Volume 1 Preparation

# Part 2 Physical preparation

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 |
|---|---|---|
| 2-01 | Lagrangians and the action | 2 |
| 2-02 | The idea of a field | 13 |
| 2-03 | Electromagnetism is a gauge theory | 24 |
| 2-04 | Special relativity and spacetime | 35 |
| 2-05 | The bones of quantum mechanics | 47 |
| 2-06 | The harmonic oscillator and creation and annihilation | 58 |
| 2-07 | The idea of the path integral | 69 |
| 2-08 | Statistical mechanics and phase transitions | 81 |
| 2-09 | Detour: the gap and the correlation length in the Ising model | 94 |
| 2-10 | The relaxation time and the dynamical exponent z | 105 |

# Lagrangians and the action — motion can be written as a stationary point of "one number attached to the whole path"

> 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** — That the equations of motion can be written as the stationarity condition of the action. That the Euler–Lagrange equations come out of it. That continuous symmetries give conserved quantities (Noether's theorem), and local symmetries give constraints.

**Prerequisite chapters** — 1-01 Vector spaces and linear maps (rotation matrices, eigenvalues of real symmetric matrices). Otherwise, multivariable calculus and Newton's equation of motion.

**Said first** — We restrict to finitely many degrees of freedom (fields are in 2-02, the Yang–Mills action in 3-03). The proofs stay within smooth paths and variations with fixed ends. We do not answer why the action principle holds.

Order of this chapter

1. Why start from the action — three advantages
2. The action, a single number — the integral of kinetic energy − potential energy
3. Shifting the path a little — discretised, a function of many variables (Figure 1)
4. The Euler–Lagrange equations — one integration by parts; no choice of coordinates
5. Not necessarily a minimum — eigenvalues and conjugate points (with computations)
6. Noether's theorem — conserved quantities from symmetries (Figure 2)
7. Detour: the discretised action has Noether too — why angular momentum is conserved exactly
8. Local symmetries give constraints — a foretaste of gauge theory

01

## Why start from the action

Newton's equation of motion $m\ddot{q} = F$ is a local rule that fixes the acceleration at each moment. In this chapter we write the same motion in another form. Fix the start and the arrival, and give the **whole path** between them a single number $S$ (the **action**). The actual motion is chosen as the path for which $S$ does not change at first order when the path is shifted slightly. This is the **action principle** (also called the principle of least action).

There are three advantages. It **does not depend on coordinates** ( § 04). **Conservation laws come out automatically** from symmetries ( § 06). **It is the language of field theory**: Yang–Mills theory is defined by an action, the integral of $\mathrm{tr}\, F^2$ (3-03), and the path integral too sums over paths with the action in the exponent (2-07).

As a metaphor, the equation of motion is "step-by-step directions", and the action principle is the view of "comparing whole routes on a map". To be exact, for systems with conservative forces both give the same equation ( § 04), and the only difference is what is easy to see.

02

## The action, a single number

For a particle on a line, let $q(t)$ be its position at time $t$, $\dot{q} = dq/dt$ its velocity, $m$ its mass and $V(q)$ the potential energy. The **Lagrangian** $L$ is the kinetic energy **minus** the potential energy, a different quantity from the energy, which is their sum. The number obtained by integrating it along the path is the **action**.

$$L(q, \dot{q}) = \tfrac{1}{2} m \dot{q}^2 - V(q)$$

$$S[q] = \int_{t0}^{t1} L\big(q(t), \dot{q}(t)\big)\, dt$$

The square brackets show that it is a map that takes a whole path and returns a number (a **functional**).

**Example.** Take $V = 0$, $m = 1$, and ends $q(0) = 0$, $q(1) = 1$. For the uniform path $q = t$, $S = 0.5$. For the detour $q = t + 0.3 \sin \pi t$ joining the same ends, $S = \tfrac{1}{2} \int_0^1 (1 + 0.3\pi \cos \pi t)^2\, dt = \tfrac{1}{2}(1 + 0.09\pi^2/2) \approx 0.722$. Taking a detour increases the action.

03

## Shifting the path a little

An extremum of a function was a point where the first-order change under a small change of the variable is 0. For the action we move the path. Taking a function $\eta(t)$ that vanishes at the ends, $S(\varepsilon) = S[q + \varepsilon\eta]$ is an ordinary function of the real number $\varepsilon$. The **action principle** is the requirement "$S'(0) = 0$ for every $\eta$ with fixed ends", and a path satisfying it is called a **stationary point**.

### Discretised, it becomes a function of finitely many variables

Dividing time into $N$ equal parts (step $h$), making the path a broken line and approximating the integral by the trapezoidal rule, the action becomes a function of many variables, the intermediate positions $q_1, \dots, q_{N-1}$ (the ends $q_0$, $q_N$ fixed).

$$S_N = \sum_{k=0}^{N-1} \Big[ \frac{m(q_{k+1} - q_k)^2}{2h} - \frac{h}{2}\big(V(q_k) + V(q_{k+1})\big) \Big]$$

$$\frac{\partial S_N}{\partial q_k} = -h\Big[ m\,\frac{q_{k+1} - 2q_k + q_{k-1}}{h^2} + V'(q_k) \Big]$$

The action principle becomes "the gradient is 0", and this is a difference version of $m\ddot{q} = -V'(q)$ (used again in § 07).

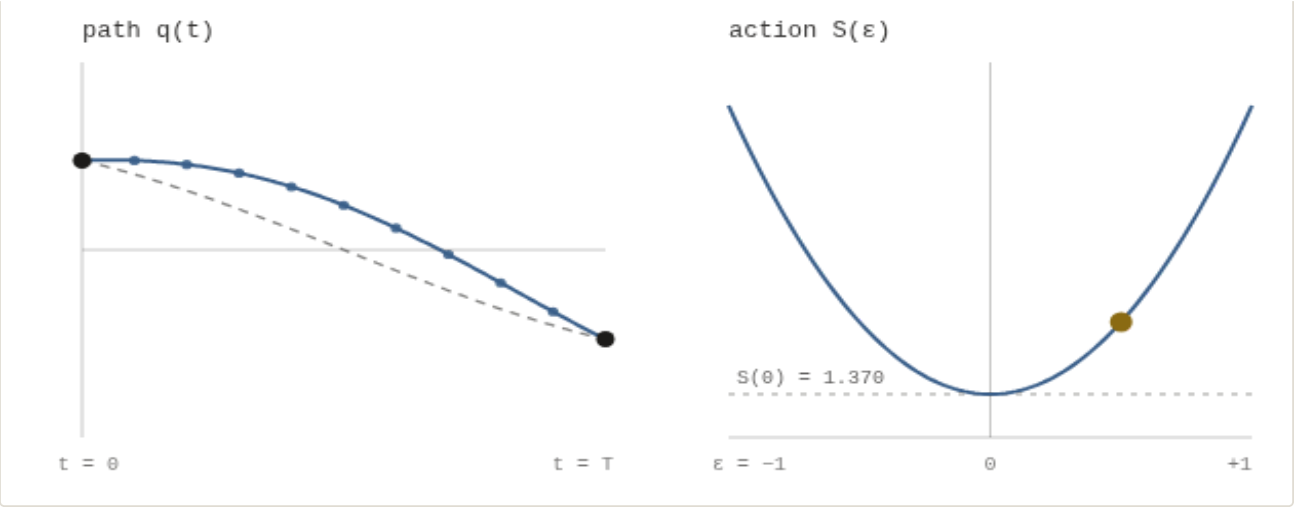
### Example: a particle on a spring

Take $V = \frac{1}{2}m\omega^2 q^2$, $m = 1$, time length $T = 1$, and ends $q(0) = 1$, $q(1) = -1$. The solution of the equation of motion is $q_{\mathrm{cl}}(t) = \sin\big(\omega(\tfrac{1}{2} - t)\big) / \sin(\omega/2)$ (cl stands for "classical path"). For the shift $\eta_n = \sin n\pi t$ ($n = 1, 2, 3$), the action is a quadratic in $\varepsilon$.

$$S(\varepsilon) = \omega \cot\frac{\omega}{2} + \frac{\varepsilon^2}{4}\big((n\pi)^2 - \omega^2\big)$$

The absence of a first-order term is what "stationary" means. The second-order term comes from $\frac{1}{2}\int(\dot{\eta}^2 - \omega^2\eta^2)\,dt$. Moving $\varepsilon$ in Figure 1 moves the action along a parabola, with slope 0 at $\varepsilon = 0$.

Figure 1 Moving ε, the shape n and ωT changes the shifted path (left) and the action S(ε) (right)S(ε) = 1.764, coefficient of the second-order term +1.579, opens upwards = a minimum in this direction




Computed The curves draw the formula above in the browser, and the formula was checked against numerical integration on this machine (Table 1). The dotted line is the classical path, and the dots are the variables when discretised with $N = 10$.

Table 1 Values of the action ($m = T = 1$, $q(0) = 1$, $q(1) = -1$)

| $\omega T/\pi$ | $n$ | $S(0)$ | $S(1)$ | Coefficient of the second-order term |
|---|---|---|---|---|
| 0.5 | 1 | 1.5708 | 3.4213 | +1.851 |
| 1.0 | 1 | 0.0000 | 0.0000 | 0 |
| 1.5 | 1 | −4.7124 | −7.7966 | −3.084 |
| 1.5 | 2 | −4.7124 | −0.3944 | +4.318 |

Computed The difference between numerical integration by Simpson's rule (20000 intervals) and the formula is at most $4 \times 10^{-15}$. The first-order coefficient is also at most $5 \times 10^{-13}$.

04

## The Euler–Lagrange equations

We compute the condition of § 03 without discretising. Differentiate with respect to $\varepsilon$ and integrate the second term by parts (the boundary term vanishes because $\eta = 0$ at the ends).

$$S'(0) = \int_{t0}^{t1} \left[ \frac{\partial L}{\partial q}\,\eta + \frac{\partial L}{\partial \dot{q}}\,\dot{\eta} \right] dt$$

$$= \int_{t0}^{t1} \left[ \frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} \right] \eta \, dt$$

If this is 0 for every $\eta$, the bracket is 0 at each time (the bracket is continuous, so if it is positive somewhere it is positive nearby, and taking a $\eta$ positive only nearby makes the integral positive: the **fundamental lemma of the calculus of variations**). This gives the **Euler–Lagrange equations**. With several degrees of freedom there is one for each component.

$$\frac{d}{dt}\frac{\partial L}{\partial \dot q} - \frac{\partial L}{\partial q} = 0 \quad \text{(EL)}$$

**Example.** If $L = \tfrac{1}{2}m\dot q^2 - V(q)$, then $m\ddot q = -V'(q)$: Newton's equation for a conservative force itself.

### No choice of coordinates

This derivation does not use what $q$ represents. It may be an angle (**generalised coordinates**), and equations of the same form hold Known . This is because the same path gets the same action whatever the coordinates (1-05). Writing a pendulum with the angle $\theta$, $L = \tfrac{1}{2}ml^2\dot\theta^2 + mgl\cos\theta$ gives $ml^2\ddot\theta = -mgl\sin\theta$. The tension does not appear.

Adding $dG(q,t)/dt$ to $L$ does not change the equations (the action changes only by a quantity fixed by the end values) Known . The **θ term** of Yang–Mills theory is the four-dimensional version of this, with the form of a four-dimensional divergence (a **total divergence**) instead of a total time derivative. To be exact, the θ term does not change the classical equations, but it matters in quantum theory through the topology of the field (3-05).

05

## Not necessarily a minimum

The action principle asks only that the first-order change be 0; whether it is a minimum is decided by the second-order term. In the example of § 03 it is the sign of the coefficient $((n\pi)^2 - \omega^2)/4$. If $\omega T < \pi$ it is positive for every $n$, and a general $\eta$, expanded in a sine series, splits into a sum over each $n$, so the classical path really is a minimum (because the action is quadratic).

When $\omega T$ exceeds $\pi$, the coefficient of $n = 1$ becomes negative, and in that direction the action **decreases** (a **saddle point**). Moving $\omega T/\pi$ either side of 1 in Figure 1 turns the parabola of $n = 1$ over. At exactly $\omega T = \pi$, moving in the direction of $n = 1$ does not change the action (it increases in the direction of $n \geq 2$), and the solutions joining the same ends form a family $\cos\pi t + c\sin\pi t$. In general, a time at which a family of nearby solutions leaving the same point at time 0 gathers again at one point is called a **conjugate point** of 0 Known . The conjugate points in this example are $t = \pi/\omega, 2\pi/\omega, \ldots$, and $\omega T = \pi$ is the case where the final time falls exactly on the first conjugate point.

### The second-order term is an eigenvalue problem for a matrix

Discretised, the second-order term is the Hessian matrix $H_{jk} = \partial^2 S_N / \partial q_j \partial q_k$. It is real symmetric, so it can be diagonalised (1-01), and a local minimum means all eigenvalues are positive. The number of negative eigenvalues is the number of directions in which the action can be reduced, called the **Morse index** (here a name for the discretised matrix). For the harmonic oscillator $H$ is tridiagonal (diagonal $2m/h - m\omega^2 h$, neighbours $-m/h$); its eigenvalues divided by $h$ approach $m((n\pi/T)^2 - \omega^2)$, and $\lfloor \omega T/\pi \rfloor$ of them are negative.

Table 2 Hessian matrix of the discretised action ($m = T = 1$, $N = 200$)

| $\omega T/\pi$ | Number of negative eigenvalues | Smallest eigenvalue $/\,h$ | $\pi^2 - (\omega T)^2$ |
|---|---|---|---|
| 0.50 | 0 | 7.4020 | 7.4022 |
| 1.05 | 1 | −1.0118 | −1.0116 |
| 2.50 | 2 | −51.8156 | −51.8154 |
| 3.20 | 3 | −91.1953 | −91.1951 |

Computed All eigenvalues found with numpy's `eigvalsh`. With $N = 20$ the numbers are the same, and the smallest eigenvalues differ by about 0.02.

Without discretising too, the directions that reduce the action in this example are those among $\sin n\pi t$ with $n\pi < \omega T$, and their number equals the number of conjugate points $k\pi/\omega$ in $(0, T)$. This is an instance of the **Morse index theorem** (the number of negative directions of the second-order term equals the number of conjugate points in $(0, T)$, counted with multiplicity) Known .

This view is extended to infinite dimensions in 1-10. The smallest eigenvalue here (of the second-order term of the action) and the mass gap of 4-11 (the gap in the spectrum of the Hamiltonian) are different things.

06

## Noether's theorem

If $L$ does not contain the coordinate $q$, the Euler–Lagrange equation says that $p = \partial L/\partial \dot q$ (the **conjugate momentum**) is conserved. $L = \frac{1}{2}m(\dot r^2 + r^2\dot\varphi^2) - V(r)$, a central force written in polar coordinates, does not contain $\varphi$, and the angular momentum $mr^2\dot\varphi$ is conserved. Noether's theorem is this detached from coordinates.

For transformations $q \mapsto \Phi_s(q)$ labelled continuously by a real number $s$ and equal to the identity at $s = 0$ (translations, rotations by the angle $s$), let $K(q) = \frac{d}{ds}\Phi_s(q)\big|_{s=0}$ be the direction of motion. **Noether's theorem**: if $L\big(\Phi_s(q), \frac{d}{dt}\Phi_s(q)\big) = L(q, \dot q)$ for every path, then the following $Q$ is constant along the motion Known .

$$Q = \sum_i \frac{\partial L}{\partial \dot{q}_i} K_i(q) \quad \text{(N)}$$

**Proof.** Differentiating the assumption with respect to $s = 0$ gives $\sum_i (\partial L/\partial q_i) K_i + \sum_i (\partial L/\partial \dot{q}_i)\, \dot{K}_i = 0$. Replacing $\partial L/\partial q_i$ by $\frac{d}{dt}(\partial L/\partial \dot{q}_i)$ using the Euler–Lagrange equation, the left-hand side is the derivative of a product $dQ/dt$. ∎

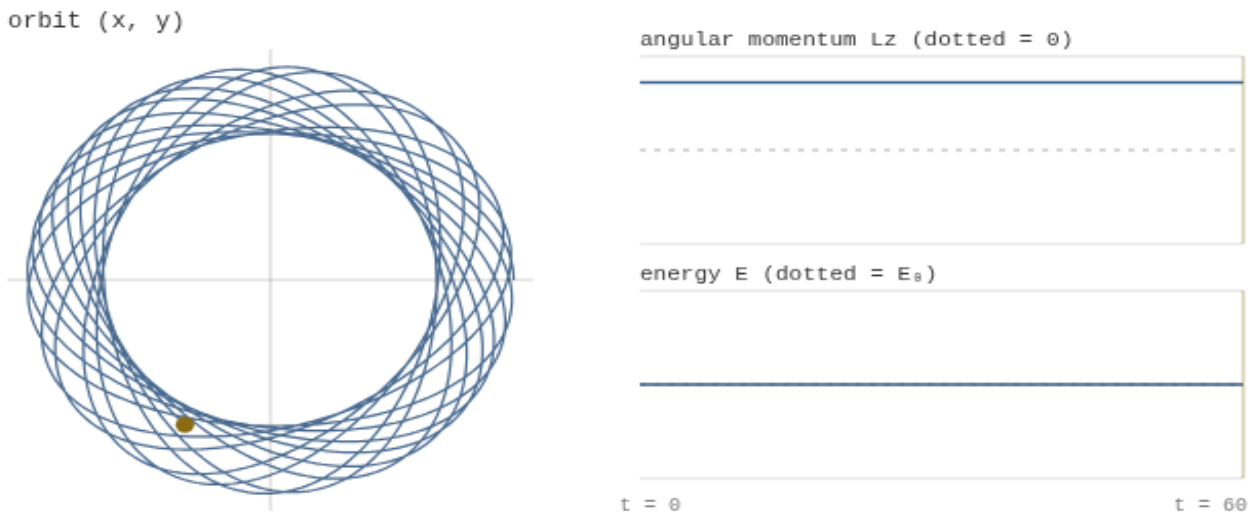
| Transformation that does not change the action | $K$ | Conserved quantity $Q$ |
|---|---|---|
| Translation $\boldsymbol{x} \mapsto \boldsymbol{x} + s\boldsymbol{e}$ | $\boldsymbol{e}$ | Component of momentum $m\dot{\boldsymbol{x}} \cdot \boldsymbol{e}$ |
| Rotation in the plane $\boldsymbol{x} \mapsto R(s)\boldsymbol{x}$ | $(-y, x)$ | Angular momentum $L_z = m(x\dot{y} - y\dot{x})$ |
| Translation in time ($\partial L/\partial t = 0$) | — | Energy $E = \sum_i \dot{q}_i\, \partial L/\partial \dot{q}_i - L$ |

The third row follows by differentiating directly, from $dE/dt = -\partial L/\partial t$ ($E = \frac{1}{2} m \dot{q}^2 + V$).

### Checking with a figure: breaking the symmetry removes the conserved quantity

Give a particle in the plane ($m = 1$) the potential $V = \frac{1}{2} r^2 + \frac{1}{4} r^4 (1 + \delta \cos 4\theta)$ in polar coordinates. If $\delta = 0$ it is rotationally symmetric, and $L_z$ is conserved. With $\delta > 0$ the rotations that leave $V$ unchanged are only multiples of 90°, and the symmetry drops to a discrete one. There are no infinitesimally small transformations for the theorem to use, so the theorem says nothing, and in fact $L_z$ is not conserved. $V$ does not depend on time, so $E$ is conserved.

Figure 2 Moving δ makes the angular momentum Lz start to fluctuate, while the energy E hardly fluctuates (about $10^{-5}$)δ = 0.00range of Lz 1.4e−14, max|E−E₀| 2.0e−5, t = 60.0




Computed Initial position $(1, 0)$, velocity $(0, 0.9)$. The orbit is computed by the browser with the velocity Verlet method ($h = 0.01$), and Table 3 gives the values from running the same procedure on this machine (1.4e−14 means $1.4 \times 10^{-14}$). The two bands on the right have the same vertical scale.

Table 3 Changes of angular momentum and energy ($0 \le t \le 60$)

| $\delta$ | $h$ | Minimum of $L_z$ | Range of $L_z$ | $\max\lvert E - E_0 \rvert$ |
|---|---|---|---|---|
| 0 | 0.01 | 0.9000 | 1.4e−14 | 2.0e−5 |
| 0 | 0.005 | 0.9000 | 1.6e−14 | 5.0e−6 |
| 0.1 | 0.01 | 0.8581 | 0.080 | 2.8e−5 |
| 0.3 | 0.01 | −1.0152 | 2.03 | 4.3e−5 |
| 0.5 | 0.01 | −1.0938 | 2.19 | 5.8e−5 |

For $\delta = 0.3$ and $0.5$ even the sign of $L_z$ changes, and the particle reverses its direction of rotation partway. The deviation of $E$ is about $10^{-5}$, and looks flat in the figure.

07

## Detour: the discretised action has Noether too

A section you may skip.

Figure 2 is computed step by step with step size $h$. Here position $q_k$ and velocity $v_k$ are vectors in the plane ($q_k \in \mathbb{R}^2$; the $q_k$ of § 03 were numbers), and they are advanced by the **velocity Verlet method** using the force $F = -\nabla V$.

$$v_{k+1/2} = v_k + \frac{h}{2m} F(q_k), \qquad q_{k+1} = q_k + h\, v_{k+1/2}, \qquad v_{k+1} = v_{k+1/2} + \frac{h}{2m} F(q_{k+1})$$

Even so, in Table 3 at $\delta = 0$, $L_z$ is constant to the level of rounding error, while the energy deviates by $10^{-5}$. There is a reason.

Eliminating the velocity gives $m(q_{k+1} - 2q_k + q_{k-1}) = h^2 F(q_k) = -h^2 \nabla V(q_k)$. This is exactly the stationarity condition of the discretised action of § 03 carried over to the plane Known (Hairer–Lubich–Wanner 2003, § 1.6; this form is called the Störmer–Verlet method, and the velocity Verlet method is one way of writing it). If $V$ is rotationally symmetric, the discretised action too is unchanged by a simultaneous rotation, the computation of § 06 goes through with differences, and the following $J$ is constant regardless of $k$.

$$J = \frac{m}{h}\,(q_k \times q_{k+1})_z$$

Here $(a \times b)_z = a_x b_y - a_y b_x$. It can also be checked directly. Multiplying the stationarity condition by $q_k \times$ on the left, for a central force $F(q_k)$ is parallel to $q_k$, so the right-hand side vanishes and $q_k \times q_{k+1} = q_{k-1} \times q_k$ remains. $J$ is $2m/h$ times the area of the triangle made by the origin and two neighbouring points, so the area swept in each step is constant.

This $J$ is exactly the $L_z = m(q_k \times v_k)_z$ of Table 3. The first two update formulas give $q_{k+1} = q_k + h\, v_k + \frac{h^2}{2m} F(q_k)$. For a central force $F(q_k)$ is parallel to $q_k$, so $q_k \times q_{k+1} = h\, q_k \times v_k$, hence $J = L_z$. Theorem 3.5 of the same paper shows in general that quadratic conserved quantities such as angular momentum are preserved exactly by this method Known . And § 1.3, citing Feynman's lectures, notes that the argument by which the "Principia" showed Kepler's second

law has the same form as one way of writing this method (half a step at constant velocity, a kick towards the centre, half a step at constant velocity).

On the other hand, time has been discretised, so it can be shifted only by integer multiples of $h$. The theorem does not work for discrete symmetries, and energy is not conserved exactly. However, if the formula for the energy (a function of position and velocity) is smooth and the numerical solution stays within some bounded region, the following has been shown for any positive integer $N$ Known (ibid. Theorem 5.1; $C$ and $C_N$ do not depend on $t$ and $h$).

$$|E_k - E_0| \le Ch^2 + C_N h^N t \qquad (0 \le t = kh \le h^{-N})$$

The second term is at most $C_N h^2$ if $t \le h^{2-N}$. Taking $N$ large, the error stays of order $h^2$ up to very long times, high powers of $h^{-1}$. It is not a theorem about infinite time. In Table 3, halving $h$ makes the deviation about 1/4.

"Keep the symmetry you want to keep exactly, as a symmetry of the discretised action" is also the starting point of Wilson's method in lattice gauge theory (5-01). What is kept exactly there is gauge symmetry, and what is broken are the continuous rotations and translations of spacetime (four dimensions turned into imaginary time; 4-02). The 90° rotations of the lattice directions and translations by the lattice spacing remain.

08

## Local symmetries give constraints

The $s$ of § 06 was a constant (a **global** symmetry). A symmetry under which the action is unchanged even when transforming by a $s(t)$ that differs from time to time is a **local** symmetry, and it is at the centre of gauge theory (3-01). We look at a toy with two variables, $x(t)$ and $\lambda(t)$.

$$L = \tfrac{1}{2} m(\dot{x} - \lambda)^2, \qquad x \mapsto x + s(t), \quad \lambda \mapsto \lambda + \dot{s}(t)$$

For any $s(t)$, $\dot{x} - \lambda$ does not change. $\lambda$ appears without time derivatives, so its equation is a **constraint**.

$$(\lambda)\ \ m(\dot{x} - \lambda) = 0, \qquad (x)\ \ \frac{d}{dt}\big[m(\dot{x} - \lambda)\big] = 0$$

The second is the derivative of the first, and only one is independent. $x$ remains as an arbitrary function that the equations do not fix. Noether's **second theorem** (1918) states that when the action is unchanged under transformations containing arbitrary functions, some of the Lagrange equations follow from the rest Known . To be exact, as many identities as there are arbitrary functions hold among the left-hand sides of the Lagrange equations and their derivatives. In this toy the arbitrary function is the one $s(t)$, and the identity is also one: "the left-hand side for $(x)$ is the time derivative of the

left-hand side for $(\lambda)$". The $Q = m(\dot{x} - \lambda)$ made by (N) from a constant $s$ is, far from conserved, constrained to be 0.

In electromagnetism the scalar potential $\phi$ plays the role of $\lambda$, the vector potential $\mathbf{A}$ the role of $x$, and Gauss's law plays the role of the equation for $\lambda$ (2-03, 4-01). To be exact, in electromagnetism $s(t)$ becomes a function at each point $\chi(t, \mathbf{x})$, $\mathbf{A}$ shifts by $\nabla\chi$ and $\phi$ by $-\partial\chi/\partial t$ (the formula for gauge transformations in 2-03; the sign of the shift of $\phi$ is opposite to the $\lambda$ of the toy). The degrees of freedom the equations do not fix are a redundancy of description, and handling them is one of the big problems of quantising Yang–Mills theory (4-08).

The toy itself has no physical content (the only solution is $\lambda = \dot{x}$, with $x$ arbitrary).

---

## What this chapter can say

| Matter | What is said | What is not said |
|---|---|---|
| The Euler–Lagrange equations | Known Derived for smooth paths with fixed ends | Variations with moving ends, the general theory of constraints |
| Stationary and minimum | Computed For the harmonic oscillator, a minimum if $\omega T < \pi$, a saddle point beyond (Tables 1 and 2) | The general theory of conjugate points (Jacobi's condition, the proof of the Morse index theorem) |
| Noether's first theorem | Known Proved for finitely many degrees of freedom when $L$ is exactly invariant | The case where $L$ changes by a total derivative, the form in field theory (2-02) |
| Numerical observation | Computed In one example ($t \leq 60$), the fluctuation and conservation of $L_z$ and the error $\propto h^2$ of $E$ were observed (Table 3) | Not a general claim. Long-time estimates rely on theorems in the literature |
| Noether's second theorem | The redundancy of equations and constraints seen in one toy | The general proof and the form in field theory (3-01) |
| The origin of the action principle | — | Physics 2-07 |

## Sources and reproduction

| Item | Kind | Source / tools |
|---|---|---|
| The Euler–Lagrange equation, generalised coordinates, (N) | Known Textbook | L. D. Landau, E. M. Lifshitz, *Mechanics*, 3rd ed., Pergamon Press, 1976. Bibliographic data only |
| Noether's two theorems | Known Primary source | E. Noether, Invariante Variationsprobleme, Nachr. Ges. Wiss. Göttingen, Math.-phys. Kl. (1918) 235–257. English translation M. A. Tavel, Invariant variation problems, Transport Theory and Statistical Physics 1(3) (1971) 186–207, doi:10.1080/00411457108231446, arXiv:physics/0503066 (Theorems I and II of § 1 checked) |
| Second variation, conjugate points | Known Textbook | D. Liberzon, *Calculus of Variations and Optimal Control Theory*, Princeton University Press, 2012, doi:10.1515/9781400842643 (Jacobi's condition and conjugate points checked in § 2.6.2 of the pre-publication version). I. M. Gelfand, S. V. Fomin, *Calculus of Variations*, 1963. Bibliographic data only |
| The Morse index theorem | Known Textbook | J. Milnor, *Morse Theory*, Annals of Mathematics Studies 51, Princeton University Press, 1963 ( § 15 Theorem 15.1 checked; what is stated is the case of geodesics) |
| On the Störmer–Verlet method of § 07 | Known Review | E. Hairer, C. Lubich, G. Wanner, Acta Numerica 12 (2003) 399–450, doi:10.1017/S0962492902000144 ( § 1.3, § 1.6, Theorems 3.5 and 5.1 checked) |
| Variational integrators | Known Review | J. E. Marsden, M. West, Acta Numerica 10 (2001) 357–514. Bibliographic data only |
| Tables 1, 2, 3 | Computed | python3 + numpy (the record of this book's study). `action_scan.py` (Simpson's rule), `hessian.py` (`eigvalsh`), `noether_verlet.py` (the velocity Verlet method; Figure 2 runs the same procedure in the browser) |

Next chapter: 2-02 The idea of a field (replacing $q_i$ by a field $\phi(x)$, the degrees of freedom become infinitely many).

# The idea of a field — when the number of beads becomes infinite, the coordinates of the system become a function

> 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** — That a field is the limit of a chain of beads as the number of beads goes to infinity.
That a string is a collection of infinitely many independent simple oscillations, and that with a mass term the frequencies have a lower bound, which remains when the box is enlarged.
The price: the value of a field in thermal equilibrium is not determined at a point when space has two or more dimensions.

**Prerequisite chapters** — 2-01 (the Euler–Lagrange equations, Noether's theorem). Eigenvalues of symmetric matrices (1-01). Fourier series are built in section 04 as far as needed.

**Said first** — Only classical fields are treated (quantisation is in 2-06). The "lower bound on frequencies" of section 06 is a property of classical waves, and is not itself the mass gap of the problem statement (a gap in the spectrum of quantum theory, 4-11). This chapter shows nothing about the mass gap.

Order of this chapter

1. A chain of beads — N degrees of freedom
2. Normal modes — splitting into independent simple oscillations
3. The continuum limit — the wave equation comes out (Figure 1)
4. Solutions as waves — travelling waves and standing waves (Figure 2)
5. Energy — conservation and distribution over modes
6. The mass term — a lower bound on frequencies (Figure 3)
7. What infinitely many degrees of freedom means — a hot field is not determined at a point (Figure 4)

8. Where this goes in the book — it split because it was linear (Detour: the string of confinement also vibrates by this equation)

---

01

## A chain of beads — a system with N degrees of freedom

The shortest way into fields is to start from finitely many beads. Place $N$ beads at spacing $a = \ell/(N+1)$ on an interval of length $\ell$, join neighbours with taut threads, and fix both ends to walls. Let $q_j(t)$ be the sideways displacement of bead $j$ (the walls are $q_0 = q_{N+1} = 0$), $m = \rho a$ the mass of a bead ($\rho$ is the linear density), and $\sigma$ the tension.

If the difference $\Delta q$ from a neighbour is small, the thread between them stretches by $\sqrt{a^2 + \Delta q^2} - a \approx \Delta q^2/(2a)$. We let the beads move only sideways, and discard the next term $-\Delta q^4/(8a^3)$. This makes the following equations linear (where linearity matters is section 08). Tension $\times$ stretch is the potential energy, so the Lagrangian (2-01) and the Euler–Lagrange equations take the following form.

$$L = \sum_{j=1}^{N} \frac{m}{2}\dot{q}_j^2 - \sum_{j=0}^{N} \frac{\sigma}{2a}(q_{j+1} - q_j)^2$$

$$m\ddot{q}_j = \frac{\sigma}{a}\left(q_{j+1} - 2q_j + q_{j-1}\right)$$

Only neighbouring beads appear on the right-hand side. This **locality** survives all the way to gauge theory. **Example.** If $N = 1$, it is a simple oscillation with $m\ddot{q} = -(2\sigma/a)q$, and when $\ell = \rho = \sigma = 1$, $\omega = \sqrt{8} \approx 2.828$.

02

## Normal modes — untangling the oscillations

Writing everything together as $\ddot{\mathbf{q}} = -K\mathbf{q}$, $K$ is the symmetric matrix with diagonal entries $2\sigma/(ma)$, $-\sigma/(ma)$ on either side, and 0 elsewhere. Being symmetric, it has $N$ orthogonal eigenvectors, and its eigenvalues are real (1-01). The eigenvalues are not negative either: $\mathbf{q} \cdot K\mathbf{q}$ is $2/m$ times the potential energy, a sum of squares. For an eigenvector $\mathbf{v}$, $\mathbf{v} \cdot K\mathbf{v} = \lambda|\mathbf{v}|^2$, so $\lambda \geq 0$. Both ends stay at the walls, so the potential energy is 0 only when $\mathbf{q} = 0$, and $\lambda = 0$ does not occur either. So we write the eigenvalues as $\omega_n^2$ ($\omega_n > 0$).

Taking the eigenvectors as new axes, each component obeys $\ddot{Q}_n = -\omega_n^2 Q_n$. **$N$ tangled beads split into $N$ mutually unrelated simple oscillations**. These are the normal modes; each way of swinging is called a "mode", and $n$ the number of the mode. For this matrix a closed formula can be written for $n = 1, \dots, N$.

$$v_n(j) = \sin \frac{jn\pi}{N+1}$$

$$\omega_n = 2\sqrt{\frac{\sigma}{ma}} \sin \frac{n\pi}{2(N+1)}$$

It can be checked by substitution using $\sin(A+B) + \sin(A-B) = 2\sin A \cos B$.
**Example.** If $\ell = \rho = \sigma = 1$ then $\omega_n = 2(N+1)\sin \frac{n\pi}{2(N+1)}$, and when $n$ is much smaller than $N$, $\omega_n \approx n\pi$: the frequencies of the string in the next section.

03

## The continuum limit — the field and the wave equation

Increase the number of beads and bring the spacing $a$ to 0. $\ell$, $\rho$, $\sigma$ are fixed, and the mass of a bead $m = \rho a$ goes to 0. **What to fix when taking the limit** returns as a question of the same form in lattice gauge theory (5-09). Reading the number $j$ as the position $x = ja$ and writing $q_j(t) = u(t, x)$, we get $(q_{j+1} - q_j)/a \to \partial_x u$, $\sum_j a(\cdots) \to \int dx$, $(q_{j+1} - 2q_j + q_{j-1})/a^2 \to \partial_x^2 u$, and

$$L = \int_0^\ell \mathcal{L}\, dx, \quad \mathcal{L} = \frac{\rho}{2}(\partial_t u)^2 - \frac{\sigma}{2}(\partial_x u)^2$$

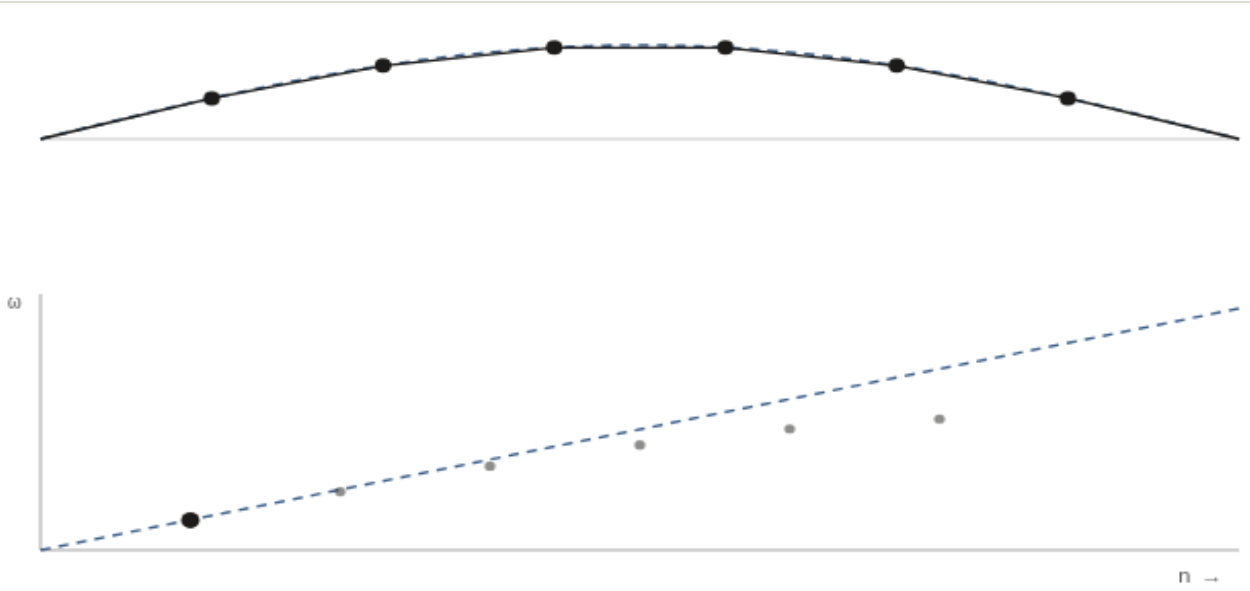
$$\partial_t^2 u = c^2 \partial_x^2 u, \quad c = \sqrt{\sigma/\rho}$$

$\mathcal{L}$ is called the **Lagrangian density**, and the formula below the **wave equation**. The same formula also comes from the Euler–Lagrange equation for fields $\partial_t \frac{\partial \mathcal{L}}{\partial(\partial_t u)} + \partial_x \frac{\partial \mathcal{L}}{\partial(\partial_x u)} - \frac{\partial \mathcal{L}}{\partial u} = 0$, and this form is used both for the electromagnetic field (2-03) and for the Yang–Mills field (3-03).

A **field** is a system in which the index $j$ has been replaced by a continuous $x$. The "coordinates" of the system are not a set of numbers but a single function $u(t, \cdot)$. The normal modes of the string are $\sin(n\pi x/\ell)\cos\omega_n t$ with frequencies $\omega_n = n\pi c/\ell$ ($n = 1$, $2, \dots$), with no upper limit.

Figure 1 Changing N and n changes the vibration of the chain (solid line and dots) and of the string (dashed), and the mismatch of their frequencies. The lower panel shows the frequencies of all modes (dashed: string ω = nπ,

dots: chain). Pressing a number in Table 1 sets that N and n




N = 6, n = 1, ω(chain) = 3.115, ω(string) = 3.142, ratio = 0.9916

Computed The numbers in the figure are computed on the spot from the closed formulas in the text.

Computed **Table 1** Ratio $\omega_n/(n\pi)$ of the frequencies of the chain and the string ($\ell = \rho = \sigma = 1$). Values found numerically as eigenvalues of the matrix $K$; the relative difference from the closed formula was at most $3 \times 10^{-12}$.

| N | n = 1 | n = N/2 | n = N |
|---|---|---|---|
| 4 | 0.9836 | 0.9355 | 0.7568 |
| 16 | 0.9986 | 0.9114 | 0.6735 |
| 64 | 0.99990 | 0.9033 | 0.6464 |
| 256 | 0.999994 | 0.9011 | 0.6391 |

**With $n$ fixed the ratio approaches 1**, but the column $n = N/2$ only approaches $\sin(\pi/4)/(\pi/4) \approx 0.900$, and the column $n = N$ only $2/\pi \approx 0.637$. The continuum limit is an approximation for **waves whose wavelength is much longer than the spacing**.

04

## Solutions as waves — travelling waves and standing waves

The solutions can be read in two ways. **As travelling waves**: for any twice differentiable $f, g$, $u = f(x - ct) + g(x + ct)$ is a solution. $f$ travels to the right and $g$ to the left, keeping their shape, at speed $c$. With initial shape $u_0$ and initial velocity $0$, $u = \frac{1}{2}[u_0(x - ct) + u_0(x + ct)]$, and the hump splits in two (with both ends fixed, extend $u_0$ to an odd function of period $2\ell$). Solutions of this form are said to go back to d'Alembert's paper of 1747.

