University of Rzeszów
Doctoral School

STER NAWA - internationalization of doctoral schools
„UR an international PhD student"
(BPI/STE/2023/1/00001/U/00001)



Modern Theoretical Physics and Machine Learning

Lecture notes and slides for doctoral students

Roman Pasechnik, PhD, Associate Professor (Docent)
Lund University (Lund, Sweden)

https://orcid.org/0000-0003-4231-0149

DOI:10.15584/ste.2026.1

This publication has been produced as part of project "UR an international PhD student" run by Rzeszów University Doctoral School. The project has been supported by Polish National Agency for Academic Exchange under the STER NAWA program - Internationalization of Doctoral Schools; No BPI/STE/2023/1/00001/U/00001

HR EXCELLENCE IN RESEARCH

Teaching materials

Roman Pasechnik

Prepared for doctoral students
at Rzeszow University

Workshops • Quantum field theory • Machine learning • Cosmology

2026

Contents

Workshops with an expert

A calculation of diffractive charm 3

The Dirac sea in a strong field 59

Quantum fluctuations and physical memory 92

Beyond the Born–Oppenheimer approximation 122

Why forbidden molecular transitions radiate 147

Introduction to quantum field theory

Fields, symmetries and particles 174

Causality and perturbative interactions 194

From amplitudes to observables 216

Problems for independent study 241

Open lectures

Modern machine learning and AI for physicists: How machines learn 249

Evolution of our Universe: The expanding Universe 312

Evolution of our Universe: From the expansion history to the hot Big Bang and inflation 367

Contents entries refer to collection page numbers.

Each contribution retains its own equation, figure and original page numbering.

A calculation of diffractive charm

A calculation of diffractive charm

From the Dirac field to a measurable spectrum,
Wilson lines and saturation

With applications to J/ψ, h_c
and the Pomeron/Odderon decomposition

One calculation, with connected projections

Prerequisites. Special relativity, undergraduate quantum mechanics, Fourier transforms and elementary matrix algebra. No previous quantum field theory course is assumed.

Central question. When does coherent diffractive charm production cease to be accurately described by the first, two-gluon term of its target interaction?

Calculated result. A fully normalized longitudinal open-charm spectrum in a specified finite-target model, its dilute limit, and a finite-bin error estimate.

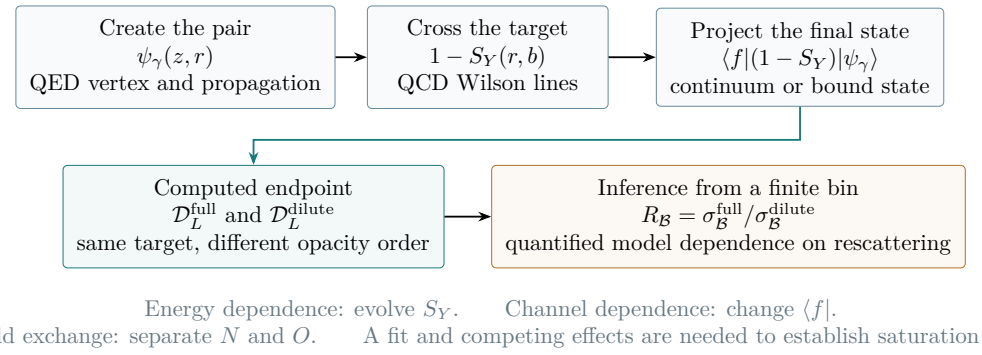


Figure 1: The course's causal chain and its computed endpoint. The benchmark answers a precise approximation question inside one specified target model. It supplies a diagnostic to carry into phenomenology; it does not by itself prove saturation in a measured process or establish a failure of a collinear factorisation theorem.

The causal chain

$$\underbrace{\mathcal{L}_{\text{QED+QCD}}}_{\text{theory}} \longrightarrow \underbrace{\psi_{\gamma\to c\bar{c}}}_{\text{splitting}} \longrightarrow \underbrace{U_x U_y^\dagger}_{\text{propagation}} \longrightarrow \underbrace{\langle p|1-\hat{S}|p\rangle}_{\text{target}} \longrightarrow \underbrace{\langle f|(1-S)|\psi_\gamma\rangle}_{\text{observable}}.$$

The final-state projector is a plane wave for open charm and a bound-state wavefunction for charmonium. Increasing energy requires evolving the same target operator. Charge conjugation separates its even and odd parts.

What is derived, imported or chosen

- **Derived here:** pair kinematics, field-theory vertex and denominator, spin/color factors, Fourier transforms, Wilson-line propagation, Gaussian-slice exponentiation, absolute open-charm normalization, the two-gluon and all-opacity spectra, and the leading-log evolution equation with its closure.
- **Stated matching input:** the dilute gluon-PDF identity, the scope of collinear factorisation theorems, and the published h_c photoproduction coefficient. Their assumptions and sources are given where used; these are not presented as newly proved theorems.
- **Chosen model:** a finite transverse disk and a quadratic dipole opacity. This allows a complete analytic/numerical calculation. It is not a precision proton fit.
- **Defined continuation:** a fully matched h_c electroproduction impact factor and a realistic phenomenological comparison. The notes teach its spin/orbital projection and known photoproduction anchor; they do not claim an unperformed new calculation.

Reading route. Sections 1–13 build the field-theory calculation and its quantitative benchmark. Sections 15–18 extend the same target operator to energy evolution, charmonium projections and exchange quantum numbers. The full covariant projector, GPD coefficient and signature-factor algebra provide supporting derivations. Appendix B collects target models, angular correlations, a numerical BK step and report tasks.

The question and the calculational endpoint

A hard charm mass makes the elementary production process calculable. Does it also guarantee that the proton is dilute? We will answer that for one specified diffractive observable, then ask how the answer changes with energy and with the quantum numbers of the final charm state.

$$e + p \longrightarrow e' + c + \bar{c} + p', \qquad \mathcal{D}_L = \left.\frac{\mathrm{d}\sigma(\gamma_L^* p \to c\bar{c}p)}{\mathrm{d}z\ \mathrm{d}p^2\ \mathrm{d}|t|}\right|_{t\simeq 0},$$

$$R_{\mathcal{B}} \equiv \frac{\displaystyle\int_{\mathcal{B}} \mathrm{d}p^2\, \mathcal{D}_L^{\text{full}}}{\displaystyle\int_{\mathcal{B}} \mathrm{d}p^2\, \mathcal{D}_L^{\text{dilute}}}.$$

The labels will be defined before they are used in a calculation. For now, "full" sums repeated target interactions in a declared model; "dilute" retains its first nonzero term. All other inputs stay the same.

Three results to obtain.

1. **One calculation with numbers.** Derive the normalized spectrum, its dilute limit, a shifted interference zero and finite-bin ratios. We will see why the same target correction can suppress one bin and enhance another.
2. **One common target description.** Derive how color transport produces a dipole Wilson operator, how repeated interactions can become important, and how an extra rapidity interval leads to BK evolution with an explicit target-correlation assumption.
3. **One map from quantum numbers to measurements.** Project the pair onto J/ψ and h_c, separate even/Pomeron and odd/Odderon exchange, and identify what an Odderon-sensitive measurement must retain.

The final claim will be precise. A hard scale, a target density and a final-state projection are all needed to assess sensitivity to repeated scattering. The worked model quantifies that statement; it does not determine a universal experimental saturation threshold or by itself invalidate a collinear factorisation theorem.

Contents

The calculation and its scope i

The question and the calculational endpoint ii

Notation iv

1 One process, one question, one calculable endpoint 1

2 Build the minimum field theory needed for this amplitude 3

3 A virtual photon makes an intermediate pair 5

4 Calculate the longitudinal spin numerator 7

5 Transform the pair into a dipole 9

6 A fast quark undergoes a color rotation 10

7 From ordered rotations to a color-singlet dipole 11

8 Resum many weak scatterings and identify the gluon density 14

9 Assemble and normalize the observable 16

10 Choose a finite target and a precise comparison 18

11 Calculate the two-gluon spectrum 19

12 Calculate the all-opacity spectrum 20

13 Calculate the first correction and the shifted zero 21

14 Quantify the effect and distinguish three approximations 22

15 Add one softer gluon 24

16 Real minus virtual: the hierarchy and BK 25

17 Project the same produced pair onto J/ψ or h_c 28

18 Which part of the target is Pomeron-like or Odderon-like? 33

19 The result we have earned 42

A Derivation sheet: the transverse photon and the J/ψ mass term 44

B Supporting calculations and the student report 46

Sources and reading route 48

Notation

Symbol	Meaning
P, P'	Initial and final proton four-momenta; $\Delta = P' - P$.
q, Q^2, W^2, t	Photon momentum; $Q^2 = -q^2$; photon–proton energy squared; $t = \Delta^2$.
K, p_1, p_2	Pair/meson momentum and charm/anticharm momenta; $K = q - \Delta$.
$z, \boldsymbol{z}$	Unbold z: charm plus-momentum fraction. Bold $\boldsymbol{z}$: a transverse gluon position in evolution.
$\boldsymbol{p}, \boldsymbol{\kappa}$	Midpoint-conjugate relative momentum; boost-invariant pair relative momentum. $\boldsymbol{\kappa} = \boldsymbol{p} + (z - 1/2)\boldsymbol{\Delta}$. They coincide in the forward benchmark.
$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{r}, \boldsymbol{b}$	Quark and antiquark positions; $r = x - y$, $b = (x + y)/2$, in the transverse plane.
$x, x_{\mathbb{P}}, Y$	Target evolution fraction (set to $x_{\mathbb{P}}$ for the open-pair benchmark); diffractive longitudinal momentum loss; $Y = \ln(x_0/x)$.
ϵ^2	$m_c^2 + z(1 - z)Q^2$; ϵ without a Lorentz index is not a polarization vector.
S, N, O	Averaged singlet dipole; $S = 1 - N + \mathrm{i}O$. A hat denotes its operator before target averaging.
$\mathcal{S}, U$	Field-theory action; fundamental color Wilson line. Neither is the scalar dipole S.
Q_s, τ, ρ, v	Saturation-scale convention; $\tau = r^2 Q_s^2/4$, $\rho = Q_s^2/\epsilon^2$, $v = p^2/\epsilon^2$. The last v is not the heavy-quark velocity.
$R, A_\perp, T_R$	Disk radius; πR^2; its transverse Fourier transform. In the radial Schrödinger equation R is a three-dimensional separation, explicitly indicated there.
$g = g_s, \alpha_s, \bar{\alpha}$	QCD coupling; $g^2/(4\pi)$; $\alpha_s N_c/\pi$. A PDF written $g(x, \mu^2)$ is a function, not this coupling.
ψ_γ, Ψ_V, R_L	Photon splitting amplitude; normalized meson valence wavefunction; normalized rest-frame radial wavefunction.
$Y_R, \varsigma, \rho_{\mathrm{el}}$	Regge energy logarithm $\ln(s/s_0)$; crossing signature; elastic phase ratio. These are distinct from Y, opacity τ, and density ratio ρ.
μ, η, ξ	Ultraviolet factorisation scale; rapidity separator; longitudinal skewness in a GPD.

