	Chatterjee's lecture material, 5-10
Exponential decay is because it is non-abelian	For finite groups, abelian or not, exponential decay at weak coupling is a theorem	Adhikari–Cao
The number of phases is decided by whether the group is abelian	Four-dimensional abelian $\mathbb{Z}_N$ (with N large) has at least three phases: confining, massless (QED-like) and screened	Elitzur–Pearson–Shigemitsu, Guth (Villain type, $N > 37$), Fröhlich–Spencer; numerics by Creutz–Jacobs–Rebbi

The "screened" of Table 3 is the “screened” of Elitzur–Pearson–Shigemitsu, the phase in which Wilson loops follow the perimeter law and excitations are massive. It is a different meaning from the Debye screening of § 05 (which produces confinement). The same paper derives this sequence of phases from the existence of the phases of the O(2) model, using self-duality and correlation inequalities (abstract). Numerically, the abstract of Creutz–Jacobs–Rebbi (1979) reports two transitions for $N \geq 5$ Physics .

What matters is not only "non-abelian" but that the group is **continuous**. Continuity matters, however, for whether massless waves remain at the end of sufficiently large β. For a finite group the edge values are discrete, and at sufficiently large β the edges freeze to the identity, so no massless waves like photons can be made Physics . On the

other hand $\mathbb{Z}_N$ (with N large) has a massless phase at intermediate β (row 4 of Table 3). Exponential decay at large β for finite groups is a different thing from the gap of the continuum theory (5-10). $U(1)$ is continuous, so it has massless waves, and since $C_A = 0$ they survive.

The precise form

For a four-dimensional continuous non-abelian compact group, (i) there is no point at finite β where the correlation length diverges ($\beta_c = \infty$), and (ii) as $\beta \to \infty$ the gap in lattice units $1/\xi$ goes to 0, but dimensionless ratios (for example $m/\sqrt{\sigma}$) stay at positive values Physics .

(ii) is the form of the lattice correlation-length problem of 5-09 (Chatterjee's Problem 5.1). Neither (i) nor (ii) has been proved. It is more accurate to say that the role of non-commutativity is not "to make a gap" but "to break the massless weak-coupling picture at large distances".

09

What remains as mathematics

On the side of theorems are the strong-coupling results of § 02, Guth and Fröhlich–Spencer (actions of Villain type) and Driver (with the conditions of the abstract) in § 04, Göpfert–Mack in § 05, rows 1 and 3 of Table 3, and Guth and Fröhlich–Spencer in row 4 Known . (7) and Polyakov in § 05, and row 2 of Table 3, are at the level of physics. On the side of belief: that for four-dimensional $SU(N)$ there is no point at finite β where the correlation length diverges ($\beta_c = \infty$, Chatterjee § 5), the area law for all β, and that the gap remains in physical units in the continuum limit Physics . Depending on the action, there can be bulk (lattice-scale) phase transitions not accompanied by a divergence of the correlation length. Lattice calculations of $SU(N)$ with large N report such a transition on the strong-coupling side, limiting the range that can be computed (abstract of Lucini–Teper–Wenger 2005) Physics . This does not contradict $\beta_c = \infty$. The lattice area-law problem of Chatterjee (2018) (Chatterjee's Problem 4.1) asks for the area law at all $\beta > 0$ for four-dimensional non-abelian compact groups, and writes that at large β it “remains elusive” (5-02 § 08).

What the $U(1)$ theorem teaches is that an argument extending strong-coupling results to weak coupling without distinguishing groups is wrong. What must be shown for $SU(N)$ includes a negative statement, "the divergence of the correlation length and the breaking of the area law do not occur at finite β", and the tools to say that over the whole range of β exist at present only on the strong-coupling side (7-10, 16-07).

What this chapter can say

Matter	Said	Not said
Leading terms (2)(3)(4), estimate (7)	Standard approximations Known . Consistent with Table 2 at both ends Computed . (7) is a physical estimate Physics	Errors at intermediate β. The coefficients of (7)
Static potential, perimeter law	A bounded Coulomb force and the perimeter law within the Gaussian approximation Computed	Estimates without the approximation
The phase transition of four-dimensional $U(1)$ (actions of Villain type)	An account of the theorems Known . Observation of hysteresis for the Wilson action on 6^4 Computed	The order and position of the transition. The form of the theorem for the Wilson action
$\beta_c = \infty$ for $SU(2)$	That no jump is seen on a small lattice Computed	A proof that the correlation length does not diverge at finite β
"Because it is non-abelian"	Sorting the phrasings by known counterexamples Known	Proofs of (i)(ii). This chapter shows nothing about the mass gap or the continuum limit

Sources and reproduction

Item	Kind	Source / tools
§ 04, Table 3	Abstract (text not obtained)	A. H. Guth, Phys. Rev. D 21 (1980) 2291
§ 04, Table 3	Abstract (text not obtained; the abstract does not give the type of action)	J. Fröhlich, T. Spencer, Commun. Math. Phys. 83 (1982) 411–454
Continuum limit in § 04	Abstract (text not obtained)	B. K. Driver, Commun. Math. Phys. 110 (1987) 479–501
Strong-coupling theorem in § 02	Bibliographic data only	K. Osterwalder, E. Seiler, Ann. Phys. 110 (1978) 440–471
§ 04, § 06, § 09	Text § 4, § 5	S. Chatterjee, arXiv:1803.01950 (2018)
Bulk transition in § 09	Abstract	B. Lucini, M. Teper, U. Wenger, “Properties of the deconfining phase transition in SU(N) gauge theories”, JHEP 0502 (2005) 033, arXiv:hep-lat/0502003

Item	Kind	Source / tools
β_T in § 04	Abstract (conference proceedings)	G. Arnold, T. Lippert, T. Neuhaus, K. Schilling (author order as on arXiv), “Finite size scaling analysis of compact QED”, Nucl. Phys. B Proc. Suppl. 94 (2001) 651–656, arXiv:hep-lat/0011058
Follow-up proceedings in § 04	Abstract	G. Arnold, B. Bunk, T. Lippert, K. Schilling, Lattice 2002, arXiv:hep-lat/0210010
Numerics in § 04, row 4 of Table 3	Abstract	M. Creutz, L. Jacobs, C. Rebbi, Phys. Rev. D 20 (1979) 1915
§ 05	Abstract	T. Banks, R. Myerson, J. B. Kogut, Nucl. Phys. B 129 (1977) 493 / A. M. Polyakov, Nucl. Phys. B 120 (1977) 429
§ 05, Table 3	Abstract	M. Göpfert, G. Mack, Commun. Math. Phys. 82 (1982) 545–606 (the year is the Crossref year of publication; INSPIRE gives 1981)
§ 06	Abstract	M. Creutz, Phys. Rev. D 21 (1980) 2308 / G. Bhanot, M. Creutz, Phys. Rev. D 24 (1981) 3212
Table 3	Abstract	A. Adhikari, S. Cao, Ann. Probab. 53 (2025) 140, arXiv:2202.10375 / S. Elitzur, R. B. Pearson, J. Shigemitsu, Phys. Rev. D 19 (1979) 3698
Row 2 of Table 3	Text of the lecture slides (the subject is the area law: “It is not expected to hold for 4D Abelian theories at large β, nor even for 4D non-Abelian theories with finite gauge groups”). The same slides mention that Cao (2020) obtained the leading term of Wilson loops at large β for every finite group in four dimensions (Cao's paper: bibliographic data only, not checked)	S. Chatterjee, “Yang–Mills on the lattice: New results and open problems” (lecture slides; year not checked, the PDF was created in 2020)
Table 2, Figure 2	Computed on this machine	Python 3, NumPy. mc_u1.py, mc_su2.py (Metropolis method, arguments 6 100 100, 1 to 2 minutes per run), approx.py
The example of ordering in § 01, the comparison of u in § 02	Computed on this machine	Python 3, SciPy. order_su2.py

Item	Kind	Source / tools
Table 1, the perimeter law of § 03, Figure 1, b_0 in § 07	Computed on this machine	Python 3, SciPy. coulomb.py (the integral for G_3, the Fourier sum on 64^4), g3table.py

Next chapter to read: 5-07 The Monte Carlo method (the method that made Table 2, and how to read its errors).

The Monte Carlo method — make configurations that follow the weight, count the correlations, and give the error

This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows

- How the integral for a lattice expectation value (tens of thousands of dimensions) is estimated by a sample mean over a Markov chain that follows the weight.
- How to count the extra error caused by correlation between samples (the autocorrelation time, binning, the Γ method).
- Critical slowing down: as the continuum limit is approached, the correlation time grows as ξ^z (for a free field, $z = 2$ by a formula).

Prerequisite chapters — 5-01 Wilson's lattice, 1-08 Probability and measure. For the mathematical side, see 18-03 § 06– § 08.

Said first — The Monte Carlo method estimates expectation values on a finite lattice with error bars; it does not prove the mass gap or the continuum limit. The calculations in this chapter are small examples: one point of $SU(2)$ on 6^4, and a two-dimensional free field.

Order of this chapter

1. Why integrate with random numbers
2. Markov chains and detailed balance
3. Moving one link at a time — the step size (Figure 1, Table 1)
4. Heat bath and overrelaxation
5. Thermalisation and autocorrelation — three update methods (Table 2)
6. Estimating the error — bins, the jackknife, the Γ method
7. Critical slowing down — $z = 2$ for a free field (Figure 2, Table 3)
8. HMC and topological freezing

9. Monte Carlo seen from mathematics — the line between theorems and observations

---

The β of this chapter is the coupling of the Wilson action (5-01), not an inverse temperature.

01

Why integrate with random numbers — an integral in fifteen thousand dimensions

A lattice expectation value is an integral over all links (5-01 § 06; dU_ℓ is the Haar measure).

$$\langle O \rangle = \frac{1}{Z} \int O(U)\, e^{-S_W(U)} \prod_\ell dU_\ell \quad (1)$$

A four-dimensional lattice of side 6 has 5,184 links, and $SU(2)$ is 3-dimensional per link, so it is an integral in 15,552 dimensions Computed . So we make configurations $U^{(n)}$ that follow the weight e^{-S_W}/Z and take the sample mean.

$$O_N = \frac{1}{N} \sum_{n=1}^{N} O\big(U^{(n)}\big), \qquad \text{if independent, } \operatorname{Var} O_N = \frac{\operatorname{Var} O}{N} \quad (2)$$

The error decreases as $1/\sqrt{N}$ whatever the dimension (1-08) Known . Drawing uniformly does not work. At $\beta = 2.3$ on 6^4, the mean of the 7,776 plaquettes is 0.6026 in equilibrium, while uniform configurations scatter around 0 with standard deviation 0.00567, so the equilibrium value is 106 standard deviations away Computed .

02

Markov chains and detailed balance — making a distribution that stands still

We make a random rule that changes the configuration one step at a time (a Markov chain, with transition probability $P(x, y)$), so that after a long time it follows the weight π. A sufficient condition for the distribution not to change in one step (stationarity) is **detailed balance**.

$$\sum_x \pi(x) P(x, y) = \pi(y), \qquad \pi(x) P(x, y) = \pi(y) P(y, x) \quad (3)$$

Summing the right-hand side over x gives the left. A chain satisfying the right-hand side is called **reversible**. If moreover one can get from anywhere to anywhere (irreducible), the time average of a finite-state chain converges to the expectation value (the ergodic theorem). If it is also free of periodicity (aperiodic), the distribution itself approaches π (the convergence theorem) Known . For continuous state spaces too, under conditions such as the transitions having a positive density, there are a law of large numbers and a central limit theorem of the same form (Tierney 1994) Known .

The **Metropolis method** (Metropolis et al. 1953; the first application was the equation of state of two-dimensional hard spheres) proposes a candidate y with a symmetric proposal $q(x,y) = q(y,x)$, accepts it with the following probability, and stays put if it is rejected.

$$A(x \to y) = \min\bigl(1,\ \pi(y)/\pi(x)\bigr) = \min\bigl(1,\ e^{-\Delta S}\bigr) \quad (4)$$

If $\pi(y) \le \pi(x)$, then $\pi(x)q \cdot \pi(y)/\pi(x) = \pi(y)q \cdot 1$, so (3) holds, and Z is not needed. Example: two states, $\pi = (0.2, 0.8)$, with the proposal "go to the other one", gives $P = \begin{pmatrix} 0 & 1 \\ 0.25 & 0.75 \end{pmatrix}$ with eigenvalues $1, -0.25$, and the distribution from state 0 is $(0.1992,\ 0.8008)$ after 5 steps Computed . The deviation decreases as powers of the eigenvalue of largest absolute value other than 1 (here -0.25), and this is the source of the autocorrelation time of § 05.

03

Moving one link at a time — Metropolis and the step size

In four dimensions 6 plaquettes contain the link U_ℓ, and the part of the action that depends on U_ℓ has the following form.

$$S_W = -\frac{\beta}{N}\,\mathrm{Re}\,\mathrm{tr}\bigl(U_\ell A_\ell\bigr) + (\,U_\ell\text{-independent terms}\,), \qquad A_\ell = \sum_{6\ \text{plaquettes}} \text{staples} \quad (5)$$

A staple is the product of the three edges of a plaquette other than U_ℓ (5-01 § 03). A unit that goes once through all links is called a **sweep**.

The Metropolis proposal is near the current value. If the step size δ is large, rejections increase; if it is small, the chain moves only a little. We look at a single $U(1)$ link (weight $e^{\kappa\cos\theta}$).

Figure 1 A single $U(1)$ link ($\kappa = 2$). The knob sets the step size δ, and "Play" advances the chain. Left: θ over the last 300 steps; right: the histogram and the target density

Computed The numbers in this figure were computed on this machine (the readings are from Table 1; the chain in the figure is an example run on the spot).

Computed Table 1 Metropolis on a single link (200,000 steps each). The acceptance rates are exact values from numerical integration (agreeing with the runs to within 0.002). Exact $\langle\cos\theta\rangle = I_1(2)/I_0(2) = 0.69777$

δ	Acceptance rate	$\tau_{\mathrm{int}}(\cos\theta)$	$\langle\cos\theta\rangle$
0.1	0.975	242(37)	0.747(17)
0.25	0.937	63.9(5.5)	0.695(10)
1	0.757	5.90(18)	0.6972(31)
2.5	0.494	1.94(4)	0.6990(18)
π	0.426	2.00(4)	0.6990(18)

Widening δ from 0.1 to 2.5 halves the acceptance rate and makes τ_{int} (§ 05) about 120 times smaller. A high acceptance rate does not indicate efficiency. At $\delta = 0.1$ the mean is off from the exact value by 2.9 standard deviations; for a slow chain, the error estimate is itself uncertain.

04

Heat bath and overrelaxation — redrawing from the conditional distribution

The **heat-bath method** redraws a single link directly from the conditional distribution $\propto e^{\frac{\beta}{2}\mathrm{tr}(U_\ell A_\ell)}$ with the rest held fixed. A sum of $SU(2)$ matrices is a real multiple of an $SU(2)$ matrix (the quaternions of 1-04), so we can write $A_\ell = kV$ ($V \in SU(2)$); $X = U_\ell V$ also moves with the Haar measure, and the weight is $e^{\beta k x_0}$ ($x_0 = \frac{1}{2}\mathrm{tr}X$). $SU(2)$ is the three-sphere $x_0^2 + |\vec{x}|^2 = 1$, and the Haar measure is the uniform measure on it. The slice at fixed x_0 is a two-sphere of radius $\sqrt{1-x_0^2}$, with area proportional to $1-x_0^2$. When x_0 moves by dx_0, the width (arc length) on the three-sphere is $dx_0/\sqrt{1-x_0^2}$. The product is $\sqrt{1-x_0^2}\,dx_0$, so

$$p(x_0) \;\propto\; \sqrt{1-x_0^2}\,e^{\beta k x_0} \qquad (-1 \le x_0 \le 1) \;\text{(6)}$$

we draw x_0 from this, choose the remaining direction uniformly, and set $U_\ell = XV^{-1}$ Known . This is the method said to have been used in the $SU(2)$ calculation of Creutz (1980). Kennedy–Pendleton (1985) is an improvement that draws (6) quickly, and Cabibbo–Marinari (1982) is a method that updates $SU(N)$ one $SU(2)$ subgroup at a time.

Overrelaxation is a move that goes far without changing the action; for $SU(2)$ it is $U_\ell \mapsto V^{-1}U_\ell^{-1}V^{-1}$. With $X' = X^{-1}$ the action does not change, two applications return to the start, and it preserves the Haar measure, so it satisfies detailed balance. On its own it is not ergodic, so it is mixed with the heat bath (Adler 1981, Creutz 1987, Brown–Woch 1987). Note that a sweep in a fixed order is in general not reversible (the chain seen backwards in time is the one updated in reverse order), and only stationarity remains Known .

05

Thermalisation and autocorrelation — forgetting the starting point and counting the correlations

A chain starts either from a **cold start**, with all links set to the identity matrix, or from a **hot start** drawn uniformly, and the first stretch is discarded (thermalisation). For the plaquette at 6^4, $\beta = 2.3$ (equilibrium 0.6026, the heat bath of Table 2), the heat bath enters the equilibrium band from both sides in 20 sweeps, while the hot side of Metropolis lags, at 0.482 after 20 sweeps and 0.593 after 150 Computed . Agreement of the two starting points is only a necessary condition (5-06 § 04; the mathematical counterpart is the mixing time of 18-03 § 06).

Even in equilibrium the configurations are not independent. We define the **autocorrelation function**, the correlation of the time series at a separation of t steps divided by the variance, and the **integrated autocorrelation time**

$$\rho(t) = \frac{\langle \delta O_n \, \delta O_{n+t} \rangle}{\langle \delta O^2 \rangle}, \qquad \tau_{\mathrm{int}} = \frac{1}{2} + \sum_{t \geq 1} \rho(t), \qquad \operatorname{Var} O_N \approx \frac{2\tau_{\mathrm{int}} \operatorname{Var} O}{N} \tag{7}$$

($\delta O = O - \langle O \rangle$, $N \gg \tau_{\mathrm{int}}$). The effectively independent samples number $N/(2\tau_{\mathrm{int}})$ (formulas (23) and (25) of Wolff 2004) Known . For the heat bath of Table 2 that is $9{,}500/(2 \cdot 5.08) \approx 935$. The time τ_{exp} of the slowest decay is set by $\lambda_\star$, the eigenvalue of the transition of largest absolute value other than 1, as $\tau_{\mathrm{exp}} = -1/\log \lambda_\star$ (in the example of § 02, $\lambda_\star = 0.25$).

On the mathematical side, slowness is measured by the **Poincaré constant** C_P. C_P is the smallest constant such that $\operatorname{Var} f \leq C_P \, \mathbb{E}|\nabla f|^2$ for every function f (1-10 § 07; for a discrete chain the right-hand side is replaced by half the mean square change in one step, $\frac{1}{2}\sum_{x,y} \pi(x) P(x,y) (f(y) - f(x))^2$). **Langevin dynamics** is a continuous-time dynamics that adds noise to motion down the gradient of the action, and there $\tau_{\mathrm{exp}} = C_P$ holds with equality. For a discrete-time chain, the following three are different quantities Known (Levin–Peres–Wilmer § 12.2, 18-03 § 06).

- $C_P = 1/\gamma$: $\gamma = 1 - \lambda_2$ is the gap between 1 and the second largest eigenvalue λ_2.
- The relaxation time $t_{\mathrm{rel}} = 1/\gamma_\star$: $\gamma_\star = 1 - \lambda_\star$ is the absolute gap.
- $\tau_{\mathrm{exp}} = -1/\log \lambda_\star$.

When $\gamma_\star$ is small, the ratio of the last two is $1 + O(\gamma_\star)$. The **Polyakov loop** in the last column of Table 2 is the quantity obtained by taking, at each spatial point, the trace of the product of links once around the time direction, and averaging over space (here we look at its absolute value). It is an example of a slowly moving quantity spread over the whole lattice.