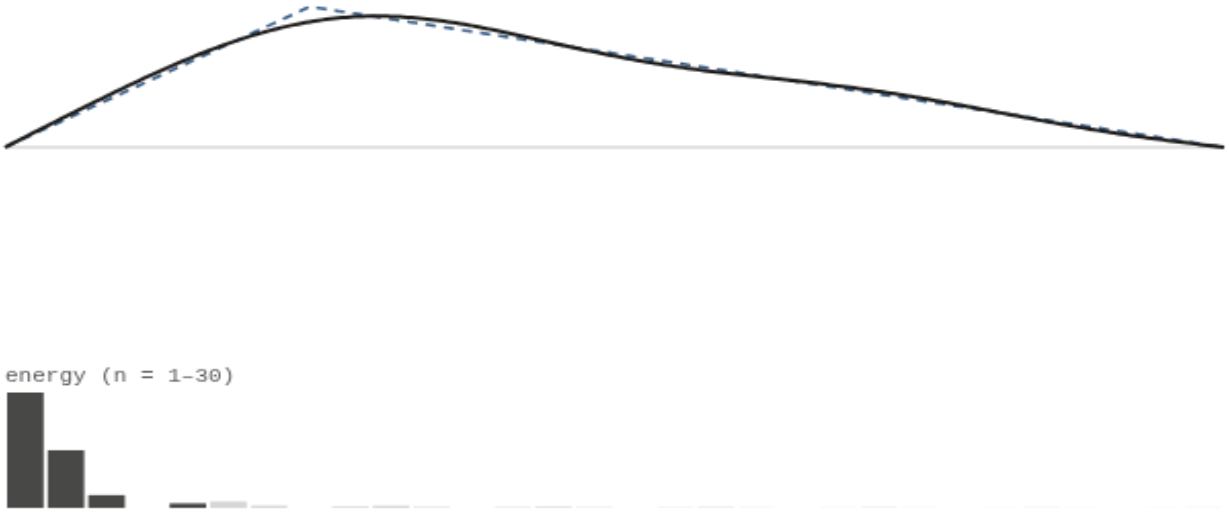
**As superpositions of standing waves**: the equation is linear, so sums of normal modes are also solutions.

$$u = \sum_{n \geq 1} b_n \sin \frac{n\pi x}{\ell} \cos \omega_n t$$

$$b_n = \frac{2}{\ell} \int_0^{\ell} u_0(x) \sin \frac{n\pi x}{\ell} \, dx$$

The formula for the coefficients comes from the orthogonality of the $\sin(n\pi x/\ell)$ (the integral of the product is 0 for $n \;/= n'$ and $\ell/2$ for $n = n'$). It is a Fourier sine series (1-11). **Example.** For the triangle obtained by pulling the position $p$ to height $h$ and releasing, $b_n = \frac{2h\ell^2}{n^2\pi^2 p(\ell - p)} \sin \frac{n\pi p}{\ell}$. The coefficients decrease only as $1/n^2$, and representing the corner needs modes up to high numbers.

Figure 2 Changing K and p changes how close the sum of standing waves (solid line) is to d'Alembert's solution (dashed), and the distribution of energy (bars). Played, the hump splits in two and turns over at the walls. Pressing a number in Table 2 sets that p and K




K = 5, p/ℓ = 0.25, energy of the first 5: 89.2%

Computed The numbers in the figure are computed on the spot from the partial sums of $b_n$ and d'Alembert's formula ($\ell = c = h = 1$).

05

## Energy — conservation, and distribution over modes

The sum $E$ of kinetic and potential energy does not depend on time. Using the wave equation, the time derivative of the integrand becomes $\sigma \, \partial_x(\partial_t u \, \partial_x u)$, and since both ends do not move the boundary terms vanish.

$$E = \int_0^\ell \Big[\frac{\rho}{2}(\partial_t u)^2 + \frac{\sigma}{2}(\partial_x u)^2\Big]\,dx$$

$$\frac{dE}{dt} = \sigma\big[\partial_t u\;\partial_x u\big]_0^\ell = 0$$

This is the field version of Noether's theorem (2-01). Splitting into normal modes, the cross terms vanish by orthogonality, $E = \sum_n E_n$, $E_n = \frac{\rho\ell}{4}(\dot b_n^2 + \omega_n^2 b_n^2)$. Each $E_n$ is conserved separately. **A string is a collection of infinitely many mutually unrelated simple oscillations**.

**Example.** For a plucked string ($s = p/\ell$), the share of mode $n$ is $f_n = \frac{2\sin^2(n\pi s)}{n^2\pi^2 s(1-s)}$.

Computed **Table 2** The fraction of the energy held by the first $K$ modes. We also checked that the sum of all terms is 1 (to 5 digits or more with 2 million terms).

| s = p/ℓ | K = 1 | K = 3 | K = 10 | K = 100 |
|---|---|---|---|---|
| 0.5 | 0.8106 | 0.9006 | 0.9596 | 0.9959 |
| **0.25** | **0.5404** | **0.8706** | **0.9508** | **0.9946** |
| 0.1 | 0.2150 | 0.5732 | 0.8920 | 0.9887 |
| 0.02 | 0.0408 | 0.1217 | 0.3882 | 0.9489 |

Plucking in the middle gives $f_1 = 8/\pi^2 \approx 0.81$. Near the end ($s = 0.02$) the first 10 do not reach 40%, and the energy escapes to the short-wavelength side.

06

## The mass term — a lower bound on frequencies

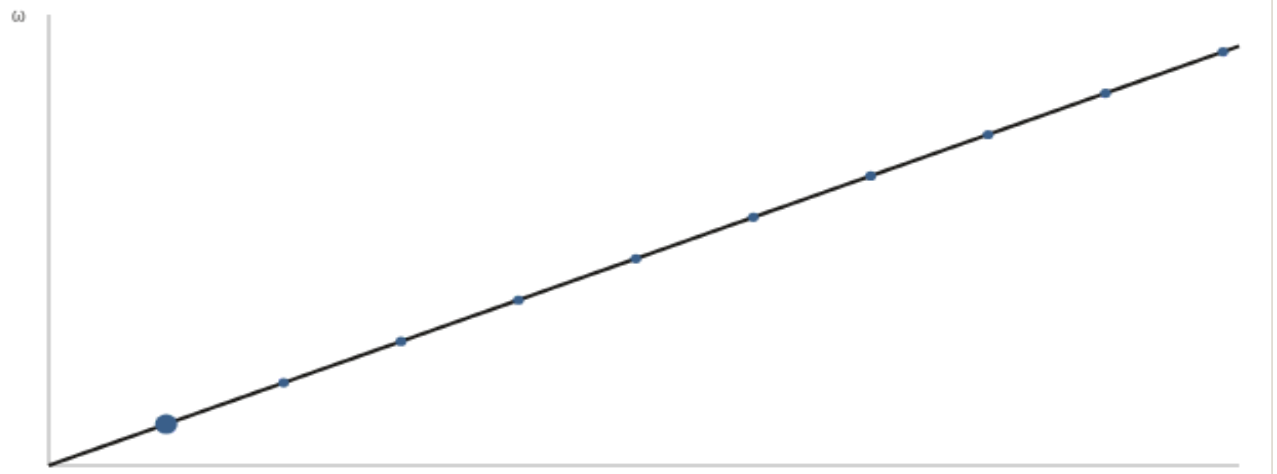
If each bead is also tied to the floor by a spring (constant $\kappa$) pulling it back to its original position, $\sum_j \frac{\kappa}{2} q_j^2$ is added to the potential energy. The continuum limit with $\Omega^2 = \kappa/m$ fixed is the Klein–Gordon equation on the left, and substituting a plane wave $\cos(kx - \omega t)$ gives the **dispersion relation** on the right. Here $k$ is the **wave number** (the advance of phase per unit length; for wavelength $\lambda$, $k = 2\pi/\lambda$), a different quantity from the mode number $n$.

$$\partial_t^2 u = c^2\partial_x^2 u - \Omega^2 u, \qquad \omega^2 = c^2k^2 + \Omega^2$$

If $\Omega = 0$ then $\omega = c|k|$, and lengthening the wavelength brings the frequency towards 0. If $\Omega > 0$ then always $\omega \geq \Omega$: **the frequencies have a lower bound.** In a box of length $\ell$ the wave numbers are restricted to $k_n = n\pi/\ell$, and the lowest frequency is $\omega_1 = \sqrt{(c\pi/\ell)^2 + \Omega^2}$. In a finite box it is positive even for $\Omega = 0$, and **what is asked is**

**whether the lower bound remains when the box is enlarged**. This is why the definition of the mass gap is written in infinite volume (4-11).

Figure 3 Changing Ω and ℓ changes the dispersion relation (curve) and the waves the box allows (dots). The large dot at the left end is the lowest frequency




ℓ = 4.0, Ω = 0.00, lowest frequency $\omega_1$ = 0.7891

Computed The numbers in the figure are computed on the spot from $\omega = \sqrt{k^2 + \Omega^2}$, $k_n = n\pi/\ell$ ($c = 1$).

Computed The lowest frequency ($c = 1$) for $\ell = 1, 10, 100, 1000$ is 3.142, 0.314, 0.0314, 0.00314 at $\Omega = 0$, and 3.297, 1.048, 1.0005, 1.000005 at $\Omega = 1$ (values of the closed formula). For $\ell \leq 100$ it agreed with the smallest eigenvalue of the matrix of a chain with spacing 0.05 to within a relative difference of 0.11% (largest at $\ell = 1$).

The classical lower bound, and the mass gap

Known After quantisation (2-06), the lowest excitation above the vacuum of a free field is $\hbar\Omega$ in infinite volume. Comparing with $E = mc^2$ we read it as $\Omega = mc^2/\hbar$, which is why this term is called the **mass term**. The Yang–Mills action has no mass term (gauge invariance does not allow it, 3-02). Physics Even so, in quantum theory a gap is thought to arise, and numerical computations on the lattice support this (5-08). Showing as mathematics, without adding a mass term, that a gap comes out of the theory itself — that is the content of the problem (0-03).

07

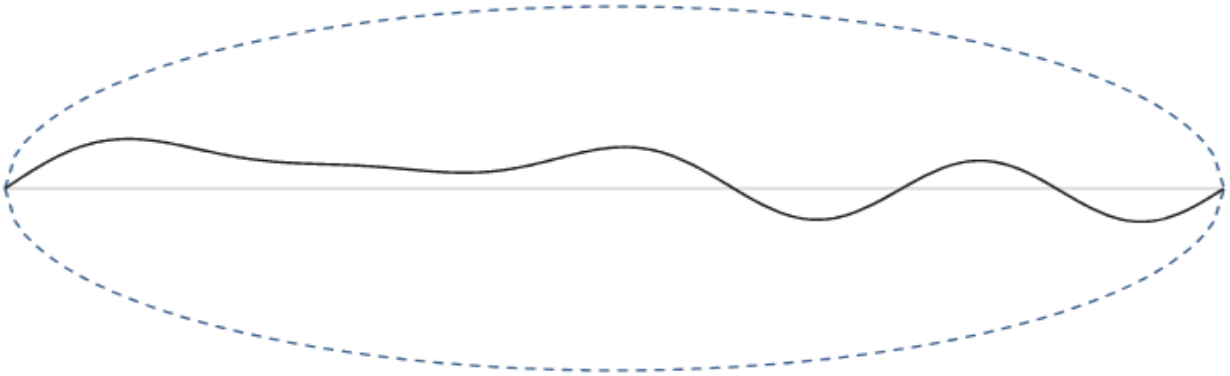
## What infinitely many degrees of freedom means

A string has countably infinitely many degrees of freedom. The trouble comes when **all modes are shaken with the same weight**, the typical case being thermal equilibrium. A minimal definition (details in 2-08): in thermal equilibrium at

temperature $T$, the probability density of a configuration is proportional to $\exp(-V/k_BT)$ ($k_B$ is Boltzmann's constant). If $V = \sum_n \frac{\kappa_n}{2} Q_n^2$, each $Q_n$ follows an independent Gaussian distribution (1-08) with variance $k_BT/\kappa_n$, and the mean potential energy is $k_BT/2$ for each (**equipartition**). For the string, $\kappa_n = \sigma n^2\pi^2/(2\ell)$ for $u = \sum_n q_n \sin(n\pi x/\ell)$, and three things happen.

1. **The total energy is infinite**, because $k_BT/2$ is given to each of infinitely many modes.
2. **The displacement at a point is finite.** $\operatorname{Var} u(x) = \sum_n \frac{2\ell k_BT}{\sigma n^2\pi^2} \sin^2\frac{n\pi x}{\ell} = \frac{k_BT}{\sigma} \cdot \frac{x(\ell-x)}{\ell}$. This random curve is called the **Brownian bridge** (Known Sheffield 2007).
3. **The slope is not determined.** $\operatorname{Var} \partial_x u = \frac{2k_BT}{\sigma\ell} \sum_n \cos^2\frac{n\pi x}{\ell}$ diverges. The curve does not break off, yet it is all corners everywhere. To be exact, it means that at no point can the slope be defined as a random variable.

Figure 4 Doubling K again and again makes the sample of the string in thermal equilibrium finer. The overall course does not change, and it stays mostly within the dashed band $\pm 2\sqrt{x(1-x)}$ (twice the standard deviation) (the probability of being outside at each point is about 5%)



```
K = 8, variance at the centre = 0.2374 k_BTℓ/σ (limit 0.25), variance of the slope at the centre
= 8 k_BT/(σℓ) (keeps growing with K)
```

Computed The figure computes $u = \sum_{n\le K} g_n \sqrt{2}\sin(n\pi x)/(n\pi)$ on the spot ($g_n$ are normal random numbers with a fixed seed, $\ell = k_BT/\sigma = 1$).

Computed Raising the dimension of space $d$ (a membrane, a field in a box), we count the variance at the centre of the box $[0,1]^d$ (in units of $k_BT/\sigma$) as a sum over modes with the number in each direction up to $K$ (doubling from $K = 10$ to 160). The contribution of mode $n = (n_1, \dots, n_d)$ is proportional to $1/|n|^2$. The table and figure of the same quantity discretised on a lattice are in Table 3 and Figure 3 of 1-09, which this book takes as authoritative. A figure setting it beside quantities that switch at four is the dimension slider of 1-12 § 10.

$d = 1$ converges to $1/4$ (the value at the centre of the Brownian bridge). $d = 2$ increases by $\ln 2/(2\pi) \approx 0.110$ each time $K$ is doubled (diverging as $\log K$); $d = 3$ doubles, and $d = 4$ quadruples. The number of modes in the shell of radius $r$ is proportional to $r^{d-1}$, with $1/r^2$ for each, so the sum has the form $\sum r^{d-3}$. The divergence comes from **the short-wavelength side**.

Known Write a field in a box of dimension $d$ as $\varphi(x)$ (the multidimensional version of the $u$ of the string). For $d \geq 2$, the value $\varphi(x)$ of a hot field at a point has no meaning, and only the value $\int \varphi f \, dx$ obtained by multiplying by a smooth weight function $f$ and integrating has meaning as a random variable. In this sense, the field takes values not in functions but in **distributions** (Sheffield 2007, 1-09, 1-11). Writing a quantum field in imaginary time treats time like one more direction of space, and a free field in four-dimensional spacetime has a sum of the same form as the $d = 4$ row (4-02).

The rougher the field, the harder it is to give meaning to products of fields (non-linear terms). To be exact, products of distributions cannot in general be defined (1-11). Divergences (4-04) are their naive appearance. Stating exactly how rough a field can be while non-linear terms can still be handled is the "subcritical / critical" of stochastic quantisation (7-09).

08

## Where this goes in the book

Uses ahead: the potentials of the electromagnetic field and their redundancy (2-03), the Lorentz invariance of $\partial_t^2/c^2 - \partial_x^2$ (2-04, the article on spacetime; whether a wave leaves a tail is decided by the dimension of space: same article), the Fock space in which the modes become quantum oscillators (2-06), probability on fields (2-07).

In 3-02 one assumption breaks down. **The splitting into normal modes worked because the equation was linear**. The Yang–Mills field takes values in a Lie algebra at each point and interacts with itself, so the modes are no longer independent. Classical waves have no lower bound either. Physics Even so, in quantum theory a gap is thought to arise.

### Detour: the string of confinement also vibrates by this equation

A subsection you may skip.

Physics When a pair of heavy charges is separated by a distance $r$, the Yang–Mills field between them is thought to become a tube like a string with tension $\sigma$ (17-03). The tube is regarded as a relativistic string, in which case the linear density is $\rho = \sigma/c^2$ and the speed of transverse waves $\sqrt{\sigma/\rho}$ equals the speed of light $c$. The transverse vibrations of the tube are taken to obey the

wave equation; summing the zero-point energy of mode $\omega_n = n\pi c/r$ (the lowest energy $\hbar\omega_n/2$ of a quantum oscillator, 2-06) over the $d-2$ directions of vibration (here $d$ is the dimension of spacetime, the same letter as in 17-03; the $d$ of section 07 is the dimension of the space the field spreads over), and formally assigning $-1/12$ to the divergent $1+2+3+\cdots$ (4-12), gives $V(r) = \sigma r - \frac{(d-2)\pi\hbar c}{24 r} + \cdots$. This Lüscher term has been compared, as a prediction of effective string theory, with numerical results on the lattice (Lüscher 1981; review Hari Dass 2023). It is an outline, not a derivation.

## What this chapter can say

| Content | Status |
|---|---|
| The classical facts of sections 01–06 (the frequencies of the chain, the continuum limit, the two solutions as waves, the conservation and distribution of energy, the lower bound from the mass term) | Said Known (numbers also checked in Tables 1 and 2 and section 06 Computed ) |
| That the variance at a point of a field in thermal equilibrium is finite for d = 1 and diverges for d $\geq$ 2 | Said Known (the rate of divergence from numbers with K $\leq$ 160 and counting shells Computed ) |
| Rigorous convergence of the continuum limit | Not said (the limit in this chapter is formal) |
| The correspondence between the classical lower bound on frequencies and the quantum gap | Said only for free fields Known (2-06) |
| The string of confinement and the Lüscher term / the Yang–Mills mass gap | The former is only an outline at the level of physics Physics . On the latter this chapter shows nothing |

## Sources and reproduction

| Item | Kind | Source / tools |
|---|---|---|
| Tables 1 and 2, the numbers of sections 06 and 07, Figures 1–4 | Computed | The tables with python3 and NumPy (eigenvalues, sums of series). The figures with JavaScript in the browser (the formulas in the text and partial sums) |
| The equation of the string and its general solution | Known | J. le R. d'Alembert, “Recherches sur la courbe que forme une corde tenduë mise en vibration”, Hist. Acad. R. Sci. Belles-Lettres Berlin 3 (for 1747) pp. 214–219. Bibliographic data only (year of publication not checked) |
| The Brownian bridge, fields becoming distributions | Known | S. Sheffield, “Gaussian free fields for mathematicians”, Probab. Theory Relat. Fields 139 (2007) 521–541, arXiv:math/0312099 ( § 2.5 checked) |

| Item | Kind | Source / tools |
|---|---|---|
| Lattice numbers supporting the gap of quantum Yang–Mills | Physics | Left to the table of sources of 5-08. One example: C. J. Morningstar, M. Peardon, “The glueball spectrum from an anisotropic lattice study”, Phys. Rev. D 60 (1999) 034509, arXiv:hep-lat/9901004 (bibliographic data only) |
| The Lüscher term | Physics | M. Lüscher, “Symmetry breaking aspects of the roughening transition in gauge theories”, Nucl. Phys. B180 (1981) 317–329 (bibliographic data only). The form of the formula checked in formula (1) of N. D. Hari Dass, “Effective String Theories (EST) of Yang-Mills Flux Tubes”, arXiv:2312.10629 (Chapter 22 of the book Strings to Strings, Springer 2023) |

Next chapter: 2-03 Electromagnetism is a gauge theory. To go ahead to quantisation, 2-06.

# Electromagnetism is a gauge theory — the freedom in writing the potentials becomes an element of U(1) chosen at each point

> 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** — That infinitely many potentials give the same electric and magnetic fields (gauge transformations).
That Gauss's law is a constraint, and that gauge invariance and the conservation of charge come as a pair.
That in quantum mechanics a gauge transformation turns the phase (an element of U(1)) point by point, and the phase factor of a closed path is observable.

**Prerequisite chapters** — 2-02, section 08 of 2-01 (local symmetries and constraints), 1-01. An introduction to vector analysis and electromagnetism is assumed; the minimum of quantum mechanics is set out in section 05.

**Said first** — The relativistic form is in 2-04, the non-abelian case in 3-02. The electromagnetic field without interactions has no mass gap (section 04). This chapter shows nothing about the mass gap of Yang–Mills.

Order of this chapter

1. The electric and magnetic fields — Maxwell's equations
2. Potentials — satisfying the source-free Maxwell equations automatically
3. Gauge transformations — the freedom in writing (Figure 1)
4. From the action — constraints and the conservation of charge
5. Quantum phase and U(1) — turning the phase point by point (phase factors)
6. The Aharonov–Bohm effect — the fringes move even with B = 0 (Figure 2, Table 1)
7. A particle on a ring — the levels are periodic functions of the flux
8. A lattice ring — putting elements of U(1) on the edges (Figure 3, Table 2)

9. From U(1) to SU(N) — towards the gauge principle

---

01

## The electric and magnetic fields — four equations

The protagonists of electromagnetism are two fields carrying a three-dimensional vector at each point, the **electric field** $\mathbf{E}(t, \mathbf{x})$ and the **magnetic field** $\mathbf{B}(t, \mathbf{x})$. A particle of charge $q$ and velocity $\mathbf{v}$ feels the **Lorentz force** $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, and the fields obey **Maxwell's equations** (SI units; $\rho$ is the charge density, $\mathbf{J}$ the current density, $c^2 = 1/\varepsilon_0 \mu_0$).

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}, \qquad \nabla \times \mathbf{B} - \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J} \quad \text{(M1)}$$

$$\nabla \cdot \mathbf{B} = 0, \qquad \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0} \quad \text{(M2)}$$

The two Maxwell equations with sources (**Gauss's law** and the **Ampère–Maxwell law**) say how charges and currents make the fields; the two source-free Maxwell equations (the absence of magnetic charge and **Faraday's law of induction**) are constraints on the form of the fields. **Example.** Where there are no sources, the four Maxwell equations give $\partial_t^2 \mathbf{E} = c^2 \nabla^2 \mathbf{E}$ (the three-dimensional version of the wave equation of 2-02), and this wave is light. The dispersion $\omega = c|\mathbf{k}|$ has no lower bound.

Since the source-free Maxwell equations contain no sources, the form of $\mathbf{E}, \mathbf{B}$ can be narrowed down before solving the equations. That is the next section.

02

## Potentials — satisfying the two source-free equations automatically

With the identities $\nabla \cdot (\nabla \times \mathbf{A}) = 0$ and $\nabla \times \nabla f = \mathbf{0}$, the source-free Maxwell equations can be satisfied automatically. Writing $\mathbf{B}$ as a curl makes the first equation hold; the second becomes $\nabla \times (\mathbf{E} + \partial_t \mathbf{A}) = \mathbf{0}$, so we write $\mathbf{E} + \partial_t \mathbf{A}$ as a gradient.

$$\mathbf{B} = \nabla \times \mathbf{A}, \qquad \mathbf{E} = -\nabla \phi - \frac{\partial \mathbf{A}}{\partial t} \quad \text{(P)}$$

We call $\mathbf{A}$ the **vector potential** and $\phi$ the **scalar potential**. On the whole of $\mathbb{R}^3$, if $\nabla \cdot \mathbf{B} = 0$ then $\mathbf{B}$ can always be written as a curl Known (1-06).

**Example 1.** For a uniform $\mathbf{B} = B_0\mathbf{e}_z$, both $\mathbf{A}_1 = \frac{B_0}{2}(-y, x, 0)$ and $\mathbf{A}_2 = B_0(0, x, 0)$ can be used. **Example 2.** A long solenoid of radius $a$, with a uniform $B_0\mathbf{e}_z$ inside, no magnetic field outside, and flux $\Phi = \pi a^2 B_0$. With the distance from the axis $r$ and the unit vector $\mathbf{e}_\theta$ pointing around the axis,

$$\mathbf{A} = \frac{B_0 r}{2}\,\mathbf{e}_\theta \quad (r < a), \qquad \mathbf{A} = \frac{\Phi}{2\pi r}\,\mathbf{e}_\theta \quad (r > a)$$

Outside, $\nabla \times \mathbf{A} = \frac{1}{r}\partial_r(rA_\theta)\mathbf{e}_z = \mathbf{0}$, and yet $\mathbf{A} \neq \mathbf{0}$. The line integral once round a circle of radius $r$ is $\Phi$, which by Stokes's theorem is the flux passing through it. Even for a path that runs only where the curl is 0, the value of one circuit is not 0. This is because the path encloses a hole (the solenoid), and section 06 is where quantum mechanics reads off this value.

03

## Gauge transformations — the freedom in writing that gives the same fields

The potential representation does not determine $(\phi, \mathbf{A})$ uniquely from $\mathbf{E}, \mathbf{B}$. With a smooth function $\chi(t, \mathbf{x})$,

$$\mathbf{A} \mapsto \mathbf{A} + \nabla\chi, \qquad \phi \mapsto \phi - \frac{\partial \chi}{\partial t} \quad \text{(G)}$$

this replacement leaves $\mathbf{E}, \mathbf{B}$ unchanged ($\nabla \times \nabla\chi = \mathbf{0}$, $\nabla\partial_t\chi = \partial_t\nabla\chi$). We call this a **gauge transformation**, and $\chi$ the **gauge function**. **Example.** The two of Example 1 are carried into each other by $\chi = B_0 xy/2$ ($\mathbf{A}_2 - \mathbf{A}_1 = \nabla\chi$).

Reducing the freedom by a condition (such as $\nabla \cdot \mathbf{A} = 0$) is called **gauge fixing** (the quantum version is in 4-08).

### Seen on a lattice

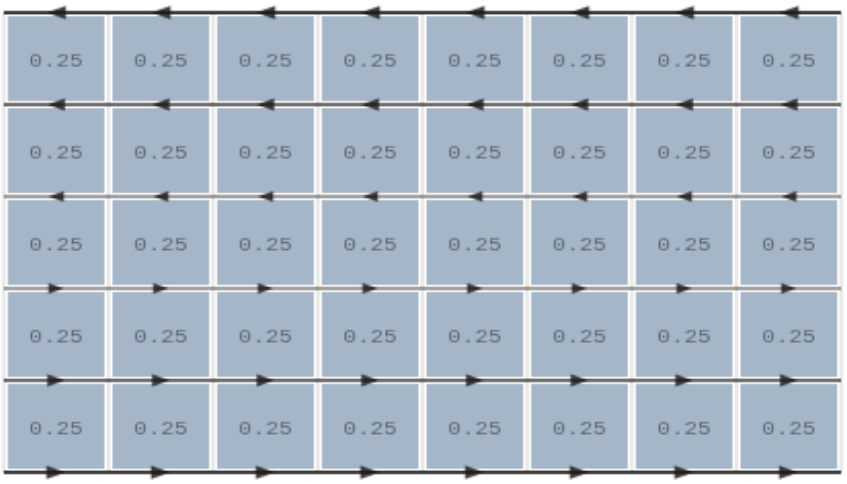
Cut the plane into a square lattice and hold $\mathbf{A}$ as a number $A_{ij} = \int_i^j \mathbf{A} \cdot d\mathbf{l}$ on each edge ($A_{ji} = -A_{ij}$). Label the corners of a square face $p$ 1 to 4 anticlockwise; by Stokes's theorem the flux $B_p$ through the face is the sum over the four edges, and the gauge function becomes the values $\chi_i$ at the lattice points.

$$B_p = A_{12} + A_{23} + A_{34} + A_{41}, \qquad A_{ij} \mapsto A_{ij} + \chi_j - \chi_i$$

The $\chi$ of each corner appears once in $+$ and once in $-$ and cancels, so $B_p$ does not change. **Example.** The sum of the edge values $(0.3,\ 0.5,\ -0.1,\ -0.2)$ is 0.5. Transforming by $\chi = (1, 0, 2, -1)$ gives $(-0.7,\ 2.5,\ -3.1,\ 1.8)$, but the sum stays 0.5.

Putting group elements on the edges gives the starting point of Wilson's lattice gauge theory (5-01).

Figure 1 Changing the size s of the gauge function, or pressing Play, changes the edge values (line thickness and arrows), but not the face values (shading and numbers)




s = 0.00, max. change on edges 0.000, max. change on faces 0.0e+0

Computed The numbers in the figure are computed on the spot from the sums over four edges and the addition of $\chi_j - \chi_i$.

"Flux through a single face only" is the lattice version of the solenoid of section 06. The row of non-zero edges changes shape under gauge transformations but does not disappear (it corresponds to the cut where the angle $\theta$ of section 06 jumps). Under "No magnetic field" only the edges fluctuate; this is a **pure gauge**.

04

## From the action — constraints and the conservation of charge

Maxwell's equations too are the stationarity conditions of an action (2-01, 2-02). The unknowns are $(\phi, \mathbf{A})$, and $\mathbf{E}, \mathbf{B}$ are understood as written by the potential representation.

$$\mathcal{L} = \frac{\varepsilon_0}{2}|\mathbf{E}|^2 - \frac{1}{2\mu_0}|\mathbf{B}|^2 - \rho\,\phi + \mathbf{J}\cdot\mathbf{A}, \qquad S = \int \mathcal{L}\, dt\, d^3x \quad \text{(L)}$$

The source-free Maxwell equations hold automatically from the potential representation, and the Maxwell equations with sources come out as the Euler–Lagrange equations Known . $\mathcal{L}$ contains $\nabla\phi$ but not $\partial_t\phi$, so the $\varepsilon_0\nabla\cdot\mathbf{E} = \rho$ obtained by varying $\phi$ is a condition at equal times containing no second time derivative: a **constraint**. It corresponds to the equation for $\lambda$ in the toy of 2-01. In this way, of the four components of $(\phi, \mathbf{A})$, one disappears through the formula for gauge transformations and one through this constraint, and the remaining $4 - 1 - 1 = 2$ is the number of polarisations of light Known (the quantum version is in 4-01).

Under the formula for gauge transformations the first two terms do not change, and the last two increase by just $\rho\,\partial_t\chi + \mathbf{J}\cdot\nabla\chi$. Integrating by parts ($\chi$ is 0 far away),

$$\delta S = -\int \chi\left(\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{J}\right) dt\, d^3x$$

That the action does not change for any $\chi$ is the same as the **conservation of charge** $\partial_t \rho + \nabla\cdot\mathbf{J} = 0$. Seen from the side of the equations, the two Maxwell equations with sources are not independent. Combining the time derivative of Gauss's law with the divergence of the Ampère–Maxwell law, the field terms vanish identically by $\nabla\cdot(\nabla\times\mathbf{B}) = 0$, and only $\partial_t\rho + \nabla\cdot\mathbf{J} = 0$ is left. If charge is conserved, part of the equations follows from the rest. This is the form of the second theorem of 2-01.

**Example: a mass term cannot be put in.** Following 2-02, adding $-\frac{\varepsilon_0\Omega^2}{2}|\mathbf{A}|^2$ makes the dispersion of transverse waves $\omega^2 = c^2|\mathbf{k}|^2 + \Omega^2$, with a lower bound $\Omega$. But this term changes under the formula for gauge transformations. Under gauge invariance, a mass term cannot be added directly to the action. For the electromagnetic field without interactions no lower bound arises, the mass of the photon is 0 in quantum theory too, and the free electromagnetic field has no mass gap Known .

Whether mass can arise from interactions while keeping gauge invariance is a separate problem. The Meissner effect in a superconductor and the Higgs mechanism (3-07) are examples Known . For Yang–Mills, the question is whether this happens without adding another field (section 09, 4-11).

05

## Quantum phase and U(1)

A classical particle feels only $\mathbf{E}, \mathbf{B}$, but in quantum mechanics the potentials enter the formulas. We use three things (2-05). (i) A state is represented by a complex-valued **wave function** $\psi(t, \mathbf{x})$, and $|\psi|^2$ is the probability density of position. (ii) For a constant $\alpha$, $\psi$ and $e^{i\alpha}\psi$ are the same state. (iii) A particle of mass $m$ and charge $q$ obeys the following Schrödinger equation ($\hbar = h/2\pi$).

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{1}{2m}\left(-i\hbar\nabla - q\mathbf{A}\right)^2\psi + q\phi\,\psi \quad \text{(S)}$$

(S) seems to break under the formula for gauge transformations, but if $\psi$ is also changed as

$$\psi \mapsto e^{iq\chi/\hbar}\,\psi \quad \text{(G')}$$

then (S) holds as it is. The key is the **covariant derivative** $D = \nabla - \frac{iq}{\hbar}\mathbf{A}$: the $\frac{iq}{\hbar}(\nabla\chi)\psi$ coming from the derivative of the product cancels the change in $\mathbf{A}$.

$$\Big(\nabla - \frac{iq}{\hbar}(\mathbf{A} + \nabla\chi)\Big)\big(e^{iq\chi/\hbar}\psi\big) = e^{iq\chi/\hbar}\Big(\nabla - \frac{iq}{\hbar}\mathbf{A}\Big)\psi$$

The time side works in the same way. Fock (1926) noticed this invariance, London (1927) connected it with the theory of scale (Eich) that Weyl had set up earlier, and Weyl (1928–29) raised it as a principle under the name of gauge invariance, it is said (Jackson–Okun 2001).

$e^{iq\chi/\hbar}$ is a complex number of absolute value 1, and the set of all of them forms a group under multiplication. This is **U(1)** (the unit circle; 1-02). **A gauge transformation is the choice of one element of U(1) at each point.** If the element is the same at every point, it is (ii) itself; if it differs from point to point, $\mathbf{A}$ makes things consistent (the argument in the opposite direction is in section 09).

### The phase factor along a path

If the reference is chosen point by point, how are phases at distant points compared? For a path $C$ from $\mathbf{x}$ to $\mathbf{y}$,

$$W(C) = \exp\Big(\frac{iq}{\hbar}\int_C \mathbf{A}\cdot d\mathbf{l}\Big), \qquad W(C) \mapsto e^{iq\chi(\mathbf{y})/\hbar}\, W(C)\, e^{-iq\chi(\mathbf{x})/\hbar}$$

Since it cancels the replacements at both ends, $W(C) \in U(1)$ is a tool that carries the phase from $\mathbf{x}$ to $\mathbf{y}$ (**parallel transport**, 1-07). For a closed path it is gauge invariant, and with the enclosed flux $\Phi_C$, $W(C) = e^{iq\Phi C/\hbar}$. On a lattice it is the product of the $e^{iqA_{ij}/\hbar}$ of the edges, and the addition of section 03 becomes multiplication in the group.

06

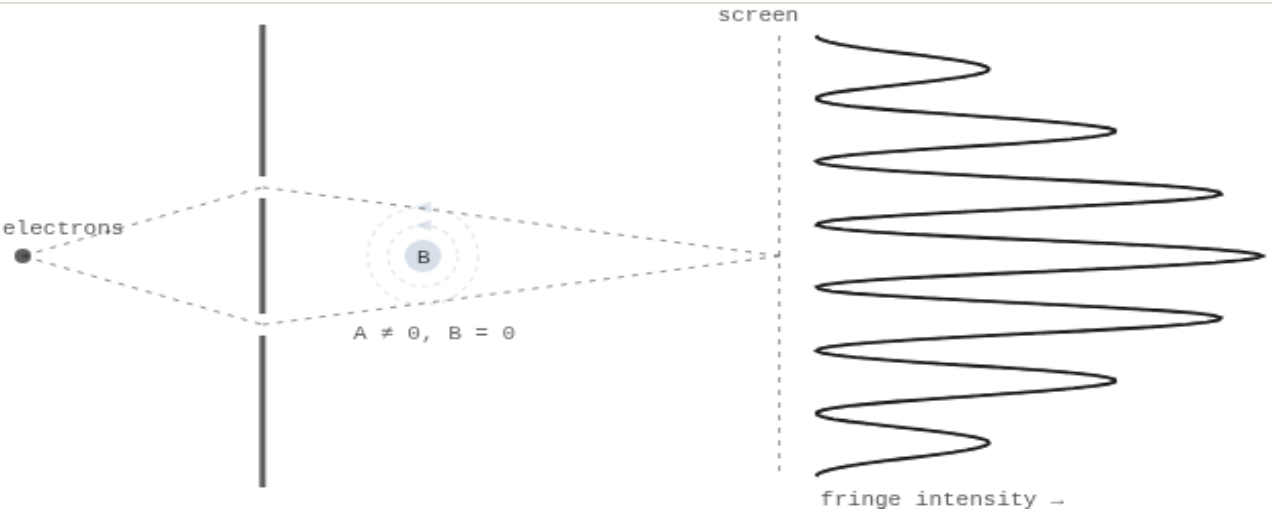
## The Aharonov–Bohm effect — the fringes move where there is no magnetic field

Consider a double slit in which electrons pass on either side of the solenoid of Example 2. Where the electrons pass, $\mathbf{B} = \mathbf{0}$, and classically nothing changes. In quantum mechanics the contributions of the path $C_1$ through slit 1 and the path $C_2$ through slit 2 are added, and the phase difference makes the fringes. Joining $C_1$ with the reversed $C_2$ goes once round the solenoid, so the phase difference due to $\mathbf{A}$ is

$$\Delta = \frac{q}{\hbar}\oint \mathbf{A}\cdot d\mathbf{l} = \frac{q\Phi}{\hbar} = \pm 2\pi\,\frac{\Phi}{\Phi_0}, \qquad \Phi_0 = \frac{h}{|q|}$$

and the fringes shift by $\Phi/\Phi_0$ of the fringe spacing. If $\Phi$ is an integer multiple of $\Phi_0$, the result cannot be told apart from the original. The prediction is Aharonov–Bohm (1959). Ehrenberg–Siday (1949) had reached the same conclusion earlier, and a clue from 1939 has also been pointed out Known (Hiley 2013).

Figure 2 Changing the flux $\Phi/\Phi_0$, or pressing Play, shifts the fringes (solid line on the right) away from the fringes for $\Phi = 0$ (dashed), and they coincide each time $\Phi/\Phi_0$ increases by 1. B = 0 on the paths of the electrons




Φ/Φ₀ = 0.00, phase difference = 0.00π, fringe shift = fringe spacing × 0.00

Computed A schematic in which the fringes $I(y) \propto \operatorname{sinc}^2(0.6y)\cos^2(3.5\pi y + \pi\Phi/\Phi_0)$ ($\operatorname{sinc} u = \sin(\pi u)/(\pi u)$) are computed on the spot (the number and width of the fringes are chosen for legibility).

**Why it cannot be removed by a gauge transformation.** Outside, $\mathbf{A} = \nabla(\Phi\theta/2\pi)$ ($\theta$ is the angle around the axis), so it looks removable by $\chi = -\Phi\theta/2\pi$, but $\theta$ increases by $2\pi$ in one circuit and is not determined as a single function. What needs to be single-valued is the phase $e^{iq\chi/\hbar} = e^{-iq\Phi\theta/2\pi\hbar}$, and that holds only when $\Phi/\Phi_0$ is an integer; only then does $\mathbf{A}$ vanish and the effect disappear. The period $\Phi_0$ comes from here.

Computed **Table 1** Numerical integration of $\oint \mathbf{A}\cdot d\mathbf{l}$ around a solenoid of radius 1 with $\Phi = 1$. The right two columns are the $\mathbf{A}$ of Example 2 and the $\mathbf{A}'$ obtained by transforming with $\chi = \sin x\cos 2y + x^2y/3$. The difference from enclosed flux × winding number is within $4\times10^{-10}$.

| Closed path | $\lvert\mathbf{B}\rvert$ on the path | $\oint \mathbf{A}\cdot d\mathbf{l}$ | $\oint \mathbf{A}'\cdot d\mathbf{l}$ |
|---|---|---|---|
| circle radius 2, centre (0, 0) | 0 | 1.000000 | 1.000000 |
| square side 5, centre (0.7, −0.4) | 0 | 1.000000 | 1.000000 |
| circle radius 1, centre (3, 0) (not enclosing) | 0 | 0.000000 | 0.000000 |
| circle radius 2, twice round | 0 | 2.000000 | 2.000000 |
| circle radius 0.5 (inside the solenoid) | 0.3183 | 0.250000 | 0.250000 |

Computed For electrons $\Phi_0 = h/e = 4.1357 \times 10^{-15}$ Wb; the magnetic field that sends this through a circle of diameter 1 μm is 5.27 mT. Known The first experimental test is said to be Chambers (1960), and Tonomura et al. (1986) measured it with the magnetic field confined in a superconductor.

07

## A particle on a ring — the levels are periodic functions of the flux

The shift of the fringes also appears in the levels. Consider a particle of charge $q > 0$ confined to a ring of radius $R$, with a flux $\Phi$ through the centre of the ring. On the ring $\mathbf{B} = \mathbf{0}$, $A_\theta = \Phi/2\pi R$, and $\nabla$ reduces to the $\frac{1}{R}\partial_\theta$ along the ring. Putting a **stationary state** $\psi = e^{-iEt/\hbar} e^{in\theta}$ (a state in which $|\psi|^2$ does not depend on time; since it returns after one circuit, $n$ is an integer) into (S), the left side becomes $E\psi$ and the right side $\frac{1}{2mR^2}(\hbar n - q\Phi/2\pi)^2\psi$, and the energy is

$$E_n = \frac{\hbar^2}{2mR^2}\left(n - \frac{\Phi}{\Phi_0}\right)^2, \qquad n \in \mathbb{Z}$$

For $\Phi \mapsto \Phi + \Phi_0$ the labels are merely renumbered as $n \mapsto n+1$, so **the set of levels is a periodic function with period** $\Phi_0$, and at $\Phi = \Phi_0/2$ the lowest two coincide. There is no magnetic field on the ring, and yet the levels know the flux. Byers–Yang (1961) is said to have discussed from this periodicity the quantisation of flux in a superconducting cylinder (period $h/2e$) Known .