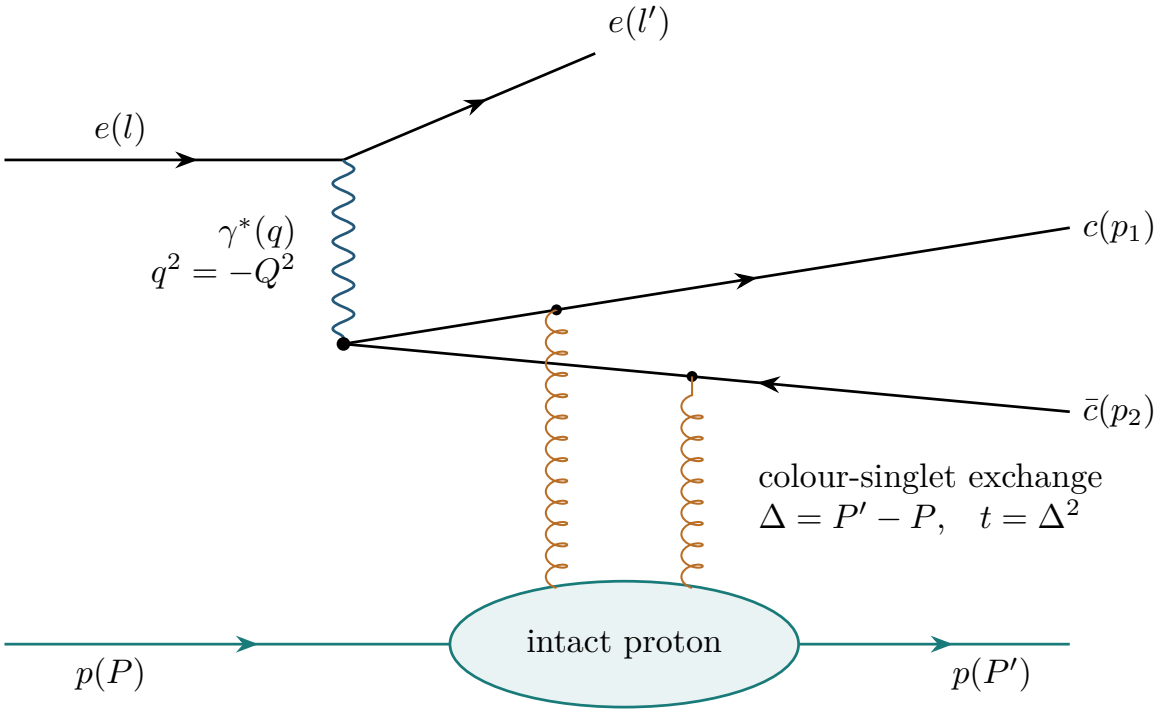
Accuracy used throughout. Leading electromagnetic splitting, leading eikonal high-energy propagation, and leading-logarithmic fixed-coupling evolution unless a different accuracy is explicitly named. "All opacity" sums target interactions within a model; it does not mean all perturbative orders in QCD. $\hbar = c = 1$; transverse lengths are in GeV^{-1}.

1 One process, one question, one calculable endpoint

Purpose. Determine when repeated target scattering changes coherent diffractive open-charm production.

Our question is whether a hard charm pair necessarily sees a dilute proton. We shall calculate one experimentally meaningful spectrum twice, changing only the treatment of repeated target scattering, and ask how much the answer changes in a specified momentum bin.

$$e(l)+p(P) \longrightarrow e(l')+c(p_1)+\bar{c}(p_2)+p(P'), \qquad q=l-l', \quad \Delta=P'-P, \quad K=p_1+p_2=q-\Delta. \tag{1.1}$$



Shown: one two-gluon topology. All required attachments are summed.

Figure 2: The measured electron fixes the virtual photon momentum; the proton remains intact. The arrows on charm and anticharm lines follow fermion-number flow. Two exchanged gluons illustrate the first nonzero singlet term; this one topology is not the completed amplitude.

Coherent diffraction: the same proton appears in the final state. The charm pair is a perturbative final state; identified charm hadrons would require fragmentation information.

$$\hbar=c=1, \qquad g_{\mu\nu}=\mathrm{diag}(1,-1,-1,-1), \qquad a\cdot b=a^0b^0-\boldsymbol{a}_{3\mathrm{D}}\cdot\boldsymbol{b}_{3\mathrm{D}},$$
$$a^{\pm}\equiv\frac{a^0\pm a^3}{\sqrt{2}}, \qquad \boldsymbol{a}\equiv(a^1,a^2), \qquad a\cdot b=a^+b^-+a^-b^+-\boldsymbol{a}\cdot\boldsymbol{b},$$
$$a^2=2a^+a^--\boldsymbol{a}^2, \qquad a^2=m^2 \implies a^-=\frac{m^2+\boldsymbol{a}^2}{2a^+}.$$

The photon moves mainly in the + direction; the proton mainly in the − direction. The evolution

time will be x^+; its conjugate energy is p^-.

$$Q^2 \equiv -q^2 > 0, \qquad W^2 \equiv (P+q)^2, \qquad t \equiv \Delta^2 < 0,$$
$$2P\cdot q = W^2 + Q^2 - m_p^2, \qquad M_{c\bar{c}}^2 \equiv K^2, \qquad x_B \equiv \frac{Q^2}{2P\cdot q},$$
$$M_{c\bar{c}}^2 = (q-\Delta)^2 = -Q^2 + t - 2q\cdot\Delta,$$
$$x_{\mathbb{P}} \equiv \frac{q\cdot(P-P')}{q\cdot P} = \frac{Q^2 + M_{c\bar{c}}^2 - t}{W^2 + Q^2 - m_p^2}, \qquad \beta_D \equiv \frac{x_B}{x_{\mathbb{P}}} = \frac{Q^2}{Q^2 + M_{c\bar{c}}^2 - t}.$$

Separate pair motion from relative motion. Choose $\boldsymbol{q} = 0$:

$$z \equiv \frac{p_1^+}{K^+}, \qquad 1 - z = \frac{p_2^+}{K^+}, \qquad \boldsymbol{K} = -\boldsymbol{\Delta}.$$

$$\begin{aligned}
\boldsymbol{\kappa} &\equiv (1-z)\boldsymbol{p}_1 - z\boldsymbol{p}_2, \\
\boldsymbol{p}_1 &= \boldsymbol{\kappa} + z\boldsymbol{K}, \qquad \boldsymbol{p}_2 = -\boldsymbol{\kappa} + (1-z)\boldsymbol{K}, \\
M_{c\bar{c}}^2 &= 2K^+(p_1^- + p_2^-) - \boldsymbol{K}^2 \\
&= \frac{(\boldsymbol{\kappa} + z\boldsymbol{K})^2 + m_c^2}{z} + \frac{[-\boldsymbol{\kappa} + (1-z)\boldsymbol{K}]^2 + m_c^2}{1-z} - \boldsymbol{K}^2 \\
&= \frac{\kappa^2 + m_c^2}{z(1-z)}.
\end{aligned}$$

Working approximation and promised result.

$$x_{\mathbb{P}} \ll 1, \qquad W^2 \gg Q^2, M_{c\bar{c}}^2, m_p^2, |t|, \qquad K^+ \simeq q^+,$$
$$\epsilon^2 \equiv m_c^2 + z(1-z)Q^2, \qquad \epsilon^2 \gg \Lambda_{\mathrm{QCD}}^2, \qquad 0 < z < 1.$$

Use the leading electromagnetic splitting, leading high-energy propagation, and a specified target model. Keep arbitrarily many target interactions. Select the longitudinal photon response: its spin algebra is simplest. Charm supplies a short-distance scale even when Q^2 is moderate.

$$\mathcal{D}_L(\kappa^2; x_{\mathbb{P}}, Q^2, z) \equiv \left.\frac{\mathrm{d}\sigma(\gamma_L^* p \to c\bar{c}p)}{\mathrm{d}z\ \mathrm{d}\kappa^2\ \mathrm{d}|t|}\right|_{t\simeq 0}, \tag{1.2}$$

$$R_{\mathcal{B}} \equiv \frac{\displaystyle\int_{\mathcal{B}} \mathrm{d}\kappa^2\, \mathcal{D}_L^{\mathrm{full}}(\kappa^2)}{\displaystyle\int_{\mathcal{B}} \mathrm{d}\kappa^2\, \mathcal{D}_L^{\mathrm{dilute}}(\kappa^2)}. \tag{1.3}$$

Both calculations use the same target input: "dilute" retains its first nonzero interaction term. Forward $t \simeq 0$ is the high-energy approximation to the physical forward boundary. Hold $x_{\mathbb{P}}, Q^2, z$ fixed in this benchmark: varying κ then varies W.

Student check. Derive $M_{c\bar{c}}^2$ and $x_{\mathbb{P}}$. Why does a small QCD coupling not by itself guarantee that repeated scattering is negligible?

Keep. We will calculate one spectrum and its finite-bin change, then identify what that change does and does not establish about QCD factorisation.

Check. At fixed Q^2 and pair momentum, does a small α_s alone guarantee a small interaction with the target?

Next. *What field-theory amplitude creates the charm pair whose scattering we want to measure?*

2 Build the minimum field theory needed for this amplitude

Purpose. Turn relativistic one-particle waves into a pair-creation matrix element.

We need field theory because the incoming current creates two particles. The only new machinery needed now is a field whose operators create a quark and an antiquark; contracting those operators will give the vertex, rather than asking us to memorize a diagrammatic rule.

Quantum-mechanical postulate: amplitudes and time evolution.

$$|\Psi\rangle = \sum_n a_n |n\rangle, \qquad \mathcal{P}(n) = |a_n|^2, \quad \langle n|m\rangle = \delta_{nm},$$

$$U_I(t_f, t_i) = \mathcal{T}\exp\left[-\mathrm{i}\int_{t_i}^{t_f} \mathrm{d}t\, H_I(t)\right] = 1 - \mathrm{i}\int \mathrm{d}t\, H_I(t) + \mathcal{O}(H_I^2).$$

Field theory permits the states $|n\rangle$ to contain different particle numbers. We adopt a local Dirac field and its canonical anticommutation relations as the theory's postulates; the splitting amplitude will be derived from them.

Relativistic free waves. A spinor has four complex components:

$$\mathcal{L}_0 = \bar{c}(\mathrm{i}\gamma^\mu\partial_\mu - m_c)c, \qquad \bar{c} \equiv c^\dagger\gamma^0, \qquad \not{a} \equiv \gamma^\mu a_\mu,$$

$$(\mathrm{i}\not{\partial} - m_c)c = 0, \qquad (\not{k} - m_c)u_h(k) = 0, \qquad (\not{k} + m_c)v_h(k) = 0,$$

$$\gamma^0 = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}, \qquad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}, \qquad \{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} I_4,$$

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

$$(\not{k} + m_c)(\not{k} - m_c) = (k^2 - m_c^2) I_4.$$

Thus each free spinor component satisfies the relativistic mass-shell equation. There are two independent spin labels $h = \pm\frac{1}{2}$ for each particle species.

$$\bar{u}_h u_{h'} = 2m_c\delta_{hh'}, \qquad \bar{v}_h v_{h'} = -2m_c\delta_{hh'},$$

$$\sum_h u_h(k)\bar{u}_h(k) = \not{k} + m_c, \qquad \sum_h v_h(k)\bar{v}_h(k) = \not{k} - m_c.$$

Quantization: promote wave coefficients to creation/annihilation operators.

$$[\mathrm{d}k] \equiv \frac{\mathrm{d}k^+\, \mathrm{d}^2\boldsymbol{k}}{2k^+(2\pi)^3}, \qquad k^+ > 0, \qquad k^- = \frac{\boldsymbol{k}^2 + m_c^2}{2k^+},$$

$$c_i(x) = \sum_h \int [\mathrm{d}k] \left[b_{hi}(k) u_h(k) \mathrm{e}^{-\mathrm{i}k\cdot x} + d_{hi}^\dagger(k) v_h(k) \mathrm{e}^{+\mathrm{i}k\cdot x} \right],$$

$$\{b_{hi}(k), b_{h'j}^\dagger(k')\} = \{d_{hi}(k), d_{h'j}^\dagger(k')\} = 2k^+(2\pi)^3\delta^{(3)}(\hat{k} - \hat{k}')\delta_{hh'}\delta_{ij},$$

$$\{b, b\} = \{d, d\} = \{b, d\} = \{b, d^\dagger\} = 0, \qquad b|0\rangle = d|0\rangle = 0,$$

$$|k, h, i\rangle_c = b_{hi}^\dagger(k)|0\rangle, \qquad |\bar{k}, \bar{h}, j\rangle_{\bar{c}} = d_{\bar{h}j}^\dagger(\bar{k})|0\rangle,$$

$$\langle k', h', j | k, h, i\rangle_c = 2k^+(2\pi)^3\delta^{(3)}(\hat{k} - \hat{k}')\delta_{hh'}\delta_{ij}.$$

Here $i, j = 1, \ldots, N_c$ are color labels; I_n is the n-dimensional identity. The braces mean $\{A, B\} = AB + BA$. They implement fermion statistics.