Computed Table 2 Four-dimensional $SU(2)$, 6^4, $\beta = 2.3$ (first 500 sweeps discarded). τ_{int} in sweeps. The cost is the number of staple computations per link

Method	Cost	Plaquette	τ: plaquette	τ: |Polyakov|
Metropolis	1	0.60291(57)	42(10)	79(24)
Heat bath	1	0.60261(21)	5.1(6)	6.0(7)
Heat bath + overrelaxation 3	4	0.60297(19)	2.1(2)	0.87(6)

The sample sizes are 9,500, 9,500 and 4,500. At the same sample size the error of Metropolis is 2.7 times that of the heat bath, consistent with $\sqrt{42/5.1} \approx 2.9$. Comparing τ multiplied by the cost, overrelaxation loses for the plaquette, 8.3 against 5.1, and

gains for the Polyakov loop, 3.5 against 6.0. What gains are slow global quantities Computed . For $SU(3)$ too, overrelaxation is reported to have shrunk the τ of the magnitude of the Polyakov loop from 28 sweeps to 9 (Brown–Woch 1987), and the tendency agrees Physics .

06

Estimating the error — bins, the jackknife, the Γ method

τ_{int} is estimated from the data. **Binning** cuts the time series into blocks of length B and treats the block means as independent. One reads off where the error levels off for $B \gg \tau_{\mathrm{int}}$ (Flyvbjerg–Petersen 1989). The **jackknife** gives the error of a non-linear quantity f from the means $\bar{b}_{(i)}$ with one block left out at a time.

$$\sigma_f^2 = \frac{n-1}{n} \sum_{i=1}^{n} \left(f(\bar{b}_{(i)}) - f(\bar{b}_{(\cdot)}) \right)^2, \qquad n = N/B \tag{8}$$

It is used for Creutz ratios and masses (5-02 § 06, 5-08). For the heat-bath plaquette of Table 2, (bin error / naive error)^2 grows as 1.00, 2.85, 6.23, 10.19 for $B = 1, 5, 20, 100$, reaching the Γ-method $2\tau_{\mathrm{int}} = 10.2$. For $B = 200$–1000 (47 to 9 blocks) it fluctuates downwards, 7.4 to 7.7. With few blocks the error of the error matters (for the square of the ratio the relative error is about $\sqrt{2/(n-1)}$) Computed .

The **Γ method** estimates the sum (7) directly, cut off at a window W.

$$\tau_{\mathrm{int}}(W) = \frac{1}{2} + \sum_{t=1}^{W} \rho(t), \qquad \langle (\delta\tau_{\mathrm{int}})^2 \rangle \approx \frac{4}{N} \left(W + \frac{1}{2} - \tau_{\mathrm{int}} \right) \tau_{\mathrm{int}}^2 \tag{9}$$

The right-hand side is formula (42) of Wolff (2004). It is derived by error propagation from the approximate formula for the error of the windowed sum given by Madras–Sokal (1988) (Wolff's formula (40)) Known . If W is short there is bias; if long, noise matters. The τ_{int} of this chapter use Wolff's (2004) automatic window ($S = 1.5$). Wolff asks one to “verify by eye” that $\tau_{\mathrm{int}}(W)$ is flat. In the lower panel of Figure 2 (§ 07) one can check by eye that it is flat around the window W of the vertical line.

07

Critical slowing down — seeing $z = 2$ by a formula for a free field

As the continuum limit is approached the correlation length ξ (in lattice units) grows (5-05 § 01), and the τ of local updates grows too. Physics writes $\tau \propto \xi^z$ and calls z the dynamical exponent (the review of Hohenberg–Halperin 1977, 18-03 § 07) Physics . The lattice free field (1-08 § 07) is where this can be checked by a formula.

$$S=\frac{1}{2}\sum_x\Big[\sum_\mu(\phi_{x+\hat\mu}-\phi_x)^2+m^2\phi_x^2\Big],\qquad \phi_x\,|\,\text{neighbours}\sim\mathcal{N}\Big(\frac{\sum_{y\sim x}\phi_y}{2d+m^2},\frac{1}{2d+m^2}\Big)\quad(10$$

In a checkerboard heat bath, let M_e be the sum over even points and M_o over odd. An odd point neighbours $2d$ even points, so $M_e'=rM_o+$ (noise), with $r=2d/(2d+m^2)$. For the zero momentum $M=M_e+M_o$

$$\rho_M(t)=\tfrac{1+r}{2}\,r^{2t-1}\ (t\ge1),\qquad \tau_{\text{int}}(M)=\frac{1}{2(1-r)}=\frac{1}{2}+\frac{d}{m^2},\qquad \tau_{\text{exp}}=\frac{-1}{\log r^2}\approx\frac{d}{m^2}\ ($$

From $\cosh(1/\xi)=1+m^2/2$, $\xi\approx1/m$, so $\tau\approx d\xi^2$, and for this dynamics $z=2$ Computed . The intuition is a random walk: a local update takes ξ^2 steps to carry information a distance ξ. Precisely, the $r=1-m^2/(2d+m^2)$ of (11) approaches 1 as $m\to0$, so the fraction that shrinks in one sweep becomes $1-O(\xi^{-2})$.

Figure 2 Two-dimensional free field, checkerboard heat bath. The knob sets m (top: $\rho_M(t)$, points measured, curves (11); bottom: $\tau_{\text{int}}(W)$, dashed line (11), vertical line the automatic window W). "Overlay" makes the horizontal axis t/ξ^2, and the four coincide

```
m = 0.25 (ξ = 4.01, 64²): τ_int measured 33.14 ± 3.91, (11) 32.5, τ_int/ξ² = 2.02
```

Computed The numbers in this figure were computed on this machine (the curves compute (11) on the spot. If JavaScript is disabled, Table 3).

Computed Table 3 Two-dimensional free field, checkerboard heat bath (side 16 to 64, 50,000 to 60,000 sweeps). τ in sweeps

m	ξ	$\tau_{\mathrm{int}}(M)$ **measured**	**(11)**	**(11)**$/\xi^2$
1	1.039	2.57(10)	2.5	2.316
0.5	2.020	8.17(53)	8.5	2.082
0.25	4.010	33.1(3.9)	32.5	2.021
0.125	8.005	125(23)	128.5	2.005

Each doubling of ξ multiplies $\tau_{\mathrm{int}}(M)$ by about 4. The cost of a calculation of the same precision is estimated to grow roughly as ξ^{d+z} (5-05) Physics . For Langevin dynamics

(§ 05) the relaxation time is $1/m^2 = C_P$ (18-03; in the free-field normalisation where the field values ϕ_x themselves are the coordinates in which gradients are measured; changing the choice of coordinates changes C_P only by a constant factor). Both are proportional to m^{-2} ($\propto \xi^2$). The τ of the heat bath is measured in sweeps and the τ of Langevin in Langevin time, so we do not compare the coefficients.

08

HMC and topological freezing — a global step, and a wall hard to cross

HMC (hybrid Monte Carlo, Duane et al. 1987) moves all links at once. Momenta P_ℓ (traceless Hermitian matrices; real linear combinations of the T^a of 1-03) are given normal distributions, and

$$H(P,U) = \frac{1}{2}\sum_\ell \operatorname{tr} P_\ell^2 + S_W(U) \tag{12}$$

the classical mechanics is run with the leapfrog method and accepted with probability $\min(1, e^{-\Delta H})$. The leapfrog method can be run backwards in time and preserves phase-space volume, so even with step errors e^{-S_W} is exactly the stationary distribution Known . Estimating the autocorrelation time of HMC by the Langevin C_P is an assumption (the generators are different).

As the continuum limit is approached, the field splits into sectors of topological charge Q (the number counting how the field is twisted, integer-valued for continuum fields; for its definition on the lattice see 3-05 § 08), and small steps find it hard to move between them. Schaefer–Sommer–Virotta (2011) reported, for HMC in pure gauge theory, an "effective dynamical critical exponent of about 5" for Q^2, and held that Wilson loops can be decoupled from this slow mode Physics . Lüscher–Schaefer (2011) showed a method that lets topological charge flow in and out through open boundary conditions in the time direction Physics . As a metaphor it is crossing passes in a landscape with many valleys. Precisely, if the measure near the boundaries of sectors is small the Poincaré constant becomes large (18-03 § 08) Known , but whether that is the cause of freezing has not been established.

09

Monte Carlo seen from mathematics — the line between theorems and observations

On a finite lattice (1) is a probability measure on a compact space (5-01 § 06). What can be said as theorems is the convergence of time averages by the ergodic theorem (for continuous state spaces, under the conditions of § 02) and the central limit theorem.

For a reversible chain, if the asymptotic variance is finite, a central limit theorem of the form (7) holds (Kipnis–Varadhan 1986) Known . What remains observation is the error bar. If there is a slow mode invisible at finite N, the error bar comes out too small ($\delta = 0.1$ in Table 1, § 08). Extrapolation from finite volume and finite a (5-05 § 07, 5-09) needs further assumptions.

The glueball masses and string tension of 5-08 are physical results that rely on that extrapolation Physics , not a proof of the existence and the gap of the problem statement (4-11). On the other hand CP is a mathematical object, and at strong coupling a Poincaré inequality for lattice Yang–Mills has been shown (Shen–Zhu–Zhu 2023, 5-04) Known . For four-dimensional Yang–Mills there is no theorem fixing the ξ-dependence of CP on the side $\xi \to \infty$.

For how this book uses this, see 23-08 § 01 and § 03 (checking numerical inequalities by interval arithmetic, under the label computer-assisted).

What this chapter can say

Matter	Said	Not said
Update methods	That Metropolis, heat bath, overrelaxation and HMC preserve the stationary distribution; (6) Known	The speed of mixing on a four-dimensional lattice
Tables 1 and 2	τ_{int} and errors in small examples Computed	Generalisation to large lattices. That the error bars are upper bounds
Critical slowing down	(11) and $z = 2$ for the free field with a checkerboard heat bath Computed . The general $\tau \propto \xi^z$ is a hypothesis Physics	The z of four-dimensional Yang–Mills
Topological freezing	The number in the abstract (effective exponent about 5) Physics	A mathematical explanation of freezing
The mass gap	—	This chapter shows nothing about the mass gap or the continuum limit

Sources and reproduction

Item	Kind	Source / tools
§ 02	Abstract	N. Metropolis et al., J. Chem. Phys. 21 (1953) 1087

Item	Kind	Source / tools
The ergodic theorem and convergence theorem in § 02, the relaxation time in § 05	Text	D. A. Levin, Y. Peres, E. L. Wilmer, *Markov Chains and Mixing Times*, 2nd ed.: Theorem C.1 (ergodic theorem, Appendix C.1), Theorem 4.9 (convergence theorem, § 4.3), § 12.2, Theorems 12.4–12.5 (relaxation time and mixing time)
Continuous state spaces in § 02 and § 09	Abstract	L. Tierney, Ann. Statist. 22 (1994) 1701
§ 04, § 05 (Brown–Woch's 28→9 sweeps)	Abstract (the attribution of the heat bath to Creutz not checked)	M. Creutz, PRD 21 (1980) 2308 / N. Cabibbo, E. Marinari, PLB 119 (1982) 387 / A. D. Kennedy, B. J. Pendleton, PLB 156 (1985) 393 / S. L. Adler, PRD 23 (1981) 2901 / M. Creutz, PRD 36 (1987) 515 / F. R. Brown, T. J. Woch, PRL 58 (1987) 2394
(7)(9) and the automatic window in § 05 and § 06	Text (formulas (23), (25), (40), (42), (50)–(52); (42) is error propagation from (40))	U. Wolff, CPC 156 (2004) 143 (erratum: CPC 176 (2007) 383), arXiv:hep-lat/0306017
§ 06 (the origin of formula (40) for the error of the windowed sum, binning)	Abstract	N. Madras, A. D. Sokal, J. Stat. Phys. 50 (1988) 109 / H. Flyvbjerg, H. G. Petersen, J. Chem. Phys. 91 (1989) 461
The dynamical exponent z in § 07	Abstract	P. C. Hohenberg, B. I. Halperin, Rev. Mod. Phys. 49 (1977) 435
§ 08	Abstract	S. Duane, A. D. Kennedy, B. J. Pendleton, D. Roweth, PLB 195 (1987) 216 / S. Schaefer, R. Sommer, F. Virotta, NPB 845 (2011) 93, arXiv:1009.5228 / M. Lüscher, S. Schaefer, JHEP 07 (2011) 036, arXiv:1105.4749
§ 09	Bibliographic data only / abstract	C. Kipnis, S. R. S. Varadhan, CMP 104 (1986) 1 / H. Shen, R. Zhu, X. Zhu, CMP 400 (2023) 805, arXiv:2204.12737
Tables 1–3, Figures 1 and 2, the numbers in § 01, § 02 and § 05–§ 07	Computed on this machine	Python 3, NumPy, SciPy. one_link_metropolis.py, su2_algos.py, analyze_su2.py, free_field.py, exact_tint.py, free_field_ext.py (extension for Figure 2), gamma_method.py (checked with AR(1)), small_numbers.py

Next chapter to read: 5-08 Measuring glueballs and the string tension (reading masses and the string tension from configurations made by the methods of this chapter).

Measuring glueballs and the string tension — read masses from how correlations decrease, and form the ratio $m/\sqrt{\sigma}$

This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — A lattice mass is read from how fast the correlation of time-slice operators decreases in imaginary time, and the value is an upper bound approached from above. Smoothing and the variational method speed up the approach. The string tension is extracted from periodic strings by assuming a string correction, and the ratio $m(0^{++})/\sqrt{\sigma}$ is 3.78–3.84 for $SU(2)$ and 3.41–3.61 for $SU(3)$ in primary sources (3.35–3.61 including Meyer's value from secondary information).

Prerequisite chapters — 5-07 The Monte Carlo method. We use the transfer matrix (5-03), the definition of the mass gap (4-11) and the variational principle (1-10). β is the inverse lattice coupling of 5-01, a different quantity from the β function $\beta(g)$ of 4-06.

Said first — Every number in this chapter is at the level of numerics, not a theorem. They are measured at finite box, β and time window, and pass through extrapolation and the string picture. The demonstration on this machine is three-dimensional $\mathbb{Z}_2$ gauge theory, not four-dimensional Yang–Mills. This chapter shows nothing about the existence of the mass gap.

Order of this chapter

1. A mass is how a correlation decreases — the effective mass is an upper bound
2. Name tags for glueballs — the rotation group of the cube
3. The noise does not decrease — the relative error grows exponentially
4. Fattening the operators — smoothing in space only
5. The variational method — optimising the upper bound
6. Measuring the string tension — Creutz ratios and periodic strings
7. A demonstration on this machine — three-dimensional $\mathbb{Z}_2$

8. Numbers from primary sources — a table of $m/\sqrt{\sigma}$
9. What the numbers say, and what they do not — upper bounds, windows, extrapolation, the string assumption

---

01

A mass is how a correlation decreases — the effective mass is an upper bound

Advancing time by one step on the lattice is applying the transfer matrix $T = e^{-aH}$ once (5-03). The eigenvalues of H are read from how the correlation of an operator $O(t)$, built from quantities on one time slice t, decreases. O is summed over space to give zero momentum, and its vacuum expectation value is subtracted (lattice units $a = 1$).

$$C(t) = \langle O(t)\, O(0)\rangle - \langle O\rangle^2 = \sum_{n\geq 1} |\langle n|O|\Omega\rangle|^2\, e^{-E_n t}, \qquad m_{\text{eff}}(t) = \ln \frac{C(t)}{C(t+1)} \quad (1)$$

$|\Omega\rangle$ is the vacuum, and E_n are the eigenvalues measured from the vacuum. The weights $|\langle n|O|\Omega\rangle|^2$ are squared absolute values, so they are 0 or more. The eigenvalues e^{-E_n} are positive reals because link-reflection positivity ($\beta \geq 0$) gives $T \geq 0$, and for $\beta > 0$, T does not have 0 as an eigenvalue (5-03 § chain, § transfer matrix). A sum of positive exponentials with weights of 0 or more is log-convex (by Cauchy–Schwarz, $C(t)^2 \leq C(t-1)C(t+1)$), so the effective mass $m_{\text{eff}}(t)$ decreases monotonically in t and approaches from above the energy of the lightest state that O touches Known . A value read at finite t is an **upper bound**, and no lower bound comes out (for an example with a finite window, 4-11 § window). On a lattice periodic in the time direction (length N_t), each term gets a backward $e^{-E_n(N_t - t)}$. For $N_t = 32$, $t \leq 7$ of § 07 it is negligible.

Example. For $C(t) = 0.8\, e^{-0.77t} + 0.2\, e^{-1.8t}$, m_{eff} goes down as 0.908, 0.824, 0.790, 0.777, 0.773, 0.771 for $t = 0, \ldots, 5$ Computed .

02

Name tags for glueballs — J^{PC} and the rotation group of the cube

Gauge-invariant states are called glueballs and are classified by spin J, parity P and charge conjugation C. P is the operation reversing the spatial coordinates $\vec{x} \to -\vec{x}$ (with the links mapped accordingly), and C is the operation mapping links to their

complex conjugates $U \to U^*$. For the trace of a loop, $\operatorname{tr} U^* = \operatorname{tr} U$, which is the trace of the loop with the opposite orientation, so C reverses the orientation of loops. The lightest is thought to be 0^{++}, and the next 2^{++} Physics .

The rotation group of the cube and its characters

The only rotations left on the lattice are the 24 that carry a cube onto itself, and the irreducible representations of this group O are A_1, A_2, E, T_1, T_2 (dimensions 1, 1, 2, 3, 3). "Name tag" is a metaphor; precisely, it is the pair of the irreducible representation of O to which a state belongs and the eigenvalues (± 1) of P and C.

The representation of spin J splits on the lattice into a sum of these. Characters decide how it splits. The character $\chi(g)$ of a representation is the trace of the matrix representing the element g. The 24 elements of O fall into five sets on which characters take the same value (conjugacy classes): the identity (1), 90° rotations about axes through face centres (6), 180° rotations about the same axes (3), 120° rotations about diagonal axes (8), and 180° rotations about axes through edge midpoints (6). The number of times a representation Γ appears is the following sum, which comes from a theorem, the orthogonality of characters of finite groups (here we only use it) Known .

$$n_\Gamma = \frac{1}{24} \sum_{g \in O} \chi_J(g)\, \chi_\Gamma(g), \qquad \chi_J(\theta) = \frac{\sin\big((2J+1)\theta/2\big)}{\sin(\theta/2)}$$

θ is the rotation angle of g. For example, for $J = 2$, $\chi_2 = (5, -1, 1, -1, 1)$, and $n_E = \frac{1}{24}(5 \cdot 2 + 3 \cdot 1 \cdot 2 + 8 \cdot (-1) \cdot (-1)) = 1$.

Known **Characters of O** The columns are the five conjugacy classes above (number of elements).

	Identity (1)	90° (6)	180° (3)	120° (8)	Edge 180° (6)
A_1	1	1	1	1	1
A_2	1	−1	1	1	−1
E	2	0	2	−1	0
T_1	3	1	−1	0	−1
T_2	3	−1	−1	0	1

Computed Known **Table 1** How the representation of spin J splits under the rotation group O of the cube (computed from the characters of the five conjugacy classes; we checked that the dimensions add up to $2J+1$).

J	0	1	2	3	4
Splitting	A_1	T_1	$E + T_2$	$A_2 + T_1 + T_2$	$A_1 + E + T_1 + T_2$

0^{++} is measured in A_1^{++}, but A_1 also contains $J = 4$. Lucini–Teper 2001 themselves raise the possibility that the excitation 0^{++*} is the lightest 4^{++}. 2^{++} appears in E^{++} and T_2^{++}, and the two agreeing is a sign that rotational symmetry is restored (4.582(14) and 4.578(11) for $SU(\infty)$ in Athenodorou–Teper 2021) Physics .