From an examination of the AB effect, Wu–Yang (1975) summarised that the field strength describes electromagnetism too little, the phase $\frac{q}{\hbar}\oint \mathbf{A}\cdot d\mathbf{l}$ describes it too much, and the phase factor $e^{\frac{iq}{\hbar}\oint \mathbf{A}\cdot d\mathbf{l}}$ is what describes it neither more nor less Known . These levels are one example.

08

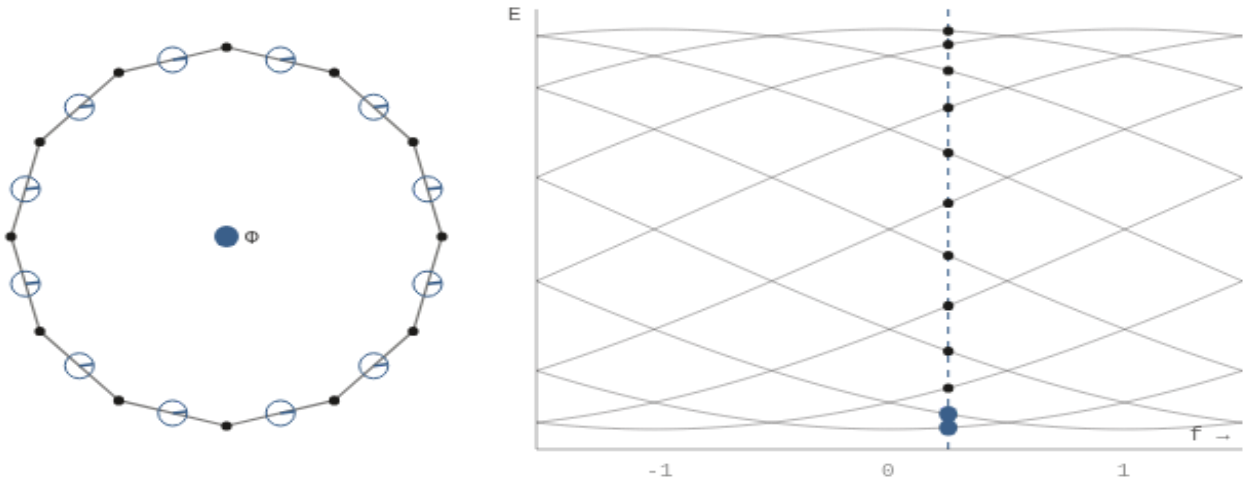
## A lattice ring — putting elements of U(1) on the edges

On a ring of $N$ points, consider the Hermitian matrix $H_{j+1,j} = -t\,e^{i\theta_j}$, $H_{j,j+1} = H_{j+1,j}$ (indices modulo $N$), in which the amplitude $-t$ for moving to a neighbour ($t > 0$) is multiplied by an element of U(1) for each edge. A gauge transformation is $\theta_j \mapsto \theta_j + \chi_{j+1} - \chi_j$, and the only thing that does not change is the product once round, $\prod_j e^{i\theta_j} = e^{2\pi i f}$ ($f = \Phi/\Phi_0$). For a uniform $\theta_j = 2\pi f/N$ the plane waves are the eigenvectors, and

$$E_n = -2t\cos\frac{2\pi(n-f)}{N}, \qquad n = 0, \dots, N-1$$

For $|n-f| \ll N$ this is $E_n \approx -2t + t(2\pi/N)^2(n-f)^2$, the same form as the continuous ring of section 07.

Figure 3 Changing f = $\Phi/\Phi_0$ and the number of points N (or pressing a row of Table 2) moves the levels (dots on the right). Pressing "Gauge" changes how the edge phases (the clocks on the left) are distributed, but not the levels




N = 12, f = 0.25, $E_0$ = -1.98289, $E_1$ = -1.84776, $E_1 - E_0$ = 0.13513, sum of phases once round ÷ 2π = 0.25 (mod 1)

Computed The levels are computed on the spot from the closed formula ($t = 1$).

Computed **Table 2** The lowest two levels (eigenvalues of the matrix) for $N = 12$, $t = 1$. The rightmost column is the maximum of the differences among three gauges (uniform / gathered on one edge / with a random $\chi$ added) and of the difference from the closed formula. Pressing a row moves to Figure 3.

| f | $E_0$ | $E_1$ | $E_1 - E_0$ | Max. difference |
|---|---|---|---|---|
| 0 | −2.000000 | −1.732051 | 0.267949 | 1.0e−15 |
| 0.25 | −1.982890 | −1.847759 | 0.135131 | 6.7e−16 |
| 0.5 | −1.931852 | −1.931852 | 0.000000 | 7.8e−16 |
| 0.75 | −1.982890 | −1.847759 | 0.135131 | 1.3e−15 |
| 1.25 | −1.982890 | −1.847759 | 0.135131 | 8.9e−16 |

The levels agree among the three gauges within rounding error, and are the same for $f$, $1-f$, $f+1$. For $N = 192$, $(E+2t)/t(2\pi/N)^2$ approaches the $(n-f)^2$ of the continuous ring, and for the lowest two levels the relative difference was at most $5.0 \times 10^{-5}$ ($E_1$ of $f = 0.25$) Computed .

This "gap" and the mass gap

$E_1 - E_0$ is a statement about a finite ring, and it approaches 0 as the ring is made larger. It is a different thing from the mass gap of the problem statement (4-11). What we want you to see is that the levels are determined by the phase factor once round alone. The observables of Yang–Mills on the lattice are also made from the phase factors of closed paths (Wilson loops, 5-01).

09

## From U(1) to SU(N) — where this book is going

Yang–Mills theory replaces the element chosen at each point, from U(1) to SU(N) (the $N \times N$ unitary matrices of determinant 1, 1-04) (3-02). Because the result depends on the order of multiplication, a commutator term is added to the field strength, the field itself carries colour charge (the counterpart of electric charge), and the equations become non-linear (3-03). Classical waves have no lower bound, as with light, but in the quantum theory a gap is thought to arise Physics (5-08). The difference from U(1) is in 5-06.

The next bridge is 3-01. This chapter put **A** first and found (G') afterwards. 3-01 goes the other way: it demands that (S) not change even if the phase is turned arbitrarily at each point, and derives **A** as the field that makes things consistent. Making the same demand with SU(N) gives the Yang–Mills field. This is the **gauge principle**, and the parallel transport of section 05 and the edge elements of section 08 are its parts.

## What this chapter can say

| Content | Status |
|---|---|
| Sections 01–05 (gauge transformations, constraints, conservation of charge, that a mass term cannot be put directly into the action, (G'), phase factors) | Said Known . The lattice is Figure 1 Computed |
| Sections 06–08 (the phase difference of the AB effect and the period $\Phi_0$, the periodicity and gauge invariance of the levels on the ring) | Said Known (as a comparison of phases along paths; exact solutions of scattering are not treated). Tables 1 and 2 Computed |
| That the free electromagnetic field has no mass gap / mechanisms by which mass arises while keeping gauge invariance, and the mass gap of Yang–Mills | The former is known. For the latter, examples are only named, and this chapter shows nothing |

## Sources and reproduction

| Item | Kind | Source / tools |
|---|---|---|
| Tables 1 and 2, the numbers in sections 06 and 07, Figures 1–3 | Computed | python3, NumPy (midpoint rule for line integrals, eigenvalues of Hermitian matrices). The figures by JavaScript in the browser (the formulas in the text) |
| Maxwell's equations, the history of gauge invariance (Fock, London, Weyl) | Known | J. D. Jackson, L. B. Okun, Rev. Mod. Phys. 73 (2001) 663, arXiv:hep-ph/0012061 (abstract and section III checked). J. C. Maxwell, Phil. Trans. R. Soc. Lond. 155 (1865) 459, V. Fock, Z. Phys. 39 (1926) 226, F. London, Z. Phys. 42 (1927) 375, H. Weyl, Gruppentheorie und Quantenmechanik (1928), Z. Phys. 56 (1929) 330 (bibliographic data only) |
| Examples in which mass arises while keeping gauge invariance | Known | P. W. Anderson, Phys. Rev. 130 (1963) 439 (title and abstract checked). For details see 3-07 |
| The prediction, history and experiments of the AB effect | Known | Y. Aharonov, D. Bohm, Phys. Rev. 115 (1959) 485, A. Tonomura et al., Phys. Rev. Lett. 56 (1986) 792 (abstracts checked for both). B. J. Hiley, arXiv:1304.4736 (text checked). W. Ehrenberg, R. E. Siday, Proc. Phys. Soc. B 62 (1949) 8 (bibliographic data only). R. G. Chambers, Phys. Rev. Lett. 5 (1960) 3 (bibliographic data only; title checked; "first" not checked) |
| Description by phase factors | Known | T. T. Wu, C. N. Yang, Phys. Rev. D 12 (1975) 3845 (conclusion of § II checked) |
| Periodicity in the flux and superconductivity | Known | N. Byers, C. N. Yang, Phys. Rev. Lett. 7 (1961) 46 (bibliographic data only; the subject checked from the title, content not checked) |

Next chapter: 2-04 Special relativity and spacetime. To go on to the gauge principle, 3-01.

# Special relativity and spacetime — preserving a quadratic form with one sign different narrows down the form of the action

> 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** — That Lorentz transformations are the transformations that preserve the form of the wave equation, and that the sign of the interval divides the relation between two points into three.
That mass is the invariant length of the four-momentum, and that the action of a free particle is the proper time.
That the form of the action of a field is narrowed down, and that the action of pure Yang–Mills in four dimensions has no length scale.

**Prerequisite chapters** — 2-01 (the action, Euler–Lagrange, Noether) and the product of matrices (1-01). The wave equation and the mass term of 2-02 are also used (the formulas are restated on the spot). The groups of section 04 anticipate 1-02, "why it is called the mass gap" in section 05 anticipates 2-05, the electromagnetic field of section 07 anticipates 2-03, and its Yang–Mills anticipates 1-03 and 1-04. On a first reading you can take only the conclusions and move on.

**Said first** — Only the classical story in flat spacetime without gravity (quantum theory starts in 2-05). The imaginary time of section 09 is a formal substitution, and its justification is the work of 4-02 and 4-10. Including "the classical action has no scale" of section 08, this chapter shows nothing about the mass gap.

Order of this chapter

1. The wave equation changes form under Galilean transformations — making the speed of light the same for everyone
2. Lorentz transformations — preserving the interval (Figure 1, Table 1)
3. The light cone and the relativity of simultaneity — three cases by sign
4. Four-vectors and the Lorentz group — indices, non-compactness

5. Proper time and four-momentum — mass is an invariant length (why a "mass" gap)
6. The action of a free particle — the straight line maximises proper time (Figure 2, Table 2)
7. Lorentz-invariant actions of fields — scalar, electromagnetic, Yang–Mills
8. Counting dimensions — in four dimensions the coupling is dimensionless (Tables 3 and 4)
9. A bridge to imaginary time — mass becomes a rate of decay

---

01

## The wave equation changes form under Galilean transformations

We write the wave equation built in 2-02 with one dimension of space and wave speed $c$.

$$\left(\frac{1}{c^2}\partial_t^2 - \partial_x^2\right)u = 0 \quad (1)$$

In Newtonian mechanics the coordinates of an observer moving to the right with speed $v$ are $t' = t,\ x' = x - vt$ (a Galilean transformation). Substituting $\partial_t = \partial_{t'} - v\partial_{x'}$, $\partial_x = \partial_{x'}$, the operator of (1) changes form.

$$\frac{1}{c^2}\partial_{t'}^2 - \frac{2v}{c^2}\partial_{t'}\partial_{x'} - \left(1 - \frac{v^2}{c^2}\right)\partial_{x'}^2$$

**Example.** Substituting $u = f(x' - st')$ gives $s^2 + 2vs - (c^2 - v^2) = 0$, that is, $s = c - v$ and $-(c + v)$. For $v = c/2$ the waves look like $0.5c$ to the right and $1.5c$ to the left Computed .

For sound this is correct, because the observer at rest relative to the air is special. But the wave equation of light comes from Maxwell's equations (2-03) and needs no medium. Einstein (1905) took as premises that the laws have the same form in every inertial frame and that the speed of light does not depend on the motion of the source Known . We turn the question round. **Which transformations do not change the form of (1)?**

02

## Lorentz transformations — the transformations that preserve the interval

For linear transformations, preserving the operator (1) as it is is equivalent to preserving the following **interval** (the two coefficient matrices change in mutually inverse ways; scaling up or down preserves equation (1) but not the interval).

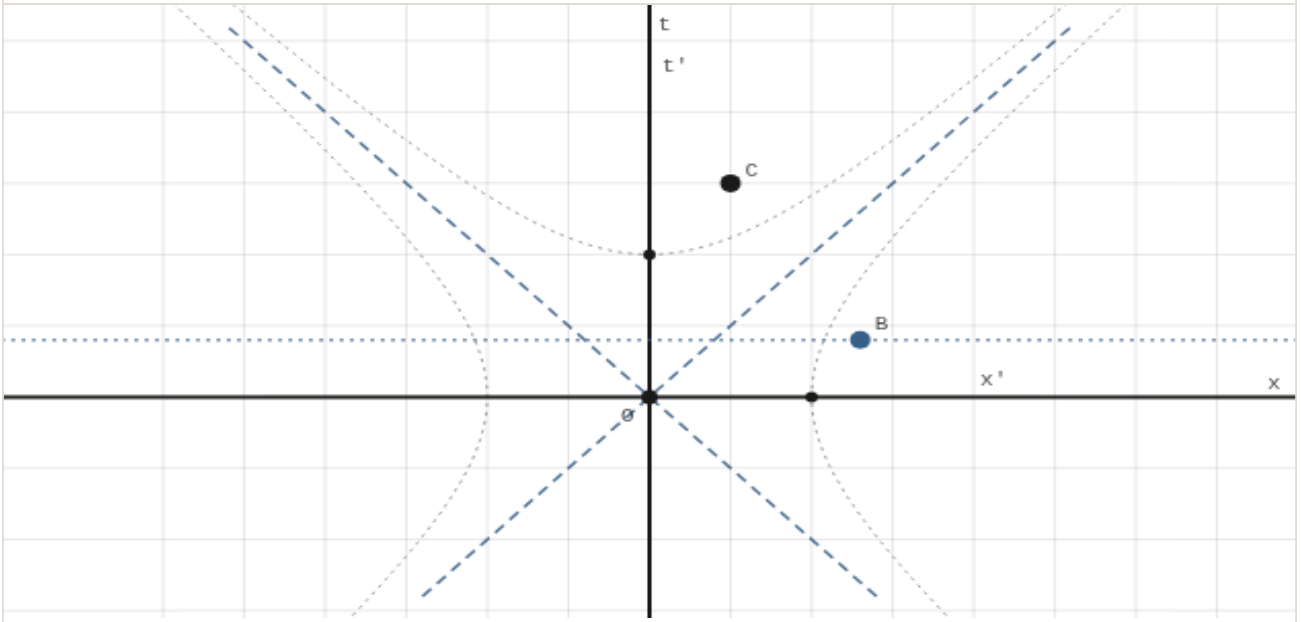
$$s^2 = -c^2t^2 + x^2 \quad (2)$$

Rotations preserve $x^2 + y^2$ and could be written as $\cos, \sin$. The transformations that preserve $s^2$ can be written, with $\cosh^2\varphi - \sinh^2\varphi = 1$, as $\cosh, \sinh$.

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh\varphi & -\sinh\varphi \\ -\sinh\varphi & \cosh\varphi \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix} \quad (3)$$

We call $\varphi$ the **rapidity** ($\tanh\varphi = v/c$), and this transformation a **boost**. Putting $\gamma = \cosh\varphi = 1/\sqrt{1 - v^2/c^2}$ gives $t' = \gamma(t - vx/c^2)$, $x' = \gamma(x - vt)$, and for $c \to \infty$ we return to the Galilean transformation ( § 3 of Einstein 1905 has the same form, except that the speed of light is written $V$) Known . The difference from rotations is in § 01 and § 04 of the spacetime article. **Example.** For $v = 0.5c$, $\gamma = 1.1547$. The time $t$ of a clock at $x = 0$ is $t' = \gamma t$ for the moving observer, and each tick looks longer.

Figure 1 Changing the rapidity ϕ tilts the axes t' and x' of the moving observer (solid) towards the light cone (dashed). The hyperbolas $s^2 = \pm 1$ (dotted) do not move. The coloured dotted line is the line of events simultaneous with B for the moving observer




φ = 0.00, v/c = 0.000, γ = 1.0000, B: t' = 0.400 (after O), C: t' = 1.500, s²: B 1.53, C −2 (invariant)

Computed The numbers in the figure are computed on the spot from (3). Pressing a row of Table 1 (or selecting it with Tab and pressing Enter) sets the figure to that speed.

Computed **Table 1** $\gamma$, rapidity, and the ratio of kinetic energies ($(\gamma-1)mc^2 \div mv^2/2$). $\gamma = \cosh\varphi$ is cross-checked.

| v/c | γ | φ | Energy ratio |
|---|---|---|---|
| 0.1 | 1.005038 | 0.1003 | 1.0076 |
| 0.5 | 1.154701 | 0.5493 | 1.2376 |
| 0.9 | 2.294157 | 1.4722 | 3.1955 |
| 0.999 | 22.366272 | 3.8002 | 42.8181 |

03

## The light cone and the relativity of simultaneity — three cases by the sign of the interval

The $s^2$ of the difference between two events does not change under boosts, and its **sign** is the same for whoever looks.

$$s^2 < 0:\ \text{timelike}, \quad s^2 = 0:\ \text{lightlike}, \quad s^2 > 0:\ \text{spacelike}$$

Two timelike-separated points can be connected by a signal slower than light, and under transformations that do not reverse the direction of time (section 04) their order is the same for whoever looks. Two spacelike-separated points cannot be connected by any signal, and **their order is exchanged depending on the observer**. The view of time and space as a single four-dimensional geometry became widely known through Minkowski's lecture of 1908. The treatment that introduces an imaginary time coordinate $t\sqrt{-1}$ and sees Lorentz transformations as rotations in four dimensions is already in Poincaré (1906) Known .

**Example (Figure 1).** $c = 1$. The origin O and B $= (t, x) = (0.4, 1.3)$ are spacelike-separated, with $s^2 = 1.53 > 0$. $t'_B = \gamma(0.4 - 1.3v)$ becomes negative at $v > 0.308$ ($\varphi > 0.318$), and B moves from "after" O to "before" it. C $= (1.5, 0.5)$ is timelike-separated, with $s^2 = -2$, and stays $t'_C = \gamma(1.5 - 0.5v) > 0$.

If two points whose order is exchanged were causally connected, there would be a contradiction. In quantum field theory this becomes the axiom "field operators at spacelike-separated points commute" (Jaffe–Witten § 3, 4-09). For the light cone and "the degeneration of distance", see § 03 of the spacetime article.

04

## Four-vectors and the Lorentz group — writing with indices makes the invariants visible

We return space to three dimensions and write $x^\mu = (ct, x, y, z)$. When the same index appears once up and once down, it is summed.

$$\eta = \mathrm{diag}(-1,1,1,1), \quad s^2 = \eta_{\mu\nu}\Delta x^\mu \Delta x^\nu \\ \Lambda^{\mathsf{T}}\eta\Lambda = \eta \tag{4}$$

To lower an index we multiply by $\eta$ ($x_\mu = (-ct, x, y, z)$). $\partial_\mu = \partial/\partial x^\mu$, $\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu = -\partial_t^2/c^2 + \nabla^2$. The signature $(-,+,+,+)$ is the same as in the spacetime article; some textbooks use the opposite convention.

A **Lorentz transformation** is a $4\times 4$ real matrix $\Lambda$ satisfying the right-hand formula of (4), and these form a group $\mathrm{O}(1,3)$ (1-02). $\mathrm{O}(1,3)$ splits into four parts according to $\det\Lambda = \pm 1$ and the direction of time (whether the top-left component is $\geq 1$ or $\leq -1$; by (4) its square is at least 1). The part containing the identity matrix, $\mathrm{SO}^+(1,3)$ ($\det\Lambda = 1$, preserving the direction of time), is generated by 3 rotations and boosts in 3 directions, and the group obtained by adding the 4 translations is the **Poincaré group** Known . The quantum field of the problem statement is placed on a Hilbert space carrying a representation of this group (Jaffe–Witten § 3).

**Example.** $\mathrm{SO}(4)$ is compact, but $\mathrm{SO}^+(1,3)$ is not, since $\varphi$ can be taken as large as we like. The "compact simple gauge group" of the problem statement concerns the internal symmetry (1-04); the Lorentz group of spacetime is non-compact from the start.

05

## Proper time and four-momentum — mass is a length that does not depend on the observer

We call the time of a clock carried by the particle its **proper time** $\tau$. From the invariance of $s^2$, $c^2 d\tau^2 = c^2 dt^2 - |d\underset{\sim}{x}|^2$, that is, $d\tau = dt/\gamma$. Since $\tau$ is the same for everyone, the **four-momentum** $p^\mu = m\,dx^\mu/d\tau$ changes by $\Lambda$, just as $x^\mu$ does.

$$p^\mu = \left(\frac{E}{c}, \underset{\sim}{p}\right) = (\gamma mc,\ \gamma m\underset{\sim}{v}) \\ \eta_{\mu\nu}p^\mu p^\nu = -m^2c^2 \tag{5}$$

The right-hand formula is $E^2 = |\underset{\sim}{p}|^2c^2 + m^2c^4$. Although $E$ and $\underset{\sim}{p}$ differ from observer to observer, **mass is the invariant length of the four-momentum**. **Example.** $E = \gamma mc^2$

$= mc^2 + \frac{1}{2}mv^2 + \frac{3}{8}mv^4/c^2 + \cdots$, and the rightmost column of Table 1 is this correction.

### Why it is called the "mass" gap

The dispersion relation $\omega^2 = c^2k^2 + \Omega^2$ of 2-02, read as $E = \hbar\omega,\ p = \hbar k$ (2-05), is (5) itself, with $\Omega = mc^2/\hbar$. The problem statement defines the mass gap as "the spectrum of the Hamiltonian $H$ (the set of values the energy can take; defined in 2-05 and 1-10) has no points in $(0, \Delta)$" (Jaffe–Witten § 4). When three things hold together, Lorentz invariance, positive energy ($H \geq 0$ holds for every observer) and a unique vacuum (Jaffe–Witten § 3), this is the same as "the pairs of energy and momentum $(E, \underline{p})$ of states other than the vacuum satisfy $E \geq \sqrt{|\underline{p}|^2c^2 + \Delta^2}$". For if a pair violates it, then if it is spacelike ($E < |\underline{p}|c$) a boost makes the energy negative, contradicting positive energy; if it is timelike, in its rest frame, and if it is lightlike, by a boost that makes it as small as we like, it enters $(0, \Delta)$. That only the vacuum has the pair $(0, \underline{0})$ uses the premise that the vacuum is unique Known . $\Delta$ can be read as the $mc^2$ of the lightest state (4-11).

06

## The action of a free particle — the straight path maximises proper time

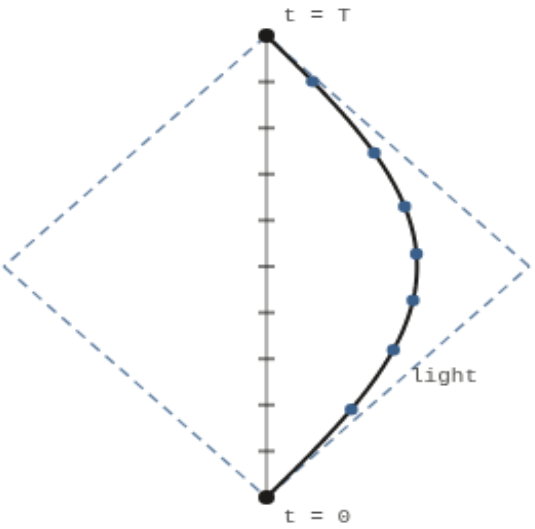
If the action (2-01) is the same for everyone, so is the form of the equation of motion. The simplest invariant of a world line is the proper time, so

$$S = -mc^2 \int d\tau$$
$$= -mc^2 \int \sqrt{1 - |\underline{v}|^2/c^2}\ dt \quad (6)$$

The coefficient was chosen so that the slow-motion expansion $L = -mc^2 + \frac{1}{2}m|\underline{v}|^2 + \cdots$ matches Newton. The equation of motion is $\frac{d}{dt}(\gamma m\underline{v}) = 0$, so the world line is straight, and the Noether conserved quantity of time translation (2-01) is $E = \gamma mc^2$. From the sign, **minimising the action and maximising the proper time are the same thing**, and among timelike world lines with fixed endpoints the straight line maximises the proper time. Seen in the frame of an observer for whom both ends are at the same place (such a frame exists because the ends are timelike-separated; call the time difference there $T$), $d\tau = \sqrt{1 - |\underline{v}|^2/c^2}\,dt \leq dt$, with equality only when $\underline{v} \equiv 0$. Hence $\tau \leq T$, and the maximum is attained only by the straight line ($\tau$ is the same in every frame). We call this the **reverse triangle inequality** Known . A stationary point need not be a minimum (2-01), but here it really is a minimum, and the twin paradox is a restatement of this.

**Example.** The proper time of $T = 1$, $x(t) = (u/\pi)\sin \pi t$ (top speed $u$) is $\tau = \frac{2}{\pi}\mathcal{E}(u)$ with $\mathcal{E}(k) = \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2\theta}\, d\theta$ (the complete elliptic integral of the second kind). For $u \to 1$, $2/\pi \approx 0.637$.

Figure 2 Changing the top speed u bends the world line with fixed endpoints, and the proper time (coloured dots: every 0.1 T) decreases. Pressing Play shows the two clocks at the same coordinate time




u = 0.90, proper time of the bent path 0.7459 T (the straight path is 1 T)

Computed The numbers in the figure are computed on the spot by the trapezoidal rule for $d\tau = dt/\gamma$ (400 subdivisions).

Computed **Table 2** $\tau/T$ of world lines with fixed endpoints. For the curve, the numerical integral and $\frac{2}{\pi}\mathcal{E}(u)$ agree to $2 \times 10^{-16}$. The broken line is the path that goes out and back at speed $u$ (the twins).

| u/c | Curve | Approx. 1 − u²/4 | Broken line √(1 − u²) |
|---:|---:|---:|---:|
| 0.3 | 0.977105 | 0.977500 | 0.953939 |
| 0.6 | 0.902780 | 0.910000 | 0.800000 |
| 0.9 | 0.745926 | 0.797500 | 0.435890 |
| 0.99 | 0.654748 | 0.754975 | 0.141067 |

07

## Lorentz-invariant actions of fields — scalar fields, the electromagnetic field, Yang–Mills fields

The action of a field is $S = \frac{1}{c}\int \mathcal{L}\, d^4x$. Since $d^4x$ does not change under $|\det \Lambda| = 1$, the action is invariant if the Lagrangian density $\mathcal{L}$ is a scalar. If we also impose the condition of being **local** (only the field at that point and its first derivatives), the form of the terms is strongly narrowed down.

For a **scalar field**, $\partial_\mu\phi$ is contracted with $\eta$.

$$\mathcal{L} = -\tfrac{1}{2}\eta^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi - \tfrac{1}{2}\mu^2\phi^2$$
$$(\Box - \mu^2)\phi = 0, \quad \mu = \frac{mc}{\hbar} \qquad (7)$$

The right-hand one is the Klein–Gordon equation. **Example.** In one dimension, multiplying the whole by the tension $\sigma$ gives the string with a mass term of 2-02, with the correspondence $c^2 = \sigma/\rho$, $\Omega = c\mu$.

**The electromagnetic field.** Gather the potentials $\phi, \boldsymbol{A}$ of 2-03 (this $\phi$ is different from the field of (7)) as $A_\mu = (-\phi/c, \boldsymbol{A})$, and put $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The components are $F_{0i} = -E_i/c$, $F_{ij} = \varepsilon_{ijk}B_k$ ($\varepsilon_{ijk}$ is the symbol with $\varepsilon_{123} = 1$ that changes sign when indices are exchanged), and it gathers the electric and magnetic fields into one antisymmetric matrix. It does not change under the gauge transformation $A_\mu \to A_\mu + \partial_\mu\chi$ (the formula for gauge transformations in 2-03). Since $F^{0i} = -F_{0i}$, $F^{ij} = F_{ij}$, we have $F_{\mu\nu}F^{\mu\nu} = -2|\boldsymbol{E}|^2/c^2 + 2|\boldsymbol{B}|^2$, and

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} = \tfrac{1}{2}\left(|\boldsymbol{E}|^2/c^2 - |\boldsymbol{B}|^2\right) \qquad (8)$$

In SI units the whole is multiplied by $1/\mu_0$, and this becomes the part of (L) of 2-03 without $\rho, \boldsymbol{J}$. The equation of motion $\partial_\mu F^{\mu\nu} = 0$ is half of Maxwell's equations in vacuum, and the other half holds as soon as $F$ is built from $A$.

**Yang–Mills fields.** Make $A_\mu = A^a_\mu T^a$ a field with values in a Lie algebra (1-03, 1-04), and add a term with the structure constants $f^{abc}$ to $F$ (3-02).

$$\mathcal{L} = -\frac{1}{4g^2}\sum_a F^a_{\mu\nu}F^{a\,\mu\nu}$$
$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu \qquad (9)$$

In matrices, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]$ ($T^a$ Hermitian with $[T^a, T^b] = if^{abc}T^c$, the convention of 1-03; $g$ absorbed into $A$). The last term is the only difference from Maxwell, and the field becomes its own source. Even so, the classical equations describe "massless waves travelling at the speed of light" (Jaffe–Witten § 1) Known .

If we also impose gauge invariance (3-02) and the group is simple (the invariant quadratic form is unique up to a constant multiple; 1-03), the Lorentz-invariant quadratic expressions in $F$ are only the linear combinations of $\sum_a F^a_{\mu\nu}F^{a\,\mu\nu}$ and $\sum_a \varepsilon^{\mu\nu\rho\sigma}F^a_{\mu\nu}F^a_{\rho\sigma}$ (the four-dimensional version of the symbol $\varepsilon^{0123} = 1$) Known . For the electromagnetic field the second is proportional to $\boldsymbol{E}\cdot\boldsymbol{B}$; it changes sign under spatial reflection and, being a total derivative, does not change the equations of motion (the $\theta$

term of 3-05). Without the condition, the $\sum c_{ab} F^a_{\mu\nu} F^{b\,\mu\nu}$ built from an arbitrary symmetric matrix $c_{ab}$ is also invariant.

08

## Counting dimensions — ħ = c = 1, and the dimension of the coupling constant

With $c = 1$, length and time are put on the same footing, and with $\hbar = 1$ energy and inverse time, so one unit remains. If mass has dimension 1, length has $-1$. The only conversion is $\hbar c = 197.3269804\ldots$ MeV fm ($1\,\text{fm} = 10^{-15}\,\text{m}$). The gap $\Delta$ corresponds to a length $\hbar c/\Delta$, a rough measure of how far correlations reach. If there is a gap, then for a local quantity $O$ whose vacuum expectation value $\langle O\rangle$ is 0, the correlation $|\langle O(\underline{x})O(\underline{y})\rangle|$ at two points at the same time and far enough apart is at most $\exp(-C|\underline{x}-\underline{y}|)$ for any $C$ satisfying $C < \Delta/(\hbar c)$ (Jaffe–Witten § 5; for $\hbar = c = 1$, $C < \Delta$). If the expectation value is not 0, the correlation approaches $\langle O\rangle^2$.

Computed **Table 3** Converting masses into lengths ($\hbar c$ computed from the defined SI values $h, c, e$).

| Item | mc² (MeV) | ħ/(mc) (fm) |
|---|---|---|
| Electron | 0.51099895069 | 386.159267 |
| Proton | 938.27208943 | 0.210309 |

The action is dimensionless (measured in $\hbar$), so in a spacetime of dimension $d$, $[\mathcal{L}] = d$, $[\partial] = 1$. Since $(\partial\phi)^2$ and $\lambda\phi^4$ have dimension $d$, and since $F$ of (9) is the sum of $\partial A$ and $AA$ with $[A] = 1$, $[F^2] = 4$,

$$[\phi] = \frac{d-2}{2}, \quad [\lambda] = [g^2] = 4 - d \tag{10}$$

Computed **Table 4** Mass dimensions ($\hbar = c = 1$, computed with fractions).

| d | [φ] | [λ] | [g²] | Reading |
|---|---|---|---|---|
| 2 | 0 | 2 | 2 | coupling has dimension mass² |
| 3 | 1/2 | 1 | 1 | coupling has dimension mass |
| 4 | 1 | 0 | 0 | coupling is dimensionless |
| 5 | 3/2 | −1 | −1 | coupling has dimension length |

**In four dimensions, the action (9) contains not a single constant with dimensions.** Neither a mass nor a length can be built from the coefficients, and classical Yang–Mills theory has no scale Known . If in the quantum theory there is $\Delta > 0$, the scale must come from outside the classical action; in physics it is thought to enter in the process of renormalisation, and this is called dimensional transmutation (4-07) Physics . The

exposition too gives the fact that in four-dimensional gauge theory the mass gap is not classically visible ('the mass gap is not classically visible') as a reason why a non-classical change of variables may be needed (Jaffe–Witten § 6.6). In two and three dimensions $g^2$ itself is a scale (4-04, 7-06). For $\phi^4$ see 7-08; for other ways four dimensions is special, see the four-dimensions article. A figure that moves the $d$ of Table 4 continuously and sets it beside other quantities that switch at four is the dimension dial of 1-12 § 10.

09

## A bridge to imaginary time — putting one sign back, mass becomes a rate of decay

Substituting $t = -i\tau$ ($\tau$ real) into (2) gives $-c^2t^2 = c^2\tau^2$, and the interval becomes the square of the Euclidean distance $c^2\tau^2 + |\underline{x}|^2$. This is the **Wick rotation** (the name is said to come from Wick 1954), which formally puts back the difference of sign; it is the same operation as the $ict$ of § 02 of the spacetime article (the authoritative treatment of its form for actions and correlations of fields is 4-02 § 02). Since $\partial_t^2 = -\partial_\tau^2$, $\Box$ becomes the four-dimensional Laplacian, (7) becomes $(-\partial_\tau^2/c^2 - \nabla^2 + \mu^2)\phi = 0$, and oscillation turns into decay. The one-dimensional version, dropping the dependence on space and putting $c = 1$, is (11). $\delta(\tau)$ is the thing concentrated at $\tau = 0$ with $\int f(\tau)\delta(\tau)\,d\tau = f(0)$ (strictly a distribution rather than a function; 1-11).

$$\left(-\frac{d^2}{d\tau^2} + \mu^2\right) G = \delta(\tau)$$
$$G(\tau) = \frac{e^{-\mu|\tau|}}{2\mu} \qquad (11)$$

**Example.** For $\tau \neq 0$, $G'' = \mu^2 G$; at $\tau = 0$ the slope $G'$ jumps by $-1$, so $G''$ contains $-\delta$, and $-G'' + \mu^2 G = \delta$. In three dimensions, $e^{-\mu r}/(4\pi r)$ Known . **On the Euclidean side, mass appears as a rate of decay (an inverse length).** This is the starting point of 4-11, which ties the mass gap to the correlation length. Writing the Euclidean theory as a probability measure (4-02) and the conditions for reconstructing the Minkowski side (Osterwalder–Schrader) are the subject of 4-10. This chapter has only shown the substitution.

## What this chapter can say

| Content | Status |
|---|---|
| Lorentz transformations, the three classes of intervals, four-momentum, the action of a free particle and the maximum of proper time, the forms of the actions of three kinds of fields | Said Known (numbers also checked in Tables 1 and 2 Computed ) |
| That under Lorentz invariance, positive energy and a unique vacuum, the definition of the gap becomes the same as $E \geq \sqrt{\lvert p\rvert^2 c^2 + \Delta^2}$ | Said Known (section 05; the axiomatic framework is in 4-09 and 4-11) |
| That the classical action of pure Yang–Mills in four dimensions has no constant with dimensions | Said Known (Table 4 Computed ) |
| That the scale of the gap comes from renormalisation (dimensional transmutation) | At the level of physics Physics (4-07) |
| Justification of the Wick rotation, curved spacetime, spin | Not treated |
| The mass gap of Yang–Mills | This chapter shows nothing |

## Sources and reproduction

| Item | Kind | Source / tools |
|---|---|---|
| The coefficients of section 01, Tables 1–4, Figures 1–2 | Computed | Tables by python3 (NumPy, SciPy). Figures by JavaScript in the browser ((3) and the trapezoidal rule). The static images shown when JavaScript is disabled are SVGs made by python3 from the same formulas |
| The two premises and Lorentz transformations | Known | A. Einstein, “Zur Elektrodynamik bewegter Körper”, Ann. Phys. 17 (1905) 891–921. Introduction and § 3 checked (the speed of light is written V) |
| The view as the geometry of spacetime | Known | H. Minkowski, “Raum und Zeit”, Phys. Z. 10 (1909) 104–111 (lecture in Cologne on 21 September 1908). Opening checked |
| The imaginary time coordinate, seeing Lorentz transformations as rotations in four dimensions | Known | H. Poincaré, “Sur la dynamique de l’électron”, Rend. Circ. Mat. Palermo 21 (1906) 129–176. Bibliographic data only (only the relevant passage checked in the Wikisource transcription) |
| The definition of the gap, the Poincaré group, positive energy and uniqueness of the vacuum, locality, the cluster property, classical waves | Known | A. Jaffe, E. Witten, “Quantum Yang–Mills theory” (the Clay Mathematics Institute's problem description). § 1, § 3– § 5 and § 6.6 checked |
| The name of the Wick rotation | Known | G. C. Wick, Phys. Rev. 96 (1954) 1124–1134. Bibliographic data only (attribution not checked) |

| Item | Kind | Source / tools |
|---|---|---|
| ħc, the masses of the electron and the proton | Known | CODATA 2022 (checked at NIST). ħc agrees with the calculation from the defined SI values |
| The maximum of proper time, the four parts of $\mathrm{O}(1,3)$, the two invariants of $F$ (under gauge invariance and a simple group), the components of the electromagnetic field, counting dimensions | Known | Standard facts (can be checked with the formulas in the text) |

Next chapter: 2-05 The bones of quantum mechanics. Imaginary time continues in 4-02, and spacetime itself in One sign is different.

# The bones of quantum mechanics — states are vectors, energies are a spectrum, and the gap is the space at the very bottom

> 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** — That a state is a unit vector, an observable is a self-adjoint operator, and the values of the energy are a spectrum.
That the gap $E_1 - E_0$ can be read in three ways: as "the price of an excitation", "the speed of oscillation" and "the rate of decay in imaginary time".
Gaps that close and gaps that remain, a gap invisible to perturbation theory, and the agreement with the Poincaré constant.

**Prerequisite chapters** — 1-10 (Hilbert spaces, spectra, the variational principle, the Poincaré inequality). For the case of matrices, 1-01.

**Said first** — Only quantum mechanics with finitely many degrees of freedom is treated. The mass gap of quantum field theory is treated in 4-11, and this chapter shows nothing about the mass gap of Yang–Mills.

Order of this chapter

1. States and observables — states are vectors, measured quantities are operators
2. Time evolution — the gap between levels sets the speed of the back-and-forth
3. The spectrum — values that are not eigenvalues also enter (Detour: a well whose valleys go on for ever)
4. The ground state — trial functions give only upper bounds (Table 1)
5. Three readings of the gap — price, oscillation, decay (Figure 1, Table 2)
6. Gaps that close, gaps that remain — comparing as the box is widened (Table 3)
7. The double well — a gap invisible to perturbation theory (Figure 2, Table 4)
8. The gap and the Poincaré inequality — the gap is the inverse of a weighted constant

01

## States and observables — vectors and self-adjoint operators

In quantum mechanics a state is a vector $\psi$ of length 1 in a **complex Hilbert space** $\mathcal{H}$ (1-10). A sum $\alpha\psi + \beta\phi$ rescaled to length 1 is also a state (**superposition**), and $z\psi$, multiplied by $z$ with $|z| = 1$, is the same state (the phase of 2-03). The inner product is taken to be conjugate-linear in the left argument.

$$\mathcal{H} = \mathbb{C}^2 : \ |\alpha|^2 + |\beta|^2 = 1; \qquad \mathcal{H} = L^2(\mathbb{R}) : \ \int_{\mathbb{R}} |\psi(x)|^2 \, dx = 1 \quad (1)$$

On the left is a system with two choices (spin, or "the left valley or the right valley"); on the right is a particle on a line, and we read $|\psi(x)|^2$ as the probability density of position (this is said to go back to Born 1926). Measurable quantities correspond to **self-adjoint operators** $A$ (Hermitian matrices in finite dimensions, 1-01); the value of a measurement is one of the eigenvalues $a_k$, and its probability, using orthonormal eigenvectors $e_k$, is $|\langle e_k, \psi\rangle|^2$.

$$\langle A \rangle_\psi = \sum_k a_k \, |\langle e_k, \psi\rangle|^2 = \langle \psi, A\psi \rangle \quad (2)$$

**Example.** Measuring $\sigma_z = \mathrm{diag}(1, -1)$ in $\psi = (1, 1)/\sqrt{2}$ gives $\pm 1$ half and half. For $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_x \psi = \psi$, and $\psi$ is an eigenvector with eigenvalue $+1$, so a measurement certainly gives $+1$ ($\sigma_x, \sigma_z$ are the Pauli matrices, 1-04). In $L^2(\mathbb{R})$ the position $X$ is the operator of multiplication by $x$, the momentum is $P = -i\hbar \, d/dx$, and from $XP - PX = i\hbar$ comes the uncertainty relation $\Delta X \, \Delta P \geq \hbar/2$ Known .