Local electromagnetic and color interactions: the action we use.

$$\mathcal{S} = \int \mathrm{d}^4x\, \mathcal{L}, \qquad \mathcal{L} = -\tfrac{1}{4}F^{\mathrm{em}}_{\mu\nu}F^{\mu\nu}_{\mathrm{em}} - \tfrac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} + \bar{c}(\mathrm{i}\gamma^\mu D_\mu - m_c)c,$$

$$D_\mu = \partial_\mu + \mathrm{i}ee_cA^{\mathrm{em}}_\mu - \mathrm{i}gA^a_\mu t^a, \qquad e_c = \tfrac{2}{3}, \quad \alpha_{\mathrm{em}} = \frac{e^2}{4\pi}, \quad \alpha_s = \frac{g^2}{4\pi},$$

$$F^{\mathrm{em}}_{\mu\nu} = \partial_\mu A^{\mathrm{em}}_\nu - \partial_\nu A^{\mathrm{em}}_\mu, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - \mathrm{i}g[A_\mu, A_\nu],$$

$$A_\mu \equiv A^a_\mu t^a, \qquad \mathcal{L}_{\mathrm{int}} = -ee_c\bar{c}\gamma^\mu cA^{\mathrm{em}}_\mu + g\bar{c}\gamma^\mu t^a cA^a_\mu.$$

The Hermitian color generators t^a will be specified on Section 6. The displayed action omits the electron and other quark flavors; the electron supplies the electromagnetic source.

Derive the pair-creation vertex from the first interaction. For $A^{\mathrm{em}}_\mu(x) = \epsilon_\mu \mathrm{e}^{-\mathrm{i}q\cdot x}$, contraction of the mode operators gives

$$\begin{aligned}
|c\bar{c}\rangle &= b^\dagger_{hi}(k)d^\dagger_{\bar{h}j}(\bar{k})|0\rangle,\\
\langle c\bar{c}|\bar{c}(x)\gamma^\mu c(x)|0\rangle &= \delta^{ij}\bar{u}_h(k)\gamma^\mu v_{\bar{h}}(\bar{k})\mathrm{e}^{\mathrm{i}(k+\bar{k})\cdot x},\\
\int \mathrm{d}x^-\ \mathrm{d}^2\boldsymbol{x}\, \mathrm{e}^{\mathrm{i}[(k^+ + \bar{k}^+ - q^+)x^- - (\boldsymbol{k}+\bar{\boldsymbol{k}}-\boldsymbol{q})\cdot\boldsymbol{x}]} &= (2\pi)^3\delta^{(3)}(\hat{k} + \hat{\bar{k}} - \hat{q}),\\
\langle c\bar{c}|P^-_{I,\mathrm{em}}(x^+)|0\rangle &= (2\pi)^3\delta^{(3)}(\hat{k} + \hat{\bar{k}} - \hat{q})V^{ij}_{h\bar{h}}\mathrm{e}^{\mathrm{i}(k^- + \bar{k}^- - q^-)x^+},\\
V^{ij}_{h\bar{h}} &= ee_c\delta^{ij}\bar{u}_h(k)\not{\epsilon}\, v_{\bar{h}}(\bar{k}),\\
U_I(0,-\infty) &= 1 - \mathrm{i}\int_{-\infty}^0 \mathrm{d}x^+\, P^-_I(x^+) + \cdots .
\end{aligned}$$

$P^-_{I,\mathrm{em}} = ee_c\int \mathrm{d}x^-\ \mathrm{d}^2\boldsymbol{x}\, \bar{c}\gamma^\mu cA^{\mathrm{em}}_\mu$ is the current-insertion interaction. Spatial integrals conserve $p^+, \boldsymbol{p}$; the remaining time integral produces Section 3's energy denominator.

Gauge invariance is a constraint on the completed amplitude.

$$J^\mu \equiv \bar{c}\gamma^\mu c, \qquad \partial_\mu J^\mu = 0,$$

$$\mathcal{M}_{ep} \propto \frac{e}{Q^2}[\bar{u}(l')\gamma_\mu u(l)]\mathcal{H}^\mu, \qquad q_\mu\mathcal{H}^\mu = 0.$$

The Dirac equations imply current conservation; all electromagnetic attachments must be retained in $\mathcal{H}^\mu$. An isolated light-front intermediate state need not conserve p^-: Section 4 resolves the resulting longitudinal contact term.

Student check. Use the anticommutator to obtain the one-particle normalization. Identify the $b^\dagger d^\dagger$ term that creates the pair.

Keep. The field operators create the pair; the vertex supplies its spin and charge; the time integral supplies its propagation denominator.

Check. Which term in $\bar{c}\gamma^\mu c$ creates the pair from the vacuum?

Next. *What does the remaining time integral tell us about the pair's lifetime?*

3 A virtual photon makes an intermediate pair

Purpose. Derive the pair's lifetime and fix the wavefunction measure before calculating its numerator.

The virtual photon does not decay into a free on-shell pair in empty space. It supplies an intermediate fluctuation with a calculable light-front energy mismatch; that mismatch sets both the denominator of the amplitude and the time available for the pair to meet the target.

Keep the measured virtuality; put the intermediate fermions on shell.

$$q = \left(q^+, -\frac{Q^2}{2q^+}, \mathbf{0}\right), \qquad k^+ = zq^+, \qquad \bar{k}^+ = (1-z)q^+,$$
$$\boldsymbol{k} = -\bar{\boldsymbol{k}}, \qquad k^- = \frac{\boldsymbol{k}^2 + m_c^2}{2zq^+}, \qquad \bar{k}^- = \frac{\boldsymbol{k}^2 + m_c^2}{2(1-z)q^+}.$$

$$\begin{aligned} E_D &\equiv q^- - k^- - \bar{k}^- \\ &= -\frac{1}{2q^+}\left[Q^2 + \frac{\boldsymbol{k}^2 + m_c^2}{z} + \frac{\boldsymbol{k}^2 + m_c^2}{1-z}\right] \\ &= -\frac{\boldsymbol{k}^2 + \epsilon^2}{2q^+ z(1-z)}, \qquad \epsilon^2 \equiv m_c^2 + z(1-z)Q^2. \end{aligned} \tag{3.1}$$

The same perturbation theory as in quantum mechanics, with time x^+. For a vertex V_{fi} at $x^+ < 0$ and scattering at $x^+ = 0$,

$$\begin{aligned} F_{fi} &= -\mathrm{i} \lim_{\eta\to 0^+} \int_{-\infty}^{0} \mathrm{d}x^+ \, \mathrm{e}^{\mathrm{i}(E_f - E_i)x^+ + \eta x^+} V_{fi} \\ &= \lim_{\eta\to 0^+} \frac{-\mathrm{i}V_{fi}}{\eta + \mathrm{i}(E_f - E_i)} = \frac{V_{fi}}{E_i - E_f + \mathrm{i}0}, \\ \Delta x^+ &\sim \frac{1}{|E_D|} = \frac{2q^+}{Q^2 + M_{c\bar{c}}^2}, \qquad M_{c\bar{c}}^2 = \frac{\boldsymbol{k}^2 + m_c^2}{z(1-z)}. \end{aligned}$$

The electron supplies a virtual electromagnetic current. γ_L^* is its longitudinal component, not an asymptotic photon state. The "wavefunction" below is a current-induced splitting amplitude.[1, 2]

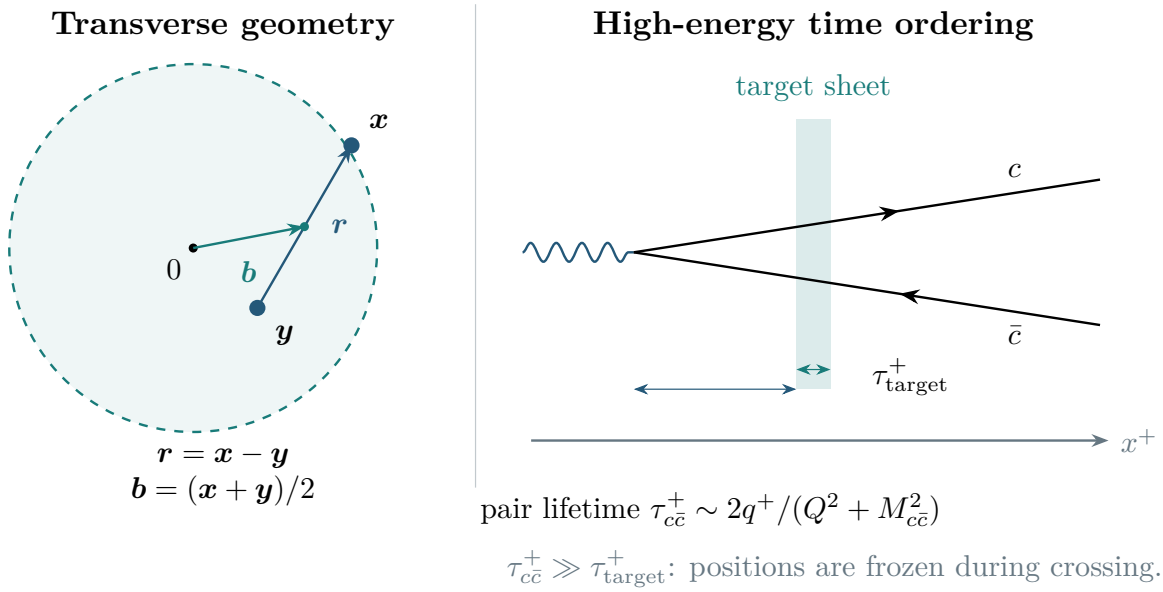


Figure 3: Two descriptions of the same fluctuation. The transverse midpoint and separation are independent coordinates. On the right, the horizontal coordinate is light-front time, not a spatial separation; the arrows specify particle flow and the drawing only sketches the lifetime hierarchy. The disk is a schematic target profile.