Example. The sum of spatial plaquettes in the three orientations $P_{12} + P_{23} + P_{31}$ is A_1^{++}, and the difference $P_{12} - P_{23}$ is E^{++}. For $SU(2)$, $U^* = gUg^{-1}$ ($g = i\sigma_2 \in SU(2)$, where σ_2 is the Pauli matrix of 1-04), so C is the same as making the same gauge transformation at every point. Hence every gauge-invariant state has $C = +$.

03

The noise does not decrease — the signal falls exponentially, and the fluctuation stays

$C(t)$ is estimated by an average over N configurations (how configurations are made and their errors: 5-07). The variance of the product $O(t)O(0)$, which sets the error, approaches a constant of about $\langle O^2 \rangle^2$ as t grows, and does not decrease. The signal falls as $e^{-E_{\mathrm{min}} t}$, so

$$\frac{\delta C(t)}{C(t)} \approx \frac{\text{const.}}{\sqrt{N}} e^{+E_{\mathrm{min}} t} \quad (2)$$

This argument is standard in lattice spectroscopy and is said to go back to Parisi 1984 Known . Glueballs have a large E_{min}, and 0^{++} also needs the vacuum subtracted, so it bites especially hard.

Example. In the demonstration of § 07 ($E_{\mathrm{min}} \approx 0.77$, about 35,000 configurations), the relative errors of the plaquette $C(t)$ were 0.004, 0.011, 0.025, 0.050, 0.121, 0.214, 0.379, 0.657 for $t = 0, \ldots, 7$ Computed . The factor 1.8–2.9 per step is of the same order as $e^{0.77} \approx 2.2$.

04

Fattening the operators — smoothing and blocking in space only

A plaquette has size a, but the physical length $1/\sqrt{\sigma}$ is about $7.5a$ for $SU(2)$ at $\beta = 2.60$ (Table 2). So we "fatten" the operators. Precisely, we replace links by averages over nearby paths, so that the operator overlaps strongly with extended states.

$$U_i(x) \longmapsto \mathrm{Proj}\Big[U_i(x) + \alpha \sum_{j \neq i} \big(U_j(x)\, U_i(x+\hat{j})\, U_j(x+\hat{i})^\dagger + (\text{reversed}) \big) \Big] \quad (3)$$

i, j run over spatial directions only, and Proj is the operation that brings the sum back to the group. There is also blocking, which builds links of length $2a$. Lucini–Teper 2001 cite Teper 1987 and Albanese et al. 1987 (APE smoothing) for these as standard methods Known . Averaging only in space is the key: the operator stays on one time slice, so the positive weights of (1) and "approaching from above" are kept.

Quality is measured by the fraction c_0^2 of the weight in the ground state. For the best operators of Lucini–Teper 2001, for 0^{++} it was from about 0.95 ($SU(2)$) to about 0.90 ($SU(5)$) on coarse lattices, and from about 0.92 to about 0.85 on fine lattices. For periodic strings it is from about 0.90 to about 0.75 on both. The larger N, the worse the overlap Physics .

For the three-dimensional $\mathbb{Z}_2$ of § 07, the rough fraction $C(2)e^{2m}/C(0)$ ($m = 1/1.296$; it comes out on the large side) was 0.57 for the plaquette and 0.85 for the 3×3 loop. With a smoothing in which Proj is a majority vote of the link and two staples, 0.59 rose to 0.66 in one step Computed .

05

The variational method — the generalised eigenvalue problem, and optimising the upper bound

We line up candidates $O_1, \dots, O_N$ and let the correlations themselves choose the best linear combination. For the correlation matrix $C_{ij}(t) = \langle O_i(t) O_j(0) \rangle - \langle O_i \rangle \langle O_j \rangle$ we solve

$$C(t)\, v_n = \lambda_n(t, t_0)\, C(t_0)\, v_n, \qquad E_n^{\mathrm{eff}}(t) = \ln \frac{\lambda_n(t, t_0)}{\lambda_n(t+1, t_0)} \quad (4)$$

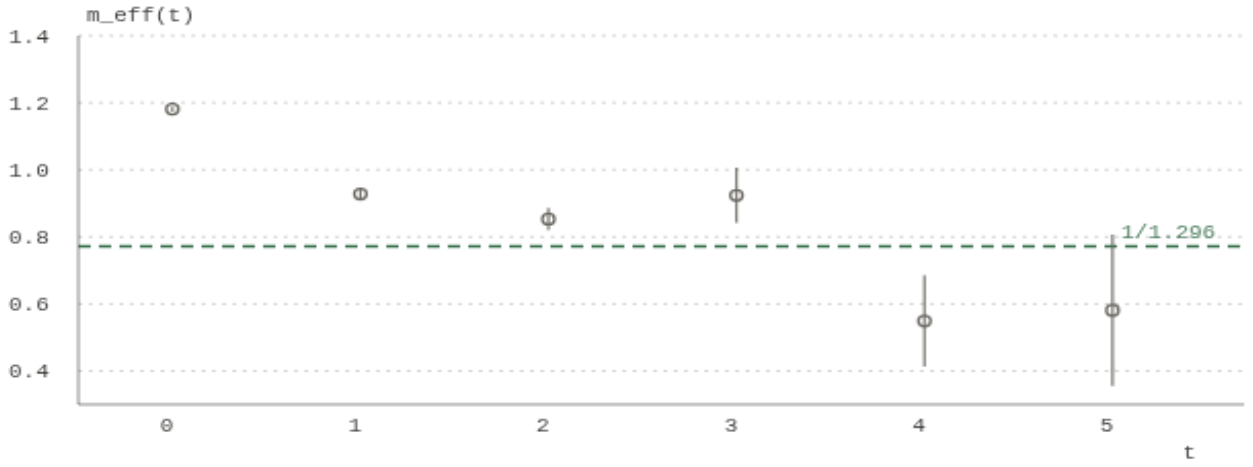
The variational idea began with Wilson's 1981 talk, and is said to have been used for glueballs by Berg 1982 and Michael–Teasdale 1983. The form of the generalised eigenvalue problem with t_0 and the estimate of its corrections are attributed to Lüscher–Wolff 1990 (according to the introduction of Blossier et al. 2009) Known . Blossier et al. 2009 show that taking $t_0 \geq t/2$ makes the corrections vanish as $\exp\!\big(-(E_{N+1} - E_n)t\big)$. The $t_0 = 0$ of Figure 1 is outside this condition, and is used only in the sense of the upper bound of formula (5).

For $t_0 = 0$, $t = 1$ it is the Rayleigh–Ritz variational principle itself (1-10, 2-05). For $\psi = \sum_i v_i O_i |\Omega\rangle$ (with the vacuum component subtracted)

$$\lambda_{\max}(1, 0) = \max_{v} \frac{v^\dagger C(1)\, v}{v^\dagger C(0)\, v} = \max_{\psi} \frac{\langle \psi | T | \psi \rangle}{\langle \psi | \psi \rangle} \leq e^{-E_1} \quad (5)$$

so $-\ln \lambda_{\max}$ is the best upper bound within the space spanned by the candidates Known . The variational method improves the upper bound, but does not change the fact that it is an upper bound.

Figure 1 The effective mass of 0^{+} in three-dimensional $\mathbb{Z}_2$. Choosing an operator with the buttons changes the points. The dashed line is $1/1.296$ from the primary source




```
1×1: m_eff(0) = 1.181, m_eff(3) = 0.924±0.082 (reference 0.772)
```

Computed The numbers in this figure were computed on this machine.

Computed The numbers of Figure 1 (jackknife errors, 39 bins. The variational basis is 1×1, $1 \times 2 + 2 \times 1$, 2×2, 3×3, with $t_0 = 0$).

t	0	1	2	3	Fraction
1×1	1.181(7)	0.928(17)	0.853(33)	0.924(82)	0.57
2×2	0.940(6)	0.858(14)	0.841(29)	0.842(70)	0.78
3×3	0.883(6)	0.818(13)	0.831(29)	0.795(61)	0.85
Variational (4)	0.860(6)	0.824(13)	0.831(29)	0.803(63)	0.87

How to read it. At $t = 0$ the variational value is lowest (formula (5)); every column goes down to 0.80–0.85 at $t = 2, 3$ and approaches the reference value 0.772 from above. For $t \geq 4$ the errors exceed 0.1 (Figure 1). The rightmost column is the rough fraction (§ 04).

06

Measuring the string tension — Creutz ratios and periodic strings

The string tension σ is the slope of the energy $V(R)$ of two separated static charges. An $R \times T$ Wilson loop (5-01) decreases like $e^{-V(R)T}$, and at long distances the following form is used (17-03) Physics .

$$V(R) = \sigma R + \mu - \frac{\pi(d-2)}{24R}, \qquad \chi(R,R) = -\ln \frac{W(R,R)\,W(R-1,R-1)}{W(R,R-1)\,W(R-1,R)} \quad (6)$$

d is the spacetime dimension. In the Creutz ratio χ (attributed to Creutz 1980) the perimeter and corner contributions cancel. Assuming the form (6) and ignoring finite-T corrections, $\chi(R,R) \approx V'(R-\tfrac{1}{2}) = \sigma + \pi(d-2)/\big(24(R-\tfrac{1}{2})^2\big)$ approaches σ from above Physics .

Periodic strings

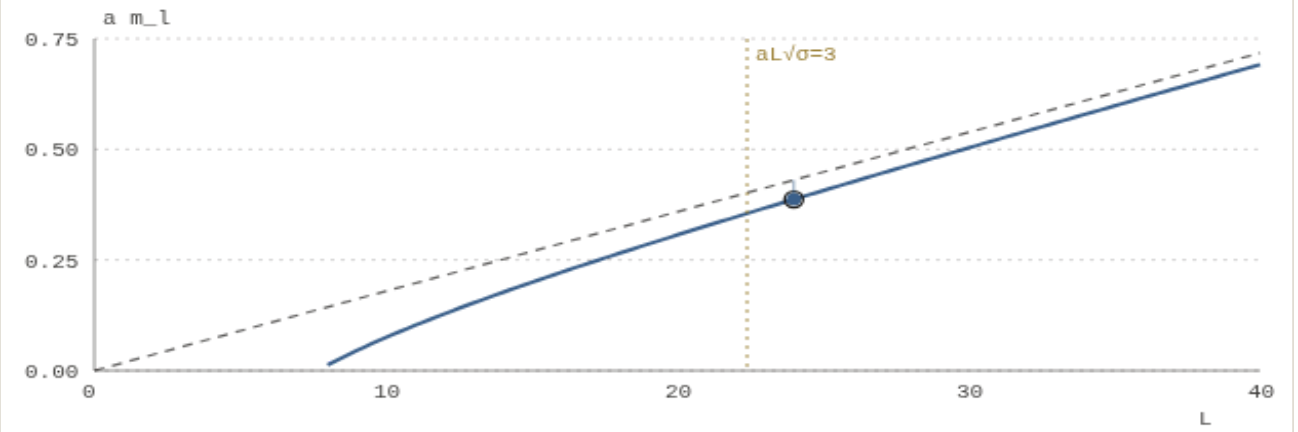
The other way is to read, by the same method as § 01, the mass m_l of a closed flux tube going once around the spatial period L (a periodic string). Lucini–Teper 2001 extracted σ in four dimensions from the following formula (the correction is in general $\pi(d-2)/(6L)$; in three dimensions $\pi/(6L)$; formula (7) of 17-03 § closed strings).

$$a\,m_l = a^2\sigma L - \frac{\pi}{3L} \quad (7)$$

The authors write 'in using eqn(4) we are assuming that we have linear confinement … Nambu-Gotto universality class' (Nambu–Goto is the approximation that treats the string as a membrane with action proportional to its area; 17-03), and caution that with two values of L 'we cannot claim to have evidence for both the 1/L functional form and for its coefficient'. Only lattices with $aL\sqrt{\sigma} \geq 3$ are used. The authors write that to study the string correction itself, periodic strings are much better than Wilson loops: the coefficient of the correction is larger ($\pi/3$ against $\pi/12$), and the Coulomb term of static charges (which has the same $1/r$ form and dominates at short distances) does not mix in.

Periodic strings do not mix with glueballs because of centre symmetry. Multiply all links in direction i crossing the plane $x_i = c$ of spatial direction i (at all times) by an element of the centre of the group (an element commuting with every element; -1 for $SU(2)$; 1-04). Every plaquette either does not contain these links or contains them as a there-and-back pair, so the action does not change. A loop going once around in direction i crosses the plane once and picks up one centre element (for $SU(2)$ its sign changes), while contractible closed loops do not change. In the confining phase this symmetry is not broken, so the correlation between the two is 0 Known .

Figure 2 Moving L changes the difference between the string mass (solid) and σL (dashed), and the deviation of the naive $\sqrt{m_l/L}$. $SU(2)$, $\beta = 2.60$. The circle is the measured value at $L = 24$ in the same paper



L = 24: a m_l = 0.3870, correction fraction 10.1%, naive a√σ = 0.1270 (94.8% of the true value)

Computed The numbers in this figure were computed on this machine (formula (7)).

Computed **Table 2** $a\sqrt{\sigma}$ recomputed by formula (7) from the am_l of Table 1 of Lucini–Teper 2001 ($SU(2)$). The master values of the six β and masses are in 16-05 Table 3.

β	L	am_l	$a\sqrt{\sigma}$	Correction fraction	$L\,a\sqrt{\sigma}$
2.25	8	1.301(17)	0.4231(25)	9.1%	3.38
2.30	10	0.861(11)	0.3108(18)	10.8%	3.11
2.40	12	0.745(9)	0.2634(14)	10.5%	3.16
2.475	16	0.585(8)	0.2016(12)	10.1%	3.23
2.55	20	0.453(4)	0.1590(6)	10.4%	3.18
2.60	24	0.387(4)	0.1340(6)	10.1%	3.21

How to read it. The $a\sqrt{\sigma}$ of the paper are reproduced by formula (7) within rounding (within 1 in the error of the last digit). $L\,a\sqrt{\sigma}$ is 3.1–3.4 at every β, so the box is almost constant in physical length. The correction is about 10% of σL, and leaving it out underestimates $\sqrt{\sigma}$ by about 5% Computed .

07

A demonstration on this machine — three-dimensional $\mathbb{Z}_2$ gauge theory

We check the same way of measuring in the simplest gauge theory: three-dimensional $\mathbb{Z}_2$ gauge theory, with ± 1 on the links and the Wilson action $S = -\beta \sum_p U_p$. It is known to map by duality to the three-dimensional Ising model (attributed to Wegner 1971; for how the correspondence is used, Agostini et al. 1997). The mass of 0^+ corresponds to the inverse correlation length of the Ising model, and the string tension to the interface tension (the free energy per unit area carried by the boundary between up and down regions; 2-09 § low temperature) Known .

Agostini et al. 1997 give, at $\beta = 0.72484$, $\xi_{0^+} = 1.296(3)$ (stable for $N_s \geq 16$, Table 2 of that paper), $\sigma = 0.05558(10)$ and $\sigma\xi^2 = 0.0934(6)$ (Table 6, from the literature on interface tension), so $m/\sqrt{\sigma} = 3.27$ Physics .

On this machine we made about 35,000 heat-bath updates on a $16^2 \times 32$ lattice (three independent runs), measuring the operators of § 04 and § 05 and the periodic string every time, and Wilson loops once every 5 (majority-vote smoothing in two other runs).

Computed **Table 3** Demonstration and reference values for three-dimensional $\mathbb{Z}_2$ at $\beta = 0.72484$. σ is the $m_{\mathrm{eff}}(2)$ of the periodic string put into $m_l = \sigma L - \pi/(6L)$. The rough guide for Creutz ratios is the crude approximation $\chi \approx V'(R - \frac{1}{2})$, without corrections for finite T (the loops being square).

Quantity	This machine	Reference value
m_{0^+} (variational, $t = 3$)	0.803(63)	0.7716(18)
m_l ($L = 16$)	0.872(39)	0.857 (from the formula)
σ	0.0566(24)	0.05558(10)
$m/\sqrt{\sigma}$	3.38(27)	3.27
$\chi(R,R)$, $R = 4, 5, 6$	0.0698(6), 0.0645(11), 0.0664(26)	Guide $\sigma + \frac{\pi}{24(R-1/2)^2}$: 0.066, 0.062, 0.060

How to read it. Both the mass and the string tension agree with the reference values within errors. The central value of the mass lies above the reference value, but the difference is about 0.5 times the error and cannot be distinguished. Being an upper bound is a property of the exact expectation value $C(t)$; an estimate from finitely many configurations can also come out below because of noise ($t \geq 4$ in Figure 1). The value at $t = 3$ with $t_0 = 0$ is a ratio of two largest eigenvalues, and is not a strict upper bound like formula (5) either. The Creutz ratios are 16–19% larger than σ even at $R = 5, 6$, slightly above the guide. The periodic string gave σ to 4% accuracy from the single point $L = 16$.

The ratio differs from theory to theory. The continuum value of Agostini et al. 1997 is $m_{0^+}/\sqrt{\sigma} = 3.08(3)$, and the value for three-dimensional $SU(2)$ that Table 9 of that paper cites from other literature is 4.763(31) Physics .

08

Numbers from primary sources — a table of $m/\sqrt{\sigma}$

These are continuum-limit values from primary sources for four-dimensional pure Yang–Mills theory (for the extrapolation procedure, 5-09 § ratios).

Physics Computed **Table 4** Continuum limits of $m/\sqrt{\sigma}$, and $2^{++}/0^{++}$ divided on this machine (errors combined as independent. Only the Meyer row is the ratio from the source's table; divided on this machine it is 1.47(4)). LT = Lucini–Teper 2001, AT = Athenodorou–Teper 2020 ($SU(3)$) and 2021 ($SU(2)$, $SU(\infty)$; 2^{++} is E^{++}). $SU(\infty)$ is the $1/N^2$ extrapolation.

Source, group	0^{++}	2^{++}	$2^{++}/0^{++}$
LT $SU(2)$	3.844(61)	5.59(15)	1.45(5)
LT $SU(3)$	3.607(87)	5.13(22)	1.42(7)
LT $SU(4)$	3.49(14)	5.21(21)	1.49(9)
LT $SU(5)$	3.38(16)	4.88(38)	1.44(13)
LT $SU(\infty)$	3.37(15)	4.93(30)	1.46(11)
AT $SU(2)$	3.781(23)	5.349(20)	1.415(10)
AT $SU(3)$	3.405(21)	4.894(22)	1.437(11)
AT $SU(\infty)$	3.072(14)	4.582(14)	1.492(8)
Meyer $SU(3)$ (via GKN II)	3.347(68)	4.916(91)	1.46(5)

Morningstar–Peardon 1999 and Chen et al. 2006 report for $SU(3)$, in units of the Sommer scale r_0 (the distance at which the force F between static charges satisfies $r_0^2 F(r_0) = 1.65$), $r_0 m(0^{++}) = 4.21(11)(4)$ and $4.16(11)(4)$; dividing by the $r_0\sqrt{\sigma} = 1.160(6)$ of AT 2020 gives 3.63(10) and 3.59(10) Computed . It is a conversion combining numbers from different sources and is only a rough guide.

How to read it. The 0^{++} ratio goes down with N, and $2^{++}/0^{++}$ is 1.4–1.5 for every N. Even for the same group the sources differ: the difference 0.063 between LT and AT for $SU(2)$ is about 1 times the combined error, and the difference 0.20 for $SU(3)$ about 2.3 times Computed . This chapter does not identify the cause of the differences. Conversion to GeV is treated in 5-09 § detour, and large N in 17-04.

09

What the numbers say, and what they do not — upper bounds, windows, extrapolation, the string assumption

Table 4 is numerical support for the physical view that the positive gap is about 3 to 4 times $\sqrt{\sigma}$. To make it a mathematical statement, four things remain.

1. **Upper bounds.** The $a\,m$ at each β is an upper bound at finite t or a fit to a plateau (§ 01, § 05). No lower bound comes out of a finite window (4-11 § window).