02

## Time evolution — stationary states and the back-and-forth of two levels

We call the operator of energy the **Hamiltonian** $H$. For a particle on a line it is $H\psi = -\frac{\hbar^2}{2m}\psi'' + V\psi$. States obey the Schrödinger equation.

$$i\hbar \frac{d\psi}{dt} = H\psi, \qquad \psi(t) = e^{-itH/\hbar}\psi(0) = \sum_n c_n \, e^{-iE_n t/\hbar} \, \phi_n \quad (3)$$

The right-hand end is the expansion in eigenvectors $H\phi_n = E_n \phi_n$, and each component only has its phase turning. $e^{-itH/\hbar}$ is unitary and preserves total probability (1-01; the converse is Stone's theorem Known ). A state with a single

eigenvector is a **stationary state** whose distribution does not change, and change arises only from the phase differences $(E_n - E_m)t/\hbar$.

**Example (two levels).** A system with the left valley $e_L$, the right valley $e_R$ and the tunnelling amplitude $g > 0$.

$$H = \begin{pmatrix} \varepsilon & -g \\ -g & \varepsilon \end{pmatrix}, \qquad E_{\mp} = \varepsilon \mp g, \qquad P_R(t) = \sin^2 \frac{gt}{\hbar} \quad (4)$$

Starting from $e_L$, the probability $P_R$ of being on the right goes back and forth with angular frequency $2g/\hbar = (E_+ - E_-)/\hbar$. **The gap between levels sets the speed of the back-and-forth** (the $x = 0$ of Figure 2 of 1-01). Section 07 gives an example where $g$ is exponentially small.

03

## The spectrum — eigenvalues alone are not enough

In infinite dimensions, values without eigenvectors also appear, so we use the **spectrum** $\sigma(H)$ (the set of $\lambda$ for which $H - \lambda$ has no bounded inverse, 1-10). For a self-adjoint operator, $\lambda \in \sigma(H)$ is the same as "there is a sequence with $\|\psi_k\| = 1$, $\|(H - \lambda)\psi_k\| \to 0$" (Weyl's criterion) Known .

$$\text{box } [0, L]: \ E_n = \frac{\hbar^2 \pi^2 n^2}{2mL^2} \ (n \geq 1); \qquad \text{line}: \ \sigma\Big(\frac{P^2}{2m}\Big) = [0, \infty) \quad (5)$$

The spectrum of a box with walls where $\psi = 0$ consists only of eigenvalues (**discrete spectrum**). On a line the plane wave $e^{ipx/\hbar}$ is not a state, since $\int |\psi|^2 = \infty$. **Example.** For a wave packet $\ell^{-1/2}\chi(x/\ell)e^{ipx/\hbar}$ of width $\ell$ ($\chi$ a smooth bump), $\|(H - p^2/2m)\psi\|$ approaches 0 like $1/\ell$, so every $E \geq 0$ is in the spectrum without being an eigenvalue (**continuous spectrum**). Widening the box makes the spacing on the left tighter, approaching the right (section 06).

### Detour: a well whose valleys go on for ever

A subsection you may skip.

For $V = \frac{1}{2}x^2y^2$ ($\hbar = m = 1$) on the plane, valleys with $V = 0$ run on for ever along the coordinate axes, and a classical particle can escape as far as it likes. In quantum mechanics the valley narrows the further in it goes, and the lowest energy $|x|/2$ of the oscillator in the $y$ direction (2-06) acts as a wall. Strictly, this is an estimate that works only deep in the valleys. Computed Solving on the square $[-R, R]^2$, the lowest eigenvalue stops at $0.5671,\ 0.55413,\ 0.55411$ for $R = 3, 5, 9$, while the value for a free particle $\pi^2/4R^2$ goes to 0. For a uniform $A_1 = x\sigma_1,\ A_2$

$= y\sigma_2$ field SU(2), the potential energy $-\operatorname{tr}[A_1, A_2]^2$ of Yang–Mills is proportional to $x^2y^2$ (3-02 § 07). What the numbers on this machine show goes only as far as the tendency of the lowest eigenvalue to stop as the box is widened. The exposition of the problem (Jaffe–Witten § 6.6) cites Simon (1983, titled "quantum operators with discrete spectrum but classically continuous spectrum") as an example in which bound states arise even when there are flat directions Known .

04

## The ground state — estimates from above are easy

We call $E_0 = \inf \sigma(H)$ the **ground-state energy**, and, if it is an eigenvalue, its eigenvector $\phi_0$ the **ground state**. The variational principle (1-10) says the following.

$$E_0 = \inf_{\psi \neq 0} \frac{\langle \psi, H\psi \rangle}{\langle \psi, \psi \rangle} \tag{6}$$

Every trial function gives an **upper bound**, but a lower bound requires control of every $\psi$ (this asymmetry appears throughout the book, so it is summarised in the box in 1-10 § 06). Three remarks. (i) The bottom need not be an eigenvalue (the free particle on a line). (ii) It does not fall to the bottom of the potential. For $V = \tfrac{1}{2}m\omega^2x^2$, $\langle X^2\rangle\langle P^2\rangle \geq \hbar^2/4$ and the inequality of arithmetic and geometric means give $\langle H\rangle \geq \hbar\omega/2$ (2-06). (iii) In one dimension, if $V$ grows far away, the ground state is non-degenerate and can be taken positive, and the $n$-th eigenfunction has $n$ zeros (Sturm's oscillation theorem) Known .

**Example.** Putting a Gaussian $e^{-x^2/4s^2}$ into $\hbar = m = 1$, $H = -\tfrac{1}{2}d^2/dx^2 + x^4$ gives $\langle H\rangle_s = 1/8s^2 + 3s^4$, and its minimum $\tfrac{3}{16}48^{1/3} = 0.68142$ is an upper bound for $E_0 = 0.667986$. The odd trial function $xe^{-x^2/4s^2}$ is orthogonal to the even ground state, so it gives an upper bound for $E_1$.

Computed **Table 1** The best upper bounds from Gaussian trial functions, and the eigenvalues of three-point differences.

| V | Upper bound on $E_0$ | $E_0$ | Upper bound on $E_1$ | $E_1$ | Difference of upper bound |
|---|---|---|---|---|---|
| $x^4$ | 0.681420 | 0.667986 | 2.423739 | 2.393644 | 1.74231 |

**The difference between two upper bounds is neither an upper nor a lower bound on the gap** (depending on the trial functions it can be off in either direction). To bound it from below, we need estimates for every state orthogonal to the ground state. Showing the mass gap in quantum field theory is likewise, by definition, bounding from below the energy of the states orthogonal to the vacuum (4-11).

05

## Three readings of the gap — price, oscillation, decay

When the ground state is non-degenerate and $E_0$ is an isolated point of the spectrum, we call $\Delta = \inf(\sigma(H) \setminus \{E_0\}) - E_0 > 0$ the **gap**. The exposition of the problem sets the energy of the vacuum to 0 and defines the mass gap as "the spectrum of $H$ has no points in $(0, \Delta)$" (Jaffe–Witten § 4) Known . There are three readings. (i) **Price**: precisely, for states orthogonal to $\phi_0$, $\langle H \rangle \geq E_0 + \Delta$. (ii) **Oscillation**: a superposition of $\phi_0$ and $\phi_1$ oscillates with angular frequency $\Delta/\hbar$ ((4)). (iii) **Decay**: formally putting $t = -i\tau$ into (3) (2-04), oscillation turns into decay.

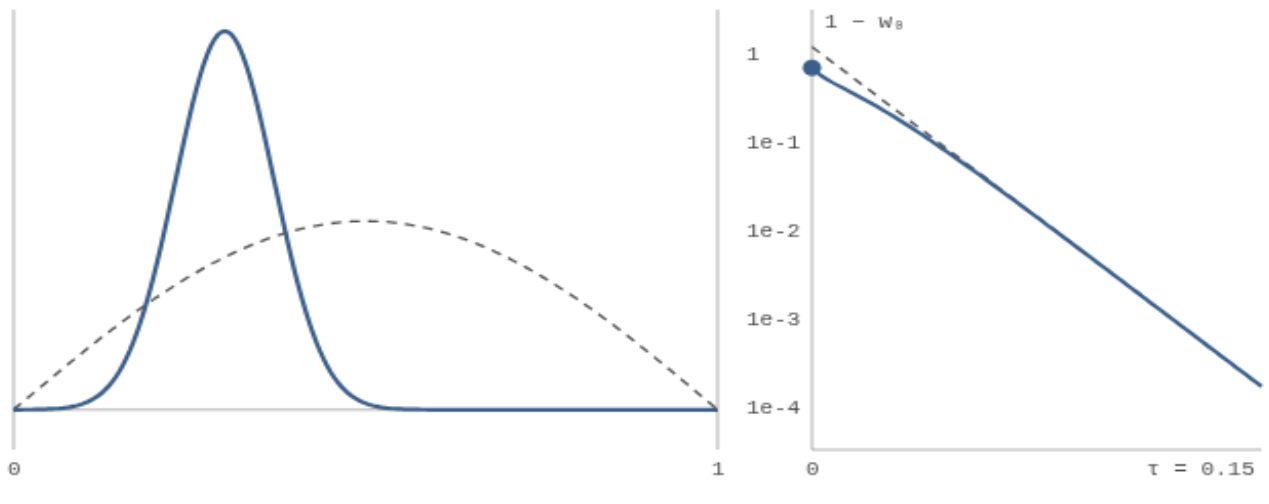
$$e^{-\tau(H-E_0)/\hbar}\psi = c_0\phi_0 + \sum_{n\geq 1} c_n\, e^{-(E_n-E_0)\tau/\hbar}\phi_n \quad (7)$$

For $c_0 \neq 0$ it approaches the ground state, and the rest becomes at most $e^{-\Delta\tau/\hbar}$ times as large. Write the left side of (7) as $\psi(\tau)$, and the **weight** of the ground state as $w_0(\tau) = |\langle\phi_0, \psi(\tau)\rangle|^2/\|\psi(\tau)\|^2$. $1 - w_0$ is the probability of being outside the ground state, and Figure 1 and Table 2 follow it. For the **imaginary-time correlation** of a self-adjoint $A$ ($\langle\phi_0, A\phi_0\rangle = 0$) it takes the following form.

$$0 \leq \langle A\phi_0,\, e^{-\tau(H-E_0)/\hbar}A\phi_0\rangle \leq e^{-\Delta\tau/\hbar}\,\|A\phi_0\|^2 \quad (8)$$

From (8), the rough time for the correlation to vanish in the imaginary-time direction is at most $\hbar/\Delta$. This is a quantity with the dimension of time; in relativity it is multiplied by $c$ to give a length $\hbar c/\Delta$. Decay in the spatial direction (the cluster property) and the correlation length (2-08) are treated in 4-11, and the view as a sum over paths in 2-07.

Figure 1 As imaginary time τ advances, the bump centred at x = 0.3 (colour) approaches the ground state $\phi_0$ (dashed). On the right is $1 - w_0$ (common logarithmic scale), which lines up with the prediction keeping only the next level (dashed; slope $-2\Delta$ in natural logarithms, $\hbar = 1$)



```
τ = 0.000, w₀ = 0.30949, 1 − w₀ = 6.91e-1, prediction 1.20e+0
```

Computed The numbers in the figure were computed on this machine (box $[0,1]$, $H=-d^2/dx^2$ ($\hbar=1$, $\hbar^2/2m=1$), $\Delta=3\pi^2$; the coefficients were found with python3, and the partial sums of (7) are computed on the spot).

Computed **Table 2** The weight outside the ground state $1-w_0$, and the prediction keeping only the next level $(c_1/c_0)^2 e^{-2\Delta\tau}$.

| τ | 1 − w₀ | Prediction |
|---|---|---|
| 0 | 6.905e-1 | 1.195e+0 |
| 0.05 | 5.831e-2 | 6.189e-2 |
| 0.10 | 3.194e-3 | 3.204e-3 |
| 0.15 | 1.659e-4 | 1.659e-4 |

The weight is the square of an amplitude, so the exponent is $2\Delta$. For a state symmetric about the centre, $c_1=0$, and the decrease is set by $E_2-E_0=8\pi^2$. **What the gap guarantees is the slowest decrease, valid for every state.**

06

## Gaps that close, gaps that remain — widening the box

The mass gap is asked about in infinitely wide space. A gap that is positive in a finite box can close when the box is widened. The exposition of the problem also says that a **uniform** gap, independent of the volume, could play a role in proving the infinite-volume limit, and that the proof would need new ideas (Jaffe–Witten § 5, § 6.5) Known . With $\hbar=m=1$ and box $[-L/2, L/2]$, we compare three $V$.

Computed **Table 3** The length of the box $L$ and the gap $E_1-E_0$ (three-point differences; $V=0$ agrees with the exact value $3\pi^2/2L^2$).

| L | V = 0 | V = −sech²x | V = x²/2 |
|---|---|---|---|
| 10 | 0.14804 | 0.57518 | 1.00000 |
| 40 | 0.00925 | 0.50342 | 1.00000 |
| 160 | 0.00058 | 0.50020 | 1.00000 |

The gap of the free particle closes like $1/L^2$. $-\operatorname{sech}^2 x$ has exactly one bound state $\operatorname{sech} x/\sqrt{2}$ ($E_0=-1/2$, which can be checked by substitution), and the next level drops to the bottom 0 of the continuous spectrum, so the gap tends to $1/2$ Known . **Above the gap there may be continuous spectrum.** Even in a field theory with a mass gap, above the vacuum is the continuous spectrum of particles, starting from $\Delta$ (2-04).

**Example.** For $N$ independent oscillators, the ground energy $N\hbar\omega/2$ grows with $N$ and diverges as $N\to\infty$. Even so, the gap stays $\hbar\omega$ (2-06). In quantum field theory the

energy of the vacuum is set to 0 and the difference is what is asked about. Jaffe–Witten § 4 states $H\Omega = 0$ from the fact that the vacuum vector $\Omega$ is invariant under the Poincaré group (2-04).

07

## The double well — a gap invisible to perturbation theory

Put two identical wells side by side and make the hill between them higher ($\hbar = m = 1$).

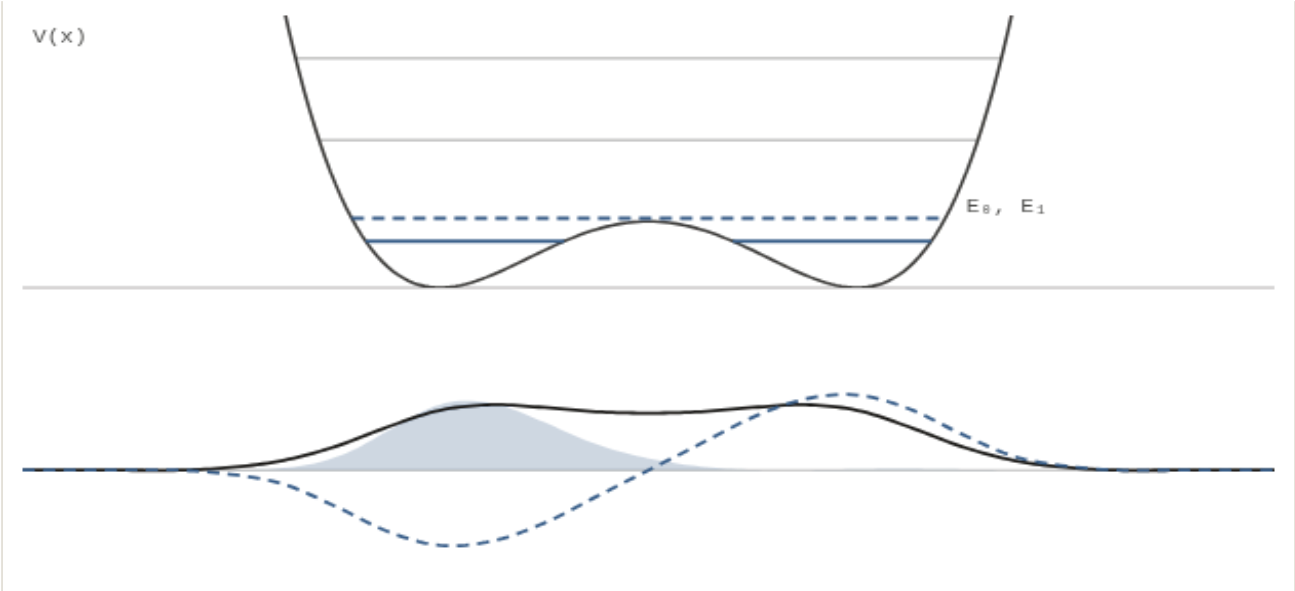
$$V_a(x) = \frac{(x^2 - a^2)^2}{8a^2}, \qquad V_a''(\pm a) = 1, \qquad S_0 = \int_{-a}^{a} \sqrt{2V_a}\, dx = \frac{2a^2}{3} \quad (9)$$

With a single valley the lowest level is about $1/2$, and with two valleys it is doubly degenerate. Tunnelling between the valleys plays the role of the $g$ of (4), the levels split into an even $\phi_0$ and an odd $\phi_1$, and when the hill is high the splitting is as follows Known .

$$E_1 - E_0 \approx \frac{4a}{\sqrt{\pi}}\, e^{-2a^2/3} = 2\sqrt{\frac{6S_0}{\pi}}\, e^{-S_0} \quad (10)$$

This is equation (2.8) of Zinn-Justin–Jentschura (2004), $\frac{2}{\sqrt{\pi g}} e^{-1/6g}$, with $g = 1/4a^2$ (shifting the potential of that paper by $a$ gives (9); the height of the hill is $a^2/8 = 1/32g$). **Perturbation theory** splits $V_a$ around one of the valleys into an oscillator term and a remainder that is small when $g$ is small, expands the levels in powers of $g$, and finds the coefficients of each order in turn (4-03). That perturbation series does not depend on parity (evenness or oddness under $x \to -x$) and is the same for the two levels (the explanation of (2.13)–(2.14) in that paper) Known . The $e^{-1/6g}$ of the gap, expanded as a power series in $g$, has all coefficients 0. **However high the order to which perturbation theory is computed, this gap does not appear.** $S_0$ is the action of the classical solution (instanton) that rolls from hilltop to hilltop of $-V_a$ in imaginary time. Its counterpart in Yang–Mills is in 3-05, and quantities of the form $e^{-1/g^2}$ appear in 4-07 and 4-12.

Figure 2 Increasing a moves the valleys (±a) apart and raises the hill, and the levels $E_0$ (coloured solid line) and $E_1$ (coloured dashed line) draw together and coincide. The lower panel shows $\phi_0$ (solid), $\phi_1$ (dashed), and the probability density of a state starting from the left valley (shaded). "Play" moves it to the right valley (time shown compressed)




```
a = 2, S₀ = 2.667, E₀ = 0.35024, E₁ = 0.52319, Δ = 0.1729 ((10) gives 0.3136), π/Δ = 18.17,
probability of being on the right at t = 0.000: 0.041
```

Computed The numbers in the figure were computed on this machine (levels and wave functions by three-point differences in python3; the shading is $\frac{1}{2}(\phi_0^2 + \phi_1^2 - 2\phi_0\phi_1 \cos \Delta t)$).

Computed **Table 4** Comparison of the gap of (9) (three-point differences) with (10). The rightmost column is the correction factor $1 - \frac{71}{12}g - \frac{6299}{288}g^2$ from the coefficient (equation (2.19a)) of the expansion (2.13) that the same paper states as a conjecture Physics .

| a | Δ (numerical) | Formula (10) | Ratio | Correction factor |
|---|---|---|---|---|
| 2 | 1.729e-1 | 3.136e-1 | 0.551 | 0.545 |
| 3 | 1.355e-2 | 1.678e-2 | 0.807 | 0.819 |
| 4 | 1.895e-4 | 2.104e-4 | 0.901 | 0.902 |
| 5 | 6.118e-7 | 6.519e-7 | 0.9384 | 0.9386 |

The ratio approaches 1, and also the correction factor (difference $2 \times 10^{-4}$ at $a = 5$). The gap at $a = 5$ is 6 orders of magnitude smaller than the level spacing (about 1), and the time $\pi/\Delta$ to move to the right is about $5 \times 10^6$.

08

## The gap and the Poincaré inequality — weighting by the ground state

In one dimension when $\phi_0 > 0$ (section 04), writing $\psi = f\phi_0$ with a real-valued $f$ and integrating by parts, the $V - E_0$ term vanishes by $H\phi_0 = E_0\phi_0$.

$$\langle \psi, (H - E_0)\psi \rangle = \frac{\hbar^2}{2m} \int |f'|^2 \, d\mu, \qquad d\mu = \phi_0^2 \, dx \tag{11}$$

$\psi \perp \phi_0$ means $\int f\,d\mu = 0$, and then $\|\psi\|^2 = \mathrm{Var}_\mu(f)$. Applying the variational principle to $\psi$ orthogonal to $\phi_0$, the following holds with the best constant.

$$\mathrm{Var}_\mu(f) \le C_P \int |f'|^2\,d\mu, \qquad C_P = \frac{\hbar^2}{2m\Delta} \quad (12)$$

**The gap is inversely proportional to the Poincaré constant of the ground-state distribution** $\mu$ Known (the Poincaré inequality is in 1-10 § 07, the weighted form in § 08). Here $\mu$ is the distribution of the ground state at a single time. Matching with $\psi = f\phi_0$, $H - E_0$ corresponds to $\frac{\hbar^2}{2m}L$ ($L$ is the operator of 1-10 § 08 for this $\mu$), and the gap of $L$ is $2m\Delta/\hbar^2$. The $C_P \approx \xi^2$ of 1-10 § 09 is a statement about the measure on the whole of the paths in imaginary time (1-08 (10)), and is a different quantity. **Example.** For the oscillator ($\hbar = m = \omega = 1$), $\phi_0 = \pi^{-1/4}e^{-x^2/2}$ (which can be checked by substitution), so $\mu$ is the normal distribution with variance $1/2$, and the gap 1 (the right column of Table 3, 2-06) gives $C_P = 1/2$. "The Poincaré constant of a normal distribution equals its variance" has the same content as the gap $\hbar\omega$ of the oscillator (with equality at $f = x$). For the double well $a = 3$, $C_P = 1/2\Delta = 36.9$. Taking $f$ close to "$-1$ on the left, $+1$ on the right", the variance is about 1, but the slope acts only on the hill between, where $\mu$ is almost absent. **A waist in the distribution makes the gap small.** The same form is also used in one-link integration (5-11, this site's papers).

---

## What this chapter can say

| What is said | What is not said (and where it is treated) |
|---|---|
| Known The framework of states, spectra and the variational principle, the levels and gaps of Tables 1–3, the decrease in imaginary time (numbers in Tables 1–3 and Figure 1 Computed ) | Stone's theorem, Weyl's criterion and Sturm's oscillation theorem are cited, not proved |
| Computed The tendency of the lowest eigenvalue of $\frac{1}{2}x^2y^2$ to stop as the box is widened to $R = 3, 5, 9$ | A proof that it stops (discrete spectrum is the subject of Simon (1983); its text has not been read) |
| Known The leading term (10) of the double-well gap, and that it appears at no order of the perturbation series (the tendency of the ratio towards 1 is Table 4 Computed ) | The derivation of (10). The expansion (2.13) including the correction coefficients is a conjecture of that paper Physics (instantons are in 3-05, divergent series in 4-12) |

| What is said | What is not said (and where it is treated) |
|---|---|
| Known That the gap is inversely proportional to the Poincaré constant of the ground-state distribution (one dimension, $\phi 0 > 0$) | The justification of (11) with infinitely many variables. The relation to $C_P \approx \xi^2$ of the path measure (1-10, 23-01) |
| — | The mass gap with infinitely many degrees of freedom and the requirements of relativity (4-11). This chapter shows nothing about the mass gap of Yang–Mills |

## Sources and reproduction

| Item | Kind | Source / tools |
|---|---|---|
| Tables 1–4, the numbers in the detour, Figures 1–2 | Computed | python3 (three-point differences, sparse matrices and numerical integration in NumPy and SciPy). The figures by JavaScript in the browser |
| The definition of the mass gap, $H\Omega = 0$, the uniform gap, the remark on flat directions | Known | A. Jaffe, E. Witten, "Quantum Yang–Mills Theory" (the official problem statement of the Clay Mathematics Institute; the problem was announced in 2000). In: J. Carlson, A. Jaffe, A. Wiles (eds.), *The Millennium Prize Problems*, CMI / AMS, pp. 129–152 (year of publication of the book not checked). § 4, § 5, § 6.5 and § 6.6 checked |
| The leading term (10) of the double-well gap, the independence of the perturbation series from parity | Known | J. Zinn-Justin, U. D. Jentschura, Ann. Phys. 313 (2004) 197–267, arXiv:quant-ph/0501136. Equations (2.1)–(2.8) of § 2.1 and the explanation of (2.13)–(2.14) checked |
| The correction factor of Table 4 | Physics | The coefficient (2.19a) of the expansion (2.13) that the same paper states as a conjecture (cited there without proof). On this machine the tendency was checked numerically Computed |
| The Schrödinger equation | Known | E. Schrödinger, "Quantisierung als Eigenwertproblem", Ann. Phys. 384 (1926) 361–376 (first communication; the time-independent eigenvalue problem), and the fourth communication of the same title, Ann. Phys. 386 (1926) 109–139. The time-dependent form (3) is said to be in the fourth communication (not checked). Bibliographic data only for both |
| The probabilistic interpretation / Stone's theorem / flat directions and discrete spectrum | Known | M. Born, Z. Phys. 37 (1926) 863–867 (attribution from secondary sources) / M. H. Stone, Ann. of Math. 33 (1932) 643–648 / B. Simon, "Some quantum operators with discrete spectrum but classically continuous spectrum", Ann. Phys. 146 (1983) 209–220 (reference [43] of Jaffe–Witten; the passage treating $x^2 y^2$ not checked). Bibliographic data only for all |

| Item | Kind | Source / tools |
|---|---|---|
| The variational principle, the Poincaré inequality (weighted form), the Poincaré constant of the normal distribution | Known | 1-10 § 06, § 07, § 08 |
| Weyl's criterion, Sturm's oscillation theorem, the uncertainty relation, (11) | Known | Standard facts (not in 1-10; no references given) |

Next chapter: 2-06 The harmonic oscillator and creation and annihilation. Imaginary time continues in 2-07, and the main treatment of the definition of the gap is in 4-11.

# The harmonic oscillator and creation and annihilation — the levels are equally spaced, and the quanta of a field are counted as particles

> 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** — That the levels $\hbar\omega(n+\frac{1}{2})$ are determined by the algebra of creation and annihilation alone.
That the gap sets the speed of oscillation and of the decay of correlations, and that with two valleys it is invisible to perturbation theory.
That quantising a field makes particles appear, and that the gap of a free field of mass $m$ is $m$.

**Prerequisite chapters** — 2-05 (states, spectra, the gap, the double well), 2-02 (normal modes, the mass term), 2-04 (the $E^2 = p^2 + m^2$ used in section 08, the imaginary time used in section 04), 1-03 (commutators). The ladder of 1-04 appears only for comparison.

**Said first** — Only single-particle quantum mechanics and fields without interactions are treated. The gap $m$ of section 08 is a value put in by hand, and is a different thing from the mass gap of Yang–Mills (4-11). This chapter shows nothing about the mass gap.

Order of this chapter

1. Why the oscillator — near an equilibrium point everything is quadratic
2. Creation and annihilation — algebra determines the levels
3. The ground state and zero-point energy — a Gaussian (Figure 1)
4. Time evolution — the gap is the speed of oscillation, and in imaginary time the rate at which correlations decrease (Figure 2)
5. Distorting the oscillator — the gap moves (Table 1)
6. Detour: two valleys — seen from the side of the perturbation series (Figure 3, Table 2)

7. Many oscillators and Fock space — counting by occupation numbers
8. Quantising a field — particles appear, and the gap is m (Figure 4, Table 3)
9. The sum of zero-point energies, and a caution about infinitely many — only differences can be measured

---

01

## Why the oscillator — near an equilibrium point everything is quadratic

Quantum field theory begins as a collection of harmonic oscillators. Near an equilibrium point $x_0$ any potential energy can be approximated as $V(x_0) + \frac{1}{2}V''(x_0)(x - x_0)^2$, and the string too was a collection of independent simple oscillations (2-02). The position $x$ and the momentum $p = -i\hbar\, d/dx$ are self-adjoint operators on $L^2(\mathbb{R})$, with $[x, p] = i\hbar$ (2-05).

$$H = \frac{p^2}{2m} + \frac{m\omega^2}{2}x^2 \quad (1)$$

Measuring length in $\ell = \sqrt{\hbar/m\omega}$ and energy in $\hbar\omega$, and putting $\xi = x/\ell$, $\eta = p\ell/\hbar$, we get $H = \frac{\hbar\omega}{2}(\xi^2 + \eta^2)$, $[\xi, \eta] = i$. **Example.** Computed For a spring with 1 g and 1 Hz, $\ell \approx 1.3 \times 10^{-16}$ m and $\hbar\omega \approx 6.6 \times 10^{-34}$ J, a value 13 orders of magnitude smaller than $k_B T \approx 4.1 \times 10^{-21}$ J at room temperature.

02

## Creation and annihilation — algebra determines the levels

The eigenvalues of (1) can be found by algebra alone, without solving a differential equation ($a^\dagger$ is the adjoint of $a$).

$$a = \frac{\xi + i\eta}{\sqrt{2}}, \qquad a^\dagger = \frac{\xi - i\eta}{\sqrt{2}} \quad (2)$$

Since $[a, a^\dagger] = 1$, $a^\dagger a = \frac{1}{2}(\xi^2 + \eta^2 - 1)$,

$$H = \hbar\omega\Big(N + \frac{1}{2}\Big), \qquad N = a^\dagger a \quad (3)$$

From $[N, a] = -a$, $[N, a^\dagger] = a^\dagger$: if $N\psi = \nu\psi$, then $N(a\psi) = (\nu - 1)a\psi$, $N(a^\dagger\psi) = (\nu + 1)a^\dagger\psi$; $a$ lowers by 1 and $a^\dagger$ raises by 1. Since $\|a\psi\|^2 = \langle\psi, N\psi\rangle = \nu\|\psi\|^2 \geq 0$, if $\nu$ is not an

integer we keep going down and reach a contradiction; if it is an integer we stop at $a\psi_0 = 0$ with $\nu = 0$. Going up never stops, by $\|a^\dagger\psi\|^2 = (\nu+1)\|\psi\|^2 > 0$.

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right), \qquad n = 0, 1, 2, \ldots \tag{4}$$

Known The levels are **equally spaced**, and the gap is $\hbar\omega$. Normalising, $a^\dagger\psi_n = \sqrt{n+1}\,\psi_{n+1}$, $a\psi_n = \sqrt{n}\,\psi_{n-1}$. This is the same argument as the ladder of 1-04, but that was finite-dimensional, stopping at both ends. No finite matrices satisfy $[a, a^\dagger] = 1$ (the trace, 0 against the dimension, does not match), so **this algebra is realised only in infinite dimensions.**

03

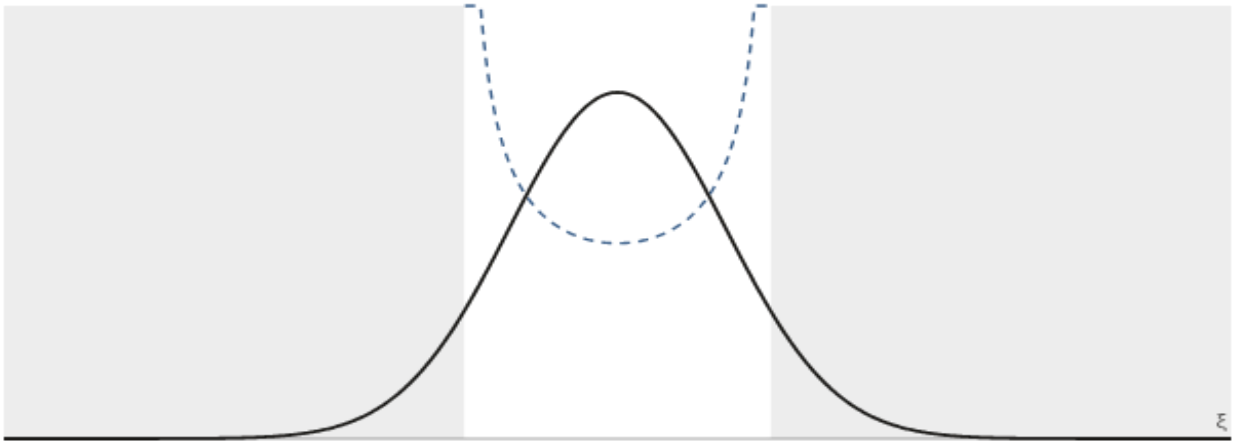
## The ground state and zero-point energy — a Gaussian

In coordinates $a\psi_0 = 0$ is the first-order equation $(\xi + \frac{d}{d\xi})\psi_0 = 0$, and its solution is unique up to a constant multiple (no degeneracy).

$$\psi_0(\xi) = \pi^{-1/4} e^{-\xi^2/2}, \quad \psi_n = \frac{(a^\dagger)^n}{\sqrt{n!}}\,\psi_0 \tag{5}$$

The ground state is a **Gaussian**, and $\langle x^2\rangle = \hbar/2m\omega$, $\Delta x\,\Delta p = \hbar/2$ is exactly the lower limit of the uncertainty relation. We call $E_0 = \hbar\omega/2$ the **zero-point energy**; it cannot be made 0 because position and momentum cannot both be brought towards 0 at the same time. Precisely, $\langle H\rangle \geq \omega\sqrt{\langle p^2\rangle\langle x^2\rangle} \geq \omega\,\Delta x\,\Delta p \geq \hbar\omega/2$. $\psi_n$ is the product of a Hermite polynomial and $e^{-\xi^2/2}$.

Figure 1 Changing n changes the probability density $|\psi_n|^2$ (solid) and the fraction of time spent by a classical particle of the same energy (dashed). Grey is the region a classical particle cannot enter




n = 0, E = 0.5 ħω, ⟨ξ²⟩ = 0.5, probability in the forbidden region = 0.1573

Computed The curves are computed on the spot from the recurrence $\psi_{n+1} = \sqrt{\frac{2}{n+1}}\,\xi\psi_n - \sqrt{\frac{n}{n+1}}\,\psi_{n-1}$ (which follows from (2) and the ladder relations) and the classical density $1/(\pi\sqrt{2n+1-\xi^2})$.

Computed The probability of being in the forbidden region, found by numerical integration, is 0.1573, 0.1116, 0.0951, 0.0740, 0.0602, 0.0425 for $n = 0, 1, 2, 5, 10, 30$ ($n = 0$ is $\operatorname{erfc}(1) = \frac{2}{\sqrt{\pi}}\int_1^\infty e^{-t^2}\,dt$). For large $n$, $|\psi_n|^2$ with its wiggles averaged out approaches the classical density.

04

## Time evolution — the gap is the speed of oscillation

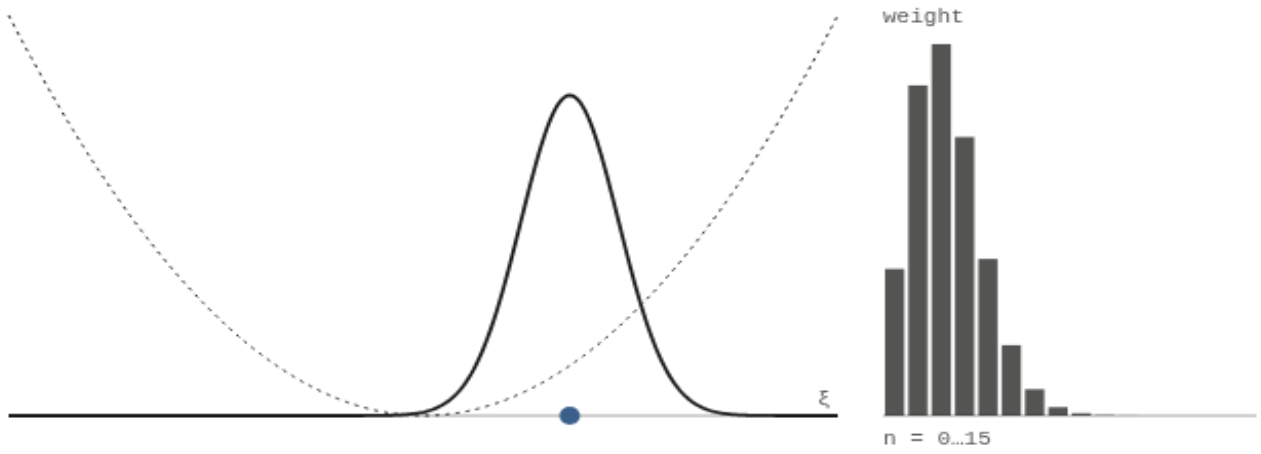
An eigenstate only has its phase turning, and its density does not move. For $\psi = c_0\psi_0 + c_1\psi_1$, with $\Delta = E_1 - E_0$ (the diagonal terms are 0 by parity),

$$\langle x\rangle(t) = 2\,\mathrm{Re}\big(c_0 c_1 e^{-i\Delta t/\hbar}\big)\langle\psi_0, x\psi_1\rangle \quad (6)$$

**The gap is the speed at which observable quantities oscillate**, and for the oscillator it is the classical $\omega$. The weights of a **coherent state** $\psi_\alpha = e^{-|\alpha|^2/2}\sum_n \frac{\alpha^n}{\sqrt{n!}}\psi_n$ ($a\psi_\alpha = \alpha\psi_\alpha$) follow a Poisson distribution with mean $|\alpha|^2$, and if $\alpha$ is real the density keeps its shape and goes back and forth along the classical orbit.

$$|\psi_\alpha(\xi,t)|^2 = \pi^{-1/2}e^{-(\xi-\sqrt{2}\alpha\cos\omega t)^2} \quad (7)$$

Figure 2 Changing α and pressing "Play", the density of the coherent state (solid) goes back and forth with the same period as the classical particle (dot). The dotted line is the potential energy, and the bars on the right are the weights of the levels n




α = 1.5, mean number of quanta = 2.25, E = 2.75 ħω

Computed Computed on the spot from formula (7) and the Poisson distribution.

### In imaginary time, the gap is the rate at which correlations decrease

We compute (8) of 2-05 with the $A = x$ of the oscillator. From (2), $x = \ell\,(a + a^\dagger)/\sqrt{2}$, so $x\psi_0 = \frac{\ell}{\sqrt{2}}\psi_1$, and there are no other components.

$$G(\tau) = \langle x\psi_0,\ e^{-\tau(H-E_0)/\hbar}\,x\psi_0\rangle = \frac{\ell^2}{2}\,e^{-\omega\tau} = \frac{\hbar}{2m\omega}\,e^{-\omega\tau} \quad (8)$$

The rate of decrease is the gap $(E_1 - E_0)/\hbar = \omega$, and (8) of 2-05 is an equality for the oscillator. For a general $V$ it becomes $G(\tau) = \sum_{n\geq 1} |\langle \psi_n, x\psi_0\rangle|^2 e^{-(E_n-E_0)\tau/\hbar}$, and the term $n = 1$ remains at $\tau \to \infty$ (when $\langle \psi_1, x\psi_0\rangle \ /= 0$). **Correlations in imaginary time decrease at the rate of the gap.** Measuring glueball masses on the lattice (5-08) and restating the mass gap as the decay of correlations (4-11) both have this form (for the meaning of $t = -i\tau$ see 2-04; for the path integral, 2-07).

05

## Distorting the oscillator — the gap moves

We add a quartic term ($\hbar = m = \omega = 1$).

$$H = \frac{p^2}{2} + \frac{x^2}{2} + \lambda x^4 \quad (9)$$

$x^4 = \frac{1}{4}(a + a^\dagger)^4$ moves $n$ by up to 4, so the ladder no longer closes. At first order of perturbation (replacing by the expectation value), $E_n \approx n + \frac{1}{2} + \frac{3\lambda}{4}(2n^2 + 2n + 1)$, and the gap is $1 + 3\lambda$.