Fix the measure. Write $\hat{k} = (k^+, \boldsymbol{k})$ and

$$[\mathrm{d}k] \equiv \frac{\mathrm{d}k^+\ \mathrm{d}^2\boldsymbol{k}}{2k^+(2\pi)^3}, \qquad \langle k'|k\rangle = 2k^+(2\pi)^3\delta^{(3)}(\hat{k}-\hat{k}'),$$

$$|\gamma_L^*(\hat{q})\rangle_{c\bar{c}} \equiv \sum_{i,j,h,\bar{h}} \int [\mathrm{d}k][\mathrm{d}\bar{k}]\,(2\pi)^3\delta^{(3)}(\hat{q}-\hat{k}-\hat{\bar{k}}) \times F_{h\bar{h}}^{ij}|k,h,i;\bar{k},\bar{h},j\rangle.$$

This is notation for the current insertion. Divide the pair norm by the reference external normalization $2q^+(2\pi)^3\delta^{(3)}(0)$:

$$\mathrm{d}\mathcal{W}_{c\bar{c}} = \sum_{i,j,h,\bar{h}} \frac{[\mathrm{d}k][\mathrm{d}\bar{k}]}{2q^+}(2\pi)^3\delta^{(3)}(\hat{q}-\hat{k}-\hat{\bar{k}})|F_{h\bar{h}}^{ij}|^2.$$

$$\begin{aligned}\frac{[\mathrm{d}k][\mathrm{d}\bar{k}]}{2q^+}(2\pi)^3\delta^{(3)}(\hat{q}-\hat{k}-\hat{\bar{k}}) &= \frac{1}{2q^+}\frac{\mathrm{d}k^+}{2k^+(2\pi)}\frac{\mathrm{d}\bar{k}^+}{2\bar{k}^+(2\pi)}(2\pi)\delta(q^+-k^+-\bar{k}^+) \\ &\quad\times \frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2}\frac{\mathrm{d}^2\bar{\boldsymbol{k}}}{(2\pi)^2}(2\pi)^2\delta^{(2)}(\boldsymbol{k}+\bar{\boldsymbol{k}}) \\ &\xrightarrow{\int \mathrm{d}\bar{k}^+\ \mathrm{d}^2\bar{\boldsymbol{k}}} \frac{1}{2q^+}\frac{\mathrm{d}k^+}{4k^+(q^+-k^+)(2\pi)}\frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2} \\ &\xrightarrow{k^+=zq^+} \frac{\mathrm{d}z\ \mathrm{d}^2\boldsymbol{k}}{(2\pi)^2}\frac{1}{16\pi(q^+)^2z(1-z)}.\end{aligned}$$

$$\widetilde{\psi}_{L,h\bar{h}}^{ij}(z,\boldsymbol{k}) \equiv \frac{F_{h\bar{h}}^{ij}}{4\sqrt{\pi}\,q^+\sqrt{z(1-z)}}, \qquad \mathrm{d}\mathcal{W}_{c\bar{c}} = \mathrm{d}z\frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2}\sum_{i,j,h,\bar{h}}|\widetilde{\psi}_{L,h\bar{h}}^{ij}|^2. \tag{3.2}$$

Student check. Reproduce the factor $16\pi(q^+)^2z(1-z)$. Which momentum components are conserved at an individual light-front vertex, and which is not?

Keep. The denominator is a calculable lifetime; the prefactor is fixed by state normalization.

Check. Why is there no $\delta(q^- - k^- - \bar{k}^-)$ in this intermediate-state calculation?

Next. *What spin states multiply this propagation denominator?*

4 Calculate the longitudinal spin numerator

Purpose. Evaluate the electromagnetic current explicitly and identify the gauge contact term.

Now calculate the numerator with explicit spinors. The useful simplification of a longitudinal photon is that the completed conserved current can be written using J^+; the apparent extra term from an isolated intermediate state is a contact term, whose zero-size dipole does not scatter.

Matrices and the two independent spin components. Use the Dirac matrices displayed on Section 2.

$$\alpha^i \equiv \gamma^0\gamma^i, \qquad \beta \equiv \gamma^0, \qquad \gamma^\pm \equiv (\gamma^0 \pm \gamma^3)/\sqrt{2},$$
$$\Lambda_\pm \equiv \frac{1 \pm \alpha^3}{2}, \qquad \Lambda_+ = \frac{\gamma^-\gamma^+}{2}, \qquad \gamma^0\gamma^+ = \sqrt{2}\Lambda_+.$$

$$\begin{aligned}
(\not{k} - m_c)u = 0 \quad &\Longleftrightarrow \quad k^0 u = (\alpha^3 k^3 + \boldsymbol{\alpha}_\perp \cdot \boldsymbol{k} + \beta m_c)u, \\
u = u_+ + u_-, \quad u_\pm = \Lambda_\pm u, \qquad &\{\alpha^3, \alpha^{1,2}\} = \{\alpha^3, \beta\} = 0, \\
(k^0 + k^3)u_- &= (\boldsymbol{\alpha}_\perp \cdot \boldsymbol{k} + \beta m_c)u_+, \\
u_- = \frac{\boldsymbol{\alpha}_\perp \cdot \boldsymbol{k} + \beta m_c}{\sqrt{2}k^+}u_+, \qquad &v_- = \frac{\boldsymbol{\alpha}_\perp \cdot \bar{\boldsymbol{k}} - \beta m_c}{\sqrt{2}\bar{k}^+}v_+.
\end{aligned}$$

The subscript + denotes a projected component, not positive helicity. The equation determines the "bad" component from the independent "good" component.

$$\begin{aligned}
w_+ &= \frac{1}{\sqrt{2}}(1,0,1,0)^T, & w_- &= \frac{1}{\sqrt{2}}(0,1,0,-1)^T, \\
\Lambda_+ w_s &= w_s, & w_s^\dagger w_{s'} &= \delta_{ss'}, \\
u_+(k,h) &= 2^{1/4}\sqrt{k^+}\, w_h, & v_+(\bar{k},\bar{h}) &= 2^{1/4}\sqrt{\bar{k}^+}\, w_{-\bar{h}}.
\end{aligned}$$

Here $h, \bar{h} = \pm\frac{1}{2}$ label light-front helicities; the displayed equations fix antiquark phases. They give $\bar{u}u = 2m_c$ and $\bar{v}v = -2m_c$.

$$\begin{aligned}
J^+_{h\bar{h}} &\equiv \bar{u}_h(k)\gamma^+ v_{\bar{h}}(\bar{k}) = \sqrt{2}\, u_+^\dagger v_+ \\
&= 2\sqrt{k^+\bar{k}^+}\, w_h^\dagger w_{-\bar{h}} = 2q^+\sqrt{z(1-z)}\,\delta_{h,-\bar{h}}.
\end{aligned} \tag{4.1}$$

Check without choosing spinors.

$$\sum_h u_h\bar{u}_h = \not{k} + m_c, \qquad \sum_{\bar{h}} v_{\bar{h}}\bar{v}_{\bar{h}} = \bar{\not{k}} - m_c,$$
$$\begin{aligned}
\sum_{h,\bar{h}} |J^+_{h\bar{h}}|^2 &= \mathrm{Tr}[(\not{k} + m_c)\gamma^+(\bar{\not{k}} - m_c)\gamma^+] \\
&= 4\big(k^+\bar{k}^+ + k^+\bar{k}^+ - k\cdot\bar{k}\, g^{++}\big) = 8k^+\bar{k}^+.
\end{aligned}$$

The mass term vanishes because $(\gamma^+)^2 = 0$; odd traces vanish. Equation (4.1) gives the same result: two nonzero helicity pairs, each of modulus squared $4k^+\bar{k}^+$.

Which longitudinal polarization can we use?

$$\epsilon_L^\mu = \left(\frac{q^+}{Q}, \frac{Q}{2q^+}, \mathbf{0}\right), \qquad q\cdot\epsilon_L = 0, \qquad \epsilon_L^2 = +1,$$
$$n^\mu = (0,1,\mathbf{0}), \qquad \epsilon_L^\mu = \frac{q^\mu}{Q} + \frac{Q}{q^+}n^\mu, \qquad \not{n} = \gamma^+.$$

Current conservation constrains the *complete* photon–target amplitude:

$$q_\mu \mathcal{M}^\mu = 0 \quad \Longrightarrow \quad \epsilon_{L\mu}\mathcal{M}^\mu = \frac{Q}{q^+} n_\mu \mathcal{M}^\mu.$$

It does not set the isolated intermediate-pair current contraction to zero:

$$\begin{aligned} q^\mu &= k^\mu + \bar{k}^\mu + E_D n^\mu, \\ \bar{u}\not{q}v &= \bar{u}(\not{k} + \not{\bar{k}})v + E_D \bar{u}\gamma^+ v = (m_c - m_c)\bar{u}v + E_D J^+ = E_D J^+, \\ \frac{\bar{u}\not{\epsilon}_L v}{E_D} &= \frac{QJ^+}{q^+E_D} + \frac{J^+}{Q}. \end{aligned}$$

The last term has no $\boldsymbol{k}$ dependence:

$$\int \frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \frac{J^+}{Q} = \frac{J^+}{Q}\delta^{(2)}(\boldsymbol{r}),$$
$$S_{\boldsymbol{xx}} = 1 \quad \Longrightarrow \quad [1 - S_{\boldsymbol{xy}}]\delta^{(2)}(\boldsymbol{x}-\boldsymbol{y}) = 0.$$

$S_{\boldsymbol{xy}}$ will be the color-singlet pair's scattering matrix; coincident quark and antiquark color rotations cancel exactly. Thus this contact term does not scatter. Canonical light-front theory includes the corresponding instantaneous lepton–quark interaction and directly supplies the effective vertex[1, 2]

$$V^{ij}_{L,h\bar{h}} = \delta^{ij} e e_c \frac{Q}{q^+} J^+_{h\bar{h}}. \tag{4.2}$$

Insert the numerator and the denominator.

$$\begin{aligned} F^{ij}_{h\bar{h}} &= \frac{\delta^{ij} e e_c (Q/q^+)\, 2q^+ \sqrt{z(1-z)}\delta_{h,-\bar{h}}}{-(\boldsymbol{k}^2 + \epsilon^2)/[2q^+ z(1-z)]} \\ &= -\frac{4 e e_c q^+ Q [z(1-z)]^{3/2}}{\boldsymbol{k}^2 + \epsilon^2} \delta^{ij}\delta_{h,-\bar{h}}. \end{aligned}$$

$$\boxed{\widetilde{\psi}^{ij}_{L,h\bar{h}}(z,\boldsymbol{k}) = -\frac{e e_c}{\sqrt{\pi}} \frac{Qz(1-z)}{\boldsymbol{k}^2 + \epsilon^2} \delta^{ij}\delta_{h,-\bar{h}}.} \tag{4.3}$$

Student check. Verify $\bar{u}u = 2m_c$ using the good/bad components. Then show explicitly why $q_\mu \bar{u}\gamma^\mu v \neq 0$ here.

Keep. Longitudinal production selects opposite light-front helicities; gauge invariance applies to the complete amplitude.

Check. Why can we use the Ward identity for the completed amplitude but not simply impose it on the isolated light-front vertex?

Next. *Which transverse separations are present in this momentum-space fluctuation?*

5 Transform the pair into a dipole

Purpose. Calculate the transverse-size distribution that will weight the target's scattering amplitude.