2. **Boxes and quantum numbers.** In a finite box, two glueballs or a pair of periodic strings also appear with the same quantum numbers (Morningstar–Peardon 1999 say in their abstract that they distinguished them). Spin is identified through the representations of the cube (§ 02).
3. **Extrapolation.** Continuum-limit values are extrapolations that assume the form of the a^2 corrections (5-09, 5-05). That the limits of the ratios of masses and string tensions exist has not been shown (the construction published in 2026-11 built limits of correlations of local fields, but does not treat the mass gap or Wilson loops; 6-04).
4. **The string assumption.** σ is extracted by assuming linear confinement and corrections of Nambu–Goto type (§ 06). Confinement at large β for four-dimensional non-abelian groups has not been proved, so using this σ as evidence of confinement is circular (17-03).

What the mass gap problem (4-11) asks for is the continuum theory and a **lower bound** on the gap, while every method in this chapter approaches from above. What to hold fixed when taking the limit is treated in 5-09.

What this chapter can say

	Content
Said	That the effective mass decreases monotonically and gives an upper bound on the lowest energy. That the generalised eigenvalue problem at $t_0 = 0, t = 1$ is the Rayleigh–Ritz upper bound. That smoothing in space only keeps the operators on one time slice. How J splits under the cube group (Table 1) Known .
At the level of numerics	The $m/\sqrt{\sigma}$ of primary sources (Table 4) and the reproduction of $a\sqrt{\sigma}$ by formula (7) (Table 2) Physics Computed . That the demonstration with three-dimensional $\mathbb{Z}_2$ agreed with the reference values within errors (Figure 1, Table 3) Computed .
Not said	That the continuum limit of four-dimensional Yang–Mills and the limit of $m/\sqrt{\sigma}$ exist; a lower bound on the gap. An independent check of the string correction. The cause of the spread in the $SU(3)$ values. This chapter shows nothing about the existence of the mass gap.

Sources and reproduction

Item	Kind	Source / tools
am_l, $a\sqrt{\sigma}$ and am of $SU(N)$ (Table 1 of the paper), formula (4), continuum values (Table 5 of the paper), the operators and the variational procedure, c_0^2, the question about 4^{++}, verbatim quotations	Primary (text)	B. Lucini, M. Teper, JHEP 0106 (2001) 050, arXiv:hep-lat/0103027
$m/\sqrt{\sigma}$, $r_0\sqrt{\sigma}$, E^{++} and T_2^{++}	Primary (text checked against the checked material)	A. Athenodorou, M. Teper, JHEP 11 (2020) 172, arXiv:2007.06422; JHEP 12 (2021) 082, arXiv:2106.00364
$r_0 m$, abstract	Primary (text checked against the material)	C. Morningstar, M. Peardon, Phys. Rev. D 60 (1999) 034509, arXiv:hep-lat/9901004; Y. Chen et al., Phys. Rev. D 73 (2006) 014516, arXiv:hep-lat/0510074
Meyer's $SU(3)$ values and the ratio 1.46(5)	Secondary (original not obtained)	The values of H. B. Meyer, arXiv:hep-lat/0508002, from the table of U. Gürsoy, E. Kiritsis, F. Nitti, JHEP 0802 (2008) 019, arXiv:0707.1349
ξ_0+ of three-dimensional $\mathbb{Z}_2$ (Table 2 of the paper), σ and $\sigma\xi^2$ (Table 6), 3.08(3), how the dual correspondence is used	Primary (text)	V. Agostini, G. Carlino, M. Caselle, M. Hasenbusch, Nucl. Phys. B 484 (1997) 331, arXiv:hep-lat/9607029
4.763(31) for three-dimensional $SU(2)$	Secondary (via Table 9 of Agostini et al. 1997; original not checked)	Table 9 of the same
The corrections of the generalised eigenvalue problem and the condition $t_0 \geq t/2$, the history of the variational method (introduction of the paper)	Primary (text)	B. Blossier et al., JHEP 0904 (2009) 094, arXiv:0902.1265
Starting points of the variational method (Wilson's 1981 talk, Berg 1982)	Bibliographic data only (from the introduction of Blossier et al. 2009; original not checked)	Quotation in the same introduction

Item	Kind	Source / tools
The variational method, smoothing, blocking, Creutz ratios, the noise argument, the string correction, duality	Bibliographic data only (attributions are "said to be")	Michael–Teasdale, Nucl. Phys. B 215 (1983) 433 (bibliographic data from the citation in Blossier et al. 2009); M. Lüscher, U. Wolff, Nucl. Phys. B 339 (1990) 222; M. Albanese et al., Phys. Lett. B 192 (1987) 163; M. Teper, Phys. Lett. B 183 (1987) 345; M. Creutz, Phys. Rev. D 21 (1980) 2308; G. Parisi, Phys. Rept. 103 (1984) 203; M. Lüscher, Nucl. Phys. B 180 (1981) 317; F. J. Wegner, J. Math. Phys. 12 (1971) 2259
Definition of the Sommer scale $r0$	Bibliographic data only (the defining formula checked against the checked material)	R. Sommer, Nucl. Phys. B 411 (1994) 839, arXiv:hep-lat/9310022
Table 1, Table 3 and Figure 1 (heat bath, jackknife, generalised eigenvalue problem), Table 2 and Figure 2, the ratios and conversions in Table 4, the example in § 01	Computed	python3 (numpy, scipy). Scripts and output are kept in the record of this book's investigation. The majority-vote smoothing was run as three runs; one was cut off and its record was damaged, so it is not used

Next chapter to read: 5-09 What to hold fixed when taking the limit (turning numbers in lattice units into physical quantities, and what to hold fixed in going to $\beta \to \infty$).

What to hold fixed when taking the limit — what gives lattice numbers a unit is a scale we choose

This chapter organises established knowledge. § 05 and § 06 are a map of open questions.

What this chapter shows — Every number a lattice calculation returns is a dimensionless number "in units of the lattice spacing", and physical units are attached only once one quantity is held fixed. What has meaning independently of how one fixes it are dimensionless ratios (for example $m/\sqrt{\sigma}$). The lattice correlation-length problem (Chatterjee's Problem 5.1) asks for the correlation length in lattice units to diverge, and "the gap in lattice units is uniformly positive" is the opposite of that.

Prerequisite chapters — 5-05 Weak coupling and the continuum limit ($a(\beta)$ and the lattice version of asymptotic freedom), 4-11 The definition of the mass gap (the gap, the correlation length, uniformity in the volume). We use the notation of 5-01 (β, links, Wilson loops) and "a critical point is where the lattice disappears" from 2-08 § continuum limit.

Said first — This chapter is about how to pose the question. Both the lattice correlation-length problem and the continuum string-tension problem (Chatterjee's Problem 5.2) are unsolved, and this chapter shows nothing about the existence of the continuum limit or about the mass gap. All numbers are lattice calculations from primary sources (at the level of numerics), and recalculations of them on this machine.

Order of this chapter

1. Lattice numbers have no unit — all one can measure are numbers in units of the lattice spacing
2. Choosing one scale — hold fixed one of σ, r_0, the correlation length or Λ (Detour: MeV is borrowed)
3. Dimensionless ratios — quantities from which the choice of scale drops out, and extrapolation in a^2
4. Misreading "a uniform gap" — hold the gap in lattice units fixed, and the physical mass goes to infinity

5. Reading the lattice correlation-length problem — the correlation length finite at each β, and divergent as $\beta \to \infty$
6. The continuum string-tension problem and where the ratio comes from — another formulation, fixing the scale with Wilson loops
7. Uniform in what — three kinds of "uniform": volume, lattice units, physical units
8. Fixing alone is not enough — three types of trivial limit
9. The order of limits — first the volume, then β. The shortcut on the numerical side

---

01

Lattice numbers have no unit — all one can measure are numbers in units of the lattice spacing

Pure Yang–Mills lattice theory (5-01 § action) has only one parameter: the β multiplying the action. The lattice sits on the integer points $\mathbb{Z}^4$, and nowhere is a "metre" or an "MeV" written.

So every number a lattice calculation returns is a number in units of one lattice step. A mass comes out only as am (the product of the lattice spacing a and the mass m), a string tension as $a^2\sigma$, and a correlation length as $\xi = 1/(am)$ (how many lattice spacings). In this book we write m_{lat} for the mass in lattice units and ξ for the correlation length in lattice units.

$$m_{\text{lat}}(\beta) = a\,m, \qquad \xi(\beta) = \frac{1}{m_{\text{lat}}(\beta)}, \qquad \sigma_{\text{lat}}(\beta) = a^2\sigma \tag{1}$$

Each left-hand side is a function of β alone, and has no dimension. The a and m on the right are not yet fixed separately at this stage. What is fixed on the lattice is only the product.

As a metaphor, it is like measuring things with a ruler on which the spacing of the marks is not written. One can say "4 marks", but not "so many centimetres". Precisely, given β, lattice theory determines a set of dimensionless functions such as m_{lat} and σ_{lat}, and a is defined only once a physical value is assigned to one of them.

Example. We copy the numbers of a lattice calculation from a primary source (Lucini–Teper 2001, $SU(2)$) into Table 1. At $\beta = 2.60$, $a\sqrt{\sigma} = 0.13395$ means "the square root of the string tension is 0.134 times the inverse lattice spacing", and $m_{\text{lat}} = 0.514$ means

"the mass of the lightest scalar glueball is 0.514 times" (we write this glueball 0^{++}: the state with spin 0, positive parity and positive charge conjugation; notation in 5-08); both are unitless numbers Physics .

Computed **Table 1** The numbers used in this chapter ($a\sqrt{\sigma}$, $m_{\text{lat}} = a\,m(0^{++})$ and lattice sizes are from Table 1 of Lucini–Teper. The six m_{lat} are the same as in 16-05 Table 3 (the master copy). $\xi = 1/m_{\text{lat}}$, and $L\,a\sqrt{\sigma}$, the side of the box measured in units of $1/\sqrt{\sigma}$, are computed on this machine).

β	Lattice	$a\sqrt{\sigma}$	m_{lat}	ξ	$L\,a\sqrt{\sigma}$
2.25	8^4	0.4231(25)	1.390(30)	0.72	3.38
2.30	10^4	0.3108(17)	1.090(33)	0.92	3.11
2.40	12^4	0.2634(14)	0.953(19)	1.05	3.16
2.475	16^4	0.2016(13)	0.754(10)	1.33	3.23
2.55	20^4	0.15896(63)	0.586(10)	1.71	3.18
2.60	24^4	0.13395(62)	0.514(8)	1.95	3.21

02

Choosing one scale — hold fixed one of σ, r_0, the correlation length or Λ

To attach units, one chooses a quantity with a dimension and fixes its value by convention. This is called **setting the scale**. If the chosen quantity is Q (of mass dimension) and the agreed value Q_*, the lattice spacing is fixed as follows.

$$aQ(\beta) := \frac{(aQ)(\beta)}{Q_*} \quad (2)$$

The numerator is a lattice number, the denominator a conventional number. Other quantities can be converted to physical units as $m = m_{\text{lat}}(\beta)/aQ(\beta)$. There are many ways to choose.

- **The string tension** σ: Creutz 1980, with 'hold fixed the string tension', saw numerically that holding the string tension fixed makes the bare coupling decrease logarithmically (abstract) Physics .
- **The Sommer scale** r_0: the length fixed by $F(r_0)\,r_0^2 = 1.65$ for the force F between static charges (Sommer 1994) Physics .
- **The correlation length itself**: Chatterjee 2018 § 5 explains sending the lattice spacing ϵ to 0 so that the correlation length approaches a finite positive value Known .

- **The Λ of asymptotic freedom**: the two-loop formula gives $a\Lambda_{\mathrm{L}}$ as a function of β, and Λ_{L} is held fixed (5-05, 4-06 § two loops).

The ratio of the lattice spacings made with two choices Q and Q' can be written with lattice numbers alone.

$$\frac{aQ(\beta)}{aQ'(\beta)} = \frac{(aQ)(\beta)}{(aQ')(\beta)} \cdot \frac{Q'_*}{Q_*} \quad (3)$$

The first factor on the right is the lattice value of the dimensionless ratio Q/Q'. If it settles to a constant as $\beta \to \infty$, the two scales differ only by a constant factor in the limit. That it settles is not a theorem but a matter supported by numerics (§ 03).

Example. We compare the scale fixed by the string tension with the scale fixed by the two-loop Λ_{L}. For $SU(2)$, with $g^2 = 4/\beta$, $b_0 = 11N/(48\pi^2)$ and $b_1 = \frac{34}{3}\left(\frac{N}{16\pi^2}\right)^2$,

$$a\Lambda_{\mathrm{L}} = (b_0 g^2)^{-b_1/(2b_0^2)}\, e^{-1/(2b_0 g^2)} \quad (4)$$

Dividing the $a\sqrt{\sigma}$ of Table 1 by (4) gives Table 2 (higher-order corrections to (4) are not included).

Computed **Table 2** The ratio $\sqrt{\sigma}/\Lambda_{\mathrm{L}}$ of the two scales for $SU(2)$.

β	**2.25**	**2.30**	**2.40**	**2.475**	**2.55**	**2.60**
$a\Lambda_{\mathrm{L}} \times 10^3$	6.704	5.914	4.600	3.809	3.152	2.777
$\sqrt{\sigma}/\Lambda_{\mathrm{L}}$	63.1	52.6	57.3	52.9	50.4	48.2

At $\beta = 2.25$ and 2.30 the lattice is coarse ($a\sqrt{\sigma} > 0.3$) and the ratio goes up and down, so comparisons are made at 2.40 and above. Between $\beta = 2.40$ and 2.60 the ratio moves from 57 to 48, by about 16%. In this range of β, "holding σ fixed" and "holding the Λ_{L} of the bare coupling fixed" do not give the same scale. This distinction (scaling and asymptotic scaling) is treated in 18-04 § the continuum limit and a(β) and in 5-05 Physics .

Detour: MeV is borrowed

A subsection you may skip.

There is no experiment corresponding to pure $SU(2)$ Yang–Mills theory, and the values of $\sqrt{\sigma}$ or r_0 written in "MeV" are conventions brought in from outside. The same holds for $SU(3)$. Athenodorou–Teper 2020 apply $r_0 = 0.472(5)$ fm, which comes from lattice QCD with quarks, to the continuum-limit $r_0\sqrt{\sigma} = 1.160(6)$, obtain $\sqrt{\sigma} = 485(6)$ MeV, and write 0^{++} as 1.653(26) GeV Physics . In the same place the authors themselves caution 'Strictly speaking this is of course not possible: the pure gauge theory and QCD … are two different theories'.

As a metaphor, it is like measuring with a ruler borrowed from the country next door. Precisely, what pure Yang–Mills theory predicts are only the dimensionless ratios of the next section, and the values in GeV are those multiplied by one convention from outside.

03

Dimensionless ratios — quantities from which the choice of scale drops out, and extrapolation in a^2

Dividing two lattice numbers cancels the lattice spacing.

$$\frac{m\mathrm{lat}(\beta)}{\sqrt{\sigma\mathrm{lat}(\beta)}} = \frac{a\,m}{a\sqrt{\sigma}} = \frac{m}{\sqrt{\sigma}} \quad (5)$$

This ratio has the same value whichever scale is chosen, since the Q_* of formula (2) appears nowhere. Pure Yang–Mills theory has no dimensionless parameter left to choose freely after renormalisation (dimensional transmutation, 4-07 § one parameter disappears). So what has meaning as a prediction of the theory are only the limiting values of ratios such as (5) Physics .

At finite β lattice corrections remain in the ratios too. For the Wilson action their leading term is thought to be proportional to a^2, and Lucini–Teper extrapolate in the following form (formula (5) of the paper; the coefficient c is really a power series in g^2, but they state that they treat it as a constant) Physics .

$$\frac{m(a)}{\sqrt{\sigma(a)}} = \frac{m(0)}{\sqrt{\sigma(0)}} + c\,a^2\sigma \quad (6)$$

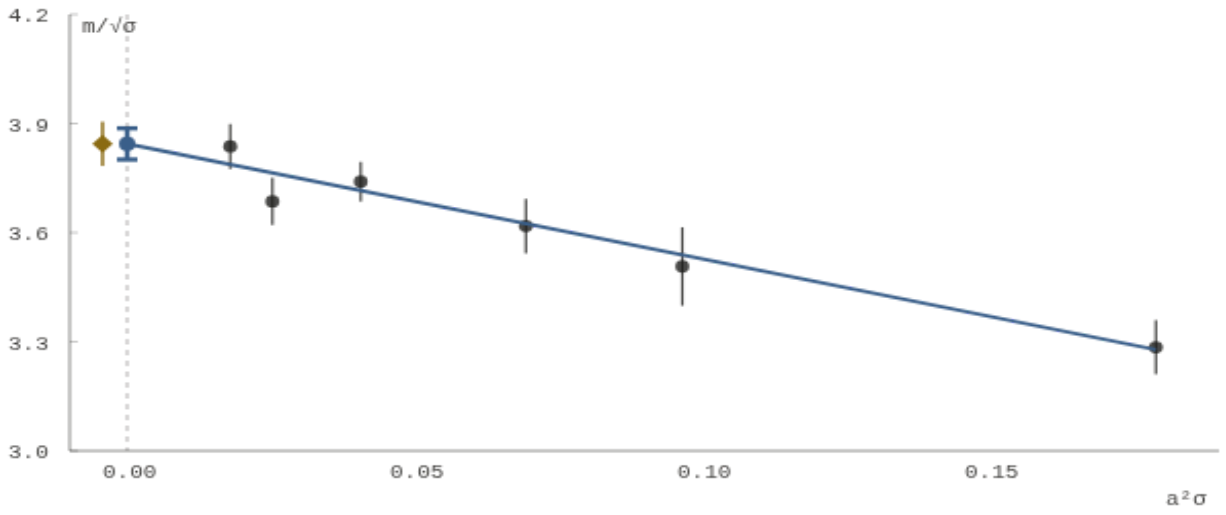
The procedure of the paper is to fit all points and, if the fit is poor, to drop points one at a time starting from the coarsest. Tables 3 and 4 redo the fits on this machine from the numbers of Table 1.

The fit is least squares with weights $1/\delta_i^2$ from the error δ_i of each point. The goodness of fit is measured by the following quantity.

$$\chi^2 = \sum_i \left(\frac{y_i - (\text{value on the line})}{\delta_i} \right)^2$$

Here y_i is the $m/\sqrt{\sigma}$ of each point. If χ^2 divided by the degrees of freedom (number of points − 2) is around 1, the fit matches the size of the errors. The confidence level is the probability that a χ^2 larger than this would occur by chance if the line were correct; the smaller it is, the worse the fit.

Figure 1 Vertical axis $m(0^{++})/\sqrt{\sigma}$, horizontal axis $a^2\sigma$ (coarser lattices to the right). Dropping points on the coarse side with the buttons changes the line and the intercept (the value extrapolated to $a = 0$). The diamond is the value 3.844(61) of Table 5 of the paper



6 points fitted: intercept 3.844(43), slope c = −3.16(53), χ²/dof = 0.60

Computed The numbers in this figure were computed on this machine (points from Table 3, lines from the fits of Table 4).

Computed **Table 3** The ratio at each β (errors propagated treating the errors of m_{lat} and $a\sqrt{\sigma}$ as independent). The rightmost column is the "misreading" value used in § 04 (with m_{lat} held at 1.390).

β	**2.25**	**2.30**	**2.40**	**2.475**	**2.55**	**2.60**
$a^2\sigma$	0.1790	0.0966	0.0694	0.0406	0.0253	0.0179
$m/\sqrt{\sigma}$	3.285(74)	3.507(108)	3.618(75)	3.740(55)	3.686(65)	3.837(62)
Misreading	3.29	4.47	5.28	6.89	8.74	10.38

Computed **Table 4** Weighted fits of formula (6).