Computed **Table 1** The levels of formula (9) found by finite differences (the change on halving the step is within $3 \times 10^{-6}$).

| λ | Gap $E_1-E_0$ | First-order estimate 1+3λ | $E_2-E_1$ |
|---|---|---|---|
| 0 | 1.0000 | 1 | 1.0000 |
| 0.01 | 1.0284 | 1.03 | 1.0552 |
| 0.1 | 1.2104 | 1.3 | 1.3691 |
| 1 | 1.9341 | 4 | 2.4414 |
| 10 | 3.8166 | 31 | 5.0254 |

The estimate fits at $\lambda = 0.01$ and fails at $\lambda = 1$. **The spacing widens further up**, and the gap stays positive. Known In one dimension, if $V$ is continuous and $V \to \infty$ as $|x| \to \infty$, the spectrum consists only of eigenvalues (discrete spectrum, 1-10), and no eigenvalue is degenerate. Why there is no degeneracy can be said briefly. For eigenfunctions $\psi, \chi$ with the same $E$, the Wronskian $W = \psi\chi' - \psi'\chi$ is constant by the

equation, $W' = 0$, and is 0 far away, so it is identically 0; that is, $\psi$ and $\chi$ are proportional. The proof of discreteness is in the textbook in the table of sources (for the ground state see section 04 of 2-05).

06

## Detour: two valleys — seen from the side of the perturbation series

A section you may skip.

We look again at the double well of 2-05 from the side of the perturbation series of the oscillator ($\hbar = m = 1$).

$$V(x) = \frac{1}{2}x^2\left(1 - \sqrt{g}\,x\right)^2 = \frac{x^2}{2} - \sqrt{g}\,x^3 + \frac{g}{2}\,x^4 \quad (10)$$

This is (9) of 2-05 with $g = 1/4a^2$, shifted by $a$ ($V(x) = V_a(x - a)$; the valleys are at 0 and $1/\sqrt{g}$), and has the form of equation (2.4) of Zinn-Justin–Jentschura (2004). The bottom of the left valley is the oscillator of section 01, and the remainder $-\sqrt{g}\,x^3 + \frac{g}{2}x^4$ is expanded as a perturbation in powers of $g$, as in section 05. Starting from the right valley via the reflection $x \mapsto 1/\sqrt{g} - x$ gives the same computation, and the lowest two levels both have the following series as their asymptotic expansion Known (equation (2.6) of that paper).

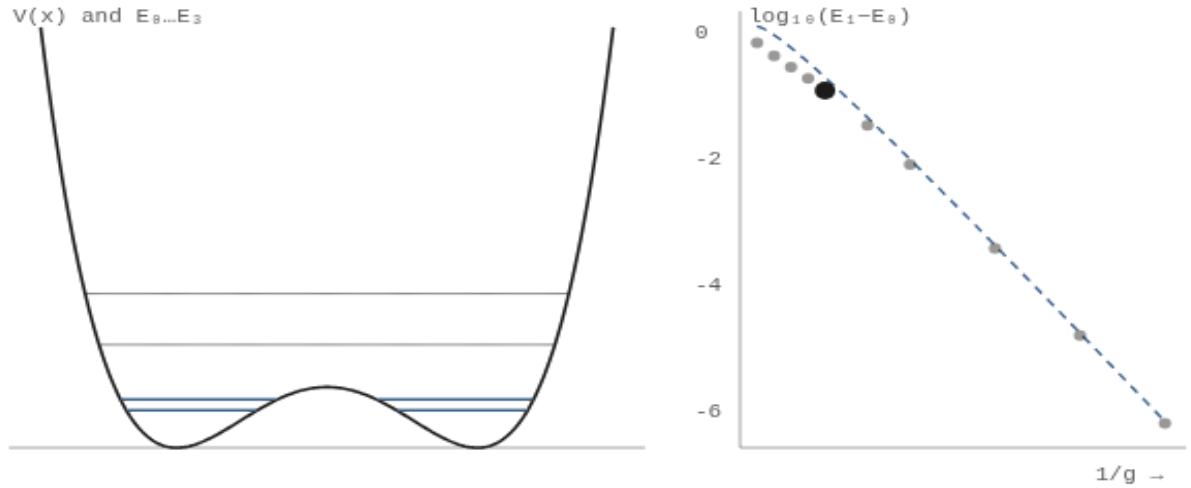
$$E \sim \frac{1}{2} - g - \frac{9}{2}g^2 - \frac{89}{2}g^3 - \frac{5013}{8}g^4 - \cdots$$

The coefficients grow factorially, and **at every order the gap is 0**. The gap lies outside the series; it is the following formula, (10) of 2-05 written in $g$ (equation (2.8) of that paper).

$$E_1 - E_0 \approx \frac{2}{\sqrt{\pi g}}\,e^{-1/(6g)} \quad (11)$$

The $\frac{1}{6g}$ in the exponent is the $S_0 = 2a^2/3$ of 2-05, and is said to be the action of the path that goes from valley to valley by passing under the hill. Precisely, it is the action of the classical solution (instanton, 3-05) that rolls from hilltop to hilltop of $-V$ in imaginary time.

Figure 3 Changing 1/g changes the two valleys (left) and the lowest four levels. On the right is the common logarithm of the gap $E_1 - E_0$ (dots: numerical, dashed: formula (11)). As the valleys move apart the gap shrinks exponentially




```
1/g = 20, height of hill = 0.625, E₀ = 0.38510, E₁ = 0.49649, gap =1.114e-1 (formula (11): 1.800e-1)
```

Computed The levels are values found on this machine; the dashed line is formula (11).

Computed The partial sums only draw towards the **mean** of the two levels, and do not separate them. The partial sums up to five terms draw, for $1/g = 40$, as 0.5, 0.475, 0.472188, 0.471492, 0.471247 towards the mean 0.471054 ($E0 = 0.467282$, $E1 = 0.474826$), and for $1/g = 100$, as 0.5, 0.49, 0.48955, 0.4895055, 0.4894992 towards the mean 0.4894978 ($E0 = 0.4894975$, $E1 = 0.4894981$).

Computed **Table 2** The ratio of the gap to formula (11) on the low-hill side. The rightmost column is the ratio to the formula multiplied by the next correction $(1 - \frac{71}{12}g)$ of the same paper (the coefficient of equation (2.19a)). $1/g = 16$–$100$ are in Table 4 of 2-05.

| 1/g | Height of hill | Gap E₁−E₀ | Ratio | Ratio with correction |
|---|---|---|---|---|
| 20 | 0.625 | 1.114e−1 | 0.619 | 0.879 |
| 30 | 0.938 | 3.153e−2 | 0.757 | 0.943 |
| 40 | 1.250 | 7.544e−3 | 0.831 | 0.975 |

Physics The gap of Yang–Mills too is held to be proportional to the scale $\Lambda$, and $\Lambda$ depends on the coupling constant in the form $e^{-c/g^2}$ ($c > 0$; $g$ is the coupling constant at the scale $\mu$, with a power of $g$ in front), so it is invisible to perturbation theory (4-07, 4-12).

07

## Many oscillators and Fock space — counting by occupation numbers

If there are $K$ independent oscillators, $[a_k, a_l^\dagger] = \delta_{kl}$, $[a_k, a_l] = 0$, and

$$H = \sum_{k=1}^{K} \hbar\omega_k \left( a_k^\dagger a_k + \frac{1}{2} \right) \quad (12)$$

The eigenstates are labelled by the tuple of **occupation numbers** $(n_1, \dots, n_K)$, and $n_k$ is "the number of quanta in mode $k$". All zeros is the **vacuum** $|0\rangle$, and $a_k^\dagger|0\rangle$ is a one-quantum state. Since $a_k^\dagger a_l^\dagger|0\rangle = a_l^\dagger a_k^\dagger|0\rangle$, quanta carry no name tags; only "how many in which mode" has meaning. Quantising the chain of 2-02 gives this form.

The form that does not depend on the number of oscillators is **Fock space**. Let the one-quantum space be $\mathfrak{h}$ (here $\mathbb{C}^K$; for a field, a space of functions). For $\mathfrak{h} = \mathbb{C}^K$, $\mathrm{Sym}^r \mathfrak{h}$ may be read as the space whose orthonormal basis is the tuples of occupation numbers above summing to $r$, $(n_1, \dots, n_K)$ (precisely, it is the symmetrisation of the tensor product of $r$ copies). If there are infinitely many modes, we take as basis the tuples with finitely many non-zero $n_k$ summing to $r$.

$$\mathcal{F}(\mathfrak{h}) = \bigoplus_{r=0}^{\infty} \mathrm{Sym}^r \mathfrak{h}, \quad \mathrm{Sym}^0 \mathfrak{h} = \mathbb{C}\,|0\rangle \tag{13}$$

The direct sum is the set of all sequences whose sums of squared norms are finite, and the total number of quanta $\hat{N} = \sum_k a_k^\dagger a_k$ has value $r$ on $\mathrm{Sym}^r \mathfrak{h}$. **Example.** The number of ways to distribute $r$ quanta over $K$ modes is $\binom{r+K-1}{r}$, and $\dim \mathrm{Sym}^2 \mathbb{C}^3 = 6$. Known The method of treating light with creation and annihilation is said to go back to Dirac (1927), and this space to Fock (1932).

08

## Quantising a field — particles appear, and the gap is m

On a ring of length $L$ (periodic boundary), we quantise a field with a mass term $m$ (2-02 § 06) ($\hbar = c = 1$). Placing an oscillator for each mode $e^{ikx}$ ($k = 2\pi n/L$, $\omega_k = \sqrt{k^2 + m^2}$; the dispersion relation $\omega^2 = c^2k^2 + \Omega^2$ of 2-02 with $c = 1$, $\Omega = m$),

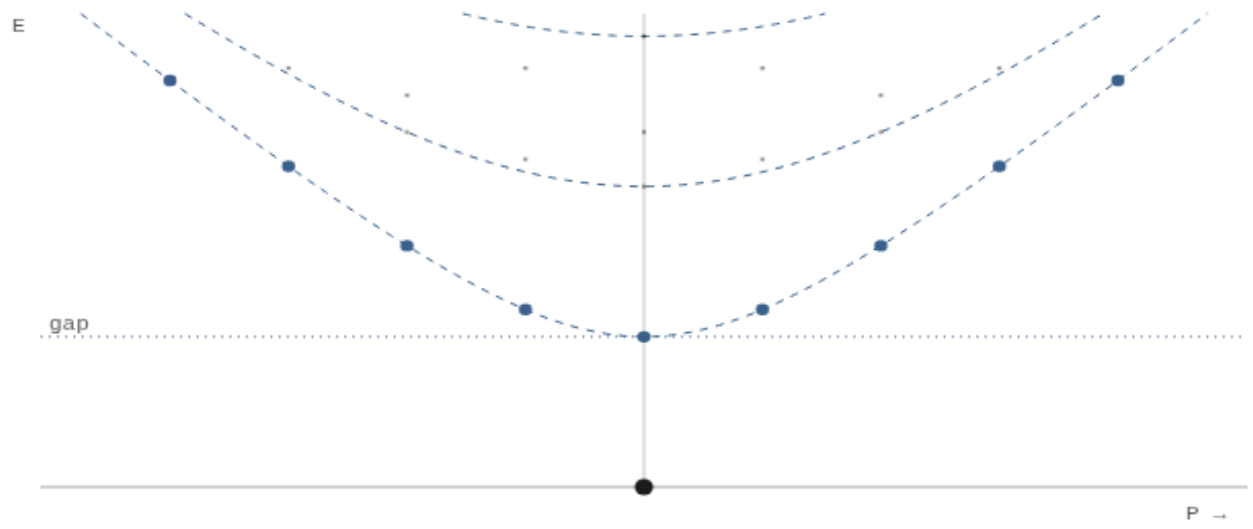
$$\phi(x) = \sum_k \frac{a_k e^{ikx} + a_k^\dagger e^{-ikx}}{\sqrt{2L\omega_k}} \tag{14}$$

$$H = \sum_k \omega_k \left( a_k^\dagger a_k + \frac{1}{2} \right) \tag{15}$$

The total momentum is $P = \sum_k k\, a_k^\dagger a_k$. One quantum of mode $k$ carries energy $\sqrt{k^2 + m^2}$ and momentum $k$, which is exactly the relation $E^2 = p^2 + m^2$ of a particle of mass $m$ (2-04). **A particle is a quantum of a normal mode of the field.** In the $(P, E)$ plane (Figure 4) a single particle lies on the hyperbola $E = \sqrt{P^2 + m^2}$, and two or more particles fill the side $E \geq \sqrt{P^2 + 4m^2}$. Next after the vacuum is the single particle with $k = 0$, and **even when the box is widened the gap is** $m$. Known In infinite volume $k$ ranges over all real numbers, the single-particle energies $\sqrt{k^2 + m^2}$ cover $[m,$

$\infty)$, and multi-particle states lie in $[2m, \infty)$, so the spectrum of the Hamiltonian is $\{0\} \cup [m, \infty)$. For $m = 0$ the gap $2\pi/L$ shrinks to 0 (excluding $k = 0$, which is not an oscillator).

Figure 4 Changing L and m changes the momentum P (horizontal) and energy E (vertical) of all states (E $\leq$ 3) of the free field in a box. Large dots are single particles. The dashed lines are $E = \sqrt{P^2 + (jm)^2}$ (j = 1, 2, 3; for m = 0, $E = |P|$), and the dotted line is the gap. When there are more than 10,000 states, the dots for two or more particles are omitted




L = 10.0, m = 1, number of states with E ≤ 3 = 23, gap =1.0000

Computed The dots are tuples of occupation numbers enumerated on the spot (the same procedure as Table 3).

Computed **Table 3** The number of states with $E \leq 3$ and the gap, for the free field in a box of period $L$.

| L | m = 1: gap | Number of states | m = 0: gap | Number of states |
|---|---|---|---|---|
| 5 | 1.0000 | 10 | 1.2566 | 8 |
| 10 | 1.0000 | 23 | 0.6283 | 38 |
| 20 | 1.0000 | 69 | 0.3142 | 734 |
| 40 | 1.0000 | 219 | 0.1571 | 55535 |

The gap of a free field, and the gap of the problem statement

The $m$ here is a value put in by hand. The action of Yang–Mills has no mass term (3-02), and self-interaction makes the modes no longer independent. Physics Even so, a gap is thought to arise, and showing it as mathematics is the content of the problem (0-03, 4-11). The way of saying that the uncertainty relation of § 03 is the origin of the gap is corrected in 16-08 § 09. What the theorem gives is only the step that converts a length into an energy, and the length side is this problem.

09

## The sum of zero-point energies, and a caution about infinitely many — only differences can be measured

The vacuum energy $\sum_k \omega_k/2$ of (15) diverges. Cutting off at $|k| \le \Lambda$ and putting $\Lambda \gg m$,

$$\sum_{|k|\le\Lambda} \frac{\omega_k}{2} \approx \frac{L}{2\pi}\int_{-\Lambda}^{\Lambda} \frac{\omega_k}{2}\,dk \approx \frac{L\Lambda^2}{4\pi} \quad (16)$$

It is proportional to the volume and grows like $\Lambda^2$. Since what can be measured is differences, we discard the constant and put $H = \sum_k \omega_k a_k^\dagger a_k$ (**normal ordering**). The gap is a difference, so it does not change. Physics The **difference** of vacuum energies when the box is changed can remain finite, and the Lüscher term is held to be an example (17-03).

Providing operators satisfying $[a_k, a_l^\dagger] = \delta_{kl}$ on a Hilbert space is called a **representation** of the commutation relations. Known For finitely many oscillators, when the commutation relations are written in exponential form (the Weyl relations; for one oscillator $e^{isx}e^{itp} = e^{-i\hbar st}e^{itp}e^{isx}$), the unitary irreducible representations (those whose only invariant closed subspaces are trivial) are unique up to unitary equivalence (being carried into each other by a unitary map) (Stone–von Neumann). For infinitely many, representations not unitarily equivalent to the Fock representation can be built. For example, the state with one quantum in every mode has an infinite sum of occupation numbers and is not in Fock space, and another representation with it as the vacuum can be considered. Fields with interactions are treated together with Haag's theorem in 4-09.

---

## What this chapter can say

| Content | Status |
|---|---|
| Sections 01–04: the levels and states of the oscillator, the decay of correlations | Said Known |
| Section 05: the levels and gap of the distorted oscillator | Said as numbers Computed . Discreteness and non-degeneracy are Known |
| Section 06: the perturbation series common to the two levels and formula (11) | Formulas from the literature Known . Partial sums and the gap checked against numbers Computed . Not derived |
| Sections 07–08: Fock space, the spectrum {0} ∪ [m, ∞) | Said Known (Table 3 Computed ) |

| Content | Status |
|---|---|
| Representations of infinitely many oscillators | Introduced only (4-09) |
| The mass gap of Yang–Mills | This chapter shows nothing |

## Sources and reproduction

| Item | Kind | Source / tools |
|---|---|---|
| Tables 1–3, Figures 1–4, the numbers in the text | Computed | python3, NumPy, SciPy (eigenvalues by finite differences, numerical integration, enumeration). The figures by JavaScript in the browser |
| The method of treating radiation with creation and annihilation | Known | P. A. M. Dirac, Proc. R. Soc. Lond. A 114 (1927) 243–265, doi:10.1098/rspa.1927.0039. Bibliographic data only |
| Fock space | Known | V. Fock, Z. Phys. 75 (1932) 622–647, doi:10.1007/BF01344458. Bibliographic data only |
| Two valleys: formulas (10)(11), the series, the correction 71/12 | Known | J. Zinn-Justin, U. D. Jentschura, Ann. Phys. 313 (2004) 197–267, doi:10.1016/j.aop.2004.04.004, arXiv:quant-ph/0501136 (equations (2.4)(2.6)(2.8)(2.19a)(2.20) of § 2.1 checked). The introduction and numbers for the same system are in 2-05 |
| That the spectrum is discrete in one dimension when $V \to \infty$ | Known | M. Reed, B. Simon, *Methods of Modern Mathematical Physics IV: Analysis of Operators* (Academic Press), chapter XIII. Bibliographic data only (year of publication and the relevant section not checked) |
| The spectrum $\{0\} \cup [m, \infty)$ of the free field | Known | M. Reed, B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness* (Academic Press, 1975), the section on the free relativistic field (said to be § X.7; section number not checked). Bibliographic data only |
| Stone–von Neumann uniqueness | Known | J. von Neumann, Math. Ann. 104 (1931) 570–578, doi:10.1007/BF01457956. Bibliographic data only. Infinitely many oscillators and Haag's theorem are treated in 4-09 |

Next chapter: 2-07 The idea of the path integral.

# The idea of the path integral — add up every path, and making time imaginary turns it into probability

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** — That an amplitude can be written as the limit of discretised multiple integrals, "the sum of $e^{iS/\hbar}$ over all paths".
That making time imaginary makes the weight $e^{-S_E/\hbar}$ positive, giving a probability measure, the Wiener measure.
That the partition function becomes an integral over paths on a ring, and quantum mechanics takes the same form as one-dimensional classical statistical mechanics.

**Prerequisite chapters** — 2-05 (decay in imaginary time), 1-08 (probability, the Gaussian distribution). For the oscillator, 2-06.

**Said first** — We treat a single particle on a line. The real-time path integral is not a measure, and many of the formulas are read as limits of discretised integrals. We do not go into the construction of path integrals for fields (4-02, Part 7). This chapter shows nothing about the mass gap of Yang–Mills.

Order of this chapter

1. The idea of a sum over paths — add the amplitudes, then square
2. Gaussian integrals — the value is set by the determinant, the correlations by the inverse matrix
3. Discretising time — building discretised integrals from the propagator and the product formula
4. Oscillating integrals and the classical path — cancellation selects the stationary point (Figure 1)
5. Imaginary time — oscillation turns into a weight, and the Wiener measure appears
6. The partition function is paths on a ring — quantum mechanics takes the same form as a one-dimensional classical chain (Figure 2)

7. Computing the oscillator by discretising — the correlation length grows like 1/a (Figure 3)
8. Finding the levels of $x^4$ with the transfer matrix — even without Gaussians we can compute with products of matrices
9. To fields — the fluctuation at a point diverges for d $\geq$ 2 (tables and figures in 1-09)

---

01

## The idea of a sum over paths — add the amplitudes, then square

The path integral rewrites quantum mechanics as the sum of contributions from **every path** the particle could take. Feynman (1948) wrote the amplitude (the **propagator**) for a particle leaving $a$ at time 0 to be found at $b$ at time $t$ in the following form.

$$K(b,t;a,0) = \langle b|e^{-itH/\hbar}|a\rangle = \text{“}\sum_{\text{paths } x(\cdot)}\text{”}\, e^{iS[x]/\hbar}, \qquad S[x] = \int_0^t \Big(\frac{m}{2}\dot{x}^2 - V(x)\Big)\,ds \quad (1)$$

$S$ is the action of 2-01, and the paths are all those satisfying $x(0) = a$, $x(t) = b$. Every path brings a complex number of absolute value 1, differing only in phase, and the probability is the square of the absolute value of the sum Known . The idea is said to go back to Dirac (1933). As a metaphor: just as light passes through both slits and interferes, the particle passes along every path and interferes. Precisely, the "sum" in (1) is not defined as it stands, and is read as the limit of the discretised integrals of section 03.

02

## Gaussian integrals — the foundation of every computation

Many computable path integrals reduce to Gaussian integrals. Let $A$ be a positive-definite symmetric matrix; diagonalising it and using the one-variable $\int e^{-ax^2/2+Jx}\,dx = \sqrt{2\pi/a}\,e^{J^2/2a}$ for each eigenvalue,

$$\int_{\mathbb{R}^n} e^{-\frac{1}{2}x\cdot Ax + J\cdot x}\,d^n x = \frac{(2\pi)^{n/2}}{\sqrt{\det A}}\, e^{\frac{1}{2}J\cdot A^{-1}J} \quad (2)$$

Normalised, this is the Gaussian distribution with mean 0 and covariance matrix $A^{-1}$ (1-08). Differentiating with respect to $J$ and putting $J = 0$,

$$\langle x_i x_j\rangle = (A^{-1})_{ij}, \qquad \langle x1x2x3x4\rangle = \langle x1x2\rangle\langle x3x4\rangle + \langle x1x3\rangle\langle x2x4\rangle + \langle x1x4\rangle\langle x2x3\rangle \quad (3)$$

The expectation of a product of $2k$ factors is the sum over the $(2k-1)!!$ ways of pairing them off (Isserlis 1918) Known . In physics the same combinatorics appears in operator form in Wick (1950), and goes by the name of Wick's theorem. Drawing these pairings as lines gives Feynman diagrams (4-03). **Example.** For $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ the integral is $2\pi/\sqrt{3}$, $\langle x_1^2 \rangle = \frac{2}{3}$, $\langle x_1 x_2 \rangle = \frac{1}{3}$, $\langle x_1^2 x_2^2 \rangle = \frac{4}{9} + 2 \cdot \frac{1}{9} = \frac{2}{3}$ Computed . **The value is set by** $\det A$**, and the correlations by** $A^{-1}$**.**

03

## Discretising time — the propagator and the product formula

There are three ingredients. **Composition**: from $e^{-i(t_1+t_2)H/\hbar} = e^{-it_2H/\hbar} e^{-it_1H/\hbar}$ comes the formula $K(b, t_1+t_2; a, 0) = \int K(b, t_1+t_2; c, t_1) K(c, t_1; a, 0)\, dc$ that sums over all intermediate places $c$. **The short-time kernel**: for a free particle, exactly

$$K_0(b, t; a, 0) = \sqrt{\frac{m}{2\pi i\hbar t}}\, \exp\Big( \frac{i}{\hbar} \frac{m(b-a)^2}{2t} \Big) \quad (4)$$

and the exponent is the action of the uniform-velocity path. **The product formula**: split the Hamiltonian as $H = T + V$. $T = P^2/2m$ is the operator of kinetic energy ($P = -i\hbar\, d/dx$ is the momentum of 2-05), and $V$ is the operator of multiplication by the potential energy $V(x)$. $T$ and $V$ do not commute, but if, for example, $T+V$ is self-adjoint on the intersection of the domains of $T$ and $V$, then $e^{-it(T+V)/\hbar} = \lim_N \big( e^{-i\varepsilon T/\hbar} e^{-i\varepsilon V/\hbar} \big)^N$, $\varepsilon = t/N$ holds Known . It is a special case of Trotter's (1959) product theorem for contraction semigroups, and this form is in chapter VIII of Reed–Simon I. Putting these together, with $x_0 = a$, $x_N = b$,

$$K(b, t; a, 0) = \lim_{N\to\infty} \Big( \frac{m}{2\pi i\hbar\varepsilon} \Big)^{N/2} \int dx_1 \cdots dx_{N-1}\, \exp\Big( \frac{i}{\hbar} \sum_{j=0}^{N-1} \varepsilon \Big[ \frac{m}{2} \Big( \frac{x_{j+1} - x_j}{\varepsilon} \Big)^2 - V(x_j$$

The sum in the exponent is a Riemann sum for the action of a broken-line path, that is, the discretised action of 2-01. (1) is shorthand for this limit, also written $\int \mathcal{D}x\, e^{iS/\hbar}$.

Caution — $\mathcal{D}x$ is not a measure

Cameron (1960) is said to have shown that the real-time $e^{iS/\hbar}\mathcal{D}x$ cannot be constructed as a measure in the form of the Wiener measure (section 05) with complex variance (such a measure would have infinite total variation even on bounded sets of the path space) Known . The total variation is the supremum, over fine partitions of a set, of the sum of the absolute values of the measures of the

parts; for a probability measure it is 1. Measures in infinite dimensions in general are treated in 1-09.

04

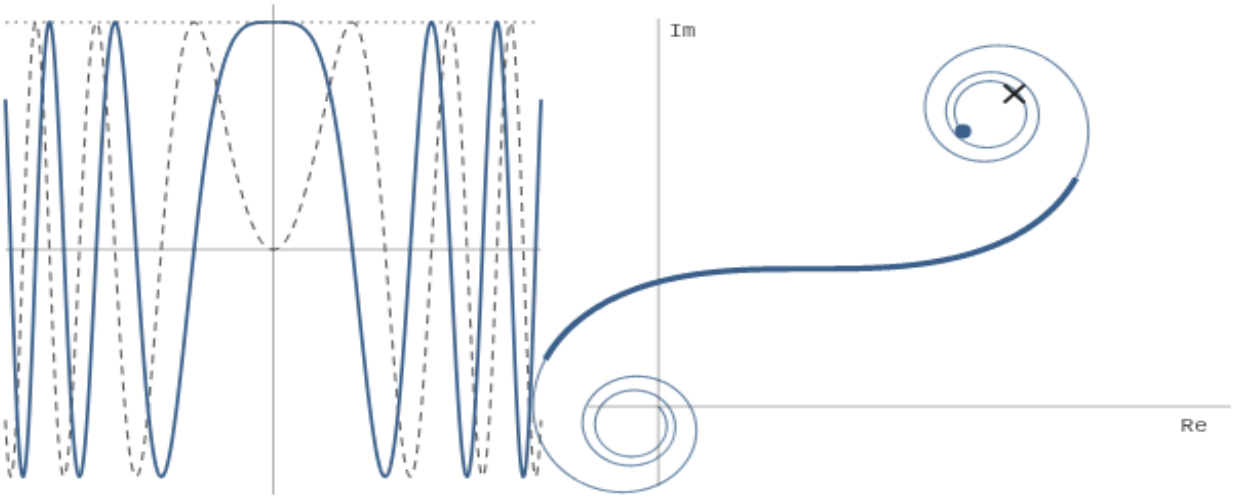
## Oscillating integrals and the classical path — cancellation and the stationary point

The integrand of (5) has absolute value 1 and oscillates, and does not converge absolutely. Figure 1 shows the model case $\int e^{ix^2/2}dx = \sqrt{2\pi}\,e^{i\pi/4}$ (the Fresnel integral). On the right is the trace of the partial integral (the integral with a moving upper limit) $F(x) = \int_{-6}^{x} w\,dx'$; for $\theta = 0$ both ends spiral and cancel, and the net progress happens only near the **stationary point** $x = 0$ where the slope of the phase is 0. For $\hbar \to 0$ (the sign is the sign of $S''(x_c)$) Known

$$\int g(x)\,e^{iS(x)/\hbar}\,dx \approx g(x_c)\sqrt{\frac{2\pi\hbar}{|S''(x_c)|}}\;e^{iS(x_c)/\hbar \pm i\pi/4}, \qquad S'(x_c) = 0 \tag{6}$$

In path space too, the contribution comes from near the **classical path** with $\delta S = 0$ (2-01). If the action is quadratic, (6) is exact, and (4) has that form. **Example.** A 1 g ball moving at 1 cm/s for 1 second gives $S/\hbar \approx 4.7 \times 10^{26}$ Computed . For displaced paths the phase turns round many times, and they cancel.

Figure 1 Changing the angle θ changes $w(x) = e^{ie^{i\theta}x^2/2}$ (left; solid is the real part, dashed the imaginary part, dotted the absolute value) and the trace of the partial integral $F(x)$ (right). θ = 0 is real time and 90° is imaginary time (the $t = e^{-i\theta}\tau$ that joins the $t = -i\tau$ of 2-04 by an angle; see also the spacetime article). The thick line is |x| < 1.5, and × is the exact value




θ = 0°, F(6) = 1.517 + 1.560i, exact value = 1.772 + 1.772i, difference 3.3e-1

Computed The numbers in the figure were computed on this machine (the trapezoidal rule on the spot). Table 1 gives values from the complex error function in python3.

Computed **Table 1** The absolute value of the difference between $\int_{-X}^{X} w\,dx$ and the exact value $\sqrt{2\pi}\,e^{i(\pi/2-\theta)/2}$.

| X | θ = 0 | θ = 22.5° | θ = 90° |
|---|---|---|---|
| 2 | 9.3e-1 | 4.1e-1 | 1.1e-1 |
| 8 | 2.5e-1 | 1.2e-6 | < 1e-14 |
| 32 | 6.25e-2 | < 1e-14 | < 1e-14 |

In real time the error decreases only like $2/X$; rotating even slightly makes it vanish at the rate $e^{-\sin\theta\,X^2/2}$.

05

## Imaginary time — oscillation turns into a weight

Replacing the step $\varepsilon$ in (5) by $-i\delta$ ($\delta>0$), the exponent becomes $-\frac{\delta}{\hbar}\big[\frac{m}{2}(\Delta x/\delta)^2+V\big]$, and the sign of the potential energy changes. With $\tau=N\delta$,

$$\langle b|e^{-\tau H/\hbar}|a\rangle=\lim_{N\to\infty}\Big(\frac{m}{2\pi\hbar\delta}\Big)^{N/2}\int dx_1\cdots dx_{N-1}\;e^{-S_E/\hbar},\qquad S_E=\sum_{j=0}^{N-1}\delta\Big[\frac{m}{2}\Big(\frac{x_{j+1}}{\delta}$$

We call $S_E$ the **Euclidean action**. The weight is positive, and if $V$ is bounded below the integral converges absolutely. The $V=0$ part is a random walk whose increments independently follow normal distributions with variance $\hbar\delta/m$; for $N\to\infty$ it becomes a probability measure on continuous paths, the **Wiener measure** (the distribution of Brownian motion), and $V$ enters as a weight.

$$\big(e^{-\tau H/\hbar}\psi\big)(x)=\mathbb{E}_x\Big[\exp\Big(-\frac{1}{\hbar}\int_0^\tau V(X_s)\,ds\Big)\,\psi(X_\tau)\Big]\qquad (8)$$

$X_s$ is Brownian motion leaving $x$ and spreading with variance $\hbar s/m$. (8) is called the Feynman–Kac formula (Kac 1949). It is a theorem when, for example, $V$ is bounded and continuous, and it holds for quite general $V$ bounded below Known . **In real time there is no measure; in imaginary time there is.** This is one reason to construct field theories on the Euclidean side (the rotation for fields is in 4-02 § 02 and 4-10).
**Example.** The stationary point $\delta S_E=0$ is motion in $-V$, and the $S_0$ of the double well of 2-05 was the action of that path (the instanton, 3-05).

### The roughness of paths and the measure

The paths themselves are nowhere differentiable with probability 1 Known . Since $\Delta x\sim\sqrt{\delta}$, the expectation $\langle S_{\text{kin}}\rangle$ of the kinetic term $S_{\text{kin}}=\sum(\Delta x)^2/2\delta$ ($\hbar=m=1$) of $S_E$ grows in proportion to the number of steps Computed (for the ring of section 07, 6.97,

126.96, 2046.96 for $\beta = 4$, $N = 16, 256, 4096$; almost $N/2$). The potential-energy part $\sum \delta V(x_j)$ on its own approaches $\int_0^\tau V(X_s)\,ds$, but the weight of the kinetic term $e^{-S_{\text{kin}}/\hbar}$ and $\mathcal{D}x$ have no limits separately; only the product survives, as the Wiener measure (both in 1-09).

06

## The partition function is paths on a ring — the agreement with statistical mechanics

In thermal equilibrium a state $n$ appears with weight $e^{-\beta E_n}$ ($\beta = 1/(k_B \times \text{temperature})$, $k_B$ is Boltzmann's constant), and their sum is the **partition function** (2-08). Taking the trace means putting $a = b$ and integrating over $a$, so by (7) it becomes an integral over **closed paths on a ring** of length $\beta\hbar$.

$$Z(\beta) = \operatorname{Tr} e^{-\beta H} = \sum_n e^{-\beta E_n} = \int_{x(0)=x(\beta\hbar)} \mathcal{D}x\; e^{-S_E[x]/\hbar} \tag{9}$$

Read in discretised form, it is the partition function of a **classical chain** in which $x_0, \ldots, x_{N-1}$ arranged on a ring are joined by springs $\frac{m}{2\delta}(x_{j+1} - x_j)^2$ and each point carries $V$. Quantum mechanics takes the same form as one-dimensional classical statistical mechanics, with $\hbar$ playing the role of temperature and $\beta\hbar$ that of the length of the chain. The exposition of the problem attributes to Symanzik the view of reading the Euclidean expectations of a field theory as a classical statistical-mechanical ensemble with density proportional to $e^{-S}$ (Jaffe–Witten § 3) Known . The chain of this section is its one-dimensional version.

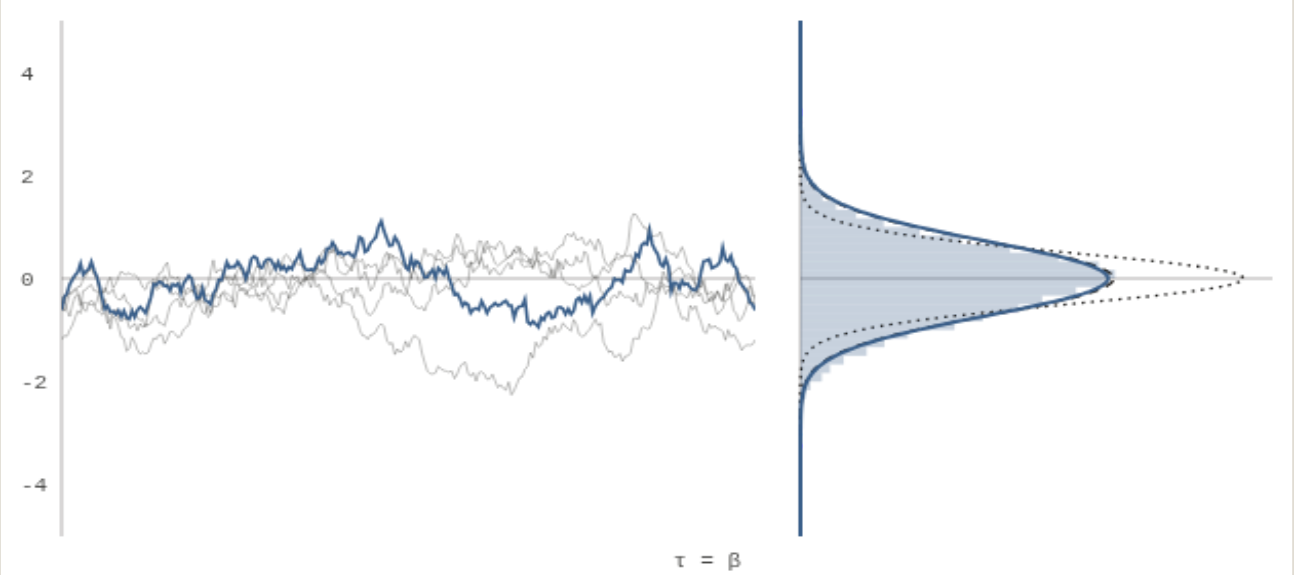
### Example: the oscillator ring

For the oscillator, $Z = 1/(2\sinh\frac{\beta\hbar\omega}{2})$; the distribution of one point $x(0)$ of the ring is the distribution of position in thermal equilibrium, with variance

$$\langle x^2\rangle_\beta = \frac{\hbar}{2m\omega}\coth\frac{\beta\hbar\omega}{2} \longrightarrow \begin{cases} \hbar/2m\omega & (\beta\to\infty) \\ 1/m\omega^2\beta & (\beta\to 0) \end{cases} \tag{10}$$

and at low temperature it approaches the $|\psi_0|^2$ of the ground state (2-06), at high temperature the classical Boltzmann distribution.

Figure 2 Changing $\beta\hbar\omega$ (the length of the ring) changes the paths on the ring (left; colour is the latest) and the distribution of positions visited by the paths (shading on the right). The solid line is (10), the dashed line the ground state, the dotted line the classical case. "Play" adds paths ($\hbar = m = \omega = 1$)




βħω = 4, paths: 120, sample ⟨x²⟩ = 0.557, discretised system 0.5186, (10) 0.5187, classical 0.2500

Computed The numbers in the figure were computed on this machine. The paths are exact samples of (7) with the ring cut into 256 pieces (the real Fourier components are independent normal variables, drawn with random numbers from a fixed seed). Table 2 gives python3 values from the same procedure.

Computed **Table 2** $\langle x^2 \rangle$ at one point of the ring (unit $\hbar/m\omega$; the ground state gives 0.5).

| βħω | Sample | Exact value for the discretised system | Quantum (10) | Classical |
|---|---|---|---|---|
| 0.25 | 4.05 ± 0.09 | 4.0208 | 4.0208 | 4.0000 |
| 4 | 0.517 ± 0.006 | 0.5186 | 0.5187 | 0.2500 |
| 16 | 0.504 ± 0.003 | 0.4998 | 0.5000 | 0.0625 |

At high temperature, on a short ring, the path shrinks almost to a point (a classical particle); at low temperature, **quantum fluctuations are seen as fluctuations of a path stretched in the imaginary-time direction**.

07

## Computing the oscillator by discretising — determinant, correlations, correlation length

$\hbar = m = \omega = 1$: cut a ring of length $\beta$ into $N$ pieces. From here on $a = \beta/N$ is the step width (the lattice spacing), a different symbol from the starting point $a$ of sections 01–06. The $A$ of $S_E = \frac{1}{2} x \cdot Ax$ is $(Ax)_j = \frac{1}{a}(2x_j - x_{j+1} - x_{j-1}) + a x_j$ (indices counted modulo $N$), a **circulant matrix** connected only to neighbours. Its eigenvectors are $v_j = e^{2\pi i j k/N}$ (discrete Fourier, $k = 0, \dots, N-1$), and substituting $v_{j\pm1} = e^{\pm 2\pi i k/N} v_j$ gives the eigenvalues $\frac{1}{a}(2 - 2\cos\frac{2\pi k}{N}) + a$. From (2) and the normalisation $(2\pi a)^{-N/2}$,

$$Z_N = \prod_{k=0}^{N-1} \left(2 - 2\cos\frac{2\pi k}{N} + a^2\right)^{-1/2} = \frac{1}{2\sinh(N\kappa/2)}, \qquad \cosh\kappa = 1 + \frac{a^2}{2} \quad (11)$$

Compared with the continuous $1/(2\sinh\frac{\beta}{2})$, the discretised system is the same as an oscillator with frequency $\Omega_a = \kappa/a = 1 - a^2/24 + \cdots$. The correlations are set by $A^{-1}$, and for $N \to \infty$,

$$\langle x_0 x_j \rangle = (A^{-1})_{0j} = \frac{a\, e^{-\kappa|j|}}{2\sinh\kappa} \longrightarrow \frac{e^{-\tau}}{2} \qquad (\tau = ja,\ a \to 0) \tag{12}$$

As in (8) of 2-06, correlations decrease at the rate of the gap. The **correlation length** counted in lattice units is $\xi = 1/\kappa = 1/(a\Omega_a)$.