The Fourier transform changes the description from relative momentum to the distance between the two color charges. This makes the target interaction simple: a fast pair keeps approximately the same transverse size while crossing the proton, and the photon tells us how strongly each size is weighted.

$$\psi^{ij}_{L,h\bar{h}}(z,\boldsymbol{r}) \equiv \int \frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\widetilde{\psi}^{ij}_{L,h\bar{h}}(z,\boldsymbol{k}), \qquad \boldsymbol{r}\equiv\boldsymbol{x}-\boldsymbol{y},$$
$$\frac{1}{\boldsymbol{k}^2+\epsilon^2} = \int_0^\infty \mathrm{d}\tau\,\mathrm{e}^{-\tau(\boldsymbol{k}^2+\epsilon^2)}, \qquad \epsilon^2>0.$$

The Fourier transform is a Gaussian integral followed by one ordinary integral.

$$\begin{aligned} G(\boldsymbol{r}) &\equiv \int \frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2}\frac{\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}}{\boldsymbol{k}^2+\epsilon^2} \\ &= \int_0^\infty \mathrm{d}\tau\,\mathrm{e}^{-\epsilon^2\tau}\prod_{a=1}^{2}\left[\int\frac{\mathrm{d}k_a}{2\pi}\mathrm{e}^{-\tau k_a^2+\mathrm{i}k_a r_a}\right] \\ &= \frac{1}{4\pi}\int_0^\infty\frac{\mathrm{d}\tau}{\tau}\mathrm{e}^{-\epsilon^2\tau-r^2/(4\tau)} = \frac{1}{4\pi}\int_0^\infty\frac{\mathrm{d}t}{t}\mathrm{e}^{-t-(\epsilon r)^2/(4t)}, \\ K_0(a) &\equiv \frac{1}{2}\int_0^\infty\frac{\mathrm{d}t}{t}\mathrm{e}^{-t-a^2/(4t)} \quad\Longrightarrow\quad G(\boldsymbol{r}) = \frac{K_0(\epsilon r)}{2\pi}. \end{aligned}$$

$$\psi^{ij}_{L,h\bar{h}} = -\frac{ee_c}{2\pi\sqrt{\pi}}Qz(1-z)K_0(\epsilon r)\delta^{ij}\delta_{h,-\bar{h}}. \tag{5.1}$$

Project onto the normalized color singlet.

$$|1\rangle = \frac{1}{\sqrt{N_c}}\sum_{i=1}^{N_c}|i\bar{i}\rangle, \qquad \langle 1|1\rangle = 1,$$
$$\psi^{(1)}_{L,h\bar{h}} = \frac{1}{\sqrt{N_c}}\sum_{i,j}\delta^{ij}\psi^{ij}_{L,h\bar{h}} = -\frac{e_c\sqrt{N_c\alpha_{\mathrm{em}}}}{\pi}Qz(1-z)K_0(\epsilon r)\delta_{h,-\bar{h}}, \qquad \alpha_{\mathrm{em}}\equiv\frac{e^2}{4\pi}.$$

$$\boxed{w_L(z,r)\equiv\sum_{h,\bar{h}}|\psi^{(1)}_{L,h\bar{h}}|^2 = \frac{2N_c\alpha_{\mathrm{em}}e_c^2}{\pi^2}Q^2z^2(1-z)^2K_0^2(\epsilon r).} \tag{5.2}$$

$$\int \mathrm{d}z\frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2}\sum_{i,j,h,\bar{h}}|\widetilde{\psi}^{ij}_{L,h\bar{h}}|^2 = \int \mathrm{d}z\;\mathrm{d}^2\boldsymbol{r}\,w_L(z,r),$$

$$K_0(a)\underset{a\to\infty}{\sim}\sqrt{\frac{\pi}{2a}}\mathrm{e}^{-a}, \qquad r_{\mathrm{typical}}\sim\epsilon^{-1},$$
$$Q\to 0: \quad \epsilon\to m_c, \qquad \psi_L\propto Q\to 0.$$

Student check. Check the factor $2N_c$ in w_L and derive G from the two one-dimensional Gaussian integrals. Why is longitudinal photoproduction zero?

Keep. The photon supplies a calculable dipole-size weight; the unknown target structure enters next.

Check. Which scale suppresses large dipoles, and what happens to this longitudinal component at $Q = 0$?

Checkpoint I: the probe is now known.

$$\underbrace{\mathcal{L}_{\text{em}}}_{\text{create the pair}} \longrightarrow \underbrace{\frac{\bar{u}\not{\epsilon}v}{\Delta E^-}}_{\text{spin and lifetime}} \longrightarrow \underbrace{\psi_L^{(1)}(z,r) \propto Qz(1-z)K_0(\epsilon r)}_{\text{weight of each size}}.$$

Nothing about the proton has entered this photon wavefunction. We have calculated which dipoles the probe supplies. We now need the amplitude for each of those dipoles to cross the proton. **Check.** What is known, and what is still missing?

Next. *How does a quark of fixed transverse position interact with the target field?*

6 A fast quark undergoes a color rotation

Purpose. Derive the target interaction from the same Dirac equation used for photon splitting.

We now use the Dirac equation a second time, for propagation rather than pair creation. At very high energy the leading target interaction rotates color along an almost straight path; transverse deflection and spin changes are subleading in this approximation.

Color adds an internal three-dimensional space. The charm field c_i carries color $i = 1, 2, 3$; spinor and color matrices act on different indices.

$$N_c = 3, \qquad a = 1, \dots, N_c^2 - 1, \qquad t^{a\dagger} = t^a, \qquad \operatorname{Tr} t^a = 0,$$
$$[t^a, t^b] = \mathrm{i} f^{abc} t^c, \qquad \operatorname{Tr}(t^a t^b) = \tfrac{1}{2}\delta^{ab}, \qquad \sum_a t^a t^a = C_F I_{N_c},$$
$$C_F N_c = \sum_a \operatorname{Tr}(t^a t^a) = \frac{N_c^2 - 1}{2}, \qquad C_F = \frac{N_c^2 - 1}{2N_c} = \frac{4}{3}.$$

Here f^{abc} are the structure constants; I_{N_c} is the color identity.

$$c'(x) = \Omega(x) c(x), \qquad \Omega(x) = \exp[\mathrm{i}\theta^a(x) t^a],$$
$$\partial_\mu c' = \Omega \partial_\mu c + (\partial_\mu \Omega) c, \qquad D_\mu \equiv \partial_\mu - \mathrm{i} g A_\mu, \qquad A_\mu = A_\mu^a t^a,$$
$$A'_\mu = \Omega A_\mu \Omega^\dagger - \frac{\mathrm{i}}{g}(\partial_\mu \Omega)\Omega^\dagger, \qquad D'_\mu c' = \Omega D_\mu c,$$
$$[D_\mu, D_\nu] = -\mathrm{i} g F_{\mu\nu}, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - \mathrm{i} g [A_\mu, A_\nu],$$
$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu} + \bar{c}(\mathrm{i}\gamma^\mu D_\mu - m_c)c, \qquad \alpha_s \equiv \frac{g^2}{4\pi}.$$

Other quark flavors are omitted here. Treat the target field A_μ as prescribed while solving the projectile's propagation; take its proton matrix element afterwards.

Extract the rapid free oscillation.

$$(\mathrm{i}\not{\partial} + g\not{A} - m_c)\psi_p = 0, \qquad \psi_p(x) = \mathrm{e}^{-\mathrm{i}p\cdot x} \sum_s u_s(p) w_s(x),$$
$$(\not{p} - m_c) u_s = 0, \qquad \bar{u}_{s'} u_s = 2m_c \delta_{s's},$$
$$2m_c \bar{u}_{s'} \gamma^\mu u_s = \bar{u}_{s'}(\not{p}\gamma^\mu + \gamma^\mu \not{p}) u_s = 2p^\mu \bar{u}_{s'} u_s,$$
$$\bar{u}_{s'} \gamma^\mu u_s = 2p^\mu \delta_{s's}.$$

The w_s are color vectors. Projection onto the fast spinors gives, at leading high-energy order,

$$0 = 2ip^\mu \partial_\mu w_s + 2gp^\mu A_\mu w_s \simeq 2p^+ \left(\mathrm{i}\frac{\partial}{\partial x^+} + gA^- \right) w_s,$$
$$\frac{\partial w_s}{\partial x^+} = \mathrm{i}gA^- w_s. \tag{6.1}$$

$$p^+ \gg m_c, |\boldsymbol{p}|, |\boldsymbol{\Delta}|, \qquad \frac{L^+(m_c^2 + \boldsymbol{p}^2)}{2p^+} \ll 1, \qquad \frac{2q^+}{Q^2 + M_{c\bar{c}}^2} \gg L^+.$$

L^+ is the target crossing time. Neglected components describe suppressed spin changes and transverse deflections. Neither $g\int A^- \,\mathrm{d}x^+$ nor the number of interactions must be small.

Student check. Verify $D'_\mu c' = \Omega D_\mu c$. Which term would prevent local color invariance if A_μ were absent?

Keep. The fast quark follows an approximately fixed path while its color evolves.

Check. What changes along the leading eikonal trajectory, and what is held approximately fixed?

Next. *How do we solve this equation when successive color matrices do not commute?*

7 From ordered rotations to a color-singlet dipole

Purpose. Sum repeated scattering and derive the first nonzero color-singlet interaction.

The Wilson line is the solution of that color-transport equation, not an additional particle or a new interaction. Combining quark and antiquark transport and projecting back onto a color singlet gives the target operator we will use in every later observable.

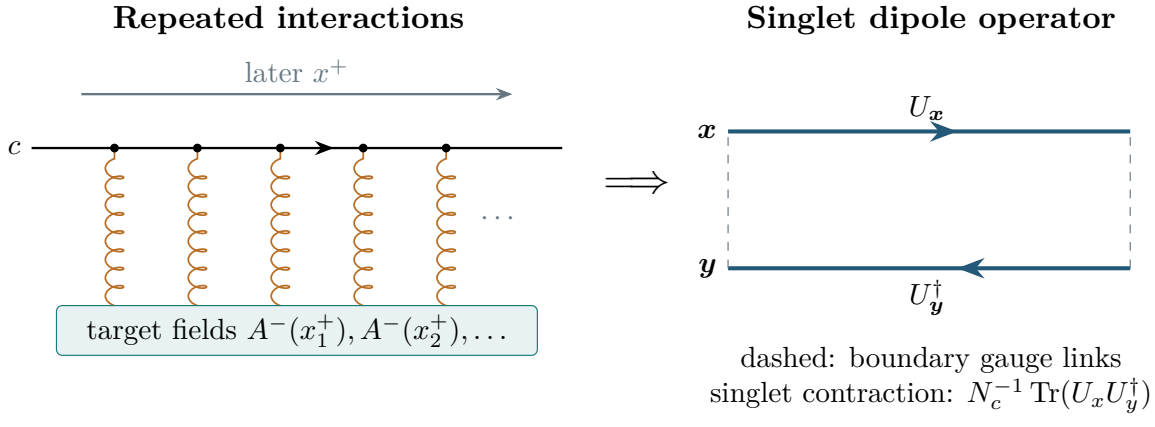


Figure 4: The eikonal approximation keeps the fast trajectory fixed while summing arbitrarily many ordered colour rotations. The conjugate antiquark line completes the singlet dipole. Boundary links are required for an explicitly closed gauge-invariant operator; they can become trivial under the boundary convention used in the text. The expansion on the left depicts only one of the two lines.