Points used	**Intercept** $m(0)/\sqrt{\sigma(0)}$	**Slope** c	χ^2**/dof**	**Confidence level**
6 points (2.25–2.60)	3.844(43)	−3.16(53)	0.60	0.66
5 points (2.30–)	3.855(61)	−3.5(1.3)	0.77	0.51
4 points (2.40–)	3.849(72)	−3.3(1.8)	1.15	0.32
3 points (2.475–)	3.84(11)	−2.9(3.6)	2.28	0.13

How to read it. The lattice ratio moves from 3.29 to 3.84, but put on a straight line in $a^2\sigma$ the intercepts line up at 3.84–3.86. The central value of the 6-point intercept agrees with the 3.844 of Table 5 of the paper (the paper's error assignment is larger). The ratio mostly increases, approaching from below, but drops once, within errors, at $\beta = 2.475 \to 2.55$ Physics . A more recent calculation (Athenodorou–Teper 2021, 5-08 § numbers) gives 3.781(23) Physics .

What can be said here goes only as far as "assuming a^2 corrections, the numbers in this range line up at one value". It is not a proof that the limit exists. How the quantities are measured (the variational method, extracting the string tension) is treated in 5-08.

04

Misreading "a uniform gap" — hold the gap in lattice units fixed, and the physical mass goes to infinity

Here we come to an easily misread point. One is tempted to read the "mass gap" on the lattice as "the gap in lattice units $m_{\mathrm{lat}}(\beta)$ is at least some fixed positive number for all β". That the gap on a finite lattice is positive is nearly obvious (4-11 § finite boxes and uniformity), and it seems enough to make it uniform in β.

But a mass in physical units is the mass in lattice units divided by the lattice spacing.

$$m_{\mathrm{phys}}(\beta) = \frac{m_{\mathrm{lat}}(\beta)}{a(\beta)}, \qquad \inf_{\beta} m_{\mathrm{lat}}(\beta) \geq c > 0 \ \text{ and } \ a(\beta) \to 0 \implies m_{\mathrm{phys}}(\beta) \geq \frac{c}{a(\beta)} \to \infty$$

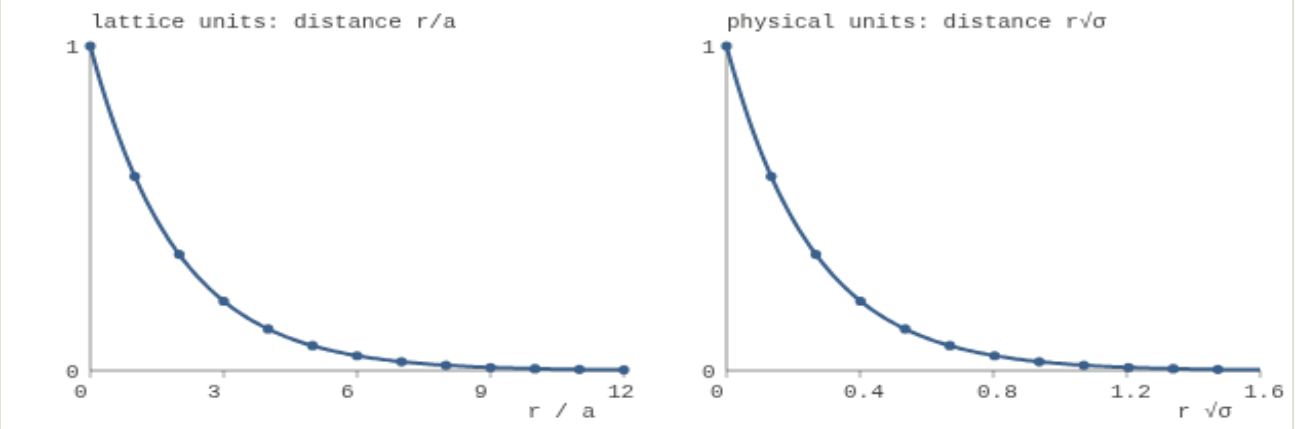
In the continuum limit $a \to 0$, so if the gap in lattice units is uniformly positive, the physical mass goes to infinity. Correlations shrink to zero width in physical distance, and the correlation between two separated points becomes 0. This is not a limit in which "the mass is positive" but one in which "nothing is transmitted".

For the mass to remain at a finite positive value in the continuum limit, m_{lat} must on the contrary go to 0 as fast as a.

$$m_{\mathrm{lat}}(\beta) \sim m_{\infty}\, a(\beta) \ \ (\beta \to \infty), \qquad \text{that is,} \quad \xi(\beta) \to \infty \ \text{(8)}$$

As a metaphor, when the marks on a ruler are made finer and finer, an object of fixed length should span more and more marks. If it always spans 2 marks, the object is shrinking. Precisely, (8) says that the correlation length in lattice units ξ diverges while the correlation length in physical units $a\xi$ is kept constant (the form seen exactly for the Gaussian chain of 2-08 § continuum limit).

Figure 2 Choosing β with the knob stretches the decay of the correlation $e^{-m_{\mathrm{lat}} r/a}$ on the left (distance in lattice units), while on the right (physical distance in units of $1/\sqrt{\sigma}$) the shape of the decay hardly changes. Pressing "Misreading" holds m_{lat} at its value for $\beta = 2.25$, and the decay on the right becomes steep



Computed The numbers in this figure were computed on this machine ($a\sqrt{\sigma}$ and m_{lat} from Table 1; the curves are exponentials drawn on the spot).

From $\beta = 2.25 \to 2.60$ the lattice spacing shrinks by a factor of about 3.2 and m_{lat} by about 2.7. The two shrink at almost the same rate, so the ratio is almost constant. The misreading column of Table 3 keeps growing in proportion to $1/a$. Note that ξ in this range does not reach 2 lattice spacings, and the numbers are not looking at the limit itself Physics .

05

Reading the lattice correlation-length problem — the correlation length finite at each β, and divergent as $\beta \to \infty$

There is a statement of the direction of the previous section, written clearly as a mathematical question: the lattice correlation-length problem (Mass gap) in § 5 of Chatterjee's survey "Yang–Mills for probabilists" (arXiv:1803.01950, 2018) Known . We translate the gist.

> Consider any compact non-abelian Lie group $G \subseteq U(N)$ $(N \geq 2)$ and any infinite-volume limit of four-dimensional lattice gauge theory. For each $x \in \mathbb{R}^4$ take the nearest plaquette p_x, and let $f_\beta(x)$ be the correlation of W_{p_0} and W_{p_x}. Show that for every $\beta > 0$ there is some $\xi(\beta) \in (0, \infty)$ such that

$$\lim_{|x| \to \infty} \frac{\log f_\beta(x)}{|x|} = -\frac{1}{\xi(\beta)} \quad (9)$$

> Show further that $\lim_{\beta \to \infty} \xi(\beta) = \infty$.

The β of the lattice correlation-length problem corresponds to $\beta' = \beta/N$ in 5-01 § action, but as far as $\beta \to \infty$ is concerned there is no difference.

Here the correlation is the connected correlation $\langle W_{\gamma 1} W_{\gamma 2} \rangle - \langle W_{\gamma 1} \rangle \langle W_{\gamma 2} \rangle$, and the same section writes that the logarithm is taken because there may be a power correction in front of the exponential. It continues 'The number ξ represents the reciprocal of the mass of the lightest glueball in the theory.'

This question has two parts.

1. **Finiteness at each** β: $\xi(\beta) < \infty$, that is, correlations decrease exponentially in lattice units. This corresponds to the gap being positive in lattice units (4-11 § correlation length).
2. **Divergence as** $\beta \to \infty$: $\xi(\beta) \to \infty$. This is the direction of (8) in § 04. The gap in lattice units must go to 0.

The first part alone does not contradict "uniformly positive in lattice units". The second part does. The lattice correlation-length problem asks for the first part at each β separately, and the second part explicitly denies uniformity.

What this question does not ask

We also note what this question **does not ask**. It does not ask for the rate of divergence. What asymptotic freedom predicts is a form like $\xi \propto e^{\beta/(4Nb0)}$ (β in the Wilson normalisation of 5-01; the reciprocal of the one-loop formula $a \propto e^{-\beta/(4Nb0)}$ of 5-05, which in Chatterjee's β' reads $e^{N\beta'/(4Nb0)} = e^{\beta'/(4b0)}$). The ratio $m/\sqrt{\sigma}$ does not appear in the wording either. That the limiting theory exists, and that it satisfies the Osterwalder–Schrader axioms (4-10), are separate questions. Even if the lattice correlation-length problem were shown, the mass gap of the Clay problem statement (0-03 § mass gap) would not follow directly.

Example. The ξ of Table 1 grows from 0.72 to 1.95 with β. These numbers agree with the direction of the second part, but ξ growing over finitely many β is not a proof of divergence in the limit Physics .

> 06
>
> **The continuum string-tension problem and where the ratio comes from — another formulation, fixing the scale with Wilson loops**

The continuum string-tension problem (Continuum limit) in the same § 5 gives another way to fix the scale Known . For the rectangular Wilson loop $W_{R,T}$ (5-01

§ plaquettes and Wilson loops), it asks to show that there are sequences with $\epsilon(\beta) \to 0$ and $c(\beta) \to \infty$ as $\beta \to \infty$, and a nonzero constant d, such that for every R, T

$$\log\langle W_{R/\epsilon,\, T/\epsilon}\rangle = -c\,(R+T) - d\,RT + o(1) \quad (10)$$

$o(1)$ is a term that goes to 0 as $\beta \to \infty$. RT is the area of the rescaled loop, and the same section writes that d corresponds to the string tension of the continuum theory. $R+T$ is a quantity proportional to the perimeter of the rescaled loop (half the perimeter; the factor 2 is absorbed into c). The d here is the coefficient of the area, not the spacetime dimension.

In (10), all that is asked is that there exist an $\epsilon(\beta)$ "in whose scale the coefficient of the area of Wilson loops is constant". That is, it is a formulation that fixes the scale with the string tension. The explanatory paragraph of § 5 fixes the scale with the correlation length instead. The continuum string-tension problem does not presuppose the lattice correlation-length problem. For neither $\epsilon(\beta)$ has existence been shown.

One more caution. About the perimeter term $c(R+T)$, Chatterjee writes in the same section 'it is not clear to me whether the renormalization term c(R + T) is indeed necessary. It seems entirely possible that the limit in (5.2) holds without the renormalization term.' Subtracting the perimeter term is one proposal for how to pose this question.

Putting the two questions together with the same scale

The ratio $m/\sqrt{\sigma}$ appears when the two questions are put together with the **same** $\epsilon(\beta)$. Suppose the continuum string-tension problem holds and moreover $\epsilon(\beta)\,\xi(\beta) \to \ell \in (0, \infty)$. The physical correlation length is $\epsilon\xi \to \ell$, the physical mass $1/\ell$ and the string tension d, so

$$\frac{m}{\sqrt{\sigma}} = \frac{1}{\ell\sqrt{d}} = \lim_{\beta\to\infty} \frac{1/\xi(\beta)}{\sqrt{\sigma_{\mathrm{lat}}(\beta)}}, \qquad \sigma_{\mathrm{lat}}(\beta) \approx d\,\epsilon(\beta)^2 \quad (11)$$

and ϵ cancels and disappears. This is (5) of § 03 written in the limit. This construction is not in Chatterjee's text; it reads the two questions together. Note that the two limits existing simultaneously with the same ϵ is itself an assumption.

Example. The lattice values of the right-hand side of (11) are the $m/\sqrt{\sigma}$ row of Table 3. At $\beta = 2.60$ it is $(1/1.95)/0.134 \approx 3.84$.

> 07

Uniform in what — three kinds of "uniform": volume, lattice units, physical units

The phrase "a uniform gap" has no definite meaning until one says uniform in what. At least three kinds of "uniform" appear in discussions of the mass gap.

Uniform in what	Form of the formula	Meaning in the continuum limit	Where it appears
In the volume L	$\inf_L m(L) \geq c > 0$	Needed. The gap does not close as the box grows	Problem statement § 6.5 (quoted below), 4-11 § finite
In β, in lattice units	$\inf_\beta m_{\text{lat}}(\beta) \geq c > 0$	By (7), the physical mass is infinite. The opposite of what is wanted	The form given by strong-coupling estimates (small β)
In β, in physical units	$\inf_\beta m_{\text{lat}}(\beta)/a(\beta) \geq c > 0$	What is wanted. Meaningful together with the choice of $a(\beta)$	The form of the lattice correlation-length problem and the continuum string-tension problem put together (§ 06)

The Clay problem statement § 6.5 writes 'New ideas are needed to prove the existence of a mass gap that is uniform in the volume of space-time.' Known . The next sentence, 'Such a result presumably would enable the study of the limit as T4 → R4', has in mind the limit from compact spacetime T^4 to $\mathbb{R}^4$. The uniformity here is that of the first row, uniformity in the volume. It is a requirement separate from the third row, and both are needed.

Theorems of the form of the second row exist. Shen–Zhu–Zhu show, for $SU(N)$ with β small enough (in four dimensions $\beta < 1/48$ in the normalisation of the paper, where the inverse coupling is $N\beta$; in the Wilson β of this book $\beta < N^2/48$: 5-01 Table 4), uniqueness of the infinite-volume measure and exponential decay of correlations (abstract) Known . Earlier, Osterwalder–Seiler 1978 proved the area law rigorously at small β (according to the text of Chatterjee 2018 § 4) Known . The abstract of that paper says it also showed that at strong coupling (small β) the infinite-volume limit exists and is analytic in β (text not read). These are results with β held in a small range, theorems about a world with the lattice spacing held fixed. They sit on the side opposite the continuum limit ($\beta \to \infty$) (5-02 § tubes, 5-04).

The constants of these estimates do not extend to $\beta \to \infty$, and do not reach the "behaviour as the scale is made finer" asked about in the continuum limit.

For gauge theories with finite groups, exponential decay of correlations of gauge-invariant quantities at large β has been proved (Adhikari–Cao, abstract) Known . This is an example in which the form of the first part of the lattice correlation-length problem

holds at large β, but since the group is not continuous it is a different setting from the question asking for $\xi \to \infty$ in the second part (5-10).

08

Fixing alone is not enough — three types of trivial limit

Fixing one scale and taking $\beta \to \infty$ looks as if it would give some limit. But there is no guarantee that the limit obtained is the theory sought. We list three types that fail.

Type 1: the correlation length does not grow

If $\xi(\beta)$ is bounded, then whatever $\epsilon(\beta) \to 0$ is chosen, $\epsilon\xi \to 0$. The physical correlation length is 0 and the mass infinite: exactly the misread world of § 04. What remains in the limit is a field in which no two points are correlated (like white noise).

Type 2: the correlation length grows, but a ratio goes to 0 or infinity

If two quantities diverge at different rates, then once the scale is fixed with one, the other escapes to 0 or infinity. For the three-dimensional compact $U(1)$ lattice theory, Göpfert–Mack proved the area law at every β (according to the text of Chatterjee 2018 § 4) Known . Even so, the abstract of that paper says, the continuum limit is a free scalar field, and the string tension in units of the mass goes to infinity (Villain action; text not read) Known . It is an example that confines at every lattice β, yet the ratio $m/\sqrt{\sigma}$ goes to 0.

Type 3: the ratio is finite, but it collapses to a Gaussian

Even if a limit can be taken with the correlation length fixed, if it is a free field (a Gaussian measure) it is not a non-trivial theory. For four-dimensional Ising-type models and φ^4 (a scalar field with a quartic self-interaction added; 7-03), the scaling limits near the critical point have been shown to be Gaussian (Aizenman–Duminil-Copin 2021) Known . This is why the problem statement asks for a "non-trivial" theory (0-03 § non-trivial, 7-08).

Putting the three together, the conditions on the limit sought are the following set.

$$\xi(\beta) \to \infty, \qquad \frac{m_{\mathrm{lat}}}{\sqrt{\sigma_{\mathrm{lat}}}} \to R_\infty \in (0, \infty), \qquad \text{the limit is not Gaussian} \ (12)$$

The second uses confinement, while the problem statement itself does not require confinement (4-11 § definition). Even if another scale such as r_0 is used instead of the string tension, the second of (12) takes the form "the ratio of two dimensionful

quantities is finite and positive". On the numerical side, Tables 3 and 4 support the second for $SU(2)$ Physics .

09

The order of limits — first the volume, then β. The shortcut on the numerical side

The theory of the problem statement needs two limits: volume $L \to \infty$, and lattice spacing $a \to 0$ (in four dimensions, tied to $\beta \to \infty$) (5-01 § the map of Part 5). The lattice correlation-length problem is written in the order of taking the infinite-volume limit **first** at each β, and considering $\beta \to \infty$ afterwards.

$$\xi(\beta) = \xi\big(\lim_{L\to\infty} \mu L, \beta\big), \qquad \text{then } \beta \to \infty \quad (13)$$

Even for the first limit, Chatterjee 2018 § 3 writes 'the uniqueness (or non-uniqueness) is in general unknown for lattice gauge theories in dimensions higher than two when β is large.' Known . This is why the lattice correlation-length problem says "any infinite-volume limit".

Numerical calculations take another road. The rightmost column of Table 1, $L\,a\sqrt{\sigma}$, is the side of the box, the lattice side L (in lattice points) measured in physical units.

Each time the lattice is made finer, the number of lattice points is increased, keeping the physical box at about $3.2/\sqrt{\sigma}$. That is, numerical extrapolation takes the road "hold the physical volume fixed and take $a \to 0$". The paper uses only lattices with $aL\sqrt{\sigma} \geq 3$, as the lengths at which the formula that obtains the string tension by subtracting the string correction $\pi/(3L)$ from the mass of a closed flux tube (formula (4) of the paper) can be used. The paper writes that finite-volume effects on glueballs also become invisible at this size Physics .

The construction published in 2026-11 also takes the road of holding the volume fixed and taking a $\to 0$, and then takes the limit along a subsequence of volumes. It does not say that the order of the two limits may be exchanged (8-05).

That this road and the road of (13) reach the same value, that is, exchanging the two limits, is something checked numerically, not a theorem. The exchange needs an estimate uniform in the size of the box, which is the requirement of the first row of the table in § 07.

Part 23 of this book uses the direction set here (the gap in lattice units should go to 0) as its starting point. The correspondence with the definitions declared there is in 23-01 The line of construction and the diffusive-relaxation-time assumption.

What this chapter can say

	Content
Said	That lattice numbers are dimensionless numbers in units of the lattice spacing, and physical units are attached by fixing one quantity (formulas (1)–(3), sorting out the definitions). That if the gap in lattice units is uniformly positive in β, the physical mass goes to infinity as $a \to 0$ (formula (7), an elementary consequence). The wording of the two problems (Chatterjee's Problems 5.1 and 5.2), Chatterjee's reservation about the perimeter term, and that uniqueness of the infinite-volume limit at large β is unknown (the text of Chatterjee 2018) Known . That putting the two questions together with the same ϵ makes the ratio $m/\sqrt{\sigma}$ appear in the form (11) (algebra under assumptions).
At the level of numerics	That $m/\sqrt{\sigma}$ for $SU(2)$ lines up at 3.84–3.86 by extrapolation in a^2, that $\sqrt{\sigma}/\Lambda_L$ moves by 16% over $\beta = 2.40$–2.60, and that the numerical box is kept at almost constant physical size (recalculations on this machine of the tables of Lucini–Teper) Physics Computed . The $O(a^2)$ form is an assumption, and extracting the string tension involves an assumption about the string correction.
Not said	Both the lattice correlation-length problem and the continuum string-tension problem are unsolved; neither $\xi(\beta) \to \infty$, nor the existence of the limit of $m/\sqrt{\sigma}$, nor the exchange of the order of limits has been shown. The statements about the continuum limit of three-dimensional compact $U(1)$ are at the level of the abstract; the text has not been checked. This chapter shows nothing about the construction of the continuum limit or about the mass gap.