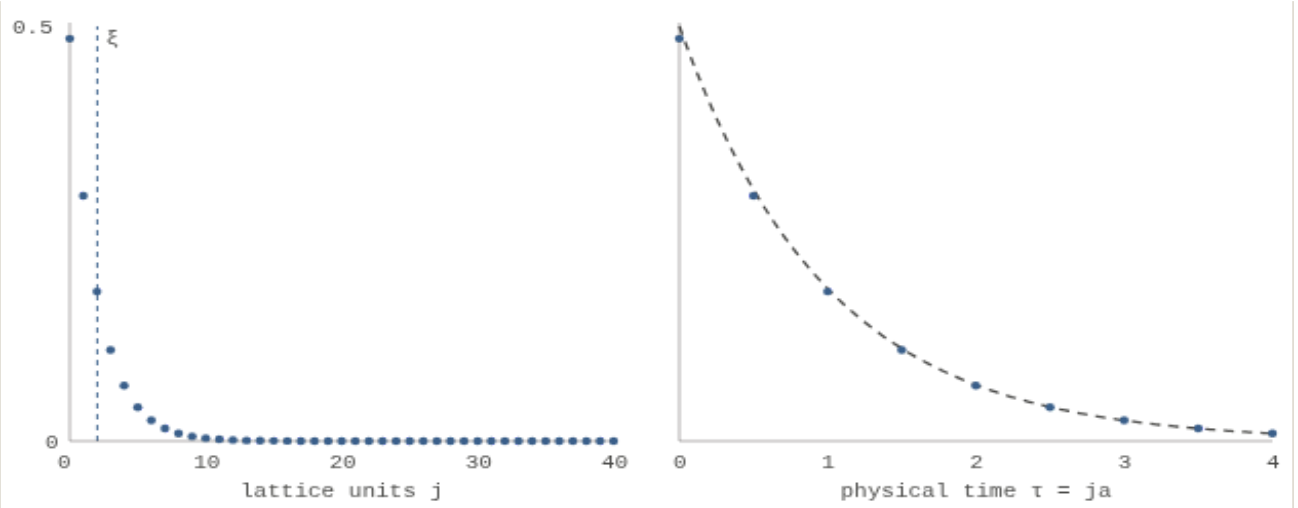
### The correlation length and the continuum limit

Computed **Table 3** The ring with $\beta = 4$ ($Z_N$: the numerical determinant and (11) agree to 10 digits).

| N | a | Z_N / Z − 1 | Ω_a | ξ |
|---|---|---|---|---|
| 4 | 1 | 8.1e-2 | 0.96242 | 1.04 |
| 16 | 0.25 | 5.4e-3 | 0.99741 | 4.01 |
| 128 | 0.03125 | 8.4e-5 | 0.99996 | 32.00 |

The discrepancy vanishes like $a^2$, and the correlation length grows like $1/a$ (Figure 3). **Approaching the continuum means bringing the lattice chain towards a state in which "the correlation length is many times the lattice spacing", in the language of statistical mechanics towards a critical point.** We hold the physical gap $1/(\xi a)$ fixed and let $\xi \to \infty$. In lattice gauge theory too, the continuum limit is thought to be taken in this form Physics . In four dimensions, the construction published in 2026-11 made the scaling limit by choosing the bare coupling to follow the running of asymptotic freedom (the committee's assessment is that it is proved in Lean; 6-00, 8-02). The form in which the correlation length grows while the physical gap is held fixed, that is, that the gap stays positive in the limit, has not been shown (6-04, 5-05, 5-09).

Figure 3 Changing the step $a$ changes how the correlation (12) $a\,e^{-\kappa j}/(2\sinh\kappa)$ ($\beta \to \infty$) looks. On the left it is counted in lattice units $j$ (the vertical dashed line is the correlation length $\xi = 1/\kappa$); on the right it is measured in physical time $\tau = ja$ (the dashed line is the continuous $e^{-\tau}/2$). "Play" moves $a$ back and forth between 1 and 1/16




Computed The numbers in the figure were computed on this machine (the closed formula (12) computed on the spot; $\kappa$, $\xi$, $\Omega_a$ at $a = 1, \frac{1}{2}, \ldots, \frac{1}{16}$ agree with the python3 values to the digits shown). On the left, the smaller $a$, the further the dots continue; on the right, the dots settle onto the same dashed line.

08

## A non-Gaussian example — finding the levels of x⁴ with the transfer matrix

Even without Gaussians, discretised integrals can be computed as products of matrices. Make one step of (7) ($\hbar = m = 1$, step $a$) into the integral kernel of $e^{-aV/2}e^{-aP^2/2}e^{-aV/2}$

$$T_a(x,y) = \frac{1}{\sqrt{2\pi a}} \exp\Big( -\frac{(x-y)^2}{2a} - \frac{a}{2}\big(V(x)+V(y)\big)\Big) \quad (13)$$

then $Z_N = \operatorname{Tr} T_a^N$; this is the **transfer matrix** of statistical mechanics. Writing the eigenvalues as $e^{-aE_n(a)}$, $E_n(a) \to E_n$ as $a \to 0$. The kernel is positive at every point ($T_a(x,y) > 0$). Moreover $V = x^4$ grows at infinity, so $\iint T_a(x,y)^2\,dx\,dy \le \frac{1}{2\pi a}\big(\int e^{-aV}\,dx\big)^2 < \infty$, and $T_a$ is **compact** (an operator that can be approximated as closely as we like by finite-dimensional matrices, and whose eigenvalues accumulate nowhere but 0). With these two together, the integral-operator version of the Perron–Frobenius theorem (said to begin with Jentzsch 1912; the form for $L^2$ is in Reed–Simon IV) shows that the largest eigenvalue is simple and its eigenfunction can be taken positive Known . Hence the ground state is non-degenerate (2-05, 5-03).

Computed **Table 4** The levels of the transfer matrix for $H = -\frac{1}{2}\frac{d^2}{dx^2} + x^4$ (lattice width for $x \le 0.02$, $|x| \le 3.2$) and their differences from the values of Table 1 of 2-05 ($E_0 = 0.667986$, $E_1 - E_0 = 1.725658$). Halving the lattice width changes nothing to 9 digits.

| a | E₀(a) | E₁(a) − E₀(a) | Difference in E₀ | Difference in gap |
|---|---|---|---|---|
| 0.1 | 0.666246 | 1.719680 | −1.7e-3 | −6.0e-3 |

| a | $E_0(a)$ | $E_1(a) - E_0(a)$ | Difference in $E_0$ | Difference in gap |
|---|---|---|---|---|
| 0.05 | 0.667546 | 1.724144 | −4.4e-4 | −1.5e-3 |
| 0.025 | 0.667876 | 1.725278 | −1.1e-4 | −3.8e-4 |

The difference is roughly 1/4 each time the step is halved. **The discretised path integral and operator quantum mechanics give the same levels and gap.**

09

## To fields — as the step is refined, the fluctuations diverge

We extend the discretisation of section 07 directly to fields. In field theory the path $x(\tau)$ is replaced by a field $\phi(x)$ ($d$ is the dimension of spacetime including imaginary time), and $S_E = \int (\frac{1}{2}|\nabla\phi|^2 + \frac{1}{2}m^2\phi^2 + \cdots)d^dx$ ($m$ is here the mass of the field; 4-02). Discretising on a lattice of spacing $a$, a field without interactions is again the Gaussian integral of (2), and the fluctuation at a point is a diagonal entry of $A^{-1}$.

$$\langle\phi(0)^2\rangle = G_a(0) = \int_{[-\pi/a,\pi/a]^d} \frac{d^dp}{(2\pi)^d} \frac{1}{\sum_{\mu=1}^{d} \frac{4}{a^2}\sin^2\frac{p_\mu a}{2} + m^2} \quad (14)$$

$d = 1$ is the chain of section 07 itself; (14) equals the value $a/(2\sinh\kappa)$ of (12) at $j = 0$, and converges to $1/2m$. For $d \geq 2$ it diverges. The divergence of the variance at a point is also treated in 1-09, from the side of Gaussian measures in infinite dimensions. Here we look, as a continuation of discretising the path integral, at what happens each time the step is halved.

Computed Each time the step is halved, $d = 1$ converges to $1/2$, $d = 2$ grows by about $\ln 2/(2\pi) \approx 0.110$ each time ($G_a(0) \approx \frac{1}{2\pi}\log\frac{1}{a} + 0.276$), $d = 3$ doubles ($\approx 0.2527/a$), and $d = 4$ quadruples ($\approx 0.1549/a^2$) ($m = 1$; (14) is turned into a one-dimensional integral of the Bessel function $I_0$ and integrated numerically). The table of values for each step and the figure for choosing the dimension are in 1-09 Table 3, Figure 3, which this book takes as authoritative. A figure setting it beside other quantities that switch at four is the dimension dial of 1-12 § 10.

### Fields in the continuum limit and Yang–Mills

A field in the continuum limit is a **distribution** without values at points (1-11). Products such as $\phi(x)^2$ need subtractions (renormalisation) (4-04, 7-03).

In four-dimensional Yang–Mills this divergence, the gauge freedom (4-08) and the interaction are handled at once. The exposition of the problem says that for a compact group and a finite (periodic) spacetime, Wilson's lattice makes the functional integral a finite-dimensional integral, and that one must then show that the limit lattice spacing

$\to 0$, volume $\to \infty$ exists for the expectations of gauge-invariant observables (Jaffe–Witten § 6.5) Known . The construction published in 2026-11 takes the limit lattice spacing $\to 0$ with the volume held fixed, and then takes the limit along a subsequence of volumes (8-05). For what is special about four dimensions, see the four-dimensions article.

## What this chapter can say

| Content | Status |
|---|---|
| The formulas (2)–(12) | Said Known ((5) and (8) under conditions; numbers also checked in Tables 2 and 3 and Figures 2 and 3 Computed ). That $\mathcal{D}x$ is not a measure is a report of the literature |
| The convergence in Table 4, the constants in the growth of the divergence in § 09 | The range checked numerically Computed |
| That the continuum limit of lattice gauge theory is also taken in the form "correlation length $\to \infty$" (section 07) | A view at the level of physics Physics . The scaling limit in four dimensions was made by the construction published in 2026-11 (at the level of the committee's assessment; 6-00). The form in which the gap stays positive has not been shown (6-04) |
| Path integrals for interacting fields and gauge fields | Not treated (4-02, Part 7). Nothing is shown about the mass gap of Yang–Mills |

## Sources and reproduction

| Item | Kind | Source / tools |
|---|---|---|
| Tables 1–4, the numbers in the examples, Figures 1–3 | Computed | python3 (NumPy, SciPy). The figures by JavaScript in the browser (the trapezoidal rule, random numbers with a fixed seed, the closed formula (12)) |
| The constants of growth in § 09 (0.2527, 0.1549, 0.276) | Computed | Computed with python3 (SciPy's $I_0$ and numerical integration): the values at the origin of the massless lattice ($d = 3$: 0.252727, $d = 4$: 0.154933), and the value of the difference $G_a(0) - \frac{1}{2\pi} \log \frac{1}{a}$ of $d = 2$ up to $a = 2^{-10}$ (0.27579). Agrees with Table 3 of 1-09 |
| (1) | Known | R. P. Feynman, “Space-time approach to non-relativistic quantum mechanics”, Rev. Mod. Phys. 20 (1948) 367–387 (bibliographic data checked with Crossref; text bibliographic data only) / P. A. M. Dirac, Physik. Z. Sowjetunion 3 (1933) 64 (bibliographic data only; that it goes back to Dirac's suggestion is from the Scholarpedia article below) |

| Item | Kind | Source / tools |
|---|---|---|
| The product formula | Known | H. F. Trotter, Proc. Amer. Math. Soc. 10 (1959) 545–551 (bibliographic data only) / M. Reed, B. Simon, Methods of Modern Mathematical Physics I (Academic Press 1972), chapter VIII (bibliographic data only; said to be Theorem VIII.31, number not checked) |
| That $\mathcal{D}x$ is not a measure, the conditions for (8) | Known | R. H. Cameron, "A family of integrals serving to connect the Wiener and Feynman integrals", J. Math. and Phys. 39 (1960) 126–140 (bibliographic data only). Content checked in S. Albeverio, S. Mazzucchi, "Path integral: mathematical aspects", Scholarpedia 6(1):8832 (2011) |
| (8), (3) | Known | M. Kac, Trans. Amer. Math. Soc. 65 (1949) 1–13 / L. Isserlis, Biometrika 12 (1918) 134–139 / G. C. Wick, Phys. Rev. 80 (1950) 268–272. Bibliographic data only for all |
| Symanzik's view, lattices and limits | Known | A. Jaffe, E. Witten, "Quantum Yang–Mills theory" (the Clay problem description), § 3 and § 6.5 checked / K. Symanzik, "Euclidean quantum field theory", in Local Quantum Theory (R. Jost ed., Academic Press 1969) 152–226 (bibliographic data only; reference [46] of Jaffe–Witten) |
| The integral-operator version of Perron–Frobenius | Known | R. Jentzsch, "Über Integralgleichungen mit positivem Kern", J. reine angew. Math. 141 (1912) 235–244 (bibliographic data checked with Crossref; text bibliographic data only) / M. Reed, B. Simon, Methods of Modern Mathematical Physics IV (Academic Press 1978) (bibliographic data only; said to be § XIII.12, not checked) |
| Stationary phase | Known | R. P. Feynman, A. R. Hibbs, Quantum Mechanics and Path Integrals (1965). Bibliographic data only |
| That paths are not differentiable | Known | 1-09 (citing Durrett's theorem) |

Next chapter: 2-08 Statistical mechanics and phase transitions.

# Statistical mechanics and phase transitions — probability weights, the length over which correlations reach, and the point where that length diverges

> 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** — Reading thermal equilibrium as a probability distribution $e^{-\beta H}/Z$, and obtaining means and fluctuations by differentiating $\log Z$.
The correlation length, and that in the Gaussian chain the correlation length grows like $1/m$.
That phase transitions occur only in infinite systems, the universality of critical exponents, and that the continuum limit is an approach to a critical point.

**Prerequisite chapters** — 1-08 (probability, variance, Gaussian measures). Inverse matrices are in 1-01. The transfer matrix of section 03 also appears in 2-07 (if you have read it; this chapter defines it on the spot too).

**Said first** — We treat classical statistical mechanics on lattices. The three-dimensional critical exponents are numbers, not theorems, and the correspondence with field theory, except for the Gaussian chain, is a reinterpretation at the level of physics. This chapter shows nothing about the mass gap of Yang–Mills.

Order of this chapter

1. The Boltzmann distribution — temperature is a way of weighting probabilities
2. The partition function is a generating function — differentiating gives means and fluctuations
3. Correlations and the correlation length — exactly for the Gaussian chain; the inverse of the correlation length is the gap of the transfer matrix
4. Phase transitions — kinks that occur only in infinite systems
5. The mean-field model — watching the tug of war (Figure 1, Table 1). Detour: zeros of the partition function (Figure 2)

6. Critical points — critical exponents (Table 2)
7. Universality — exponents are universal, coefficients are not (Figure 3, Table 3)
8. The continuum limit — the critical point is where the lattice disappears (Figure 4, Table 4)

---

01

## The Boltzmann distribution — temperature is a way of weighting probabilities

We call each single state of a system of many spins or particles a **configuration** $\sigma$, and its energy $H(\sigma)$. A system in thermal equilibrium is described by a probability distribution on configurations; the lower the energy of a configuration, the heavier it is, and the weight is exponential.

$$\mu_\beta(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z_\beta}, \qquad Z_\beta = \sum_\sigma e^{-\beta H(\sigma)}, \qquad \beta = \frac{1}{kBT} \quad (1)$$

We call $\mu_\beta$ the **Boltzmann distribution** and $Z_\beta$ the **partition function** ($T$ is the temperature; below, Boltzmann's constant $kB = 1$, and $\beta$ is the inverse temperature). For $\beta \to 0$ it becomes the uniform distribution, and for $\beta \to \infty$ it concentrates on the configurations of minimal energy. $\langle\cdot\rangle$ is the expectation under $\mu_\beta$ (1-08).

One reason for the exponential is the following variational principle Known . With the entropy $S(P) = -\sum_\sigma P \log P$,

$$-\frac{1}{\beta}\log Z_\beta = \min_P\Big(\mathbb{E}_P[H] - \frac{1}{\beta}S(P)\Big) \quad (2)$$

holds, and the minimum is attained only by $P = \mu_\beta$ (the difference between the two sides is the relative entropy $\frac{1}{\beta}\sum_\sigma P\log(P/\mu_\beta) \geq 0$). The left side is the **free energy**. It is a tug of war between the force that lowers the energy and the force that increases the number of configurations; at low temperature the former wins, at high temperature the latter. Precisely, in a finite system the minimum of (2) is always the single $\mu_\beta$. If we consider the minimisation of (2) per site in an infinite system, at low temperature the states attaining the minimum can be several rather than one. This "splitting from one into several" is one precise form of a phase transition (the coexistence of phases) (we see it in one-variable form in section 05).

**Example.** If $N$ non-interacting spins $\sigma_x = \pm 1$ are in an external field $h$ ($H = -h\sum_x \sigma_x$), then $Z = (2\cosh\beta h)^N$, the spins are independent, and $\langle\sigma_x\rangle = \tanh\beta h$.

02

## The partition function is a generating function — differentiating gives means and fluctuations

Differentiating $\log Z$ gives means and variances. In the language of probability, $\log Z\beta$ is the cumulant generating function of $-H$ (plus a constant). The cumulant generating function of a random variable $X$ is $t \mapsto \log \mathbb{E}[e^{tX}]$; differentiating once in $t = 0$ gives the mean, twice the variance.

$$\langle H \rangle = -\frac{\partial}{\partial \beta} \log Z\beta, \qquad \operatorname{Var}(H) = \frac{\partial^2}{\partial \beta^2} \log Z\beta \geq 0 \quad (3)$$

Putting in the external field as $-hM$ with the total magnetisation $M = \sum_x \sigma_x$ gives the magnetisation and the **susceptibility** $\chi$ (the response to the external field).

$$\langle M \rangle = \frac{1}{\beta} \frac{\partial}{\partial h} \log Z, \qquad \chi = \frac{1}{N} \frac{\partial \langle M \rangle}{\partial h} = \frac{\beta}{N} \operatorname{Var}(M) \quad (4)$$

The right-hand equality is the fluctuation–response relation: the response to a push is determined by the fluctuations without the push. Since variances are non-negative, $\log Z$ is convex both in $\beta$ and in $h$. The limit $f$ of $f_N = -\frac{1}{\beta N} \log Z_N$ per site as $N \to \infty$ is called the **thermodynamic limit**. For short-range interactions the limit is known to exist, and convexity survives Known (for lattice spin systems, the standard argument from the fact that when the system is divided into boxes, $\log Z$ differs from the sum over boxes only by the boundary contributions [subadditivity]; Ruelle 1969 and others).

**Example.** For independent spins, $\chi = \beta(1 - \tanh^2 \beta h) = \beta \operatorname{Var}(\sigma_x)$. For $h = 0$, $\chi = \beta$, finite at every temperature. Without interactions nothing happens.

03

## Correlations and the correlation length — how much two distant points move together

With interactions, distant points move together. The degree of this is the **connected two-point correlation** (covariance).

$$G(x, y) = \langle \sigma_x \sigma_y \rangle - \langle \sigma_x \rangle \langle \sigma_y \rangle \quad (5)$$

In a translation-invariant infinite system, $G(x, y) = G(x - y)$, and (4) is $\chi = \beta \sum_x G(x)$. **For the susceptibility to diverge, the correlations must decrease only slowly.** When correlations decrease exponentially, we call that length the **correlation length** $\xi$.

$$\frac{1}{\xi} = -\lim_{|x|\to\infty} \frac{\log G(x)}{|x|} \tag{6}$$

The logarithm is taken to allow polynomial prefactors. The correlation length of lattice gauge theory is defined in the same form, except that instead of spins it uses correlations of the variables of small square loops (plaquettes) (5-01, Chatterjee 2018 § 5, 5-09) Known . There are two examples.

**The Ising model**: at each point of $\mathbb{Z}^d$, $\sigma_x = \pm 1$, $H = -J\sum_{\langle xy\rangle}\sigma_x\sigma_y - h\sum_x \sigma_x$ ($\langle xy\rangle$ is a neighbouring pair, and $J>0$ is a ferromagnetic coupling that rewards alignment; a worked example is in 2-09).

**The Gaussian chain**: put a real number $\phi_x$ at each point of $\mathbb{Z}$, and with $\beta = 1$, $m > 0$ (the **mass**), take

$$H(\phi) = \frac{1}{2}\sum_x \Big[(\phi_{x+1}-\phi_x)^2 + m^2\phi_x^2\Big] \tag{7}$$

Restricting to an interval of length $L$ with $\phi = 0$ outside, and taking $\phi$ as a column vector, we can write $H = \tfrac{1}{2}\phi^{\mathsf T} K \phi$ ($K$ is the $L\times L$ matrix with $m^2+2$ on the diagonal and $-1$ next to it); $e^{-H}$ is a finite-dimensional Gaussian measure with covariance $K^{-1}$ (1-08). As $L\to\infty$, the $G$ at the centre of the interval converges to (8) below. (8) is obtained by putting $G = Ae^{-\kappa|x|}$ for $x \neq 0$ into the equation $\sum_y K_{xy}G(y) = \delta_{x0}$ of the infinite chain to get $m^2+2 = e^{\kappa}+e^{-\kappa}$, and fixing $A$ by the equation at $x=0$.

$$G(x) = \frac{e^{-\kappa|x|}}{2\sinh\kappa}, \qquad \cosh\kappa = 1 + \frac{m^2}{2}, \qquad \xi = \frac{1}{\kappa} \approx \frac{1}{m} \tag{8}$$

**The smaller the mass, the longer the correlation length**, and it diverges at the rate $1/m$ as $m\to 0$.

Computed For $m = 1,\ 0.2,\ 0.05$, $\xi = 1.03904,\ 5.00831,\ 20.00208$ ($\xi m \to 1$). The $G(20)$ obtained by numerically inverting $K$ at the centre of the interval (1601 to 4001 points) agreed with (8) to 13 digits.

This chain is the imaginary-time harmonic oscillator of 2-07 placed on a lattice in time. Distributing $m^2\phi_x^2$ half and half to the couplings on either side, $e^{-H}$ becomes the product of weights of neighbouring pairs

$$T(\phi,\phi') = \exp\Big(-\frac{(\phi-\phi')^2}{2} - \frac{m^2(\phi^2+\phi'^2)}{4}\Big)$$

We call the integral operator $(Tf)(\phi) = \int T(\phi,\phi')f(\phi')\,d\phi'$ with this $T$ as kernel the **transfer matrix** (the $\sum$ of a matrix product replaced by an integral). For a periodically

joined chain of $L$ points, $Z = \int T(\phi_1,\phi_2)T(\phi_2,\phi_3)\cdots T(\phi_L,\phi_1)\,d\phi_1\cdots d\phi_L$, that is, the trace of $T^L$. The ratio of the two largest eigenvalues of $T$ is $e^{-\kappa}$, and we call this $\kappa$ the **gap** of the transfer matrix. This is the simplest example in which **the inverse of the correlation length is the gap** Known (the general form is in 4-11). Computed The ratio of eigenvalues of the matrix discretising $T$ with 1500 points agreed with $e^{-\kappa}$ to 10 digits at $m = 1,\ 0.5,\ 0.2$.

04

## Phase transitions — kinks that occur only in infinite systems

We call it a **phase transition** when a slight change in temperature or the like changes the behaviour completely. In mathematical language, the $f(\beta, h)$ of the thermodynamic limit is not analytic at that point (cannot be represented by a Taylor expansion around the point). **It does not occur in finite systems.** Precisely, $Z_N$ is a finite sum of exponentials $e^{-\beta H(\sigma)}$ of $\beta, h$, positive for real $\beta, h$. So $\log Z_N$ can be differentiated any number of times and its Taylor expansion converges (this is called being **real-analytic**).

For the Ising model the following is known Known . For $d = 1$ there is no phase transition (Ising 1925, 2-09). For $d = 2$, $h = 0$,

$$\sinh(2\beta_c J) = 1, \qquad \beta_c J = \tfrac{1}{2}\log\bigl(1+\sqrt{2}\bigr) = 0.440687\ldots \quad (9)$$

a phase transition occurs. Assuming there is just one transition, Kramers–Wannier 1941 located it by duality, and Onsager 1944 found the free energy exactly and showed the $-\log|T - T_c|$ divergence of the specific heat. The argument showing that the alignment survives at low temperature is said to begin with Peierls 1936 (put into shape as a proof in later years). What states this precisely is **spontaneous magnetisation**.

$$\bar{m}_0(\beta) = \lim_{h\downarrow 0}\ \lim_{N\to\infty} \langle\sigma_0\rangle_{N,h} \quad (10)$$

**Example.** The order of limits matters. Taking $h \to 0$ first, the $H$ of the finite system is invariant under the global flip $\sigma \to -\sigma$, so $\langle\sigma_0\rangle = 0$. Going to infinity first and then switching off the external field, in two dimensions $\bar{m}_0 > 0$ remains for $\beta > \beta_c$ (going to 0 like $(\beta - \beta_c)^{1/8}$ as $\beta \downarrow \beta_c$; Yang 1952). **That the answer changes with the order of limits** is a recurring issue in Yang–Mills too (5-09, 17-10).

The view of phase transitions through the zeros of the partition function is treated in the detour after section 05.

05

## The mean-field model — watching the tug of war

The **Curie–Weiss model** $H_N = -\frac{J}{2N}M^2 - hM$ ($M = \sum_{x=1}^{N}\sigma_x$, $J>0$), in which every spin is coupled with the same strength, is the smallest example in which a phase transition can be followed by hand. Grouping by the number of up spins $k$ gives $Z_N = \sum_k \binom{N}{k} e^{-\beta H_N}$. Let $\bar m = M/N$ be the magnetisation per spin (with a bar, to distinguish it from the mass $m$ of the Gaussian chain). Using $\binom{N}{N(1+\bar m)/2} = e^{Ns(\bar m)+o(N)}$, $s(\bar m) = -\frac{1+\bar m}{2}\log\frac{1+\bar m}{2} - \frac{1-\bar m}{2}\log\frac{1-\bar m}{2}$ from Stirling's formula, the sum is determined by its largest term (Laplace's method).

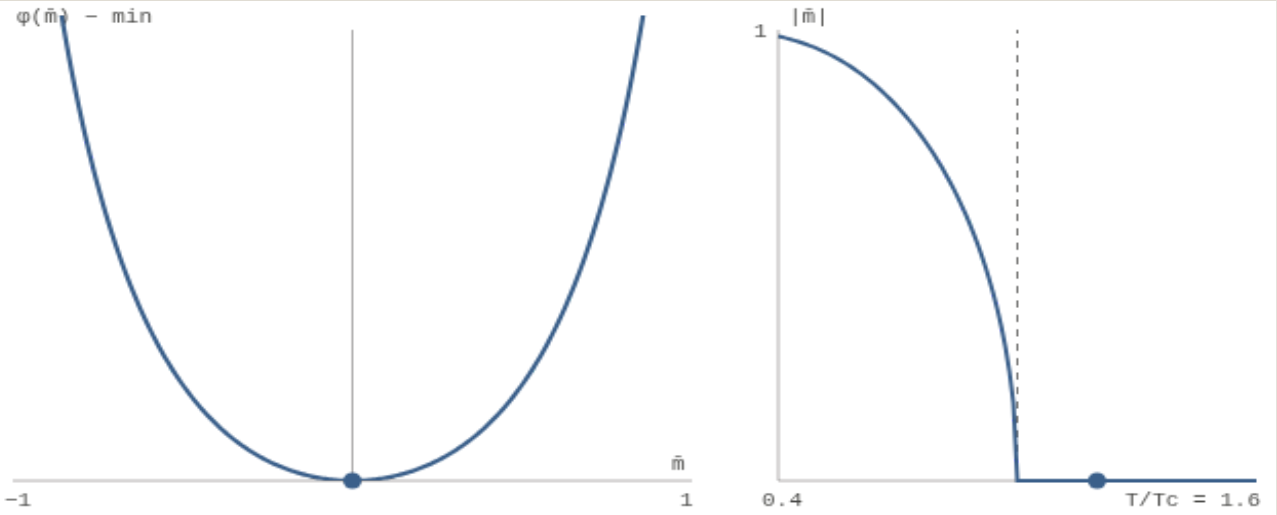
$$f = \min_{-1\le \bar m\le 1} \varphi(\bar m), \qquad \varphi(\bar m) = -\frac{J}{2}\bar m^2 - h\bar m - \frac{1}{\beta}s(\bar m) \tag{11}$$

The tug of war of (2) has shrunk to one variable (the first term is the energy, the last the entropy). $\varphi'(\bar m) = 0$ is

$$\bar m = \tanh\bigl(\beta(J\bar m + h)\bigr) \tag{12}$$

For $h=0$: if $\beta J \le 1$, the only solution is $\bar m = 0$; for $\beta J > 1$, $\pm\bar m^* \neq 0$ are the minima (critical temperature $T_c = J$). From $\tanh u \approx u - u^3/3$, $\bar m^* \approx \sqrt{3(\beta J - 1)}$, and differentiating (12) in $h$ gives $\chi = \beta/(1-\beta J)$ for $T > T_c$ (the Curie–Weiss law).

Figure 1 Moving $T/T_c$ and the external field $h$, the valley of $\varphi(\bar m)$ on the left (minimum shifted to 0) splits in two at $T_c$. On the right are the $\bar m^*$ of $h=0$ (line) and the current minimum. "Play" cools it




T/Tc = 1.20, h = 0.00, minimum m̄ = 0.00000, χ = 5.000

Computed The figure computes (11)(12) ($J=1$) on the spot. The $\bar m^*$ at $h=0$ is $0.52543,\ 0.71041,\ 0.90733$ at $T/T_c = 0.9,\ 0.8,\ 0.6$ (agreeing with python3).

Computed **Table 1** $\sqrt{\langle \bar{m}^2 \rangle}$ ($h = 0$; for finite $N$ the sum over $\binom{N}{k}$ is computed exactly) and the $\bar{m}^*$ of $N = \infty$.

| N | βJ = 0.8 | βJ = 1 | βJ = 1.25 |
|---|---|---|---|
| 1000 | 0.07003 | 0.19172 | 0.70824 |
| $10^5$ | 0.00707 | 0.06083 | 0.71039 |
| ∞ | 0 | 0 | 0.71041 |

$\beta J = 0.8$ goes to 0 like $N^{-1/2}$ ($\sqrt{N}$ times it is 2.2358, against the prediction 2.2361 from (4) and the Curie–Weiss law), $\beta J = 1.25$ goes to $\bar{m}^*$, and the critical $\beta J = 1$ goes to 0 like $N^{-1/4}$ (about 1.08 when multiplied by $N^{1/4}$). **For finite $N$ each is a smooth function of $\beta$, and the kink first appears as $N \to \infty$.**

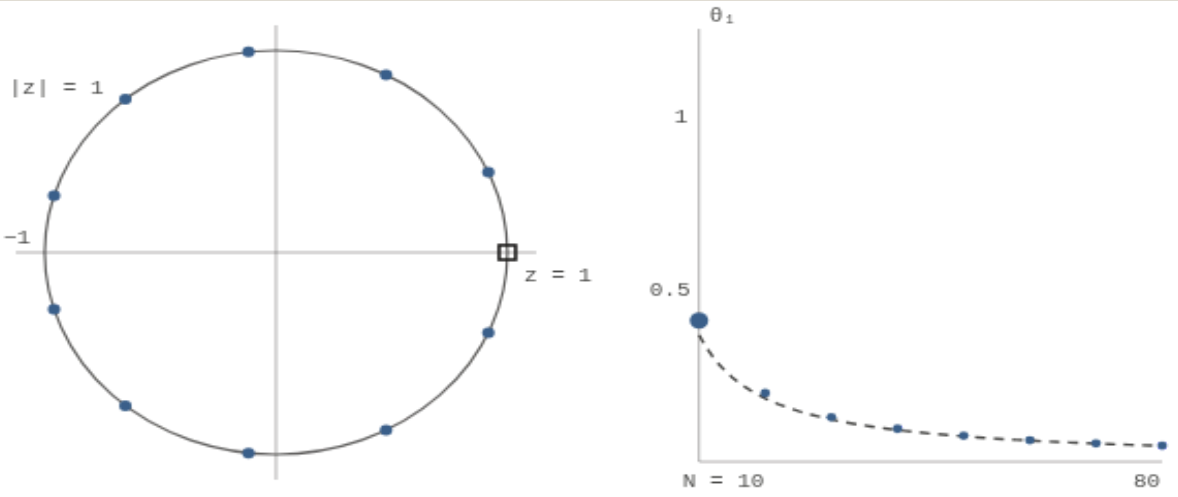
### Detour: zeros of the partition function

A subsection you may skip.

Putting $z = e^{-2\beta h}$, $Z_N$ is a polynomial in $z$ with positive coefficients (multiplied by $e^{\beta h N}$), and for a ferromagnet all its zeros lie on the unit circle $|z| = 1$ (Lee–Yang 1952) Known . Real $h$ corresponds to the positive real axis, and $h = 0$ to $z = 1$. Since the coefficients are positive, for finite $N$ no zero lies on $z = 1$, and $\log Z_N$ is smooth near $h = 0$. If the zeros pinch $z = 1$ as $N \to \infty$, $f$ can kink at $h = 0$.

Computed In the model of section 05, the angle $\theta_1$ of the zero nearest to $z = 1$, from $N = 10$ to 80, was $N\theta_1 = 4.08 \to 3.72$ for $\beta J = 1.5$ (closing in on 0; the prediction $\pi/\bar{m}^* = 3.66$ when $Z_N$ is approximated by a constant multiple of $\cosh(N\beta h \bar{m}^*)$ at low temperature), and $\theta_1 = 1.157 \to 0.722$ for $\beta J = 0.5$ (consistent with the prediction 0.571 coming from the point where, with $h$ made complex, the two solutions of (12) merge into one).

Figure 2 Moving $N$ increases the zeros on the unit circle (dots; upper and lower are complex conjugates). At $\beta J = 1.5$ (low temperature) they close in on $z = 1$ ($h = 0$, square); at $\beta J = 0.5$ (high temperature) an arc with no zeros remains around $z = 1$ (the hatching is the prediction $|\theta| < 0.571$ for $N \to \infty$). On the right (below, on narrow screens) are $\theta_1$ and the prediction (dashed). "Play" takes $N$ from 10 to 80



```
βJ = 1.5, N = 10, θ₁ = 0.40767, Nθ₁ = 4.077 (predicted limit of Nθ₁: π/m̄* = 3.659)
```

Computed The zeros in the figure are angles found with python3 and embedded. We followed the changes of sign of $z^{-N/2}Z_N$, which is real on $z = e^{i\theta}$, with 80-digit decimal arithmetic, and checked that there are exactly $N/2$ of them in $0 < \theta < \pi$ ($\theta_1$ agrees with the values above to 6 digits).

06

## Critical points — the correlation length diverges and powers appear

We call the point of a transition in which the spontaneous magnetisation rises continuously from 0 a **critical point**. Approaching it, the correlation length diverges, and many quantities behave as powers. Their exponents are the **critical exponents**. With $t = (T - T_c)/T_c$,

$$\xi \sim |t|^{-\nu}, \quad \chi \sim |t|^{-\gamma}, \quad \bar{m}0 \sim (-t)^{\beta}, \quad G(x)|_{T=T_c} \sim \frac{1}{|x|^{d-2+\eta}} \tag{13}$$

we write ($\sim$ means the ratio approaches a constant. The exponent $\beta$ is a different thing from the inverse temperature. $\eta$ is written $\eta_s$ in later chapters when it must be distinguished from the dynamical exponent: 18-03 § 07. The relaxation time and the dynamical exponent $z$ are in 2-10). From the scaling hypothesis the following relations among the exponents follow (the renormalisation group explains them; equations (1.38)–(1.40) of Pelissetto–Vicari 2002) Physics , and the right two are held to hold for $d < 4$.

$$\gamma = \nu(2 - \eta), \qquad \beta = \frac{\nu}{2}(d - 2 + \eta), \qquad \alpha = 2 - d\nu \tag{14}$$

$\alpha$ is the exponent of the specific heat. Putting in the $\nu = 1,\ \eta = 1/4$ of two-dimensional Ising gives back the exact values $\gamma = 7/4,\ \beta = 1/8,\ \alpha = 0$ Computed .

Computed **Table 2** Critical exponents. Mean-field and two-dimensional values are from Pelissetto–Vicari (Table 13 and elsewhere), the three-dimensional $\nu, \eta$ are numerical values from the conformal bootstrap (Kos et al. 2016), and * is computed from (14) (errors move it by about $\pm 1$ in the last digit).

| Exponent | Mean field (d > 4) | Two-dimensional Ising | Three-dimensional Ising Physics |
|---|---|---|---|
| ν (correlation length) | 1/2 | 1 | 0.62997 |
| η (correlation at the critical point) | 0 | 1/4 | 0.036297 |

| Exponent | Mean field (d > 4) | Two-dimensional Ising | Three-dimensional Ising Physics |
|---|---|---|---|
| γ (susceptibility) | 1 | 7/4 | 1.23708 |
| β (spontaneous magnetisation) | 1/2 | 1/8 | 0.32642 |
| α (specific heat) | 0 | 0 (logarithmic) | 0.11009 |

**Example.** In the Gaussian chain, the squared mass $m^2$ (section 03) plays the role of $t$, and $\nu = 1/2$ follows from $\xi \approx (m^2)^{-1/2}$. Together with the $\beta = 1/2,\ \gamma = 1$ of section 05, this is the same as the mean-field column. $\eta$ measures the departure from Gaussian behaviour (Pelissetto–Vicari), and for three-dimensional Ising it is small but not 0.

07

## Universality — forgetting the details

The exponents of Table 2 are thought not to depend on the shape of the lattice or the details of the couplings, but only on the dimension, the symmetry and the range of the interaction. A collection of systems with the same exponents is a **universality class** Physics . The gas–liquid critical point of simple fluids and uniaxial antiferromagnets are held to belong to the three-dimensional Ising class (Pelissetto–Vicari § 3.1), so that liquids and magnets have the same exponents.

Near a critical point $\xi$ becomes much longer than the lattice spacing. Kadanoff 1966 regrouped spins into blocks larger than the lattice spacing and smaller than the correlation length and looked again, and Wilson 1971 turned the repetition of this operation into the equations of the renormalisation group. Details fade, and systems that flow into the same fixed point have the same exponents. Precisely, universality is a principle of physics confirmed in many examples, not a general theorem (7-05, 4-04).

As the smallest example, we add to (7) a coupling $c \geq 0$ to the next-but-one neighbour.

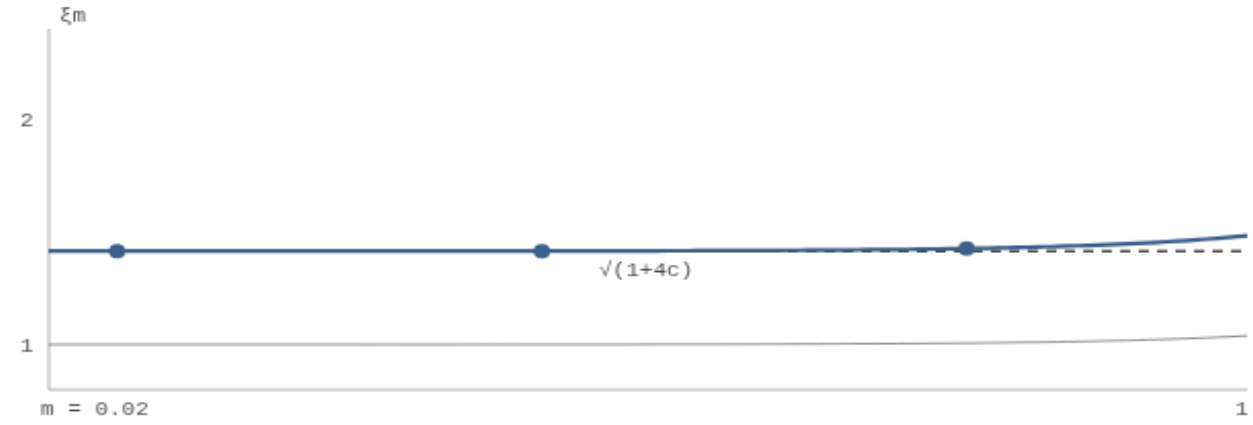
$$H_c = \frac{1}{2} \sum_x \left[ (\phi_{x+1} - \phi_x)^2 + c\,(\phi_{x+2} - \phi_x)^2 + m^2 \phi_x^2 \right] \tag{15}$$

By the same computation as section 03, $G$ becomes a sum of terms of the form $e^{-\kappa|x|}$, and the $\kappa$ of the most slowly decreasing term is the smallest positive root of $m^2 = 2(\cosh\kappa - 1) + 2c(\cosh 2\kappa - 1)$. This is the quadratic equation $4cw^2 + 2w - (2 + 4c + m^2) = 0$ in $w = \cosh\kappa$, so the root can be written by a formula. When $\kappa$ is small the right side is about $(1 + 4c)\kappa^2$, so $\xi \approx \sqrt{1 + 4c}/m$.