Solve the transport equation by iteration. Hold the transverse position $\boldsymbol{x}$ fixed:

$$
\begin{aligned}
w(b^+) &= U(b^+,a^+)w(a^+), & U(a^+,a^+) &= I_{N_c},\\
U(b,a) &= I_{N_c} + \mathrm{i}g\int_a^b \mathrm{d}\xi\, A^-(\xi)U(\xi,a)\\
&= I_{N_c} + \mathrm{i}g\int_a^b \mathrm{d}\xi\, A^-(\xi) + (\mathrm{i}g)^2\int_a^b \mathrm{d}\xi\int_a^\xi \mathrm{d}\eta\, A^-(\xi)A^-(\eta) + \cdots,\\
U_{\boldsymbol{x}} &\equiv \mathcal{P}\exp\left[\mathrm{i}g\int_{-\infty}^{\infty}\mathrm{d}x^+\, A^-(x^+,\boldsymbol{x})\right],\\
\partial_+(U^\dagger U) &= -\mathrm{i}gU^\dagger A^- U + \mathrm{i}gU^\dagger A^- U = 0, & U^\dagger U &= I_{N_c}.
\end{aligned}
$$

$\mathcal{P}$ places later interactions on the left. Antiquarks transform with U^*.

$$
\begin{aligned}
|1\rangle &\equiv \frac{\delta_{ij}}{\sqrt{N_c}}|i\bar{j}\rangle, & \hat{S}_{\boldsymbol{xy}} &\equiv \langle 1|U_{\boldsymbol{x}}\otimes U^*_{\boldsymbol{y}}|1\rangle\\
&= \frac{1}{N_c}\delta_{kl}(U_{\boldsymbol{x}})_{ki}(U^*_{\boldsymbol{y}})_{lj}\delta_{ij} = \frac{1}{N_c}\,\mathrm{Tr}(U_{\boldsymbol{x}}U^\dagger_{\boldsymbol{y}}). & & (7.1)
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{r} &\equiv \boldsymbol{x}-\boldsymbol{y}, & \boldsymbol{b} &\equiv (\boldsymbol{x}+\boldsymbol{y})/2,\\
S(x,\boldsymbol{r},\boldsymbol{b}) &\equiv \left\langle \hat{S}_{\boldsymbol{xy}}\right\rangle_x = 1 - N(x,\boldsymbol{r},\boldsymbol{b}) + \mathrm{i}O(x,\boldsymbol{r},\boldsymbol{b}),\\
\hat{S}_{\boldsymbol{yx}} &= \hat{S}^*_{\boldsymbol{xy}}, & N(-\boldsymbol{r}) &= N(\boldsymbol{r}), \qquad O(-\boldsymbol{r}) = -O(\boldsymbol{r}).
\end{aligned}
$$

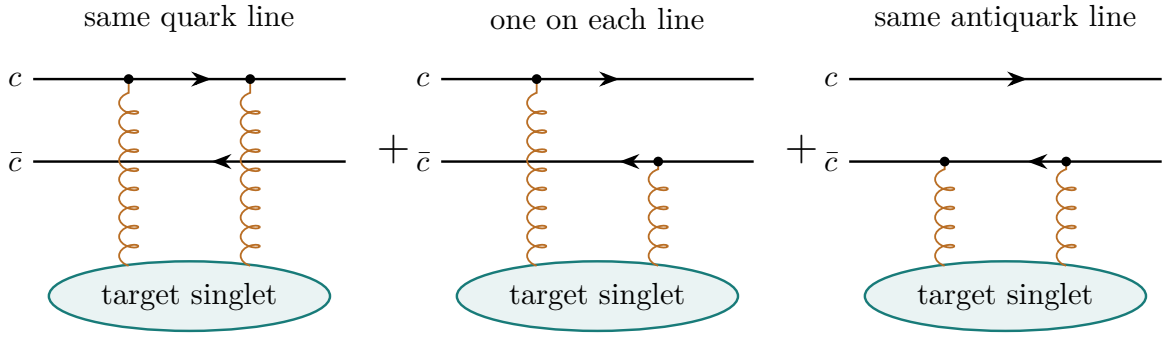
The scalar x specifies the target rapidity scale; bold $\boldsymbol{x}$ is a transverse coordinate. The average describes the proton, not an average over the projectile final state.[3, 4]

Gauge completion.

$$
\begin{aligned}
U_{\boldsymbol{x}}(b,a) &\longrightarrow \Omega(b,\boldsymbol{x})U_{\boldsymbol{x}}(b,a)\Omega^\dagger(a,\boldsymbol{x}),\\
\hat{S}^{\text{closed}}_{\boldsymbol{xy}} &= \frac{1}{N_c}\,\mathrm{Tr}[U_{\boldsymbol{x}}L_-(\boldsymbol{x},\boldsymbol{y})U^\dagger_{\boldsymbol{y}}L_+(\boldsymbol{y},\boldsymbol{x})].
\end{aligned}
$$

L_- and L_+ join the past and future endpoints. Their endpoint rotations cancel. Equation (7.1) uses boundary conditions where these links are trivial; quantum lightlike operators also require a rapidity prescription.

Two-field expansion: keep the color factors.



$$
\begin{aligned}
\delta\alpha^a\delta\alpha^a &= \alpha^a_x\alpha^a_x - 2\alpha^a_x\alpha^a_y + \alpha^a_y\alpha^a_y\\
N^{(2)}_{xy} &= \frac{g^2}{4N_c}\left\langle(\alpha^a_x-\alpha^a_y)^2\right\rangle
\end{aligned}
$$

The mixed term includes both orderings. At $\boldsymbol{x}=\boldsymbol{y}$, the sum vanishes.

Figure 5: Representative attachment classes, with the singlet colour projection understood. Crossings of a gluon with the lower fermion line without a marked vertex do not denote an interaction. Longitudinal orderings and the signs from the antiquark representation must be included. Their sum produces the difference of fields and therefore colour transparency; an isolated class does not.

Let $\alpha^a_{\boldsymbol{x}} \equiv \int \mathrm{d}x^+ A^-_a(x^+, \boldsymbol{x})$ and $\alpha_{\boldsymbol{x}} \equiv t^a \alpha^a_{\boldsymbol{x}}$. Inside the trace, ordering gives the following quadratic expression:

$$\begin{aligned}
\mathrm{Tr}(U_{\boldsymbol{x}} U^\dagger_{\boldsymbol{y}}) &= N_c + \mathrm{i}g\,\mathrm{Tr}(\alpha_{\boldsymbol{x}} - \alpha_{\boldsymbol{y}}) \\
&\quad + g^2\,\mathrm{Tr}\Big[\alpha_{\boldsymbol{x}}\alpha_{\boldsymbol{y}} - \tfrac{1}{2}\alpha^2_{\boldsymbol{x}} - \tfrac{1}{2}\alpha^2_{\boldsymbol{y}}\Big] + \mathcal{O}((gA)^3), \\
\mathrm{Tr}(\alpha_{\boldsymbol{x}} - \alpha_{\boldsymbol{y}}) &= 0, \\
g^2\,\mathrm{Tr}\Big[\alpha_{\boldsymbol{x}}\alpha_{\boldsymbol{y}} - \tfrac{1}{2}\alpha^2_{\boldsymbol{x}} - \tfrac{1}{2}\alpha^2_{\boldsymbol{y}}\Big] &= -\frac{g^2}{2}\,\mathrm{Tr}(\alpha_{\boldsymbol{x}} - \alpha_{\boldsymbol{y}})^2 \\
&= -\frac{g^2}{4}(\alpha^a_{\boldsymbol{x}} - \alpha^a_{\boldsymbol{y}})(\alpha^a_{\boldsymbol{x}} - \alpha^a_{\boldsymbol{y}}).
\end{aligned}$$

$$\boxed{N^{(2)}(x, \boldsymbol{r}, \boldsymbol{b}) = \frac{g^2}{4N_c}\left\langle (\alpha^a_{\boldsymbol{x}} - \alpha^a_{\boldsymbol{y}})^2 \right\rangle_x .} \tag{7.2}$$

Repeated color indices are summed. Superscript (2) counts target fields; it does not denote a squared amplitude.

$$\begin{aligned}
\boldsymbol{r} = 0: \quad U_{\boldsymbol{x}} U^\dagger_{\boldsymbol{x}} = I_{N_c} \quad &\Longrightarrow \quad S = 1, \quad N = O = 0, \\
\alpha^a_{\boldsymbol{b}+\boldsymbol{r}/2} - \alpha^a_{\boldsymbol{b}-\boldsymbol{r}/2} &= r^i \partial_i \alpha^a_{\boldsymbol{b}} + \mathcal{O}(r^3), \\
N^{(2)} &= \frac{g^2 r^i r^j}{4N_c}\,\langle \partial_i \alpha^a \partial_j \alpha^a \rangle + \mathcal{O}(r^4), \\
\left\langle r^i r^j \right\rangle_{\varphi_r} &= \frac{r^2}{2}\delta^{ij}, \qquad\qquad \overline{N}^{(2)} = \frac{g^2 r^2}{8N_c}\,\langle \boldsymbol{\nabla}\alpha^a \cdot \boldsymbol{\nabla}\alpha^a \rangle .
\end{aligned}$$

The gradient correlator is resolved at a finite ultraviolet scale; its logarithms can produce $r^2 \ln(1/r^2)$.

Student check. Recover $1/(4N_c)$ without referring to the answer. Explain why one target gluon cannot produce a nonzero singlet-to-singlet amplitude.

Keep. Color neutrality cancels the common rotation; the first surviving term resolves a field difference.

Check. Why does the color-singlet amplitude have no one-gluon term?

Next. *When do many weak color rotations build an appreciable dipole interaction?*

8 Resum many weak scatterings and identify the gluon density

Purpose. Derive one controlled exponentiation and its dilute matching; separate calculation from model input.

Independent weak target layers provide an explicit example in which many interactions can be summed. Their product exponentiates; a quadratic opacity then gives our chosen saturation model, while its first term can be matched to the familiar dilute gluon density under additional leading-log assumptions.

Independent Gaussian slices: an explicit target model. Slice n gives $V_{u,n} = \exp(\mathrm{i}t^a\delta\theta^a_{u,n})$ at transverse position u.

$$\left\langle \delta\theta^a_{u,n} \right\rangle = 0, \qquad \left\langle \delta\theta^a_{u,n}\delta\theta^b_{v,m} \right\rangle = \delta_{nm}\delta^{ab}\gamma_n(u,v),$$
$$\delta\theta^a_{u,n} = g\int_{\text{slice } n} \mathrm{d}x^+ A^-_a(x^+,u), \qquad \delta\Gamma_n \equiv \frac{1}{2}[\gamma_n(x,x)+\gamma_n(y,y)-2\gamma_n(x,y)].$$

Here x, y inside γ_n denote the two transverse endpoints, not Bjorken x.

$$\begin{aligned}
\left\langle V^\dagger_{y,n}V_{x,n} \right\rangle &= I_{N_c} + \left\langle t^a t^b \delta\theta^a_{y,n}\delta\theta^b_{x,n} \right\rangle - \frac{1}{2}\left\langle t^a t^b(\delta\theta^a_{x,n}\delta\theta^b_{x,n} + \delta\theta^a_{y,n}\delta\theta^b_{y,n}) \right\rangle + \cdots \\
&= [1 - C_F\delta\Gamma_n + \mathcal{O}(\delta\Gamma_n^2)]I_{N_c}, \\
W_{n-1} &\equiv U_{x,n-1}U^\dagger_{y,n-1}, \\
S_n &= \frac{1}{N_c}\left\langle \mathrm{Tr}(V_{x,n}W_{n-1}V^\dagger_{y,n}) \right\rangle = [1 - C_F\delta\Gamma_n + \mathcal{O}(\delta\Gamma_n^2)]S_{n-1}, \\
\ln S &= \sum_n \ln(1 - C_F\delta\Gamma_n) \longrightarrow -C_F\Gamma, \qquad \Gamma \equiv \sum_n \delta\Gamma_n.
\end{aligned}$$

The continuum limit holds Γ fixed and sends $\sum_n \delta\Gamma_n^2 \to 0$.

$$S = \mathrm{e}^{-C_F\Gamma}, \qquad N = 1 - \mathrm{e}^{-C_F\Gamma}, \tag{8.1}$$

$$\Gamma_{\text{model}}(x,r,b) \equiv \frac{Q_s^2(x,b)r^2}{4C_F}, \qquad N_{\text{model}} = 1 - \mathrm{e}^{-r^2Q_s^2(x,b)/4}. \tag{8.2}$$

Q_s has units of momentum. The log-free quadratic Γ_{model} is a chosen phenomenological input, not a general consequence of Gaussian sources.[5, 6]

$$\tau \equiv r^2Q_s^2/4, \qquad N_{\text{model}} = \tau - \tfrac{1}{2}\tau^2 + \tfrac{1}{6}\tau^3 - \cdots, \qquad N_{\text{dilute}} = \tau.$$

Supporting derivation. The integral matching below makes the PDF relation and its assumptions explicit.