Sources and reproduction

Item	Kind	Source / tools
The lattice correlation-length problem and the continuum string-tension problem (Problems 5.1 and 5.2), the sentence on uniqueness in § 3, the account of Göpfert–Mack in § 4, the reservation about the perimeter term	Primary (text)	S. Chatterjee, "Yang–Mills for probabilists", arXiv:1803.01950 (2018; PDF text § 3– § 5)
$a\sqrt{\sigma}$, $am(0^{++})$ and lattice sizes for $SU(2)$, the extrapolation formula (6), the continuum value 3.844(61), $aL\sqrt{\sigma} \geq 3$	Primary (text)	B. Lucini, M. Teper, JHEP 0106 (2001) 050, arXiv:hep-lat/0103027 (Table 1, formulas (4)(5), Table 5)
ξ and $L\,a\sqrt{\sigma}$ of Table 1, Tables 2–4, the numbers of Figures 1 and 2	Computed	python3 (weighted least squares, the confidence level of χ^2, the two-loop formula (4)). Scripts and output are kept in the record of this book's investigation

Item	Kind	Source / tools
'uniform in the volume of space-time'	Primary (text § 6.5)	A. Jaffe, E. Witten, "Quantum Yang–Mills theory" (the Clay problem statement) § 6.5
The prescription of holding the string tension fixed	Abstract	M. Creutz, Phys. Rev. D 21 (1980) 2308
The definition $F(r_0)r_0^2 = 1.65$ of r_0	Abstract	R. Sommer, Nucl. Phys. B 411 (1994) 839, arXiv:hep-lat/9310022
$r_0\sqrt{\sigma}$ for $SU(3)$, $\sqrt{\sigma} = 485(6)$ MeV, the authors' caution	Primary (text checked against the checked material)	A. Athenodorou, M. Teper, JHEP 11 (2020) 172, arXiv:2007.06422 (formulas (8)(9), Table 17, § 5.2)
$m(0^{++})/\sqrt{\sigma} = 3.781(23)$ for $SU(2)$	Primary (text checked against the checked material)	A. Athenodorou, M. Teper, JHEP 12 (2021) 082, arXiv:2106.00364 (Table 23)
The area law and continuum limit of three-dimensional compact $U(1)$	Abstract (continuum limit, Villain action; the INSPIRE abstract). Text not read	M. Göpfert, G. Mack, Commun. Math. Phys. 82 (1982) 545–606 (the year is the Crossref year of publication; INSPIRE gives 1981. The area law also checked in the text of Chatterjee 2018 § 4)
Scaling limits of four-dimensional Ising-type models and φ^4 are Gaussian	Abstract	M. Aizenman, H. Duminil-Copin, Ann. Math. 194 (2021) 163, arXiv:1912.07973
The area law and the infinite-volume limit at small β	Abstract (existence and analyticity of the infinite-volume limit, Wilson's confinement bound; the INSPIRE abstract). The area law also checked in the account of Chatterjee § 4	K. Osterwalder, E. Seiler, Ann. Phys. 110 (1978) 440
Uniqueness and exponential decay at strong coupling for $SU(N)$ ($\beta_{\mathrm{SZZ}} < 1/48$ in four dimensions)	Abstract	H. Shen, R. Zhu, X. Zhu, arXiv:2204.12737
Exponential decay of correlations at weak coupling for finite groups	Abstract	A. Adhikari, S. Cao, arXiv:2202.10375 (Ann. Probab. 53 (2025))

Next chapter to read: 5-10 Lattice gauge theory with finite groups (an example where correlations decrease at large β, yet there is no progress in the direction of the continuum limit).

Lattice gauge theory with finite groups — exponential decay at weak coupling is a theorem, but it lies on the side opposite the continuum limit

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — If elements of a finite group are put on the edges, at weak coupling the edges freeze, and only defects that pay the gap Δ_G of the action remain. From this come the first-order formula for Wilson loops (Chatterjee, Cao) and the theorem of exponential decay of correlations (Adhikari–Cao). That decay is about the side where the correlation length goes to 0, that is, the side opposite the continuum limit.

Prerequisite chapters — 5-02. Tools: 5-01, 1-08, 4-11. § 07 reads quickly after 5-09.

Said first — All the theorems are about theories with finite groups at fixed lattice spacing. This chapter shows nothing about the mass gap or the continuum limit for continuous groups. "No confinement at weak coupling" is a physical view, not a claim of the theorems cited (§ 08).

Order of this chapter

1. Making the group finite — only the edge values become discrete
2. Frozen edges — the smallest defect and Δ_G
3. The shape of defects is set by the dimension — area law and perimeter law
4. The line breaks on the way — duality and a first-order transition (Table 1)
5. Fixed by the perimeter alone — Wilson loops (Figure 1, Table 2)
6. Exponential decay of correlations — Adhikari–Cao (Figure 2, Table 3) / Detour: non-abelian defects become links
7. The side opposite the continuum limit — the correlation length goes to 0
8. Three "gaps" — a gap, but no confinement (Table 4)
9. Even approaching through subgroups of SU(2) — $\Delta_G \to 0$

01

Making the group finite — the same lattice, with only the edge values discrete

Replacing the group on the edges by a finite group G leaves the definitions of the lattice, plaquettes, gauge transformations and Wilson loops unchanged (5-01 § 02). The integral over each edge becomes an average over $|G|$ elements. The action uses the character $\chi = \operatorname{tr}\rho$ of a faithful unitary representation ρ (of dimension d; $\rho(g) \neq I$ if $g \neq 1$), and we write σ_p for the product once around a plaquette p.

$$S_\Lambda(\sigma) = \sum_p \operatorname{Re}\big(\chi(1) - \chi(\sigma_p)\big), \qquad \mu_{\Lambda,\beta}(\sigma) = \frac{1}{Z_{\Lambda,\beta}}\, e^{-\beta S_\Lambda(\sigma)} \quad (1)$$

(Adhikari–Cao (1.2)(1.3)) Known . $\chi(g)$ is a sum of d eigenvalues of absolute value 1, so each term is 0 or more, and it is 0 only when $\sigma_p = 1$.

Example. For $\mathbb{Z}_2 = \{\pm 1\}$ the weight is proportional to $e^{\beta \sum_p \sigma_p}$: Ising lattice gauge theory, which Chatterjee (2020) calls the first lattice gauge theory, defined by Wegner (1971) Known (2-09). $\mathbb{Z}_N$ is $U(1)$ with angles restricted to multiples of $2\pi/N$. For the finite subgroups of $SU(2)$, $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$ (the unit quaternions) and $\bar{T}$, $\bar{O}$, $\bar{I}$ (the rotation groups of the regular tetrahedron, octahedron and icosahedron, subgroups of $SO(3)$, pulled back to $SU(2)$; of orders 24, 48, 120; 1-12 § 08), $\chi = \operatorname{tr}$, and relative to Wilson's convention $\beta_W = 2\beta$.

A finite group has no Lie algebra, and no elements close to the identity. There is no classical continuum limit starting from the $U = e^{iaA}$ of 5-01 § 05; we read it as a lattice model with the same gauge structure.

02

Frozen edges — the smallest defect, and the gap Δ_G of the action

When β is large, the weight strongly favours configurations in which every plaquette is the identity. The only configurations on a box with this property are pure gauge ($\sigma_{(x,y)} = h_x h_y^{-1}$), and up to gauge there is just one, "every edge the identity". The weak-coupling theory is small fluctuations around it, and we say that "the edges freeze".

From here on we write D for the dimension of the lattice, and keep d for the dimension of the representation. The cheapest fluctuation is to change a single edge to $g \neq 1$. On a D-dimensional lattice that edge lies in $2(D-1)$ plaquettes, and each of their products is conjugate to $g^{\pm 1}$. Since $\chi(g^{-1}) = \chi(g)$,

$$\Delta S = 2(D-1)\,\mathrm{Re}\big(\chi(1)-\chi(g)\big) \;\geq\; 2(D-1)\,\Delta_G, \qquad \Delta_G := \min_{g\neq 1} \mathrm{Re}\big(\chi(1)-\chi(g)\big) \tag{2}$$

$\Delta_G > 0$ is equivalent to ρ being faithful (Cao § 1.2) Known . Example: for four-dimensional $\mathbb{Z}_2$, $\Delta_G = 2$, and the weight of this defect is $e^{-12\beta}$. This 12 appears in the theorem of § 05.

For $U(1)$, changing an edge to $e^{i\theta}$ costs $\approx (D-1)\theta^2$, which can be as small as one likes, so long waves (photons) are cheap to make (5-06 § 03). Δ_G is a different thing from the mass gap of 4-11.

03

The shape of defects is set by the dimension — points, loops, closed surfaces, and the area and perimeter laws

For $\mathbb{Z}_2$, taking a surface (a set of plaquettes) P with boundary γ, the inner edges appear twice each and cancel (a discrete Stokes; the existence of P is Lemma 3.1 of Chatterjee 2020) Known .

$$W_\gamma = \prod_{e\in\gamma} \sigma_e = \prod_{p\in P} \sigma_p = (-1)^{\#\{p\in P:\, \sigma_p = -1\}} \tag{3}$$

Here we use the dual lattice. It is the lattice whose vertices are the centres of the D-dimensional cubes, and the k-dimensional cells of the original lattice (edges, faces, cubes, ...) correspond one to one with the $D-k$-dimensional cells of the dual lattice. In Lemma 3.3 of the same paper, the negative plaquettes split into "vortices", and the dual of each vortex is closed. The dual of a plaquette ($k=2$) is a $(D-2)$-dimensional cell, so vortices are points in two dimensions, closed loops in three, and closed surfaces in four Known . Only vortices linked with γ change the sign (Lemma 3.2), and if their number N is Poisson-distributed with mean μ then $\mathbb{E}[(-1)^N] = e^{-2\mu}$ (1-08, 17-02 § 06).

Example. In two dimensions with free boundary the plaquettes are independent and $\mathbb{E}\sigma_p = \tanh\beta$, so for a loop of area A

$$\langle W_\gamma \rangle = (\tanh\beta)^A, \qquad \sigma_{\text{lat}} = -\ln\tanh\beta > 0 \quad (\text{for all } \beta) \tag{4}$$

Known . In two dimensions even finite groups have the area law at every β (for general groups, 7-02 § 03). In three and four dimensions at large β the vortices are small, only nearby ones link with γ, and their number is proportional to the perimeter (the perimeter law) Physics . The area law at small β is a theorem of the strong-coupling expansion (5-02 § 03) Known .

04

The line breaks on the way — duality and the first-order transition of four-dimensional $\mathbb{Z}_2$

Theorem 4.1 of Chatterjee 2020 writes the partition function of four-dimensional $\mathbb{Z}_2$ as the same theory on the dual lattice, moving the coupling to a λ that exchanges strong and weak Known .

$$\lambda(\beta) = -\tfrac{1}{2}\log\tanh\beta, \qquad \lambda(\beta_{\mathrm{sd}}) = \beta_{\mathrm{sd}} \iff \beta_{\mathrm{sd}} = \tfrac{1}{2}\ln\bigl(1+\sqrt{2}\bigr) = 0.44069 \quad (5)$$

The number in (5) is Computed . Duality maps transition points to transition points, so if there is exactly one transition it falls at the self-dual point Known . According to the abstract of Wegner (1971), the dual of the Ising model on the simple cubic lattice is a model of four-spin products (plaquettes), that is, three-dimensional $\mathbb{Z}_2$ gauge theory Known .

A first-order transition is one in which a quantity such as the mean plaquette jumps at one point of β. Disagreement (hysteresis) between runs raising β and runs lowering it is a sign of it. In four dimensions Creutz–Jacobs–Rebbi (1979) reported a first-order transition numerically, and reported that for $\mathbb{Z}_N$ with $N \geq 5$ there are two transitions that are not first order, one of which moves to $\beta \to \infty$ with N Physics .

On this machine too we ran the heat-bath method on a periodic 8^4 lattice (the run raising β from a random start, the run lowering it from all edges $+1$, averaging the second half of 200 sweeps at each β). The leading term $1 - 8e^{-12\beta}$ comes from the probability $\approx 4e^{-12\beta}$ that a defect of § 02 sits on one of the four edges.

Computed Table 1 The mean plaquette of four-dimensional $\mathbb{Z}_2$ on 8^4 (excerpt)

β	Run raising β	Run lowering β	$\tanh\beta$	$1 - 8e^{-12\beta}$
0.40	0.4214	0.4205	0.3799	0.9342
0.44	0.5025	0.9117	0.4136	0.9593
0.46	0.9206	0.9422	0.4301	0.9680
0.60	0.9928	0.9930	0.5370	0.9940

For $\beta = 0.43$–0.45 the two runs disagree (by 0.41 at 0.44), and that interval straddles the self-dual point Computed . These are short runs on a small lattice, with no power to decide the order of the transition (first order, or a transition without a jump) (5-07; the same comparison for $U(1)$ and $SU(2)$ is in 5-06 § 04).

05

Fixed by the perimeter alone — the first-order formula for Wilson loops at weak coupling

Theorem 1.1 of Chatterjee (2020): let ℓ be the number of edges of γ, and ℓ_0 the number of corner edges (edges lying in the same plaquette as another edge of γ; 8 for a rectangle). Then for β_0 and above, with universal constants C_1, C_2, Known

$$\left|\langle W_\gamma\rangle_\beta - e^{-2\ell e^{-12\beta}}\right| \le C_1\left(e^{-2\beta} + \sqrt{\ell_0/\ell}\right)^{C_2} \quad (6)$$

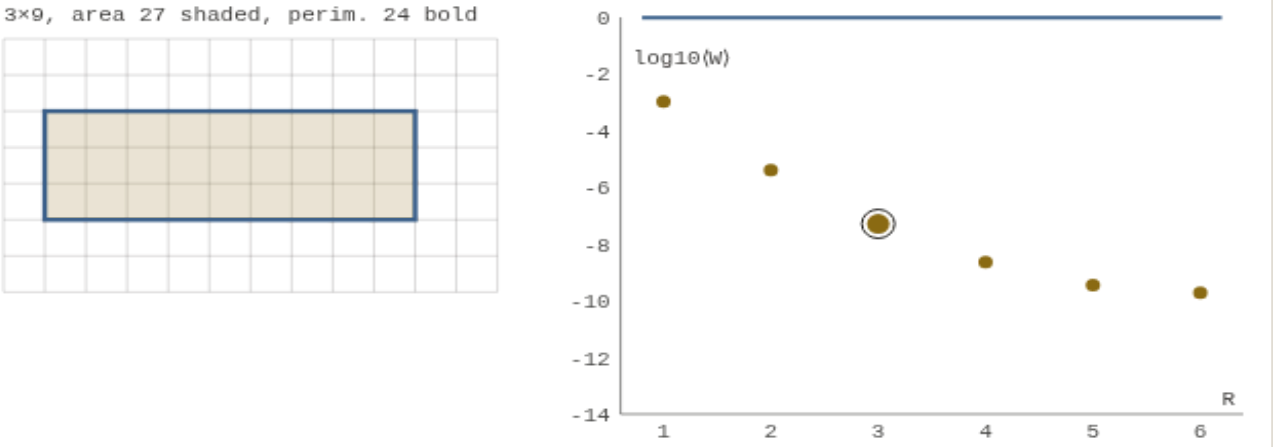
The paper reads it as "Nontrivial behavior happens if and only if ℓ is like a constant multiple of $e^{12\beta}$". Each edge becomes the centre of a defect of § 02 with probability $\approx e^{-12\beta}$ and changes the sign once; if their number is Poisson, this is the leading term of (6) (§ 03) Physics .

Cao (2020) extended this to any finite group. Its Theorem 1.2.1 is about a single self-avoiding loop γ, with conditions $\beta \ge (1000 + 14\log|G|)/\Delta_G$ and $\|A_\beta\|_{\mathrm{op}} < 1$. Under them the leading term is

$$\langle W_\gamma\rangle_\beta \approx \sum_{i=1}^{d} e^{-\ell\, r_\beta(1-\lambda_i(\beta))}, \qquad r_\beta = \sum_{g\neq 1} e^{-6\beta\,\mathrm{Re}(\chi(1)-\chi(g))} \quad (7)$$

where λ_i are the eigenvalues of the matrix A_β, the average of the $\rho(g)$ carried by defects weighted by the defect weights, and $\|A_\beta\|_{\mathrm{op}}$ is its operator norm (the largest factor by which it stretches a unit vector) Known . Because of the latter condition, $\mathbb{Z}_2$ ($A_\beta = -1$, norm 1) does not fall directly under this theorem (Cao's Remark 1.2.3) Known . Even so, putting $r_\beta = e^{-12\beta}$ and $\lambda = -1$ for $\mathbb{Z}_2$ makes the form of the leading term agree with (6) Computed . The leading term is a function of the perimeter alone: for the same perimeter, long and thin or square, it has the same value. In two dimensions, (4) was a function of the area alone.

Figure 1 Rectangles $R \times (12 - R)$ of perimeter 24. β, R and "Play" change $\log_{10}\langle W\rangle$ for the two-dimensional (4) (points) and the four-dimensional leading term (6) (horizontal line)



β = 0.60, 3×9: two dimensions 10^-7.29, four-dimensional leading term 0.9648

Computed The numbers in this figure were computed on this machine ((4)(6) computed on the spot). Example: at $\beta = 0.7$, for 1×11, 3×9, 6×6 the two-dimensional values are 3.9×10^{-3}, 1.2×10^{-6}, 1.3×10^{-8}, and the four-dimensional value is 0.989 for all. Theorem 1.1 guarantees only sufficiently large β (β_0 and above; the value is not stated explicitly in the paper). 0.4407 is a guide to the transition, and below it (dashed) the leading term does not fit even in form.

Computed Table 2 Wilson loops of four-dimensional $\mathbb{Z}_2$ on 8^4 (3000 sweeps, jackknife)

β	$R \times T$	$\langle W \rangle$	$-\ln W$ / area	$-\ln W$ / perimeter
0.40	1×1	0.42137(39)	0.864	0.216
0.40	2×3	0.00660(18)	0.837	0.502
0.50	1×1	0.97190(20)	0.0285	0.00713
0.50	3×3	0.91281(59)	0.0101	0.00760

At $\beta = 0.40$ the values divided by the area line up, and at $\beta = 0.50$ those divided by the perimeter Computed . The latter are about 1.5 times the $2e^{-6} = 0.00496$ of (6); just above the transition, corrections from corners and large vortices matter.

06

Exponential decay of correlations — the theorem of Adhikari–Cao and Δ_G

That separated observables become nearly independent (decay of correlations) had earlier results for abelian finite groups at weak coupling ($\mathbb{Z}_2$ in Chatterjee 2020 § 5, and the results of Seiler and Forsström listed in Adhikari–Cao § 1.1), and Adhikari–Cao (Ann. Probab. 2025) showed it in four dimensions including the non-abelian case. Let B_1, B_2 be rectangles separated by at least L in the ℓ^∞ distance (the largest difference of coordinates), and $|B|$ the number of plaquettes contained in a rectangle B. f_i is a function of the holonomies of loops in B_i (the product σ_γ of the edge elements along the loop; 5-01 § 02) that is conjugation-invariant in each argument ($f(h$

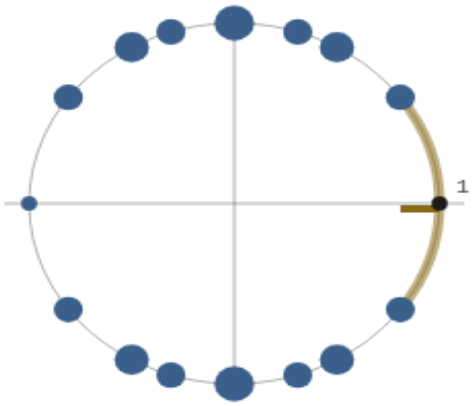
$^{-1}gh) = f(g)$). This includes any function of Wilson loops. $\|f\|_\infty = \max|f|$, and Cov is the covariance (1-08). If $\beta \geq (114 + 4\log|G|)/\Delta_G$, then in every finite box Known

$$|\mathrm{Cov}(f_1, f_2)| \;\leq\; 4\big(4\cdot 10^{24}|G|^2\big)^{|B_1|+|B_2|}\,\|f_1\|_\infty\|f_2\|_\infty\, e^{-(\beta/2)\Delta_G\,(L-1)} \quad (8)$$

It is an estimate independent of the box. The constant grows exponentially in the size of the supports of the observables (B_1, B_2), but does not depend on L. The authors write that it "can be interpreted as showing the existence of a mass gap in the setting of finite (non-Abelian) gauge groups at large β" (for small β, 5-04 § 07). The proof uses the fact that a map exchanging two independent configurations on a region preserves the measure, and bounds the covariance by "the probability that defects connect B_1 to B_2".