Computed **Table 3** $\xi m$ for the chain with $c$ added. The limit is $\sqrt{1+4c} = 1,\ 1.41421,\ 2.23607$ (also agreeing with the inverse matrix to 10 digits).

| m | c = 0 | c = 1/4 | c = 1 |
|---|---|---|---|
| 0.4 | 1.00659 | 1.42586 | 2.24613 |
| 0.1 | 1.00042 | 1.41495 | 2.23670 |
| 0.025 | 1.00003 | 1.41426 | 2.23611 |

Figure 3 Moving the next-but-one coupling $c$ changes the height of the limit $\sqrt{1+4c}$ (dashed) of the curve of $\xi m$ (horizontal axis $m$, logarithmic scale; the dots are the $m$ of Table 3). That the curve becomes flat as $m \to 0$ ($\xi \propto m^{-1}$) does not depend on $c$. The thin line is $c = 0$. "Play" takes $c$ from 0 to 1




c = 0.25, √(1+4c) = 1.41421, ξm: m = 0.4 → 1.42586, 0.1 → 1.41495, 0.025 → 1.41426

Computed The figure computes the root of the quadratic equation above on the spot. At $c = 0,\ 1/4,\ 1$ it agrees with Table 3 (and with python3's root search and inverse matrix).

**The exponent ($\xi \propto m^{-1}$) does not depend on $c$, but the coefficient $\sqrt{1+4c}$ does.** For a slowly varying field $(\phi_{x+2} - \phi_x)^2 \approx 4(\phi_{x+1} - \phi_x)^2$, so the next-but-one coupling looks merely like the neighbour coupling multiplied by $1 + 4c$, and disappears when the ruler of length is changed.

08

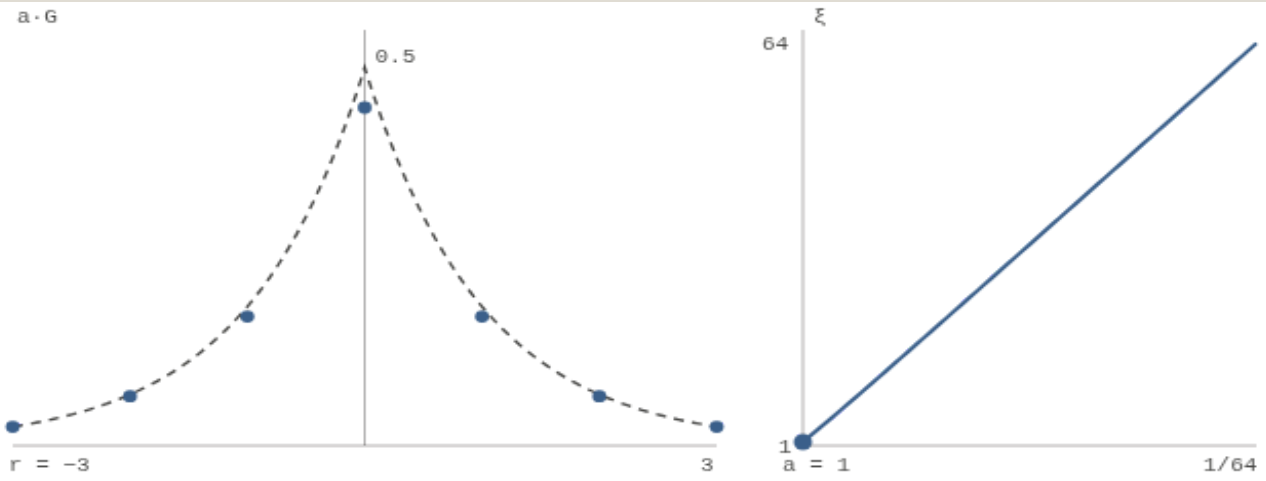
## The continuum limit — the critical point is where the lattice disappears

Introduce the lattice spacing $a$ and read the point $x$ as the position $r = xa$. If the correlation length in lattice units is $\xi$ (the $1/\kappa$ of section 03), the correlation length in physical units is $a\xi$. Keeping $a\xi$ fixed and letting $a \to 0$, $\xi$ has no choice but to go to infinity. **Taking the continuum limit is approaching a critical point** (Chatterjee 2018 § 5) Physics . For the Gaussian chain everything is explicit: fixing $M > 0$ and putting $m = Ma$, from $\kappa/a \to M$, $a/(2\sinh\kappa) \to 1/(2M)$ the following holds Known .

$$a\,G\Big(\frac{r}{a}\Big) \;\longrightarrow\; \frac{e^{-M|r|}}{2M} \qquad (a \to 0) \quad (16)$$

The right side is the kernel of the inverse of $-d^2/dr^2 + M^2$ (the propagator, 1-11), and $M$ is the mass in physical units.

Figure 4 Making the lattice spacing $a$ smaller, the $aG$ in physical units (dots) settles onto the continuous $e^{-M|r|}/2M$ (dashed, $M = 1$), and the correlation length in lattice units $\xi$ on the right (log–log) diverges. "Play" takes $a$ from 1 to 1/64




a = 1.00, ξ = 1.039, aξ = 1.0390, a·G(0) = 0.44721 (continuum 0.5)

Computed The figure computes (8)(16) on the spot. **Table 4** $\xi$ (in lattice units) and $aG$ (python3) as $a$ is shrunk with $M = 1$. The bottom row is the right side of (16).

| a | ξ (lattice units) | a·G(0) | a·G(r = 1) |
|---|---|---|---|
| 1 | 1.0390 | 0.447214 | 0.170820 |
| 1/8 | 8.0052 | 0.499026 | 0.183701 |
| 1/64 | 64.0007 | 0.499985 | 0.183936 |
| Continuum | – | 0.5 | 0.183940 |

For Yang–Mills, in many non-abelian lattice gauge theories the critical point is believed to be at $\beta = \infty$ (the $\beta$ here is the constant multiplying the action of lattice gauge theory, proportional to the inverse square of the coupling constant $g$, and plays the role of inverse temperature; 5-01), and the inverse of the correlation length corresponds to the mass of the lightest glueball (a particle made of gluons alone, 17-01) (Chatterjee 2018 § 5) Physics . One of the questions on the lattice side is "that correlations decrease exponentially for each $\beta > 0$ ($\xi(\beta)$ is determined in $(0, \infty)$), and that $\xi(\beta) \to \infty$ as $\beta \to \infty$", which Chatterjee lists in the same section as the lattice correlation-length problem (Problem 5.1) (5-09). Whether the mass stays positive in the continuum limit is a further question (5-05, 4-11). Weights, the correlation length, the critical point and universality correspond to the path integral of field theory (4-02), the inverse of the mass, the continuum limit and renormalisation, and for the Gaussian chain this is exact.

## What this chapter can say

| Content | Status |
|---|---|
| (1)–(4), analyticity of finite systems, (8) and the gap of the transfer matrix, the $\xi$ of (15), (16), (11)(12) | Said Known (Tables 1, 3, 4 and Figures 1, 3, 4 Computed ) |
| Existence of the thermodynamic limit | Theorem Known (literature bibliographic data only) |
| The facts for $d = 1, 2$ of the Ising model, the Lee–Yang theorem | Theorem Known (literature checked up to the abstracts; Ising 1925 and Peierls 1936 bibliographic data only) |
| The angles of the zeros in the detour (Figure 2) | Numbers for $N \leq 80$ Computed . The limits are predictions |
| Three-dimensional critical exponents, (14), universality in general | Numbers and the level of physics Physics (the * of Table 2 is arithmetic from (14) Computed ) |
| The correspondence between correlation length and mass, the critical point $\beta = \infty$ of lattice gauge theory | At the level of physics Physics . Nothing is shown about the mass gap or the construction of the continuum limit |

## Sources and reproduction

| Item | Kind | Source / tools |
|---|---|---|
| Tables 1–4, the numbers in section 03 and the detour, the ratio of eigenvalues of the transfer matrix, the check values for Figures 1–4 and the zeros of Figure 2 | Computed | python3 (NumPy, SciPy; the zeros with 80-digit decimal arithmetic from the standard library). The figures by JavaScript in the browser |
| The form of (6), the continuum limit and the critical point, $\beta_c = \infty$, glueballs, the lattice correlation-length problem (Chatterjee's Problem 5.1) | Known | S. Chatterjee, arXiv:1803.01950. § 5 checked |
| (13)(14), mean-field and two-dimensional exponents, universality classes | Known | A. Pelissetto, E. Vicari, Phys. Rep. 368 (2002) 549–727, arXiv:cond-mat/0012164. § 1.1, § 1.2, § 3.1 and Table 13 checked |
| The $\nu, \eta$ of three-dimensional Ising | Physics | F. Kos, D. Poland, D. Simmons-Duffin, A. Vichi, JHEP 08 (2016) 036, arXiv:1603.04436. Abstract and introduction checked |
| (9), specific heat / duality / spontaneous magnetisation / zeros | Known | L. Onsager, Phys. Rev. 65 (1944) 117 / H. A. Kramers, G. H. Wannier, ibid. 60 (1941) 252 / C. N. Yang, ibid. 85 (1952) 808 / C. N. Yang, T. D. Lee, ibid. 87 (1952) 404 / T. D. Lee, C. N. Yang, ibid. 410. Abstracts checked |

| Item | Kind | Source / tools |
|---|---|---|
| Existence of the thermodynamic limit | Known | D. Ruelle, Statistical Mechanics: Rigorous Results (first edition 1969, reissued by World Scientific 1999) (bibliographic data only) |
| The view of grouping into blocks / the renormalisation group | Known | L. P. Kadanoff, Physics 2 (1966) 263 / K. G. Wilson, Phys. Rev. B 4 (1971) 3174. Abstracts checked |
| No phase transition in one dimension / the low-temperature argument | Known | E. Ising, Z. Phys. 31 (1925) 253 (bibliographic data only) / R. Peierls, Proc. Cambridge Philos. Soc. 32 (1936) 477 (bibliographic data only, not checked) |
| (2), Laplace's method, the covariance of Gaussian measures, the gap of the transfer matrix | Known | Standard facts (1-08, 2-07, 2-09, 4-11) |

Next chapter to read: 2-09 Detour: the gap and the correlation length in the Ising model. For the relation to the mass gap see 4-11; for the relaxation time and the dynamical exponent, 2-10.

# Detour: the gap and the correlation length in the Ising model — the ratio of eigenvalues of the transfer matrix sets the length over which correlations reach

> 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** — That the ratio of the two largest eigenvalues of the transfer matrix sets the correlation length, and $1/\xi$ is exactly the gap of $-\log T$. The three behaviours of the gap on a two-dimensional strip: high temperature, the critical point, low temperature. That the gap which closes at low temperature is not a sign that there is no mass gap.

**Prerequisite chapters** — 2-08 (the correlation length, critical points). Sections 04 and 08 are easier if you have read 2-05 (the gap, two levels, the double well) and 2-07 (imaginary time, the transfer matrix).

**Said first** — This is a detour, so you may skip it. The two-dimensional numbers are computations on strips of width up to 20. This chapter shows nothing about the mass gap of Yang–Mills.

Order of this chapter

1. The overall picture — the inverse of the correlation length is the gap
2. The one-dimensional Ising model and the transfer matrix
3. Correlations decrease by the ratio of eigenvalues (Figure 1, Table 1)
4. The transfer matrix is quantum mechanics in imaginary time
5. Two dimensions: the transfer matrix of a strip (Table 2)
6. The high-temperature side — the gap remains
7. The critical point (Figure 2) — closes like $1/L$
8. The low-temperature side (Table 3) — a gap that closes and a gap that remains
9. Back to the language of Yang–Mills (Detour: $Z_2$ gauge theory)

01

## The overall picture — the inverse of the correlation length is the gap

The mass gap was the difference between the two lowest eigenvalues of the Hamiltonian (2-05), and the correlation length was the length over which the joint movement of two distant points disappears (2-08). We check in the simplest model that the two are a single number.

The key is the transfer matrix. If the eigenvalues of the matrix $T$ that advances the lattice one column at a time are $\lambda_0 > \lambda_1 \geq \cdots$, $E_n = -\log \lambda_n$, then

$$\frac{1}{\xi} = \log \frac{\lambda_0}{\lambda_1} = E_1 - E_0 \quad (1)$$

(the conditions are in section 04). The left side is a statistical quantity, the right side is the gap of the "Hamiltonian" $H = -\log T$, and reading one tick as one step of imaginary time, this is the same content as "the rate of decay is the gap" stated by (8) of 2-05.

Example. On a two-dimensional strip of width $L$, $E_1 - E_0$ settles to a positive value on the high-temperature side, is proportional to $1/L$ at the critical point, and closes exponentially on the low-temperature side (sections 05–08).

02

## The one-dimensional Ising model and the transfer matrix — the partition function is the trace of a matrix

Arrange $N$ spins $\sigma_i = \pm 1$ on a ring ($\sigma_{N+1} = \sigma_1$), coupling neighbours. With $K = \beta J$, $B = \beta h$, and distributing the external-field term half and half to neighbouring pairs, the weight is a product of factors for each neighbouring pair.

$$e^{-\beta H} = \prod_{i=1}^{N} \exp\Big( K\sigma_i\sigma_{i+1} + \frac{B}{2}(\sigma_i + \sigma_{i+1}) \Big) \quad (2)$$

Each factor is determined by $\sigma_i$ and $\sigma_{i+1}$ alone, so it becomes a $2 \times 2$ matrix (rows are $\sigma_i$, columns $\sigma_{i+1}$, $+1$ is the first row), and since the sum goes once round the ring and returns, it is a trace.

$$T = \begin{pmatrix} e^{K+B} & e^{-K} \\ e^{-K} & e^{K-B} \end{pmatrix}, \qquad Z_N = \sum_{\sigma} \prod_{i} T_{\sigma_i \sigma_{i+1}} = \operatorname{tr} T^N = \lambda_0^N + \lambda_1^N \quad (3)$$

$T$ is real symmetric, so its eigenvalues are real (1-01), and

$$\lambda_{0,1} = e^{K}\cosh B \pm \sqrt{e^{2K}\sinh^{2} B + e^{-2K}} \tag{4}$$

For large $N$, $Z_N \approx \lambda_0^{N}$. The method of finding the partition function from the largest eigenvalue of a matrix was carried out by Kramers–Wannier 1941 for the one-dimensional chain, and taken on to two dimensions Known .

Example. Computed For $N = 14$, $K = 0.7$, $B = 0.2$, the $Z_N$ obtained by summing all $2^{14}$ configurations, and the correlations, agreed with the trace formula (3) within rounding error (differences in correlations below $10^{-13}$).

03

## Correlations decrease by the ratio of eigenvalues — there is no phase transition in one dimension

Write the orthonormal eigenvectors of $T$ as $v_0, v_1$ and the spin as $S = \mathrm{diag}(1, -1)$. In $N \to \infty$ the rest of the ring is replaced by $\lambda_0^{N-r} v_0 v_0^{\mathsf{T}}$, and expanding the $T^{r}$ in between in eigenvalues gives the correlation at distance $r$.

$$m = v_0^{\mathsf{T}} S v_0, \qquad G(r) = \langle \sigma_0 \sigma_r \rangle - m^{2} = (v_0^{\mathsf{T}} S v_1)^{2}\left(\frac{\lambda_1}{\lambda_0}\right)^{r} = (1 - m^{2})\left(\frac{\lambda_1}{\lambda_0}\right)^{r} \tag{5}$$

The last equality follows because putting $v_0 = (\cos\varphi, \sin\varphi)$, $v_1 = (-\sin\varphi, \cos\varphi)$ gives $m = \cos 2\varphi$, $v_0^{\mathsf{T}} S v_1 = -\sin 2\varphi$. The correlation length is as in (1). For $B = 0$, $\lambda_0 = 2\cosh K$, $\lambda_1 = 2\sinh K$, and

$$G(r) = (\tanh K)^{r}, \qquad \xi = \frac{1}{\log\coth K} \approx \frac{e^{2K}}{2} \quad (K \to \infty) \tag{6}$$

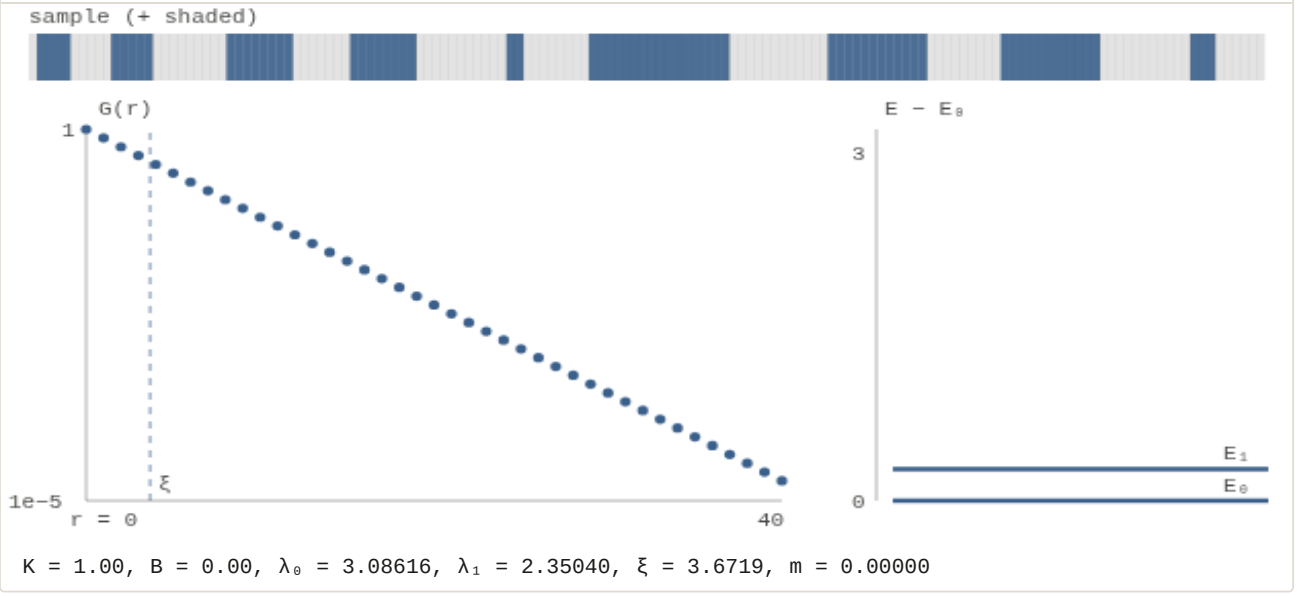
Computed **Table 1** Eigenvalues and correlation lengths for $B = 0$.

| $K$ | $\lambda_0$ | $\lambda_1$ | $\xi$ | $e^{2K}/2$ |
|---|---|---|---|---|
| 1 | 3.086161 | 2.350402 | 3.671861 | 3.6945 |
| 2 | 7.524391 | 7.253721 | 27.296022 | 27.2991 |
| 3 | 20.135324 | 20.035750 | 201.713984 | 201.7144 |

For every $K < \infty$, $\lambda_0 > \lambda_1$, so $\xi$ is finite, and switching off the external field gives $m = \sinh B / \sqrt{\sinh^{2} B + e^{-4K}} \to 0$. The quantity under the square root in (4) is positive, so $\lambda_0$ does not coincide with $\lambda_1$, and the free energy per site $f$ is $\beta f = -\lim_{N\to\infty} \frac{1}{N}\log Z_N = -\log\lambda_0$, an analytic function of $K$ and $B$. This is the reason, in terms of the transfer matrix, why there is no phase transition in one dimension (Ising 1925; for the

definition of a phase transition see section 04 of 2-08) Known . Even for a general square matrix with all entries positive, the largest eigenvalue does not coincide with another; this is called the Perron–Frobenius theorem (here we give only the name). Intuitively, the fraction of neighbours that disagree (the fraction of walls) $p = \frac{1-\tanh K}{2} \approx e^{-2K}$ does not become 0 however low the temperature, and the correlation changes sign each time it crosses a wall, so $\xi \approx 1/(2p)$.

Figure 1 Moving K and B (= βh) changes together the lengths of the clusters in the sample chain at the top, the slope of G(r) on the left (logarithmic scale), and the level spacing $E_1 - E_0 = 1/\xi$ on the right. "Play" increases K (cools)




K = 1.00, B = 0.00, λ₀ = 3.08616, λ₁ = 2.35040, ξ = 3.6719, m = 0.00000

Computed $G(r)$ and the levels are computed on the spot from (4)(5); the chain is a sample of the Markov chain built from $v_0$ (random numbers with a fixed seed). The readout agrees with python3 ($\xi = 2.9563$ at $K = 1$, $B = 0.1$).

04

## The transfer matrix is quantum mechanics in imaginary time — the two levels of one spin

The $T$ of $B = 0$ is the sum of $e^{K}$ times the identity matrix and $e^{-K}$ times $\sigma^x = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ (1-04). Comparing with $e^{K^*\sigma^x} = \cosh K^* + \sigma^x \sinh K^*$, it can be written as

$$T = \sqrt{2\sinh 2K}\; e^{K^*\sigma^x}, \qquad \tanh K^* = e^{-2K} \quad \left(\text{equivalent form } \sinh 2K \,\sinh 2K^* = 1\right)$$

We call $K^*$ the dual coupling (the quantity of Kramers–Wannier duality). $T = e^{-H}$, $H = -K^*\sigma^x + \text{const.}$ is a single spin in a transverse magnetic field, that is, a two-level system (2-05), and its gap is $E_1 - E_0 = 2K^* = \log\coth K = 1/\xi$. The direction of the chain is imaginary time, and one tick is a step of 1 (2-07). The stronger the coupling (the lower the temperature), the smaller $K^*$, the closer the two levels, and the longer the correlations persist.

In general, if $T$ is a positive symmetric matrix (in infinite dimensions, a positive self-adjoint operator) with eigenvectors $|n\rangle$, the correlation of an observable $A$ is, by the same computation,

$$\langle A_0 A_r\rangle - \langle A\rangle^2 = \sum_{n\geq 1} |\langle 0|A|n\rangle|^2\, e^{-(E_n-E_0)r} \tag{8}$$

Known . Example. A level whose coefficient $\langle 0|A|n\rangle$ is 0 by symmetry does not appear in the correlation of that $A$. The gap is "the slowest decrease, valid for every observable", and individual correlations decrease at least that fast (the remark after Table 2 of 2-05). That the transfer matrix of lattice gauge theory is positive self-adjoint follows from reflection positivity Known (said to be Osterwalder–Seiler 1978; details in 5-03).

05

## Two dimensions: the transfer matrix of a strip — a matrix of order 2^L

Put the two-dimensional Ising model on the square lattice ($B=0$, $K$ both vertically and horizontally) on a strip of width $L$. The width direction is joined periodically (a cylinder), and the length direction is read as imaginary time. One column has $2^L$ states; writing the couplings within a column as the diagonal matrix $V_1 = \exp(K\sum_j \sigma_j\sigma_{j+1})$ and the couplings to the next column as $V_2$, using (7) at each point gives

$$T = V_1^{1/2} V_2 V_1^{1/2}, \qquad V_2 = (2\sinh 2K)^{L/2} \exp\Bigl(K^* \sum_{j=1}^{L} \sigma_j^x\Bigr) \tag{9}$$

Here $\sigma_j^x$ is the matrix of order $2^L$ that flips only the $j$-th spin, and $\sigma_j^z$ is the diagonal matrix that multiplies by the sign $\pm 1$ of the $j$-th spin; neither does anything to the other spins. The exponent of $V_1$ is the sum of $\sigma_j^z\sigma_{j+1}^z$, and the exponent of $V_2$ the sum of $\sigma_j^x$. Putting half of $V_1$ on each side makes $T$ a symmetric matrix, so (8) of section 04 can be used. $T$ has the form of one step of imaginary time for a system of quantum spins in a row (the transverse-field Ising chain). Two-dimensional classical statistical mechanics has become a one-dimensional quantum system.

$T$ commutes with the global flip $\sigma \to -\sigma$, so the eigenvectors split into even and odd; the largest even one is $\lambda_0$, the largest odd one $\lambda_1$, and the correlation length of the strip is $\xi_L = 1/(E_1 - E_0)$ (equation (1) of de Queiroz 2011). The critical point $\sinh 2K_c = 1$ ((9) of 2-08) is the self-dual point where, by (7), $K^* = K$.

Computed **Table 2** Level differences on strips of width $L$ ($E_2$ is the third level). The bottom row gives the values of the formulas for $L\to\infty$ (sections 06–08).

| $L$ | $T/T_c = 1.5$<br>$E_1 - E_0$<br>To Figure 2 | **1.2**<br>$E_1 - E_0$<br>To Figure 2 | **1.0**<br>$\xi_L / L$<br>To Figure 2 | **0.8**<br>$E_1 - E_0$<br>To Figure 2 | **0.8**<br>$E_2 - E_0$<br>To Figure 2 |
|---|---|---|---|---|---|
| 8 | 0.666788 | 0.326116 | 1.264745 | 7.731e-3 | 1.122679 |
| 12 | 0.665579 | 0.314285 | 1.269548 | 1.173e-3 | 0.969005 |
| 16 | 0.665513 | 0.311692 | 1.271177 | 1.924e-4 | 0.907939 |
| 20 | – | 0.311069 | 1.271924 | 3.287e-5 | 0.878022 |
| Formula | 0.665509 | 0.310856 | 1.273240 | 0 | 0.821788 |

$L = 20$ (order $2^{20} \approx 10^6$) too was found without building the matrix, only by applying the actions of $V_1$ and $V_2$ in turn. At the three temperatures the behaviour is completely different.

06

## The high-temperature side — the gap remains, with a band of particles above it

For $T > T_c$ ($K < K^*$), $E_1 - E_0$ settles to a positive value even as $L$ is widened. The eigenvalues of the strip have a closed form (Onsager 1944; summarised in the review of Domb 1960, with the form of the formulas in equations (6)–(8) of de Queiroz 2011), and its building block is the following $\gamma(\omega)$. The right-hand form follows from $\sinh 2K \sinh 2K^* = 1$.

$$\cosh \gamma(\omega) = \cosh 2K^* \cosh 2K - \sinh 2K^* \sinh 2K \cos \omega = \cosh \gamma(0) + (1 - \cos \omega) \quad (10)$$

As $\gamma(0) = 2|K^* - K|$, the $E_1 - E_0$ of Table 2 approaches $2(K^* - K)$ Computed (difference $4 \times 10^{-6}$ at $T/T_c = 1.5$, $L = 16$). $\gamma(\omega)$ can be read as the energy of a single particle of momentum $\omega$ (the free-fermion view; Schultz–Mattis–Lieb 1964). On the strip the momentum is quantised to $2\pi n/L$, and the second odd level is close to $\gamma(2\pi/L)$, agreeing better the further from $T_c$ (at $L = 16$: 0.763461 and 0.763457 for $T/T_c = 1.5$, 0.495433 and 0.494596 for 1.2) Computed . As $L \to \infty$ these come down to $\gamma(0)$, and above the gap there is a continuous band (a real example of section 06 of 2-05).

Example. Near the critical point, when $\gamma, \omega$ is small, expanding the right-hand form of (10) to second order gives

$$\gamma(\omega)^2 \approx \gamma(0)^2 + \omega^2 \quad (11)$$

which has the relativistic form $E^2 = m^2 + p^2$ (speed of light 1, 2-04). Precisely, (11) is an approximation only when $\gamma(0)$ and $\omega$ are small (for $T/T_c = 1.05$, $\omega = 0.1$: left side 0.017234, right side 0.017263) Computed .

07

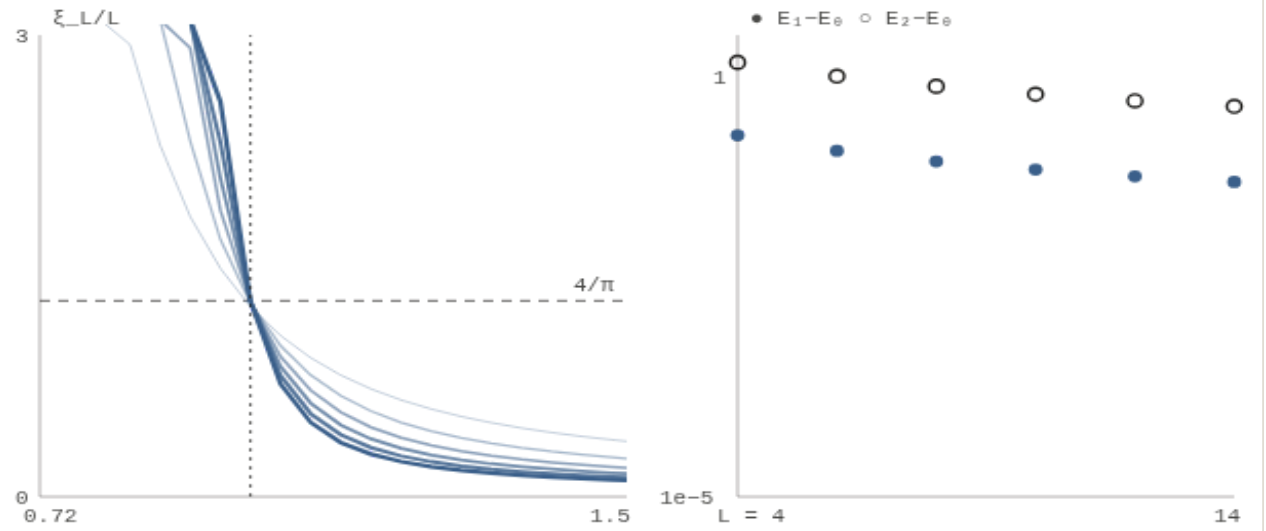
## The critical point — the gap closes like 1/L

At $T = T_c$, $\gamma(0) = 0$, and the gap of an infinitely wide strip is 0. On a finite strip $\xi L$ grows in proportion to $L$, and the constant of proportionality is a universal quantity independent of the details of the model. The first formula of (12) below (a general relation) is a prediction from conformal invariance (the property, at a critical point, of not changing form even under transformations that preserve angles; here we give only the name) Physics (Cardy 1984; the form is equations (3)(5) of de Queiroz 2011). For the two-dimensional Ising model, the same value $4/\pi$ comes out of the closed form of the eigenvalues of the strip in section 06 Known .

$$E_1 - E_0 \approx \frac{2\pi x}{L}, \qquad \frac{\xi L}{L} \longrightarrow \frac{1}{2\pi x} = \frac{4}{\pi} = 1.273240 \quad \left(x = \tfrac{1}{8}\right) \quad (12)$$

$x$ is the scaling dimension of the spin, with $G(r) \sim r^{-2x}$ at the critical point; in two dimensions it is half of $\eta = 1/4$ of 2-08. Example. Computed Extrapolating from the $\xi L/L$ of $L = 18,\ 20$, assuming the difference vanishes like $c/L^2$, gives 1.273248, differing from $4/\pi$ by $8 \times 10^{-6}$.

Figure 2 Moving $T/T_c$ changes the vertical line on the left, and $E_1 - E_0$ (●) and $E_2 - E_0$ (○) for each width L on the right (logarithmic scale). On the left are $\xi_L/L$ (L = 4–14, darker for wider) and 4/π (dashed). On the right, the dashed lines are the formulas for $L \to \infty$, and the dotted line is the low-temperature slope σ. "Play" cools. The temperature can also be chosen with "To Figure 2" in Table 2




T/Tc = 1.00, L = 14: E₁−E₀ = 0.0562 (ξ_L = 17.8), E₂−E₀ = 0.4469, L→∞: ξ_L/L → 4/π = 1.2732

Computed The numbers in the figure were computed on this machine ($L = 4$–14, $T/T_c = 0.72$–1.5 in steps of 0.04 (0.02 for the last only); python3 values embedded to 4 significant figures). The dashed and dotted lines on the right compute the values of (10)(13) on the spot.

The lines on the left gather almost at one point at $T_c$ even for different widths, and open downwards on the high-temperature side and upwards on the low-temperature

side, the larger $L$.

08

## The low-temperature side — a gap that closes, and a gap that remains

For $T < T_c$, $E_1 - E_0$ becomes exponentially small with $L$. On the strip there are two states, "almost all $+$" and "almost all $-$", and the largest even and odd eigenvectors are their sum and difference. To move from one to the other needs a wall of length $L$ across the strip, whose weight is of order $e^{-\sigma L}$. $\sigma$ is the free energy per unit length of the wall (the interface tension), and for the two-dimensional Ising model

$$\sigma = 2(K - K^*) \qquad (K > K_c) \quad (13)$$

is held to hold (the free energy of the boundary between orders of opposite direction was found by Onsager 1944). (13) becomes 0 at $T_c$ (the self-dual point $K = K^*$), and as $K \to \infty$ it approaches $2K$, one disagreeing bond per unit length of wall (the weight $e^{K}$ becomes $e^{-K}$). The relation (7) also holds with $K$ and $K^*$ exchanged ($\tanh K = e^{-2K^*}$), so an equivalent form is $\sigma = 2K + \log\tanh K$. This is the same picture as the double well of section 07 of 2-05. Precisely, what corresponds to the height of the hill is the free energy of the wall $\sigma L$, which grows in proportion to the width.

On the other hand, the difference from the third level $E_2 - E_0$ does not close, and is held to approach two particles' worth of (10), $2\gamma(0) = 4(K - K^*)$; the extrapolation of Table 3 does not contradict this Physics. It is two particles' worth because, in one column joined periodically in the width direction, a disagreement of direction returns to the original after going once round, so only an even number of them fit, and one disagreement corresponds to one particle of (10).

Computed **Table 3** $T/T_c = 0.8$. The slope is $\frac{1}{2}\log\frac{(E_1-E_0)L-2}{(E_1-E_0)L}$; the correction assumes the prefactor $L^{-1/2}$; the extrapolation is from the $E_2 - E_0$ of $L - 2$ and $L$ (Table 2), assuming the difference vanishes like $c/L^2$.

| $L$ | Slope | Corrected slope | Extrapolation of $E_2 - E_0$ |
|---|---|---|---|
| 12 | 0.46438 | 0.41880 | 0.838971 |
| 16 | 0.44854 | 0.41515 | 0.827635 |
| 20 | 0.43991 | 0.41357 | 0.824247 |
| Formula | 0.410894 | 0.410894 | 0.821788 |

The slope moves towards $\sigma$, and the extrapolation towards $4(K - K^*)$. The prefactor and the form of the correction are assumptions, so this is a check at the level of not contradicting (13). Conversely, fixing $\sigma$ at the value of (13), taking the prefactor as $L^{-p}$,

and solving back for $p$ from two neighbouring $L$, we get 0.587, 0.564, 0.551 at $L = 12$, 16, 20, approaching $1/2$ Computed . Finite-size effects of first-order transitions on a finite strip (cylinder) are said to have been discussed by Privman–Fisher 1983. Lowering $T/T_c$ below 1 on the right of Figure 2, you can see the ● line up parallel to the dotted line of slope $\sigma$.

This is an important reading. If the width $L$ is finite, every entry of $T$ is positive (the entries of $V_2$ are products of $e^{\pm K}$), so by the Perron–Frobenius theorem named in section 03, $\lambda_0 > \lambda_1$, that is, $E_1 - E_0 > 0$ (as in section 03, there is no phase transition). $E_1 - E_0 \to 0$ is the sign that in the limit where the width $L$ is widened to infinity there are two ordered states, $+$ and $-$, and above each a positive gap (the limit of $E_2 - E_0$ above) is held to remain Physics . The gap that closes is a sign that "there are two vacua", not a sign that "there is no mass gap". When discussing the mass gap, one checks separately that the vacuum is unique (4-11).

09

## Back to the language of Yang–Mills

We list the correspondences of this chapter. All are reinterpretations at the level of physics and show nothing about Yang–Mills Physics . The $\beta$ in the table and in the paragraph below is the constant multiplying the action of lattice gauge theory (proportional to the inverse of $g^2$; 2-08 § 08), a different quantity from the inverse temperature $\beta$ of this chapter.

| Ising model | Lattice gauge theory | Chapter |
|---|---|---|
| $1/\xi = E_1 - E_0$ | The mass of the lightest glueball (in lattice units) | 5-08, 4-11 |
| Widening the width $L$ of the strip | Widening the spatial box | 5-04 |
| $\xi \to \infty$ as $K \to K_c$ | $\xi(\beta) \to \infty$ as $\beta \to \infty$ (the continuum limit) | 5-05, 5-09 |
| Interface tension $\sigma$ | String tension | 5-02, 17-03 |

There are also differences. The critical point of four-dimensional non-abelian gauge theory is believed to be at $\beta = \infty$, and how it is approached is determined by asymptotic freedom (4-06). Nor is there a closed-form solution like that of the two-dimensional Ising model.

### Detour: $Z_2$ gauge theory and the Ising model

A subsection you may skip.

Lattice gauge theory with group $\mathbb{Z}_2 = \{\pm 1\}$ (link variables $\pm 1$ in the form of 5-01) is said to have been introduced by Wegner 1971. In three dimensions it is dual to the three-dimensional Ising

model, and the coefficient of the area law of its Wilson loops is held to correspond to the interface tension of the three-dimensional Ising model at the dual coupling Known (bibliographic data only). With dimension and group matched, the last row of the table above becomes an equality rather than a metaphor (this is about three dimensions, not the two-dimensional $\sigma$ of this chapter itself). Lattice gauge theory with finite groups is in 5-10.

## What this chapter can say

| Content | Status |
|---|---|
| (2)–(7) in one dimension, the two-level reinterpretation, (8) | Said Known (Table 1 and full enumeration Computed ) |
| Table 2 ($L \leq 20$), the high-temperature limit, (10)(11), the $4/\pi$ at the critical point | Formulas from the literature Known , numerical agreement Computed (the extrapolation assumes $c/L^2$). The general relation of (12) is a prediction from conformal invariance Physics |
| The interface tension (13), $E_2 - E_0 \to 4(K - K^*)$, the gap remaining above the ordered states | The form of the formulas not checked in the text of the literature ("is held to") Physics . The values at both ends of (13), Table 3 and the exponent of the prefactor are Computed (Table 3 and the exponent are checks of tendencies under assumptions) |
| The table of correspondences in section 09 | Reinterpretation at the level of physics Physics . Nothing is shown about the mass gap or the construction of the continuum limit |
| Duality between $\mathbb{Z}_2$ gauge theory and the three-dimensional Ising model | Known (bibliographic data only, not checked) |

## Sources and reproduction

| Item | Kind | Source / tools |
|---|---|---|
| Tables 1–3, full enumeration, the check of the variance, the exponent of the prefactor, the numbers in the figures | Computed | python3 (NumPy, SciPy's `eigsh`; the action of order $2^L$ applied without building the matrix). The figures by JavaScript in the browser |
| The transfer-matrix method, the one-dimensional chain | Known | H. A. Kramers, G. H. Wannier, Phys. Rev. 60 (1941) 252. Abstract checked |
| $\xi_L = 1/\log(\lambda_0/\lambda_1)$, (10), the dual coupling, the form of (12) | Known | S. L. A. de Queiroz, Phys. Rev. E 84 (2011) 031107, arXiv:1105.6248. Equations (1)–(9) of § I and § II checked. The review summarising the closed forms, C. Domb, Adv. Phys. 9 (1960) 149, bibliographic data only |
| The general relation of (12) (from conformal invariance) | Physics | J. L. Cardy, J. Phys. A 17 (1984) L385. Abstract checked |

| Item | Kind | Source / tools |
|---|---|---|
| The closed form of the eigenvalues of the strip, the free energy of the boundary (the origin of (13)) | Known | L. Onsager, Phys. Rev. 65 (1944) 117. Abstract checked (that it treats the eigenvalue problem of the cylindrical strip and the free energy of the boundary; the form of formula (13) not checked) |
| Free fermions / no phase transition in one dimension | Known | T. D. Schultz, D. C. Mattis, E. H. Lieb, Rev. Mod. Phys. 36 (1964) 856 / E. Ising, Z. Phys. 31 (1925) 253. Bibliographic data only for both |
| $Z_2$ gauge theory and duality | Known | F. J. Wegner, J. Math. Phys. 12 (1971) 2259 (bibliographic data only, not checked) |
| The transfer matrix of lattice gauge theory and reflection positivity | Known | K. Osterwalder, E. Seiler, Ann. Phys. 110 (1978) 440 (bibliographic data only, not checked) |
| Finite-size effects of first-order transitions on a finite strip | Known | V. Privman, M. E. Fisher, J. Stat. Phys. 33 (1983) 385 (bibliographic data only, not checked) |

Next chapter to read: 2-10 The relaxation time and the dynamical exponent z. To read alongside: 3-01 The gauge principle. The main treatment of the definition of the gap is in 4-11, and the transfer matrix of the lattice in 5-03.

# The relaxation time and the dynamical exponent z — by what power of the correlation length does the approach to equilibrium slow down

> This chapter organises established knowledge. § 05 is this book's own investigation.

**What this chapter shows** — Measuring the relaxation time of dynamics that produce the same equilibrium distribution (Langevin, Glauber, heat bath) by the Poincaré constant $C_P$. That for the Gaussian chain and one-dimensional Ising the relaxation time is the square of the correlation length ($z = 2$), and that at the critical point of two-dimensional Ising it is slower than that. The definitions of the dynamical exponents $z$ and $\eta_{\mathrm{dyn}} = 2 - z$, and known examples in which the pattern $z = 2$ breaks.

**Prerequisite chapters** — 2-08 (the correlation length, critical points, the Gaussian chain) and 2-09 (the transfer matrix of the Ising model). It is easier if you have read the Poincaré inequality in 1-10 § 07 and Brownian motion in 1-09 § 06.