Finite-target Fourier matching. For a localized proton field ensemble, define

$$K(\boldsymbol{s}) \equiv \int \mathrm{d}^2\boldsymbol{b}\sum_a \langle \alpha^a(\boldsymbol{b}+\boldsymbol{s}/2)\alpha^a(\boldsymbol{b}-\boldsymbol{s}/2)\rangle\,,$$
$$C(\boldsymbol{k}) \equiv \int \mathrm{d}^2\boldsymbol{s}\,\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{s}}K(\boldsymbol{s}), \qquad K(\boldsymbol{s}) = \int \frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{s}}C(\boldsymbol{k}),$$
$$\begin{aligned}
\int \mathrm{d}^2\boldsymbol{b}\left\langle (\alpha^a_{\boldsymbol{x}} - \alpha^a_{\boldsymbol{y}})^2 \right\rangle &= 2[K(0) - K(\boldsymbol{r})] \\
&= 2\int \frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2}[1 - \cos(\boldsymbol{k}\cdot\boldsymbol{r})]C(\boldsymbol{k}).
\end{aligned}$$

No infinite target area occurs. K has units GeV^{-2} and C has units GeV^{-4}.

$$\sigma_{\mathrm{dip}} \equiv 2\int \mathrm{d}^2\boldsymbol{b}\, N, \qquad \sigma^{(2)}_{\mathrm{dip}} = \frac{g^2}{N_c}\int \frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2}[1-\cos(\boldsymbol{k}\cdot\boldsymbol{r})]C(\boldsymbol{k}), \tag{8.3}$$

$$\mathcal{F}(x,k^2) \equiv \frac{k^4 C(\boldsymbol{k})}{(2\pi)^2}, \qquad \sigma^{(2)}_{\mathrm{dip}} = \frac{4\pi\alpha_s}{N_c}\int \frac{\mathrm{d}^2\boldsymbol{k}}{k^4}\mathcal{F}(x,k^2)[1-\cos(\boldsymbol{k}\cdot\boldsymbol{r})]. \tag{8.4}$$

The factor 2 in σ_{dip} is the optical-theorem normalization. Assume rotational symmetry for the following angular integral:

$$\int_0^{2\pi}\mathrm{d}\varphi\, \mathrm{e}^{\mathrm{i}kr\cos\varphi} = 2\pi J_0(kr), \qquad \mathrm{d}^2\boldsymbol{k} = \tfrac{1}{2}\,\mathrm{d}k^2\,\mathrm{d}\varphi,$$

$$\sigma^{(2)}_{\mathrm{dip}} = \frac{4\pi^2\alpha_s}{N_c}\int_0^\infty \frac{\mathrm{d}k^2}{k^4}\mathcal{F}(x,k^2)[1-J_0(kr)].$$

J_0 is the Bessel function. The QCD matching input, at dilute leading-logarithmic accuracy, is[6, 4]

$$\mathcal{F}(x,k^2) \simeq \frac{\partial[xg(x,k^2)]}{\partial\ln k^2},$$

$$xg(x,\mu^2) = xg(x,\mu_0^2) + \int_{\mu_0^2}^{\mu^2}\frac{\mathrm{d}k^2}{k^2}\mathcal{F}(x,k^2),$$

$$1-J_0(kr) = \frac{k^2r^2}{4} + \mathcal{O}(k^4r^4), \qquad \mu^2 \sim C_\mu/r^2,$$

$$\sigma^{(2)}_{\mathrm{dip}} \simeq \frac{4\pi^2\alpha_s}{N_c}\frac{r^2}{4}\left[xg(x,\mu_0^2) + \int_{\mu_0^2}^{\mu^2}\frac{\mathrm{d}k^2}{k^2}\mathcal{F}(x,k^2)\right].$$

$$\boxed{\sigma^{(2)}_{\mathrm{dip}}(x,r) \simeq \frac{\pi^2}{N_c}r^2\alpha_s(\mu^2)xg(x,\mu^2).} \tag{8.5}$$

g is the renormalized forward gluon PDF; μ_0 sets its infrared input. The order-one constant C_μ and nonlogarithmic terms require matching. This is not an exact dense-target UGD identity or an exclusive collinear-factorization calculation.

Student check. Recover every 2 and π in Eq. (8.5); check dimensions. Expand Eq. (8.1) and recover Eq. (7.2).

Keep. The same target operator contains a dilute gluon-density limit and, with additional target assumptions, multiple scattering.

Check. Which statement is derived from independent Gaussian layers, and which further statement is a model choice?

Next. *How do photon splitting and this target operator become a normalized measured spectrum?*

9 Assemble and normalize the observable

Purpose. Turn the splitting and scattering amplitudes into a measurable coherent spectrum.

We can now assemble the process: create a pair, scatter its color dipole, and project onto the observed outgoing momenta. Because the proton remains intact, average its scattering amplitude before taking the absolute square; changing that order would calculate a different final state.

The proton remains in the same quantum state. For a target configuration ω, write $\hat{S}_\omega$; the proton amplitude contains $S = \langle \hat{S}_\omega \rangle$, before squaring. At leading order the outgoing pair is a continuum state:

$$\langle \boldsymbol{r}|\boldsymbol{p}\rangle = \frac{\mathrm{e}^{\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}}}{2\pi}, \qquad \int \mathrm{d}^2p\, |\boldsymbol{p}\rangle\langle\boldsymbol{p}| = 1,$$
$$\boldsymbol{p}_1\cdot\boldsymbol{x} + \boldsymbol{p}_2\cdot\boldsymbol{y} = \boldsymbol{K}\cdot\boldsymbol{b} + \boldsymbol{p}\cdot\boldsymbol{r}, \qquad \boldsymbol{K} = -\boldsymbol{\Delta}, \quad \boldsymbol{p} = (\boldsymbol{p}_1 - \boldsymbol{p}_2)/2,$$
$$\boldsymbol{\kappa} = \boldsymbol{p} + (z - \tfrac{1}{2})\boldsymbol{\Delta}.$$

The high-energy elastic convention is $\mathcal{M} = s\mathcal{A}$, so that

$$\frac{\mathrm{d}\sigma_{2\to2}}{\mathrm{d}t} = \frac{|\mathcal{M}|^2}{16\pi\lambda(s, m_1^2, m_2^2)} \simeq \frac{|\mathcal{A}|^2}{16\pi}, \qquad \lambda(a,b,c) = a^2 + b^2 + c^2 - 2ab - 2ac - 2bc,$$
$$\mathcal{A}_{\mathrm{dip}}(\boldsymbol{r}, \boldsymbol{\Delta}) = 2\mathrm{i}\int \mathrm{d}^2b\, \mathrm{e}^{\mathrm{i}\boldsymbol{\Delta}\cdot\boldsymbol{b}}[1 - S(\boldsymbol{r}, \boldsymbol{b})], \qquad \sigma_{\mathrm{dip}} = \mathrm{Im}\,\mathcal{A}_{\mathrm{dip}}(\Delta = 0) = 2\int \mathrm{d}^2b\, N.$$

The off-shell photon cross section is the standard current-response convention; its electron flux is specified below. Applying the same normalization to each normalized continuum state gives[7]

$$\boxed{\mathcal{A}_{L,h\bar{h}}(z, \boldsymbol{p}, \boldsymbol{\Delta}) = 2\mathrm{i}\int \mathrm{d}^2b \int \frac{\mathrm{d}^2r}{2\pi}\, \mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r} + \mathrm{i}\boldsymbol{\Delta}\cdot\boldsymbol{b}} \psi^{(1)}_{L,h\bar{h}}(z, \boldsymbol{r})[1 - S(x, \boldsymbol{r}, \boldsymbol{b})].} \tag{9.1}$$

$$\boxed{\frac{\mathrm{d}\sigma_L}{\mathrm{d}z\ \mathrm{d}^2p\ \mathrm{d}^2\Delta} = \frac{1}{16\pi^2}\sum_{h,\bar{h}} |\mathcal{A}_{L,h\bar{h}}|^2.} \tag{9.2}$$

Here $\mathrm{d}^2\Delta = \frac{1}{2}\,\mathrm{d}|t|\ \mathrm{d}\varphi_\Delta$ at high energy, neglecting $|t_{\min}|$; hence its azimuthal integral is $\pi\,\mathrm{d}|t|$.

A normalization check which does not use a model. Put $F_h(\boldsymbol{r}, \boldsymbol{b}) = \psi^{(1)}_h(1 - S)$. Then

$$\begin{aligned}\int \mathrm{d}^2p\ \mathrm{d}^2\Delta\, |\mathcal{A}_h|^2 &= \frac{4}{(2\pi)^2}\int \mathrm{d}^2r\, \mathrm{d}^2b\, \mathrm{d}^2r'\, \mathrm{d}^2b'\, F_h F_h'^{*} (2\pi)^2\delta^{(2)}(r - r')(2\pi)^2\delta^{(2)}(b - b') \\ &= 16\pi^2\int \mathrm{d}^2r\, \mathrm{d}^2b\, |F_h|^2, \\ \frac{\mathrm{d}\sigma_L^{\mathrm{coh}}}{\mathrm{d}z} &= \int \mathrm{d}^2r\, \mathrm{d}^2b\, w_L(z, r)|1 - S|^2.\end{aligned}$$

This integrated completeness check holds at fixed target rapidity. At fixed experimental W, $x_{\mathbb{P}}$ varies with pair mass; it must then remain inside the momentum-dependent amplitude.

$$\frac{\mathrm{d}\sigma_{\mathrm{tot}}(r)}{\mathrm{d}^2b} = 2N, \qquad \frac{\mathrm{d}\sigma_{\mathrm{el}}(r)}{\mathrm{d}^2b} = |1 - S|^2 = N^2 + O^2,$$
$$\frac{\mathrm{d}\sigma_{\mathrm{inel}}(r)}{\mathrm{d}^2b} = 1 - |S|^2 = 2N - N^2 - O^2 \geq 0.$$

For the absorptive model $O = 0$, $0 \leq S \leq 1$: $0 \leq N \leq 1$ and $N \to 1$ is the black limit. General unitarity only requires $|S| \leq 1$; it does not require S real and positive.