The condition and the rate are set by Δ_G $(1-\cos(2\pi/N)$ for $\mathbb{Z}_N$, $2-2\cos\theta_{\min}$ for subgroups of $SU(2)$).

Figure 2 The eigenvalues of the group elements placed on the unit circle. Choosing a group with the buttons changes the angle $\theta_{\min}$ closest to 1 (arc), $1-\cos\theta_{\min}$ (thick line), and the condition of the theorem



Ī Δ_G = 0.381966

Ī (order 120) : θ_min = 36°、Δ_G = 0.3820, Adhikari–Cao condition β ≥ 348.59 (β_W ≥ 697.18)

Computed The numbers in this figure were computed on this machine (the groups were built by closing under products, and the eigenvalue angles and Δ_G counted; the same as Table 3).

Computed Table 3 Δ_G and the condition on β in the theorem ($\chi = \mathrm{tr}$ for subgroups of $SU(2)$)

Group	Order	$\theta_{\min}$	Δ_G	Adhikari–Cao	Same in β_W
$\mathbb{Z}_2$	2	180°	2	58.39	–
$\bar{T}$	24	60°	1	126.71	253.42
$\bar{O}$	48	45°	0.5858	221.04	442.09
$\bar{I}$	120	36°	0.3820	348.59	697.18

The finer the group, the smaller Δ_G, and the condition grows as $1/\Delta_G$ ($\approx 2\pi^2/N^2$ for $\mathbb{Z}_N$). Cao 2020 calls its own condition (7) "very loose" (the threshold of Adhikari–Cao is about an order of magnitude smaller).

Detour: non-abelian defects become links

A subsection you may skip.

According to Adhikari–Cao § 1.2, in the abelian case separated defects appear independently, but in the non-abelian case two linked defect loops do not decompose however far apart they are Known . This is because the element a defect carries changes as $g \mapsto hgh^{-1}$ when it goes around another defect (in Q_8, $jij^{-1} = -i$ Computed). The paper writes of the non-abelian case that it does "not (as of yet) admit a cluster expansion".

07

The side opposite the continuum limit — the correlation length goes to 0

The constant in (8) depends on the size of the supports of the observables but not on L, so (8) gives an upper bound on the correlation length (4-11 § 07).

$$\xi(\beta) \;\le\; \frac{2}{\beta\,\Delta_G} \xrightarrow{\beta\to\infty} 0 \quad (9)$$

Known (a consequence of (8), with the definition of 4-11 § 07)

At β exactly at the condition, the upper bound is 0.0171 lattice spacings for $\mathbb{Z}_2$ and 0.0150 for $\bar{I}$. The perimeter $1/\big(r\beta(1-\lambda)\big)$ at which the leading term of (7) moves away from 1 is, at β exactly at Cao's condition, $10^{2721.1}$ for $\bar{T}$ and $10^{2780.1}$ for $\bar{I}$ Computed . The conditions are loose, so what matters is the direction: a finite group at weak coupling is a frozen lattice whose correlations do not reach even one lattice spacing.

The lattice correlation-length problem of 5-09 § 05 (Problem 5.1 of Chatterjee 2018, arXiv:1803.01950) asks, for four-dimensional non-abelian compact groups, for $\xi(\beta) < \infty$ at each β, and $\xi \to \infty$ as $\beta \to \infty$. Finite groups satisfy the first part at large β, and go to $\xi \to 0$, the exact opposite of the second Known . This is "Type 1" of 5-09 § 08, and the physical mass is infinite.

08

The precise form of "a gap, but no confinement" — three "gaps"

Theories with finite groups are cited as examples of "a gap, but no confinement". This "gap" has three meanings.

Known Table 4 Three meanings of "gap", and the state of finite groups at weak coupling

Meaning	Content	Finite group at weak coupling
(a) The gap of the problem statement	The spectral gap of a continuum, infinite-volume theory (4-11)	There is no continuum theory, so it does not apply
(b) The strong form	Exponential decay under any boundary condition (the definition of strong decay [Definition 2.3 of Chatterjee 2021])	Not shown (the Theorem from Uniform Decay to the Area Law [Theorem 2.4 of Chatterjee 2021] assumes a connected group, and finite groups do not fall under it literally)
(c) The weak form	Decay of covariances in a given measure (the form of (8))	Theorem (Adhikari–Cao)

The Theorem from Uniform Decay to the Area Law of Chatterjee (2021) takes the group G to be a closed connected subgroup of $U(n)$ and derives the area law from (b) for representations acting non-trivially on the centre Known . Non-trivial finite groups are not connected, so they do not literally fall under this assumption. Even so, p.8 of the same paper cites theories with finite groups as examples “believed to be ‘gapped’ … not confining at large β”, and explains why this does not contradict the Theorem from Uniform Decay to the Area Law: the decay there is in a weaker sense than the definition of strong decay (truncated two-point correlations under certain boundary conditions) Known . The paper goes on to say that for some theories it has been rigorously shown that there are several Gibbs measures. Read with the proviso "if the proof goes through without connectedness", (b) does not hold for non-confining finite groups.

The errors of (6)(7) do not go to 0 as the loop grows, so that the string tension is 0 at fixed β is not a claim of the theorems cited here (we have not checked, in this chapter, a result that states the perimeter law directly for abelian finite groups). It is the standard physical view (consistent with Table 2) Physics . Confinement and the gap are different statements (17-02 § 02).

09

Even approaching through subgroups of SU(2) — $\Delta_G \to 0$ and uniformity

According to the abstract of Petcher–Weingarten (1980), $\bar{T}$, $\bar{O}$ and $\bar{I}$ all have first-order transitions, and as the order is raised the critical coupling goes to 0. $\bar{I}$ agreed with the numbers for $SU(2)$ over a wide range of couplings Physics .

We make a rough guide. For $SU(2)$ at $\beta_W = 2.4$ in 5-02, $\langle 2 - \operatorname{tr} U_p \rangle = 0.739 > \Delta_{\bar{I}} = 0.382$. From the leading weak-coupling term $\langle \frac{1}{2} \operatorname{tr} U_p \rangle \approx 1 - 3/(4\beta_W)$ (5-06 § 04), $\langle 2 - \operatorname{tr} U_p \rangle \approx 3/(2\beta_W)$ equals Δ_G at

$$\beta_W \approx \frac{3}{2\Delta_G} : \qquad \bar{T}\ 1.50, \qquad \bar{O}\ 2.56, \qquad \bar{I}\ 3.93 \ (10)$$

Computed . It is the crude estimate that for β_W larger than this the fluctuations of $SU(2)$ are finer than the steps of the finite group, not the value of a transition point Physics .

The point is the direction of the constants. Bringing the subgroups closer to $SU(2)$ gives $\Delta_G \to 0$, and both the condition $\propto 1/\Delta_G$ and the rate $(\beta/2)\Delta_G$ degenerate. What is needed for $SU(2)$ is that the correlation length be finite in a regime where the action has no gap, and moreover diverge as $\beta \to \infty$. That mass gap is thought to come from the coupling growing stronger with distance (5-06 § 07, 4-07) Physics .

What this chapter can say

Matter	Said	Not said
(1)–(5)	Definitions, the smallest defect, vortices, the two-dimensional area law, duality Known . Tables 1 and 3 Computed	The vortex-gas reading (an estimate). The order of the four-dimensional transition
(6)(7)	Theorems on the leading term Known . Agreement of the form of the leading term for $\mathbb{Z}_2$, Table 2 Computed	That the string tension is 0 at fixed β
(8)(9)	Exponential decay and $\xi \to 0$ for four-dimensional finite groups at weak coupling Known	The strong form (b). Other dimensions
Relation to $SU(2)$	Reports of numerics for the subgroups Physics . The arithmetic of (10) Computed	The mass gap, the continuum limit and confinement for continuous groups. This chapter shows nothing about them

Sources and reproduction

Item	Kind	Source / tools
(1)(2)(8), quotations, the detour	Text § 1	A. Adhikari, S. Cao, arXiv:2202.10375v3; Ann. Probab. 53(1) (2025), doi:10.1214/24-AOP1702. The earlier results of Seiler and Forsström listed in § 1.1: bibliographic data only

Item	Kind	Source / tools
(3)(5)(6), quotations, $\mathbb{Z}_2$ in § 06	Text § 1– § 5	S. Chatterjee, arXiv:1811.09770v3; Commun. Math. Phys. 377 (2020) 307–340, doi:10.1007/s00220-020-03738-9
(7), the note on the conditions	Text § 1.2 (Theorem 1.2.1, Remarks 1.2.2–1.2.3)	S. Cao, arXiv:2001.05627v3; Commun. Math. Phys. 380 (2020) 1439–1505, doi:10.1007/s00220-020-03912-z
The Theorem from Uniform Decay to the Area Law and the quotation in § 08	Text § 2 (Definitions 2.1 and 2.3, Theorems 2.2 and 2.4, p.8)	S. Chatterjee, Commun. Math. Phys. 385 (2021) 1007–1039, arXiv:2006.16229
Duality and the transition in § 04	Abstract	F. J. Wegner, J. Math. Phys. 12 (1971) 2259; M. Creutz, L. Jacobs, C. Rebbi, Phys. Rev. Lett. 42 (1979) 1390, Phys. Rev. D 20 (1979) 1915
The subgroups in § 09	Abstract	D. Petcher, D. H. Weingarten, Phys. Rev. D 22 (1980) 2465
Tables 1–3, Figures 1 and 2, (5)(7)(10) and the numbers in § 07	Computed on this machine	python3 (calculations on this machine: group tables, heat bath for four-dimensional $\mathbb{Z}_2$ (about 100 seconds), comparison of area and perimeter)

Next chapter to read: 5-11 Detour: Bakry–Émery and one-link integration (seeing, on a single link, how theorems come out on the strong-coupling side).

Detour: Bakry–Émery and one-link integration — a criterion that gets a gap from curvature, and an example of integrating first to move the constant

> This chapter is this book's own investigation. § 02 and § 03 are established knowledge.

What this chapter shows — The Bakry–Émery criterion, "positive curvature gives a Poincaré inequality", and how conservative it is on a single link. On lattice $SU(2)$, integrating first a family of links containing exactly one link of every plaquette lowers the bending of the action to $O(\beta^2)$ and moves the threshold from $1/12$ to $\sqrt{2/27} \approx 0.272$ (a paper on this site), and the breakdown of the constant 27.

Prerequisite chapters — 5-02, 1-10. We also use 5-04 § 08 and 1-08 § 03 (the Haar measure of $SU(2)$).

Said first — This is about the side of small β at fixed lattice spacing, the shore opposite the continuum limit, and shows nothing about the continuum limit or the mass gap of the problem statement. The Lean theorems go as far as the lower bound on the bending and the combinatorics of the family; the mass gap is not a Lean theorem. Nor do we claim to have widened the strong-coupling range (§ 07).

Order of this chapter

1. Where this chapter stands — an example on the strong-coupling side
2. Curvature makes a gap — the criterion and its proof
3. Seen on a single link — the true gap grows with κ (Figure 1, Table 1)
4. Applying it to the whole lattice — 1/12 and 1/8
5. Integrating first — complete families (Table 2)
6. The breakdown of 27 — $36 \times 6 \div 8$
7. Comparing thresholds — a line and a parabola (Figure 2, Table 3)
8. Detour: complete families stop at four dimensions — equality in the Hamming bound
9. Limits and position of the method — what is Lean

The β of this chapter is in the $SU(2)$ convention $\beta = 4/g^2$, inversely proportional to the square of the coupling g (5-01). It plays the role of the inverse temperature of statistical mechanics (5-02).

$$S(U) = \beta \sum_P \Big(1 - \tfrac{1}{2}\operatorname{tr} U_P\Big), \qquad \mu_\beta(dU) = \frac{1}{Z}\, e^{-S(U)} \prod_\ell dU_\ell \tag{1}$$

Elements of $SU(2)$ are quaternions of length 1, and the group is the unit sphere S^3 (1-02). The inner product is $\langle u, v\rangle = \frac{1}{2}\operatorname{tr}(uv^\dagger)$.

01

Where this chapter stands — an example of a uniform estimate on the strong-coupling side

The key to the Langevin road of 5-04 § 08 is the **Bakry–Émery criterion**, which says "if the roundness (curvature) of the group beats the bumps of the action, a Poincaré inequality holds with a box-independent constant". This chapter looks at it concretely on a single link (§ 02, § 03) and on the whole lattice (§ 04– § 07).

What one gets is decay of correlations in lattice units at fixed lattice spacing (4-11), the shore opposite what is wanted in the continuum limit, "the gap in lattice units closing at the right rate" (5-09).

In 5-04 § 08 we used only the conclusion of the criterion. This chapter adds three things: the line of proof of the criterion (§ 02), a comparison with the true gap on a single link (§ 03), and that integrating part of the links first lowers the bending from first to second order in β and moves the constant of the threshold (§ 05– § 07).

The shortest way to read is to see first, in Figure 1, "where the criterion becomes conservative", then the threshold of the line $2 - c\beta$ in § 04, and finally the threshold of the parabola $2 - 27\beta^2$ in § 05– § 07. The breakdown in § 06 and the detour in § 08 can be skipped without losing the thread.

02

Curvature makes a gap — the criterion and a one-line proof

In 1-10 § 08 we saw, for $e^{-V}\,dx/Z$ on the line, "if $V'' \geq \kappa > 0$ then $C_P \leq 1/\kappa$". On a group the roundness of the space itself is added to this.

As a metaphor, two people who set off due north from the equator meet at the North Pole. Precisely, the Ricci curvature $\mathrm{Ric}(v, v)$ is the sum, over each direction

perpendicular to v, of how much a bundle of geodesics leaving in the direction of the unit vector v draws together; on the n-dimensional sphere of radius 1 it is $(n-1)|v|^2$ Known . Geodesics on the group are $\tau \mapsto U e^{\tau X}$ (X is a pure imaginary quaternion, an element of the Lie algebra; 1-03), and we write $\operatorname{Hess} V(X,X)$ for the second derivative along them. It corresponds to V'' on the line.

$$\operatorname{Ric}(X,X) = 2|X|^2 \ (S^3), \qquad \operatorname{Ric} + \operatorname{Hess} V \geq K > 0 \implies \operatorname{Var}_\mu(f) \leq \frac{1}{K}\int |\nabla f|^2 d\mu \quad (2)$$

The right-hand side is the Poincaré version of the criterion of Bakry–Émery (1985) ($\mu = e^{-V}\operatorname{vol}/Z$, with the condition at every point) Known . A logarithmic Sobolev inequality also follows. **No dimension enters the constant**, so if the lower bound on curvature does not depend on the box, neither does the Poincaré constant.

The line of proof is the same as in 1-10 § 08. With $Lf = -\Delta f + \nabla V \cdot \nabla f$ (Δ is the Laplacian on the group, different from the Δ of the mass gap) and $Lf = \lambda f$, integrating by parts twice (the integrated Bochner formula) gives Known

$$\lambda^2 \int f^2 d\mu = \int (Lf)^2 d\mu = \int \|\nabla^2 f\|^2 d\mu + \int (\operatorname{Ric} + \operatorname{Hess} V)(\nabla f, \nabla f)\, d\mu \ \geq \ K\lambda \int f^2 d\mu \quad (3)$$

($\int |\nabla f|^2 d\mu = \lambda \int f^2 d\mu$), so $\lambda \geq K$. (3) is the multidimensional form of $\int |g'|^2 + \int V'' g^2 = \lambda \int g^2$ in 1-10: V'' is replaced by $\operatorname{Hess} V$, and the roundness of the space is added as Ric. Example: on uniform S^3 the criterion says "gap ≥ 2", while the truth is 3 (first-order spherical harmonics; the first row of Table 1). It is conservative to the extent that it does not use the dimension.

For how this book uses this, see 24-06 § 02 (the convexity ceiling on a closed group is in 24-06 § 03).

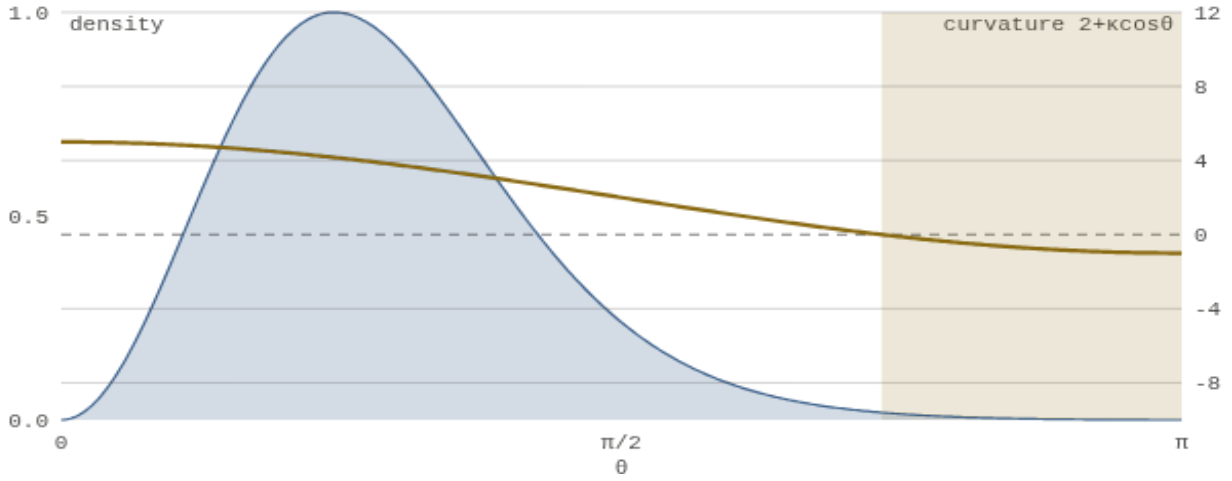
03 Seen on a single link — the true gap and the criterion

Holding fixed everything but a single link u, the weight comes from the six plaquettes containing u. The product of the remaining three links of a plaquette without u (a U shape) is called a **staple**. Using the sum $M = \sum_i V_i$ of the six staples $V_1, \dots, V_6$ with orientations aligned, the weight is $e^{\beta \langle u, M \rangle}$ (the same set-up as 23-02 Roads tried on the four-dimensional lattice, and walls; the distribution of the heat-bath method; 5-07). Taking the direction of M as the North Pole and θ as the angle from the North Pole, the Hessian of the height $\cos\theta$ is $-\cos\theta$ times the metric, so

$$\mu_\kappa(du) \propto e^{\kappa \cos\theta} du \ \ (\kappa = \beta|M| \leq 6\beta), \qquad \operatorname{Ric} + \operatorname{Hess} V = (2 + \kappa\cos\theta)|X|^2 \geq (2-\kappa)|X|^2 \ ($$

Known . All the criterion can say is a gap $\geq 2-\kappa$ for $\kappa<2$, but the true gap **grows** with κ (the measure concentrated near the North Pole is close to a Gaussian of variance $1/\kappa$). The criterion is tripped up by the negative curvature near the South Pole, where the measure carries almost no weight.

Figure 1 The knob and "Play" change κ. The fill is the density $e^{\kappa\cos\theta}\sin^2\theta$, the solid line the curvature $2+\kappa\cos\theta$ (right-hand scale), and the band the range where the curvature is negative.



κ=3.00, criterion 0.00, gap 4.1160, weight of negative curvature 0.0039

Computed The numbers in this figure were computed on this machine (the gaps are 41 points of the calculation of Table 1, the rest from formulas on the spot). If JavaScript is disabled, Table 1 holds the content.