**Said first** — Only the Gaussian chain and the one-dimensional chain are exact. The values of $z$ (2.1665 and so on) are numerical results and predictions of physics, and the computations on this machine for two-dimensional Ising are on small boxes of side up to 5. This chapter shows nothing about the mass gap of Yang–Mills.

Order of this chapter

1. Adding time — a photograph of equilibrium, and the machine that makes it
2. Langevin dynamics — the inverse of the gap is the Poincaré constant
3. Glauber dynamics (Table 1) — in one dimension $\tau \approx 2\xi^2$
4. Langevin for the Gaussian chain (Figure 1, Table 2) — $C_P = \xi^2$ exactly in the continuum limit
5. Small boxes of two-dimensional Ising (Table 3, Figure 2) — $C_P/L^2$ grows at the critical point
6. The dynamical exponents z and η_dyn — $\eta_{\mathrm{dyn}} := 2 - z$ is defined here
7. Known examples with z ≠ 2 (Table 4, Detour: z = 4 for conserving dynamics)

8. Do not identify the heat bath with Langevin — constants, units, choice of variables
9. A bridge to Yang–Mills — what is known and what is not

---

01

## Adding time — a photograph of equilibrium, and the machine that makes it

What 2-08 and 2-09 treated was the equilibrium distribution, that is, "which configurations appear with what probability". There is no time in it. But when a computer draws samples from an equilibrium distribution, or when a physical system in contact with a heat bath moves towards equilibrium, the configuration moves in time. This chapter measures the speed of that movement.

As a metaphor, the equilibrium distribution is a single photograph, and the dynamics is the machine that makes it. Among machines that make the same photograph, some are fast and some slow. Precisely, a Markov process (a random motion whose next move depends only on the current configuration) leaving a given probability measure $\mu$ invariant is not unique, and the way it approaches equilibrium depends on the choice.

The time $s$ of this chapter is a different thing from the imaginary time of 2-09 (for the contrast see 18-03 § 05).

| Imaginary time (2-07, 2-09) | The dynamical time $s$ (this chapter) |
|---|---|
| One direction **within** the configuration. The correlation $\langle \sigma_0 \sigma_r \rangle$ decays in a way that sets the correlation length $\xi$, and $1/\xi$ is the gap of the transfer matrix. | An external time in which the **whole** configuration moves. The decay of autocorrelations sets the relaxation time $\tau$, and $1/\tau$ is the gap of the generator. |

The subject is the relation $\tau \approx \xi^z$, and $z$ is called the dynamical exponent. The assumptions of the later parts (A-06 Index of names) have the form of placing $z = 2$ at the level of a power. Here we first line up examples where that form holds and examples where it breaks.

02

## Langevin dynamics — adding fluctuations to a motion down a slope

Suppose we want a single real number $X$ to follow the probability density $e^{-V(x)}/Z$. If we move it in the direction down the slope $V$ and at the same time shake it with random numbers, the distribution stops where the two effects balance. This is Langevin dynamics.

$$dX = -V'(X)\,ds + \sqrt{2}\,dW(s) \quad (1)$$

$W$ is Brownian motion (1-09 § 06), and $dW$ is a Gaussian increment with mean 0 and variance $ds$ over time $ds$. The coefficient $\sqrt{2}$ is chosen so that the distribution at which it stops is exactly $e^{-V}$.

The operator producing the time change of the expectation of a function $f$ is called the generator, written $\mathcal{L}$. Expanding $f(X+dX)$ to second order, the first-order term in $dX$ leaves $-V'f'\,ds$ in the direction of the slope, and the second-order term leaves $\frac{1}{2}f''\cdot 2\,ds$ since the mean of $dW^2$ is $ds$ (orders higher than $ds$ are dropped). Thus for (1),

$$\mathcal{L}f = f'' - V'f', \qquad \frac{d}{ds}\,\mathbb{E}\,f(X_s) = \mathbb{E}\,(\mathcal{L}f)(X_s), \qquad \int f\,(-\mathcal{L}g)\,d\mu = \int f'g'\,d\mu \quad (2)$$

($\mu = e^{-V}\,dx/Z$; the last equality is integration by parts). By the last equality $-\mathcal{L}$ is symmetric in $L^2(\mu)$ and non-negative. It is 0 only for constants, and we write the next eigenvalue (the gap) as $\mathrm{gap}(\mathcal{L})$.

The inverse of the gap is the Poincaré constant (1-10 § 07).

$$\mathrm{Var}_\mu f \;\le\; C_P \int |f'|^2\,d\mu \quad (\text{for all } f), \qquad C_P = \frac{1}{\mathrm{gap}(\mathcal{L})} \quad (3)$$

Running from equilibrium, the autocorrelation of every observable decreases faster than $e^{-s/C_P}$, and there is an observable that decreases at exactly this rate. That is, the slowest relaxation time $\tau_{\exp}$ equals $C_P$ Known . This equality is a property of reversible dynamics in continuous time; for chains in discrete time it splits into three quantities (18-03 § 06).

Example. For $V = x^2/(2\sigma^2)$, (1) is the Ornstein–Uhlenbeck process, $f = x$ is an eigenfunction, and $\mathcal{L}x = -x/\sigma^2$. Hence $C_P = \sigma^2$: the larger the variance (the wider the tails), the slower the relaxation. With many variables, we set up (1) for the field $\phi_x$ at each point,

$$d\phi_x = -\frac{\partial S}{\partial \phi_x}\,ds + \sqrt{2}\,dW_x(s) \quad (4)$$

($W_x$ independent at each point), and replace the $|f'|^2$ of (3) by the squared length of the gradient $|\nabla f|^2$. The distribution at which it stops is $e^{-S}/Z$. In field theory this is called stochastic quantisation (7-09).

03

## Glauber dynamics — flipping spins one at a time

The spins of the Ising model are $\pm 1$, so they cannot move continuously down a slope. Instead, each spin carries an independent clock, and when the clock rings the spin is redrawn from its conditional distribution. This is called the heat-bath rule, and the dynamics doing this in continuous time is called Glauber dynamics (Glauber 1963 analysed a process of this type in detail on a closed one-dimensional chain) Known .

With $K$ including the inverse temperature in the coupling (as in 2-09), let $\mu(\sigma) \propto \exp(K \sum_{\langle xy \rangle} \sigma_x \sigma_y)$. Writing the configuration with spin $j$ flipped as $\sigma^j$ and the sum of neighbours as $h_j = \sum_{y \sim j} \sigma_y$, the heat-bath flip rate is

$$w_j(\sigma) = \frac{1}{1 + e^{2K \sigma_j h_j}}, \qquad \mu(\sigma)\, w_j(\sigma) = \mu(\sigma^j)\, w_j(\sigma^j) \tag{5}$$

The right-hand formula is detailed balance: it says that "the probability flowing from $\sigma$ to $\sigma^j$" equals the flow in the opposite direction. From this, $\mu$ is the distribution at which it stops, and the dynamics is reversible. The more aligned a spin ($\sigma_j h_j > 0$), the less readily it flips. The generator is $(\mathcal{L} f)(\sigma) = \sum_j w_j(\sigma)\big(f(\sigma^j) - f(\sigma)\big)$, the $\int |f'|^2$ on the right side of (3) is replaced by $\frac{1}{2} \sum_\sigma \mu(\sigma) \sum_j w_j(\sigma)\big(f(\sigma^j) - f(\sigma)\big)^2$, and $CP = 1/\mathrm{gap}(\mathcal{L})$ is the same.

### The one-dimensional chain can be solved

In one dimension $h_j \in \{-2, 0, 2\}$, so we can write $w_j = \frac{1}{2}\big(1 - \frac{g}{2}\sigma_j(\sigma_{j-1} + \sigma_{j+1})\big)$, $g = \tanh 2K$. The expectation of a spin changes by 2 in one flip, so $\frac{d}{ds}\langle \sigma_j \rangle = -2\langle \sigma_j w_j \rangle$, and

$$\frac{d}{ds}\langle \sigma_j \rangle = -\langle \sigma_j \rangle + \frac{g}{2}\big(\langle \sigma_{j-1} \rangle + \langle \sigma_{j+1} \rangle\big) \tag{6}$$

closes. The mean of the total magnetisation $M = \sum_j \sigma_j$ decreases like $e^{-(1-g)s}$. On the other hand, the static correlation is $\langle \sigma_0 \sigma_r \rangle = (\tanh K)^r$ (2-09 § 03), and

$$\tau = \frac{1}{1 - \tanh 2K}, \qquad \xi = \frac{-1}{\log \tanh K}, \qquad \frac{\tau}{\xi^2} \xrightarrow[K \to \infty]{} 2 \tag{7}$$

The limits follow from $1-\tanh 2K \approx 2e^{-4K}$, $-\log\tanh K \approx 2e^{-2K}$. The colder it is, the longer the correlations, and relaxation slows by its square: for this dynamics $z=2$.

Computed Table 1 One-dimensional Glauber dynamics. We diagonalised the generator of the ring of $N=8,\ 10$ points on all states ($2^N$ of them) and checked that the gap agrees with $1-\tanh 2K$ to 10 digits for all of $K=0.25,\ 0.5,\ 1,\ 1.5$. On these two rings the decay of magnetisation is the slowest relaxation. For general $N$, the decay rate of magnetisation $1-\tanh 2K$ gives an upper bound on the gap (since magnetisation is an eigenfunction of $\mathcal{L}$).

| $K$ | $\xi$ | $\tau = 1/(1-\tanh 2K)$ | $\tau/\xi^2$ |
|---:|---:|---:|---:|
| 0.50 | 1.295 | 4.195 | 2.499 |
| 1.00 | 3.672 | 27.80 | 2.062 |
| 1.50 | 10.03 | 202.2 | 2.008 |
| 2.00 | 27.30 | 1491 | 2.001 |
| 3.00 | 201.7 | 81378 | 2.00002 |

04

## Langevin for the Gaussian chain — the longer the wave the slower, and the relaxation time is the square of the correlation length

We apply (4) to the Gaussian chain of (7) of 2-08 (the $H(\phi)$ there; here written as the action $S$) $S=\frac{1}{2}\sum_x[(\phi_{x+1}-\phi_x)^2+m^2\phi_x^2]$. Since $\partial S/\partial\phi_x = (2+m^2)\phi_x-\phi_{x+1}-\phi_{x-1}$ is linear, the equations separate for each Fourier component $\hat\phi_q=\sum_x e^{-iqx}\phi_x$ of wave number $q$ (1-11).

$$d\hat\phi_q = -\omega_q\,\hat\phi_q\,ds + (\text{noise}), \qquad \omega_q = m^2 + 4\sin^2\frac{q}{2} \quad (8)$$

Every component is the Ornstein–Uhlenbeck process of (1), decreasing at rate $\omega_q$. The slowest is the uniform component $q=0$, with rate $m^2$. Therefore $C_P = 1/m^2$ (the same number as "the $C_P$ of a Gaussian measure is the largest eigenvalue of the covariance" of 1-10 § 09) Known .

The correlation length was $\cosh(1/\xi) = 1+m^2/2$ by (8) of 2-08. Using $\cosh u - 1 = 2\sinh^2(u/2)$, the two are connected by a single formula.

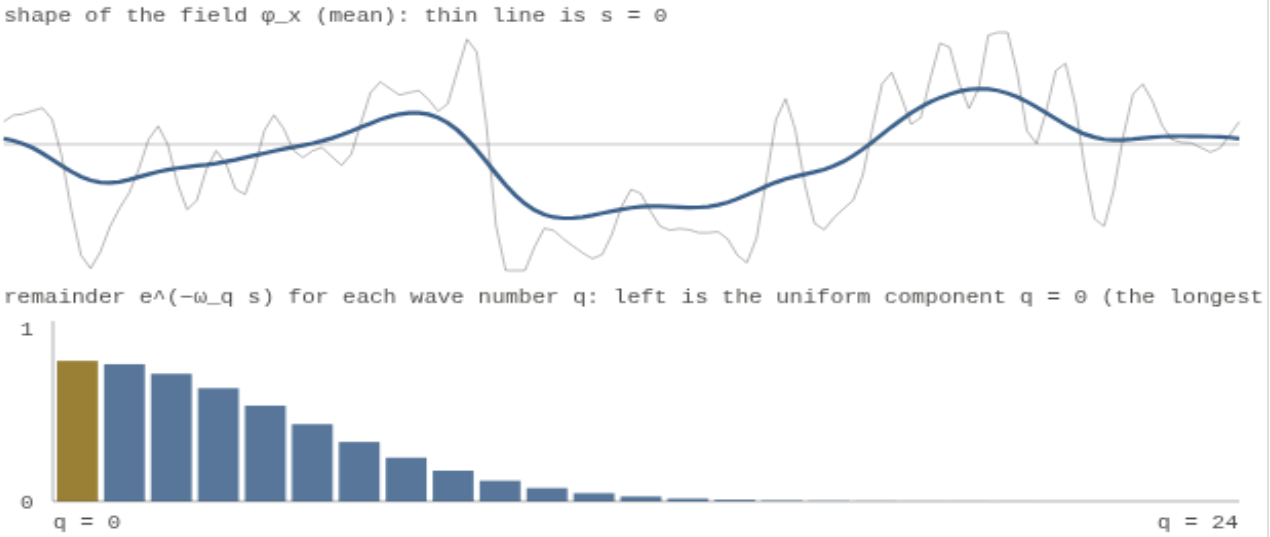
$$\frac{1}{C_P} = m^2 = 4\sinh^2\frac{1}{2\xi}, \qquad \frac{C_P}{\xi^2} = \left(\frac{1/(2\xi)}{\sinh(1/(2\xi))}\right)^2 \xrightarrow[\xi\to\infty]{} 1 \quad (9)$$

This is an exact equality; on the lattice $C_P$ is slightly smaller than $\xi^2$, and becomes $C_P/\xi^2 \to 1$ in the continuum limit ($\xi\to\infty$). For the continuous free field $C_P=\xi^2$ holds exactly Known (an elementary derivation; the numbers in Table 2 are Computed ).

### Why the square

Why the square. The mean motion of (8), with the noise removed, is the heat equation with mass $\frac{d}{ds}\phi = \Delta_{\text{lat}}\phi - m^2\phi$. Diffusion takes a time of order $\ell^2$ to smooth out bumps of length $\ell$ (as for a random walk), and the mass term removes waves longer than $\xi$ at a uniform rate $m^2$. So the time is $\xi^2$. Precisely, the content of this explanation is that the $\omega_q$ of (8) approaches $m^2 + q^2$ for small $q$, and bottoms out at $m^2$ for $q \lesssim 1/\xi$.

Figure 1 The mean motion of the Gaussian chain (a ring of 128 points). Making m smaller lengthens, like $1/m^2$, the time for which the long waves (bars on the left) survive. Moving the time s smooths out the shape above, and the bars below (the remainder $e^{-\omega_q s}$ for each wave number) disappear in order from the right




m = 0.158, ξ = 6.33, C_P = 1/m² = 40.0, s = 10 (s/C_P = 0.25), remainder of the uniform component 0.779

Computed The curves in the figure are the exact solution $\hat{\phi}_q(s) = e^{-\omega_q s}\hat{\phi}_q(0)$ of (8), computed on the spot (the initial shape was made with random numbers from a fixed seed). This is the mean motion with the noise removed, not a sample of the fluctuations.

Computed Table 2 $\xi$ and $C_P$ of the Gaussian chain. The last two columns are the heat-bath lower bound of section 08 (determined by the decrease on linear functions)

| $m$ | $\xi$ | $C_P = 1/m^2$ | $C_P/\xi^2$ | **Heat-bath lower bound** $(2+m^2)/m^2$ | **Lower bound / Lan** |
|---|---|---|---|---|---|
| 1.00 | 1.039 | 1 | 0.9263 | 3 | 3 |
| 0.50 | 2.020 | 4 | 0.9798 | 9 | 2 |
| 0.20 | 5.008 | 25 | 0.9967 | 51 | 2 |
| 0.10 | 10.004 | 100 | 0.99917 | 201 | 2 |
| 0.02 | 50.001 | 2500 | 0.999967 | 5001 | 2 |

A table set beside the gap of the transfer matrix is in Table 1 of 18-03, and a table measured with a checkerboard heat bath is in 5-07 § 07.

```
05
```

## Small boxes of two-dimensional Ising — at the critical point $C_P/\xi^2$ grows

With interactions, the formulas do not close. So on $L \times L$ boxes (periodic boundary) of the two-dimensional Ising model we diagonalise the generator of Glauber dynamics (5) numerically and find $C_P$. There are $2^{L^2}$ states, about 34 million for $L = 5$. The slowest relaxation is expected to lie in the direction of magnetisation (the subspace invariant under translations and changing sign under the global flip). For $L = 3,\ 4$ the gap of the whole space and the gap of this subspace agreed to 8 digits at all seven temperatures below Computed . For $L = 5$ the whole space cannot be handled, so we found the gap restricted to this subspace (671,104 dimensions). That value is at least the gap of the whole space, so the $C_P$ for $L = 5$ is a lower bound on the true $C_P$.

At the critical point $K_c = \frac{1}{2}\log(1+\sqrt{2})$ (2-08 § 04) the $\xi$ of the infinite system is infinite, and the length over which correlations reach inside a box levels off at the box size $L$ (2-09 § 07). So instead of $\xi$ we use $L$ as the ruler. If $\tau \approx \xi^z$ is right, then

$$C_P(L) \approx L^z, \qquad \frac{C_P(L)}{L^2} \approx L^{\,z-2}, \qquad z_{\text{eff}}(L, L') = \frac{\log\big(C_P(L')/C_P(L)\big)}{\log(L'/L)} \tag{10}$$

$z_{\text{eff}}$ is an effective exponent read from neighbouring sizes, and is expected to approach $z$ as $L \to \infty$.

Computed Table 3 Two-dimensional Ising, continuous-time heat bath (each spin redrawn at rate 1), $C_P = 1/\text{gap}$ with $L \times L$ periodic boundary. The unit of time is "one attempt per spin"

| $K/K_c$ | $L=3$ | $L=4$ | $L=5$ | $z_{\text{eff}}(3,4)$ | $z_{\text{eff}}(4,5)$ | To Figure 2 |
|---|---|---|---|---|---|---|
| 0.6 | 5.759 | 6.267 | 6.474 | 0.29 | 0.15 | Select |
| 0.8 | 14.56 | 19.34 | 22.83 | 0.99 | 0.74 | Select |
| 0.9 | 24.56 | 38.47 | 52.26 | 1.56 | 1.37 | Select |
| **1.0** | 42.89 | 82.84 | 137.4 | **2.29** | **2.27** | Select |
| 1.1 | 77.18 | 191.0 | 408.0 | 3.15 | 3.40 | Select |
| 1.2 | 142.4 | 464.6 | 1324 | 4.11 | 4.69 | Select |
| 1.5 | 997.9 | 7906 | 56656 | 7.19 | 8.83 | Select |

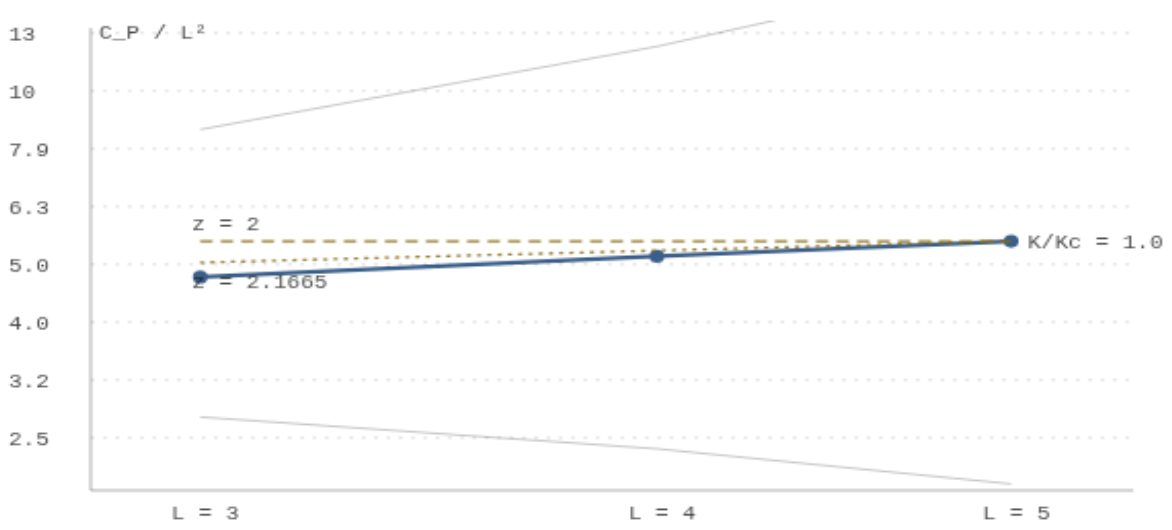
### Three behaviours

Three behaviours can be read. On the high-temperature side ($K/K_c = 0.6$), $C_P$ levels off around 6, and $z_{\text{eff}}$ tends to 0: if $\xi$ is finite, so is the relaxation time. On the low-temperature side ($K/K_c = 1.5$), $\log C_P$ increases with $L$ by an almost constant difference (the differences are 2.07, 1.97), growing more like an exponential than a power. This is thought to be because moving from "almost all +" to "almost all −"

requires making a wall across the box, whose weight becomes small like an exponential $e^{-cL}$ proportional to the length of the wall Physics . In a box with periodic boundary two walls are needed, so $c$ is expected to be about twice the interface tension $\sigma$. However, the $2\sigma \approx 1.55$ at $K/K_c = 1.5$ and the difference of about 2 for $L \le 5$ do not yet agree, and what this box can say goes only as far as "a tendency to grow exponentially". A similar picture is the closing gap of 2-09 § 08.

At the critical point $C_P/L^2$ keeps increasing as $4.77,\ 5.18,\ 5.50$, and $z_{\text{eff}}$ is $2.29,\ 2.27$. It lies on the side greater than 2. The value known from numbers on larger boxes (up to $15 \times 15$) is $z = 2.1665(12)$ (Nightingale–Blöte 1996, heat bath) Physics ; the values from the small boxes on this machine can be read as approaching it from above, but two points cannot decide this. What can be said goes only as far as "in the range $L \le 5$, $C_P/L^2$ does not settle to a constant".

Figure 2 $C_P/L^2$ for two-dimensional Ising (log–log; the vertical axis is magnified to suit the selected temperature). Moving the temperature changes the selected line (●) and the $z_{\text{eff}}$ in the readout on the right. The dashed line is z = 2 (horizontal), the dotted line the slope of z = 2.1665 (both passing through the point at L = 5). "Play" changes the temperature in turn. You can also select with "Select" in Table 3




K/Kc = 1.0, C_P = 42.89, 82.84, 137.4 (L = 3, 4, 5), z_eff = 2.29 (3→4), 2.27 (4→5)

Computed The numbers in this figure were computed on this machine (the values of Table 3 embedded). L = 2 is excluded, as it is a special box in which neighbours coincide.

06

## The dynamical exponents $z$ and $\eta_{\text{dyn}}$ — naming how far the power departs from 2

In physics one puts forward the hypothesis that as $\xi$ grows the relaxation time grows as a power of $\xi$. It is called the dynamic scaling hypothesis, and is held to be the extension of static scaling (2-08 § 06) to correlations involving time (Ferrell et al. 1967,

Halperin–Hohenberg 1967, with a detailed development in 1969) Physics . Put in mathematical language for reversible continuous-time dynamics,

$$C_P \asymp \xi^{\,z}, \qquad \eta_{\mathrm{dyn}} := 2 - z, \qquad C_P \asymp \xi^{\,2-\eta_{\mathrm{dyn}}} \quad (11)$$

$A \asymp B$ means that the ratio $A/B$ is bounded above and below by positive constants. In this book $\eta_{\mathrm{dyn}}$ is defined by this formula. The name comes from its being the quantity that measures, as a power, the departure from the $C_P \approx \xi^2$ of the free field.

It is a different quantity from the $\eta$ of (13) of 2-08. The η of 2-08 is the departure in the power of the static correlation $G(x) \sim |x|^{-(d-2+\eta)}$ at the critical point, and is written $\eta_s$ when it must be distinguished from the dynamic quantity. The two coincide only within the conventional theory (called by the name of van Hove) that holds "the relaxation time is proportional to the susceptibility"; for two-dimensional Ising, $\eta_s = 1/4 > 0$ while $\eta_{\mathrm{dyn}} \approx -0.17 < 0$, so they differ even in sign (Table 2 in 18-03 § 07).

| Pattern z = 2 | Too slow | Too fast |
|---|---|---|
| $z = 2$, $\eta_{\mathrm{dyn}} = 0$ | $z > 2$, $\eta_{\mathrm{dyn}} < 0$ | $z < 2$, $\eta_{\mathrm{dyn}} > 0$ |
| Holds for Langevin on the Gaussian chain ((9)) and one-dimensional Glauber ((7)) | Glauber for two- and three-dimensional Ising (numerical), conserving dynamics | Non-local updates (cluster methods). For local dynamics the lower bound $z \geq \gamma/\nu$ ($\gamma, \nu$ are the static exponents of the susceptibility and the correlation length) follows under two scaling hypotheses (18-03 § 07) Physics . For local Ising dynamics, $z \geq 2$ at the critical point is a theorem ( § 07) |

The diffusive-relaxation-time assumption of Part 23 (23-01) is the statement $\lim \ln C_P / \ln \xi = 2$ ($C_P = \xi^{2+o(1)}$, $\eta_{\mathrm{dyn}} = 0$) that places this left-hand form at the level of a power, for the Langevin dynamics of lattice Yang–Mills. The band statement $C_P \asymp \xi^2$ is called the band form of the diffusive-relaxation-time assumption in 23-01, and is treated separately from the diffusive-relaxation-time assumption (A-06 Index of names). Whether it holds is an unproved assumption, and the physical counterpart is set out in the form of questions in 18-05 § 03. What this chapter wants to confirm is that the form itself is not trivial: change the dynamics once and it breaks. The example is the next section.

07

## Known examples with $z \neq 2$ — three ways the pattern breaks

Table 4 lists known examples in which the pattern $C_P \asymp \xi^2$ breaks. The ways of breaking divide into three.

**Interaction shifts the power.** For Glauber dynamics of two- and three-dimensional Ising, numerically $z$ slightly exceeds 2. Even with the rule kept local, interaction moves the power. In mathematics it has been shown that for Glauber dynamics of two-dimensional critical Ising the inverse of the gap is bounded by a polynomial in $L$ (Lubetzky–Sly) Known . From below, it has been shown that at the critical point with periodic boundary (and on the low-temperature side) the heat-bath gap is at most $C(\log L/L)^2$, that is, $z \geq 2$ (Masaoka–Soejima–Watanabe 2025; extending to local reversible single-spin-flip dynamics) Known . At the two-dimensional critical point $(L/\log L)^2 \lesssim C_P(L) \leq \mathrm{poly}(L)$ is a theorem, and the "too fast" side of § 06 does not occur for local Ising dynamics. No theorem determining the value of the power $z$ itself has been found.

**Conserved quantities slow it down.** In dynamics that do not change the total magnetisation (exchanging neighbouring spins), the uniform component cannot move, and long waves are carried by diffusion, so it is slower still. The prediction of physics is held to be $z = 4 - \eta_s$ (model B in the classification of Hohenberg–Halperin) Physics .

**The way of updating speeds it up.** The Swendsen–Wang cluster method is non-local dynamics that flips whole clusters, and its abstract says it breaks dynamic universality, giving 'unusually small values of the dynamical critical exponent' Physics . Even for the same equilibrium distribution, changing the machine changes $z$.

Physics Table 4 Example values of $z$ (numerical and predicted values from the abstracts of primary sources; $\eta_{\mathrm{dyn}} = 2 - z$ computed on this machine)

| Model, dynamics | $z$ | $\eta_{\mathrm{dyn}}$ | Level |
|---|---|---|---|
| Gaussian chain, Langevin ((9)) / one-dimensional Ising, Glauber ((7)) | 2 | 0 | Known (exact) |
| Two-dimensional Ising, heat bath (Glauber) | 2.1665(12) | −0.1665 | Physics numerical |
| Three-dimensional Ising, Glauber | 2.032(3) | −0.032 | Physics numerical |
| Conserving dynamics (model B), three-dimensional Ising class | 3.972(2) | −1.972 | Physics numerical (prediction $4 - \eta_s$) |
| Conserving dynamics, Gaussian chain (detour below; measured by the relaxation time of the component of wave number $1/\xi$) | 4 | −2 | Computed (exact) |

| Model, dynamics | $z$ | $\eta_{\mathrm{dyn}}$ | Level |
|---|---|---|---|
| Pure-gauge HMC, square of the topological charge | About 5 | About −3 | Physics numerical (18-03 § 08) |

The last row is a slowness that actually appears in computations of lattice gauge theory. HMC (hybrid Monte Carlo; an update method that adds fictitious momenta and moves deterministically; 5-07 § 08) is dynamics different from Langevin (next section), and whether the same slowness occurs for Langevin is treated in 23-02 Roads tried on the four-dimensional lattice, and walls.

### Detour: $z = 4$ for conserving dynamics

You may skip this. In the Gaussian chain, if the drive $-\partial S/\partial\phi$ of (4) is passed once more through the lattice Laplacian $\Delta_{\mathrm{lat}}$, and the noise too is put in the form of differences between neighbours, then $\sum_x \phi_x$ is conserved. The decay rate of (8) is replaced by

$$\omega_q^{(\mathrm{B})} = 4\sin^2\frac{q}{2}\left(m^2 + 4\sin^2\frac{q}{2}\right) \tag{12}$$

and $q = 0$ stops (being a conserved quantity, it does not decrease). The relaxation time of the component $q = 1/\xi$ with wavelength about the correlation length is $1/\omega_{1/\xi}^{(\mathrm{B})} \approx \xi^4/2$. The computation on this machine gave $\tau/\xi^4 = 0.510,\ 0.5004,\ 0.50002$ for $m = 0.5,\ 0.1,\ 0.02$ Computed . On the infinite chain the components with $q \to 0$ are as slow as we like, so the gap is 0 and $C_P = \infty$. The pattern $C_P \asymp \xi^2$ loses its meaning here.

08

## Do not identify the heat bath with Langevin — the same photograph, different machines

"The relaxation time" is not determined until we say whose dynamics' time it is. The heat bath and Langevin have different generators for the same distribution.

We compare them on the Gaussian chain. For the continuous-time heat bath (each point, at rate 1, redrawn from the conditional distribution $\mathcal{N}\big((\phi_{x-1}+\phi_{x+1})/(2+m^2),\ 1/(2+m^2)\big)$ with its neighbours fixed), the mean of the field values moves by

$$\frac{d}{ds}\,\mathbb{E}\,\phi_x = -\phi_x + \frac{\phi_{x-1}+\phi_{x+1}}{2+m^2} = -\frac{1}{2+m^2}\,\frac{\partial S}{\partial \phi_x} \tag{13}$$

Linear functions are eigenfunctions of this generator, and the decay rate of the uniform component is $m^2/(2+m^2)$. Therefore the $C_P$ of the heat bath is at least $(2+m^2)/m^2$ Computed (checked with the matrix of a ring of 64 points). This lower bound is about twice $\xi^2$, so for the heat bath $z \geq 2$. As the right two columns of Table 2 show,

the ratio of the lower bound to the $CP$ of Langevin approaches 2 as $\xi \to \infty$. Even if the powers are the same, the constants differ.

#### Differences other than the constant

The difference is not only in the constant.

- **The choice of variables.** The heat bath is the same process even if the variable at each point is relabelled ($\psi_x = c\,\phi_x$ and so on), because the operation of drawing from a conditional distribution does not depend on coordinates. The $CP$ of Langevin is measured by the length of the gradient, so measured in $\psi = c\,\phi$ it becomes $c^2$ times as large. This is why in lattice gauge theory $CP$ changes by a factor $\beta$ depending on whether one measures by "the angle of the link $\vartheta$ or the field $A = \sqrt{\beta}\,\vartheta$". Since $\beta \sim \ln \xi$, the power statement (the diffusive-relaxation-time assumption) does not change, but the band statement does (1-10 § 09, 23-01 The line of construction and the diffusive-relaxation-time assumption).
- **The unit of time.** The time of the heat bath is "the number of attempts per point", and the time of Langevin is the $s$ of (1). The heat bath that advances one sweep at a time in discrete time (5-07 § 07) is yet another quantity, and $CP$, the relaxation time and $\tau_{\exp}$ split apart (18-03 § 06).
- **The exponent.** In physics, two local reversible dynamics with the same conserved quantities (both of model A type: not conserving magnetisation) are expected to have the same $z$ (the same dynamic universality class) Physics . The conserving dynamics of § 07 has different conserved quantities (model B), so it is not a counterexample to this expectation. For the Gaussian chain, what can be said by formulas is $z = 2$ for Langevin, and, for the heat bath, $z \geq 2$ from the lower bound from linear functions. That the gap of the heat bath is determined by linear functions, with equality, is not shown in this chapter. In general, this chapter has no theorem showing that the $z$ of two dynamics are equal.

When this book writes $CP$, unless stated otherwise it is the inverse of the gap of the generator of Langevin (4), with the variable whose gradient it is measured by stated alongside. Reading numbers from the heat bath or HMC as the $CP$ of Langevin is an assumption.

09

## A bridge to Yang–Mills — where it is used, and what is not known

Langevin dynamics (4) can be set up for lattice Yang–Mills too. Since the variables are group elements on the links, the gradient is taken as a derivative on the group (1-03).

Doing the same for a continuous field is stochastic quantisation (7-09), and drawing samples from the same distribution on a computer is the Monte Carlo method (5-07).

At strong coupling (the side where $\beta$ is small), the Poincaré inequality and the log-Sobolev inequality (an inequality stronger than (3), controlling the decrease of entropy) have been shown for lattice Yang–Mills (Shen–Zhu–Zhu 2023) Known . This is a result on the side where $\xi$ is finite. In four dimensions there is no theorem determining the power of $C_P$ as $\xi \to \infty$.

From this chapter onwards

Putting the pattern $z = 2$ for Langevin dynamics of lattice Yang–Mills as the statement at the level of powers $\lim \ln C_P / \ln \xi = 2$ ($C_P = \xi^{2+o(1)}$, $\eta_{\mathrm{dyn}} = 0$) is 23-01 The line of construction and the diffusive-relaxation-time assumption. The band statement $C_P \asymp \xi^2$ is called the band form of the relaxation-time assumption in 23-01 and treated separately. The metaphorical entrance is 23-00 (the smearing time and the ruler), the reconciliation with the vocabulary of physics is 18-05 § 02 (fixing the $\eta$ of the record to $\eta_{\mathrm{dyn}}$), and the dictionary of the two gaps $\xi$ and $C_P$ is 18-03 § 05– § 07. This chapter shows nothing about the mass gap.

## What this chapter can say

| Content | Status |
|---|---|
| The generator (2) of Langevin (1)(4), $\tau_{\exp} = C_P$ (reversible dynamics in continuous time), detailed balance (5) of the heat bath | Said Known |
| One-dimensional Glauber dynamics (6)(7), the decay rate of magnetisation $1 - \tanh 2K$, $\tau/\xi^2 \to 2$ | Said Known ((6) was derived in this chapter). The gap $= 1 - \tanh 2K$ is Computed for rings with $N = 8, 10$. For general $N$ the decay rate of magnetisation gives an upper bound on the gap |
| (8)(9) of the Gaussian chain, $C_P/\xi^2 \to 1$, (12) of conserving dynamics and $\tau \approx \xi^4/2$, (13) of the heat bath and the lower bound $C_P \geq (2+m^2)/m^2$ (for the heat bath, as far as $z \geq 2$) | Said (the formulas are exact) Known (elementary derivations). The numbers in Table 2 and (12) are Computed |
| Table 3: $C_P$ and $z_{\mathrm{eff}}$ of two-dimensional Ising with $L \leq 5$ | Computed (exact diagonalisation; the restriction to the magnetisation subspace is checked against the whole space for $L \leq 4$, and for $L = 5$ the subspace value is a lower bound on the true $C_P$). It does not determine $z$; it is a check of tendencies |

| Content | Status |
|---|---|
| Dynamic scaling $C_P \asymp \xi^z$, the values of $z$ in Table 4, the $4-\eta_s$ of model B, that two local dynamics have the same $z$ | Physics (hypotheses, numbers, predictions). For two-dimensional critical Ising, $(L/\log L)^2 \lesssim C_P(L) \le \mathrm{poly}(L)$ ($z \ge 2$ from below, polynomial from above) is Known ; the value of the power has not become a theorem |
| $\eta_{\mathrm{dyn}} := 2 - z$ | This book's definition (common to the later chapters) |
| The dependence of $C_P$ on $\xi$ for Yang–Mills | Not said. There is no theorem determining it in four dimensions, and both the relaxation-time assumption (the statement at the level of powers, 23-01 § 05) and the band statement (the band form of the relaxation-time assumption) remain assumptions. This chapter shows nothing about the mass gap or the construction of the continuum limit |

## Sources and reproduction

| Item | Kind | Source / tools |
|---|---|---|
| Tables 1, 2 and 3, the numbers in (9)(12)(13), Figures 1 and 2 | Computed | python3 (NumPy, SciPy; for Table 3 the generator was made a sparse matrix on the subspace reduced by translations and the global flip, and the smallest eigenvalue found by the Lanczos method). The figures by JavaScript in the browser. The scripts are kept in the record of computations |
| Glauber dynamics, the one-dimensional chain | Known | R. J. Glauber, "Time-dependent statistics of the Ising model", J. Math. Phys. 4 (1963) 294–307. Abstract checked (analyses the Markov process of a closed chain of N points in contact with a heat bath, and finds the expectations and correlations of the spins explicitly). The text has not been read; (6) was derived in this chapter |
| The Poincaré constant, the difference between continuous and discrete time | Known | D. A. Levin, Y. Peres, E. L. Wilmer, *Markov Chains and Mixing Times* 2nd ed., § 12.2 (the same as in the table of sources of 18-03) |
| The dynamic scaling hypothesis | Physics | R. A. Ferrell, N. Menyhárd, H. Schmidt, F. Schwabl, P. Szépfalusy, Phys. Rev. Lett. 18 (1967) 891–894. Abstract checked (derives the dispersion of second sound $\omega \propto k^{3/2}$ from the absence of a characteristic length at the lambda point). B. I. Halperin, P. C. Hohenberg, Phys. Rev. Lett. 19 (1967) 700–703. Abstract checked (generalises the Widom–Kadanoff scaling laws to dynamic phenomena under an assumption on time-dependent correlation functions). The detailed development is B. I. Halperin, P. C. Hohenberg, Phys. Rev. 177 (1969) 952 (abstract only). The classification (model A, B) is P. C. Hohenberg, B. I. Halperin, Rev. Mod. Phys. 49 (1977) 435–479. Abstract checked (identifies and characterises the main dynamic universality classes; the names model A and B are not in the abstract, and the text has not been read) |

| Item | Kind | Source / tools |
|---|---|---|
| $z = 2.1665(12)$ of two-dimensional Ising | Physics | M. P. Nightingale, H. W. J. Blöte, Phys. Rev. Lett. 76 (1996) 4548. Abstract checked (heat bath, random choice of sites, up to $15 \times 15$) |
| $z = 2.032(3)$ of three-dimensional Ising | Physics | Z. Liu, E. Vatansever, G. T. Barkema, N. G. Fytas, Phys. Rev. E 108 (2023) 034118, arXiv:2307.01837. Abstract checked |
| $z = 3.972(2)$ of conserving dynamics and the prediction $z = 4 - \eta$ | Physics | C. Chattopadhyay, J. Ott, T. Schäfer, V. Skokov, Phys. Rev. D 108 (2023) 074004, arXiv:2304.07279 (volume and article number checked with Crossref). Text checked (the prediction is attributed to Hohenberg–Halperin 1977) |
| The small $z$ of cluster methods | Physics | R. H. Swendsen, J.-S. Wang, Phys. Rev. Lett. 58 (1987) 86. Abstract checked |
| The inverse of the gap of two-dimensional critical Ising is polynomial in $L$ | Known | E. Lubetzky, A. Sly, Commun. Math. Phys. 313 (2012) 815–836, arXiv:1001.1613. Abstract checked (the inverse of the gap is polynomial for any boundary condition) |
| $z \geq 2$ for Ising in two or more dimensions (gap $\leq C(\log L/L)^2$, periodic boundary, $\beta \geq \beta_c$ ) | Known | R. Masaoka, T. Soejima, H. Watanabe, J. Stat. Phys. 192 (2025) 76, arXiv:2502.09908. Abstract checked |
| The slowness of the topological charge in pure-gauge HMC | Physics | S. Schaefer, R. Sommer, F. Virotta, Nucl. Phys. B 845 (2011) 93 (abstract only; the same as 18-03 § 08) |
| Poincaré and log-Sobolev for lattice Yang–Mills at strong coupling | Known | H. Shen, R. Zhu, X. Zhu, Commun. Math. Phys. 400 (2023) 805 (abstract only; the same as 18-03) |

Next chapter to read: 3-01 The gauge principle. This chapter continues in 5-07 § 07 (critical slowing down of the Monte Carlo method), 18-03 § 05– § 07 (the dictionary of the two times) and 23-01 (the relaxation-time assumption).