Table 1 The gap λ_1 of μ_κ and the criterion Computed (eigenvalue problems for each angular momentum $l=0,1,2$ on S^2; steps 800 and 1600 agree to 5 digits. The rightmost column is the numerical integral $\int_{\arccos(-2/\kappa)}^{\pi} e^{\kappa\cos\theta}\sin^2\theta\,d\theta \,/\, \int_0^\pi$)

κ	$l=1$ = gap	Criterion	Weight of the region of negative curvature
0	3.0000	2.0	0
1	3.1323	1.0	0
2	3.5164	0	0
4	4.8782	–	0.0027
6	6.6746	–	0.00059
10	10.580	–	1.5×10^{-5}

The gap is set by $l=1$ (fluctuations tilting perpendicular to M). The criterion is a correct but crude tool.

Note. $\lambda_1 = 1/C_P$ is the gap in Langevin time, different from the gap of the transfer matrix (5-03) (18-03).

```
04
```

Applying it to the whole lattice — thresholds 1/12 and 1/8

The configuration space of the whole lattice is a product of S^3, and the Ricci curvature stays 2 on each link. The criterion asks for the following.

$$K(\beta) = 2 + \inf_U \lambda_{\min}\big(\operatorname{Hess} S(U)\big) > 0 \quad (5)$$

The lower bound has the form $-c\beta$. Shen–Zhu–Zhu (2023), with a plaquette-by-plaquette estimate, showed for $SU(N)$ under $|\beta_{\mathrm{SZZ}}| < 1/(16(d-1))$ uniqueness of the infinite-volume measure and exponential decay of correlations (abstract: "the infinite volume measure has a strictly positive mass gap") Known . The convention for $SU(2)$ is $\beta = 4\beta_{\mathrm{SZZ}}$, so in four dimensions this is $\beta < 1/12 = 0.0833$, and in the notation of this chapter $K = 2 - 24\beta$ ($\beta < 1/12$ is the same as in 5-04 Table 4).

On the other hand, the $\mathbb{Z}_2$ configuration with every plaquette -1 gives $\lambda_{\min}(\operatorname{Hess} S) = -16\beta$ (elementary) Paper . The numerical search of the paper found nothing below it, but this is not a proof (paper § 1, § 8). The ceiling $1/8$ comes from the elementary side alone: for this configuration $K \le 2 - 16\beta$, so **as long as the criterion is used without integrating, it tops out at $\beta < 1/8$.**

05

Integrating first — complete families and one-link integration

The paper on this site integrates part of the links exactly, as equalities (rather than bounding them by inequalities), before applying the criterion. The set I^* to be integrated is required to satisfy the following.

$$I^* \text{ is a complete family} \iff \text{for every plaquette } P,\ |P \cap I^*| = 1 \quad (6)$$

No two links of the family lie on the same plaquette, so each link can be integrated independently; and since every plaquette contains a link of the family, **not a single bare plaquette remains in the action after integration**. The density of the family is $1/4$, and four dimensions has a family of period 2 (counting and figures in 23-02 Roads tried on the four-dimensional lattice, and walls; why only up to four dimensions in § 08).

The integral over one link comes out in Catalan numbers. Under the Haar measure $\operatorname{tr} u = 2\cos\theta$ has the semicircle distribution (1-08 (4)), with even moments $\mathbb{E}[(2\cos\theta)^{2k}] = \mathrm{Cat}_k$ (the detour in 1-08). Expanding the exponential and using $\mathrm{Cat}_k/(2k)! = 1/(k!(k+1)!)$ gives

$$\int_{SU(2)} e^{\beta\langle u,M\rangle}\,du = \sum_{k\ge 0}\frac{\kappa^{2k}\,\mathrm{Cat}_k}{(2k)!\,4^k} = F_0\Big(\frac{\kappa^2}{4}\Big) = \frac{2I_1(\kappa)}{\kappa}, \qquad F_0(x)=\sum_{k\ge 0}\frac{x^k}{k!\,(k+1)!} \tag{7}$$

(I_1 is a modified Bessel function) Known . In the paper this is proved in Lean from the left invariance of the Haar measure alone Lean (`IsHaarS3.one_link` ; the only assumptions are the Haar properties). The marginal density of the remaining links is $e^{-\Phi}$, a sum over the links s of the family ("stars").

$$\Phi = -\sum_{s\in I^*} G\big(|M_s|^2\big), \qquad G(t)=\ln F_0\Big(\frac{\beta^2 t}{4}\Big) = \frac{\beta^2 t}{8} - \frac{\beta^4 t^2}{384} + O(\beta^6 t^3) \tag{8}$$

For derivatives we use $F_1(x)=\sum_{k\ge 0} x^k/(k!\,(k+2)!) = F_0'(x)$ and $F_2 = F_1'$. By the chain rule

$$G'(t) = \frac{\beta^2}{4}\,\frac{F_1}{F_0}, \qquad G''(t) = \Big(\frac{\beta^2}{4}\Big)^2\,\frac{F_0F_2 - F_1^2}{F_0^2} \qquad \Big(F_i \text{ evaluated at } \frac{\beta^2 t}{4}\Big)$$

so $G'\le \beta^2/8$ is the same as $2F_1\le F_0$, and $G''\le 0$ the same as the Turán-type inequality $F_1^2 \ge F_0F_2$.

The $O(\beta)$ term disappears, and the bending starts at $O(\beta^2)$ (loops of length 6 appear with coupling $\beta^2/4$; 23-02 § 02).

Table 2 The series (7) and the Bessel closed form, and the ratio $r = 2F_1/F_0 = G'(t)/(\beta^2/8)$ Computed (also agrees with numerical integration over the angle to 12 digits)

κ	$F_0(\kappa^2/4)$	$2I_1(\kappa)/\kappa$	r
1	1.130318207985	1.130318207985	0.9608
4	4.879732576852	4.879732576852	0.6580
8	99.96828419564	99.96828419564	0.4096

06

The breakdown of 27 — 36 × 6 ÷ 8

The central theorem of the paper is a lower bound on the second derivative of Φ along geodesics. It holds for every β and every configuration, with no smallness assumed Lean (`StaplePerturbation.hess_gauge` . The assumptions are that the period L is even and $L\ge 3$ (that is, an even number 4 or more; at $L=2$ opposite staples share links and the counting changes), $\|U_e\|=1$, and X_e pure imaginary. That the marginal density is $e^{-\Phi}$ is `WilsonOneLink.integral_star_links`).

$$\frac{d^2}{d\tau^2}\Phi\left(Ue^{\tau X}\right)\Big|_0 \geq -27\beta^2\sum_e \|X_e\|^2 \quad (9)$$

÷ 8 — one-link integration	× 36 — one star	× 6 — counting
$0 \leq G' \leq \beta^2/8,\ G'' \leq 0$ $r \leq 1$ in Table 2 ($\iff 2F_1 \leq F_0$). Concavity ($\iff F_1^2 \geq F_0F_2$) from a Catalan-number identity Lean (`turan_series`)	$(\lvert M_s\rvert^2)'' \leq 36\sum \|X\|^2$ Weighted Cauchy–Schwarz and a sum of squares Lean (`StarInequality.star_S` and `StapleCurve.second_deriv_le`; equality is `StapleCurve.second_deriv_sharp`)	Each remaining link lies in exactly 6 stars Lean (`star_count`; $L \geq 3$)

Joining them takes one line. With $t = \lvert M_s\rvert^2$, the second derivative of a star term is

$$-\big(G(t)\big)'' = -G'(t)\,t'' - G''(t)(t')^2 \ \geq\ -G'(t)\max(t'',0) \ \geq\ -\frac{\beta^2}{8}\cdot 36\sum_{\text{in the star}}\|X\|^2 \quad (10$$

and thanks to $G'' \leq 0$ the $(t')^2$ term does not get in the way. Summing over stars counts each link 6 times, giving $-\frac{\beta^2}{8}\cdot 36\cdot 6 = -27\beta^2$. $G' \leq \beta^2/8$ is the claim that "the fluctuations of a tilted Haar measure do not exceed those of the Haar measure", and this is specific to $SU(2)$ (§ 09).

Example. The star ratio $t''/(36\sum\|X\|^2)$ has maximum 0.371 over 200,000 random samples, and equality occurs for the configuration with $s = (+,+,+,-,-,-)$ and the three directions aligned to $s_i\, i$ (1.000000) Computed. The slack lies on the side of joining the stars (§ 09).

07

Comparing thresholds — a line and a parabola

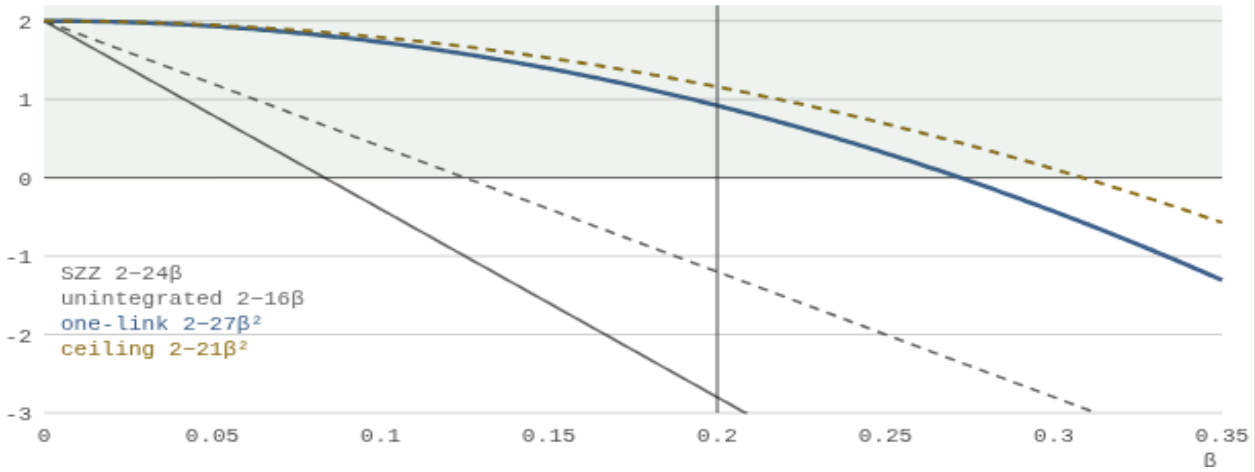
Combining (9) with $\mathrm{Ric} = 2$, the curvature of the marginal measure of the remaining links is at least $K' = 2 - 27\beta^2$, and the threshold is

$$\beta < \sqrt{2/27} = 0.27217 \qquad (\beta_{\mathrm{SZZ}} < 0.06804) \quad (11)$$

This is $4\sqrt{2/3} = 3.27$ times $1/12$, and 2.18 times the ceiling $1/8$ Computed. For d where a family exists, the threshold is $\sqrt{2/3}/(d-1)$, and the ratio 3.27 does not depend on d Paper (Lean covers only $d = 4$).

Figure 2 Four lower bounds $K(\beta)$ on the curvature. The knob and "Play" move β, and each value and whether $K >$

0 appear below.




```
β = 0.200, SZZ -2.800 (not usable), unintegrated -1.200 (not usable), one-link +0.920 (usable),
ceiling +1.160 (upper limit of this road)
```

Computed The numbers in this figure were computed on this machine (the formulas of Table 3). If JavaScript is disabled, Table 3 holds the content.

Table 3 Lower bounds on curvature and thresholds for four-dimensional $SU(2)$ Computed

Estimate	$K(\beta)$	Threshold	$\beta = 0.2$	Level
The counting of Shen–Zhu–Zhu	$2 - 24\beta$	0.0833	−2.80	Known
Best without integrating	$2 - 16\beta$	0.1250	−1.20	Paper ($\leq -16\beta$) + the numerical search of the paper
Integrating a complete family first	$2 - 27\beta^2$	0.2722	+0.92	Lean + Known
The ceiling of this road	$2 - 21\beta^2$	0.3086	+1.16	Exact computation by integer arithmetic in paper § 8 (not machine-checked; not recomputed on this machine; § 09)

The reach of the cluster expansion of Osterwalder–Seiler (1978), the strong-coupling classic, cannot be compared numerically in this chapter (source table: bibliographic data only). The crude Kotecký–Preiss estimate of paper § 9 gives $\beta \lesssim 0.014$ (Mayer form) to 0.06 (character expansion), the former of the same order as 5-04 Table 3. On the other hand, a necessary condition for the character expansion to converge absolutely allows β up to 0.4–0.8, and a careful expansion may reach further. **So the paper does not claim to have widened the strong-coupling range, and says only that "within the Bakry–Émery method the constant improves by a factor of 3.27".** The same dynamical method has also been used for the area law (Cao–Nissim–Sheffield) Known .

```
08
```

Detour: complete families stop at four dimensions — one line of coding theory

A section you may skip.

Complete families exist in $\mathbb{Z}^d$ only for $d \le 4$ Lean (`no_perfectZ_of_five`, with no periodicity assumption. On a box of period L it is equivalent to "$d \le 4$ and L even": `PerfectFamily.perfectK_one_iff`, $d \ge 2$). 23-02 § 02 shows it by counting vertices. The paper's argument on paper is coding theory. We prepare three words. A collection of strings of 0s and 1s of length n is a **binary code** and its strings are **words**; the **Hamming distance** of two strings is the number of coordinates where they differ, and the **minimum distance** is the smallest distance between two different words.

Inside a unit d-cube, the position of a family link of direction μ is fixed by the remaining $d-1$ coordinates (0 or 1), so they form a binary code C_μ of length $n = d-1$. The number of words comes from counting faces. There are 2^{d-2} faces of directions μ, ν, and each face contains exactly one family link, so $|C_\mu| + |C_\nu| = 2^{d-2}$; solving simultaneously for three directions gives $|C_\mu| = 2^{d-3}$. From the condition on the faces of three-dimensional sub-cubes, words can be told apart after deleting any two coordinates, that is, the minimum distance is at least 3. Balls of radius 1 around words at distance 3 or more do not overlap (the Hamming bound), so

$$2^{d-3}(1+n) \le 2^n \iff d \le 4 \qquad (d=4:\ 8=8, \quad d=5:\ 20>16) \ (12)$$

$d = 4$ is the case of equality: the repetition code of length 3, $\{000, 111\}$, fills $\{0,1\}^3$ exactly Computed . This "4" is a combinatorial number of the lattice, a coincidence unrelated to physical four-dimensional spacetime. For $d \ge 5$ only "at most one" families can be taken (density at most $1/d$ Lean), and the improvement drops to a factor of 1.72.

09

Limits and position of the method — 21 and 27, large β, SU(N), what is Lean

The four limits of paper § 8 are different faces of the fact that "a local curvature method can move the constant but not the shore".

(a) 27 and 21. An explicit configuration gives exactly $\lambda_{\min}(\mathrm{Hess}\,\Phi) = -21\beta^2$ (exact computation by integer arithmetic in $\mathbb{Z}[e^{i\pi/3}]$ in paper § 8; not machine-checked, not recomputed on this machine). The true value lies between 21 and 27, and the ceiling is $\sqrt{2/21} = 0.3086$.

(b) Large β. When β is large, integrating first makes things worse. $\lambda_{\min}$ at $\beta = 1$ is still -12.1 against -16 (with against without integration), but at $\beta = 2$ it reverses to $-43.1\beta^2$ (without integration -32β). At either value $K < 0$, and the criterion cannot be used (numerical calculations of the paper; not recomputed on this machine). As in § 03, the

criterion demands curvature at every configuration, and $\mathrm{Ric} = 2$ does not move with β, so the threshold does not exceed $O(1)$.

(c) It cannot be repeated. Integrating one more link makes the weight no longer of Bessel type Paper .

(d) $SU(2)$ only. "Tilting does not increase fluctuations" is already false for $U(2)$ with $M = \mathrm{diag}(s, 0)$. What the paper measured directly is mainly the star constant for $U(N)$. The threshold measured for $N = 2, 3, 4$ is nearly constant ($\beta_{\mathrm{SZZ}} \approx 0.093$), but for large N it is estimated as $\beta^* \sim 0.0962/\sqrt{N}$ (paper § 8), falling below the N-independent threshold of Shen–Zhu–Zhu for $N \gtrsim 21$. In the 't Hooft limit the gain disappears (numerical calculations and asymptotic estimates of the paper; not recomputed on this machine).

What is Lean. The Lean theorems are (7), the marginal density $e^{-\Phi}$, the three factors, (9), and $d \leq 4$ for families. $\mathrm{Ric} = 2$ and the criterion are standard facts Known , and the step to exponential decay of correlations and uniqueness (a transfer of the argument of Shen–Zhu–Zhu) is an **outline** with constants Paper . That is, the lattice-unit mass gap at $\beta < 0.2722$ is not a Lean theorem. Within the range the paper searched the same construction was not found, but the paper writes that it may be known in the context of variance reduction for Monte Carlo methods.

The curvature method works uniformly in a world with $C_P = O(1)$, where the correlation length is a few lattice spacings, the side opposite what is wanted in the continuum limit (23-01). However much this estimate is improved, it comes no closer to the mass gap of the problem statement.

What this chapter can say

Matter	Said	Not said
(2)–(4), (7)	The criterion, one-link integration Known	Details of the criterion (original: bibliographic data only)
Table 1	The single-link gap grows with κ, while the criterion covers only $\kappa < 2$ Computed	The gap of the whole lattice
§ 04	The conversion of $\beta < 1/12$ Known , the ceiling $1/8$ without integration Paper	The reach of the cluster expansion (original: bibliographic data only)

Matter	Said	Not said
(9), (11)	Lower bound $-27\beta^2$ Lean , curvature threshold 0.2722	The best constant. The mass gap is an outline on paper. It cannot be said that the range is widened
§ 08	Families for $d \leq 4$ Lean	Relation to physical four dimensions (none)
The whole chapter	An example on the strong-coupling side at fixed lattice spacing	Shows nothing about the continuum limit or the mass gap of the problem statement

Sources and reproduction

Item	Kind	Source / tools
The theorems, constants and limits of § 04– § 09	Text § 1– § 11 (draft, not submitted, not refereed)	The paper on this site “One-link integration before Bakry–Émery for SU(2) lattice Yang–Mills” (version of 2026-09-25, list of papers)
The condition and quotation in § 04	Abstract (the numbering in the text is as cited by the paper above)	H. Shen, R. Zhu, X. Zhu, Commun. Math. Phys. 400 (2023) 805–851, arXiv:2204.12737
The criterion (2)	Bibliographic data only	D. Bakry, M. Émery, Diffusions hypercontractives, Séminaire de Probabilités XIX (1985) 177–206
That there is nothing below the -16β of § 04, and the values in § 09 (a)(b)(d) ($-21\beta^2$, $-12.1\beta^2$, $-43.1\beta^2$, the counterexample for $U(2)$, the thresholds for $N = 2, 3, 4$)	Calculations of the paper (not recomputed on this machine)	§ 8 of the paper above (numerical search, integer arithmetic; not machine-checked)
The Kotecký–Preiss estimate in § 07	A crude estimate of the paper	§ 9 of the paper above
The area law in § 07	Abstract	S. Cao, R. Nissim, S. Sheffield, Dynamical approach to area law for lattice Yang–Mills, arXiv:2509.04688
The strong-coupling classic	Bibliographic data only	K. Osterwalder, E. Seiler, Ann. Phys. 110(2) (1978) 440–471
Turán type $F_1^2 \geq F_0 F_2$	Abstract	Á. Baricz, S. Ponnusamy, Proc. Amer. Math. Soc. 141(2) (2013) 523–532, arXiv:1010.3346

Item	Kind	Source / tools
Tables 1–3, Figures 1 and 2, the ratio in § 06, (11)(12)	Computed on this machine	Python 3, NumPy, SciPy: gap_s3.py (tridiagonal eigenvalue problems for $l = 0, 1, 2$, and the numerical integral of the weight of the region of negative curvature in Table 1), onelink_series.py (coefficients of $F_1^2 - F_0 F_2$ checked in rationals up to $n \leq 30$), star_check.py, thresholds.py

Next chapter to read: 6-00 Entrance — the promise of this volume. To read alongside it: 7-01 What constructive field theory is (between the two shores). The families are used again in 23-02. For Lean, the guide and the list.