Supporting derivation. The completeness identity and the repeated Fourier delta functions provide an independent normalization check.

Diffraction measures fluctuations as well as the mean. For a normalized discrete projectile channel (with its labels suppressed), target eigenstates $|n\rangle$ and completeness give

$$\sum_n |\langle n|\hat{\mathcal{A}}|p\rangle|^2 = \langle p|\hat{\mathcal{A}}^\dagger\hat{\mathcal{A}}|p\rangle,$$

$$\frac{\mathrm{d}\sigma_{\text{coherent}}}{\mathrm{d}t} = \frac{|\langle\mathcal{A}\rangle|^2}{16\pi}, \qquad \frac{\mathrm{d}\sigma_{\text{target dissociation}}}{\mathrm{d}t} = \frac{\langle|\mathcal{A}|^2\rangle - |\langle\mathcal{A}\rangle|^2}{16\pi}.$$

For the continuum, retain $\mathrm{d}z\ \mathrm{d}^2p$ and the corresponding normalization of Eq. (9.2); the same variance identity holds. The proton-intact observable in Eq. (9.2) uses the first expression.[8]

Return to the electron experiment. After averaging the lepton azimuth, neglecting the electron mass and target-mass corrections, define $y = P\cdot q/(P\cdot l)$:

$$\frac{\mathrm{d}\sigma_{ep}}{\mathrm{d}y\ \mathrm{d}Q^2\ \mathrm{d}\Phi} = \frac{\alpha_{\text{em}}}{2\pi y Q^2}\left\{[1+(1-y)^2]\frac{\mathrm{d}\sigma_T}{\mathrm{d}\Phi} + 2(1-y)\frac{\mathrm{d}\sigma_L}{\mathrm{d}\Phi}\right\}, \qquad \mathrm{d}\Phi = \mathrm{d}z\ \mathrm{d}^2p\ \mathrm{d}^2\Delta. \tag{9.3}$$

σ_T averages the two transverse photon polarizations. The present complete open-charm calculation is σ_L; real photoproduction requires σ_T.

Student check. Recover the integrated coherent probability by Parseval. Why is $\langle|\mathcal{A}|^2\rangle$ inappropriate for a strictly intact proton?

Keep. Everything now follows from one amplitude: photon splitting, a target matrix element, and a final-state projection.

Check. Does a proton-intact measurement involve $|\langle\mathcal{A}\rangle|^2$ or $\langle|\mathcal{A}|^2\rangle$?

Checkpoint II: the complete amplitude.

$$\psi_L^{(1)} = -C_L K_0(\epsilon r)\delta_{h,-\bar{h}}, \qquad S_{xy} = \frac{1}{N_c}\langle \mathrm{Tr}\, U_x U_y^\dagger\rangle,$$

$$\mathcal{A}_L = 2\mathrm{i}\int \mathrm{d}^2b\frac{\mathrm{d}^2r}{2\pi}\,\mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}+\mathrm{i}\boldsymbol{\Delta}\cdot\boldsymbol{b}}\psi_L^{(1)}(1-S), \qquad \frac{\mathrm{d}\sigma_L}{\mathrm{d}z\ \mathrm{d}^2p\ \mathrm{d}^2\Delta} = \frac{1}{16\pi^2}\sum_{h\bar{h}}|\mathcal{A}_L|^2.$$

The field-theory construction is complete at our chosen accuracy: pair creation, propagation through the target, final-state projection and normalization. Once we specify the target matrix element, there is a definite spectrum to compute. **Check.** Which factor changes if the outgoing pair is instead a meson?

Next. *With this normalization fixed, what spectrum follows from one completely specified target?*

10 Choose a finite target and a precise comparison

Purpose. Specify the entire benchmark before calculating it.

A target profile is now needed to turn the operator formula into a number. We choose a finite disk so that geometry is explicit and compare two approximations to the same target; no change in a fitted area or wavefunction can masquerade as the effect we wish to isolate.

$$N(x,r,b) = \Theta(R-b)\left(1-\mathrm{e}^{-Q_s^2 r^2/4}\right), \qquad O = 0, \tag{10.1}$$

$$N_{\mathrm{dilute}}(x,r,b) = \Theta(R-b)\frac{Q_s^2 r^2}{4}, \qquad \rho \equiv Q_s^2/\epsilon^2, \qquad v \equiv p^2/\epsilon^2. \tag{10.2}$$

Q_s and R are inputs at a fixed $x = x_{\mathbb{P}}$. The sharp disk makes the geometry integral analytic; it is not a fitted proton profile.

$$\begin{aligned}
T_R(\Delta) &\equiv \int_{b<R} \mathrm{d}^2 b\, \mathrm{e}^{\mathrm{i}\boldsymbol{\Delta}\cdot\boldsymbol{b}} = 2\pi \int_0^R b\,\mathrm{d}b\, J_0(\Delta b) = \frac{2\pi R}{\Delta} J_1(\Delta R), \\
T_R(0) &= \pi R^2 \equiv A_\perp, \\
I(p) &\equiv \int \mathrm{d}^2 r\, \mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}} K_0(\epsilon r)\left(1-\mathrm{e}^{-Q_s^2 r^2/4}\right), \\
C_L &\equiv \frac{e_c\sqrt{N_c \alpha_{\mathrm{em}}}}{\pi} Q z(1-z), \\
\mathcal{A}_{L,h\bar{h}} &= -\frac{\mathrm{i}C_L}{\pi} T_R I\, \delta_{h,-\bar{h}}, \\
\sum_{h,\bar{h}} |\mathcal{A}_{L,h\bar{h}}|^2 &= \frac{2C_L^2}{\pi^2}|T_R|^2 |I|^2, \\
I(p) &\equiv \frac{2\pi}{\epsilon^2} f(\rho, v).
\end{aligned}$$

After both transverse azimuths are integrated,

$$\boxed{\mathcal{D}_L \equiv \frac{\mathrm{d}\sigma_L}{\mathrm{d}z\ \mathrm{d}p^2\ \mathrm{d}|t|} = \frac{N_c \alpha_{\mathrm{em}} e_c^2 Q^2 z^2 (1-z)^2 |T_R(\sqrt{|t|})|^2}{2\pi^2 \epsilon^4} |f(\rho, v)|^2.} \tag{10.3}$$

This model has no relative-azimuth dependence. We compare its *forward* limit; finite recoil also changes the exact pair-mass kinematics. $[\mathcal{D}_L] = \mathrm{GeV}^{-6}$, $[T_R] = \mathrm{GeV}^{-2}$, $[f] = 1$.

Student check. Find the two independent factors of π from integrating the azimuths.

Keep. Our endpoint is now explicit: compute the dimensionless function f, with no free normalization left.

Check. Which inputs and kinematic quantities must stay fixed when comparing the full and dilute answers?

Next. *What is f when only the first, two-gluon target term is retained?*

11 Calculate the two-gluon spectrum

Purpose. Obtain an analytic dilute result, including its zero.

The dilute approximation multiplies the photon wavefunction by r^2. In momentum space this is a Laplacian, so the entire first result follows by differentiating one propagator-like function; its zero will be as informative as its magnitude.

$$
\begin{aligned}
F_0(p) &\equiv \int \mathrm{d}^2 r\, \mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}} K_0(\epsilon r) = \frac{2\pi}{p^2+\epsilon^2}, \\
\nabla_p^2 \mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}} &= -r^2 \mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}}, \\
\nabla_p^2 F(p^2) &= \frac{1}{p}\frac{\partial}{\partial p}\left(p\frac{\partial F}{\partial p}\right) = 4\left(\frac{\partial F}{\partial p^2} + p^2 \frac{\partial^2 F}{\partial (p^2)^2}\right), \\
\nabla_p^2 \frac{2\pi}{p^2+\epsilon^2} &= 8\pi\left[-\frac{1}{(p^2+\epsilon^2)^2} + \frac{2p^2}{(p^2+\epsilon^2)^3}\right] = 8\pi \frac{p^2-\epsilon^2}{(p^2+\epsilon^2)^3}, \\
I_{\text{dilute}}(p) &= -\frac{Q_s^2}{4}\nabla_p^2 F_0(p) = 2\pi Q_s^2 \frac{\epsilon^2 - p^2}{(p^2+\epsilon^2)^3}.
\end{aligned}
$$

$$
\boxed{f_{\text{dilute}}(\rho, v) = \rho \frac{1-v}{(1+v)^3}.} \tag{11.1}
$$

$$
\begin{aligned}
p^2 = 0: &\quad I_{\text{dilute}} = 2\pi Q_s^2/\epsilon^4, \\
p^2 = \epsilon^2: &\quad I_{\text{dilute}} = 0, \\
p^2 \gg \epsilon^2: &\quad I_{\text{dilute}} \sim -2\pi Q_s^2/p^4, \qquad \mathcal{D}_L \propto p^{-8}.
\end{aligned}
$$

The zero is destructive Fourier interference, not a vanishing dipole interaction: $N_{\text{dilute}} > 0$ for every $r > 0$.

$$
\begin{aligned}
I(p) &= 2\pi \int_0^\infty r\,\mathrm{d}r\, J_0(pr) K_0(\epsilon r) N(r), \\
J_0(pr) &= 1 - \frac{p^2 r^2}{4} + \cdots \quad \text{and changes sign as } pr \text{ increases.}
\end{aligned}
$$

Student check. Differentiate F_0 without using a table of integrals. Explain why a positive coordinate-space amplitude can give a momentum-space zero.

Keep. A hard scale controls typical dipoles, but cancellations can amplify corrections to a differential observable.

Check. Does the zero at $p^2 = \epsilon^2$ mean the target stops interacting with the dipole?

Next. *How does summing repeated target interactions change this Fourier amplitude?*

12 Calculate the all-opacity spectrum

Purpose. Reduce the multiple-scattering answer to one stable numerical integral.

Now keep the exponential target interaction instead of expanding it. A Schwinger parameter makes the transverse integral Gaussian, leaving one ordinary integral; this provides a full answer within our target model and an immediate check that the dilute result is recovered.

Set $a = Q_s^2/4$ and use the same Schwinger representation as Section 5:

$$
\begin{aligned}
K_0(\epsilon r) &= \frac{1}{2}\int_0^\infty \frac{\mathrm{d}u}{u}\exp\left[-u-\frac{\epsilon^2 r^2}{4u}\right],\\
I(p) &= F_0(p) - I_a(p),\\
I_a(p) &= \frac{1}{2}\int_0^\infty \frac{\mathrm{d}u}{u}\mathrm{e}^{-u}\int \mathrm{d}^2 r\,\exp\left[-\left(a+\frac{\epsilon^2}{4u}\right)r^2 - \mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}\right],\\
\int \mathrm{d}^2 r\,\mathrm{e}^{-Br^2-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}} &= \frac{\pi}{B}\exp[-p^2/(4B)], \qquad B>0,\\
I_a(p) &= \frac{2\pi}{\epsilon^2}\int_0^\infty \mathrm{d}u\,\frac{\mathrm{e}^{-u}}{1+\rho u}\exp\left[-\frac{vu}{1+\rho u}\right].
\end{aligned}
$$

$$
\boxed{f_{\mathrm{full}}(\rho,v) = \frac{1}{1+v} - \int_0^\infty \mathrm{d}u\,\frac{\mathrm{e}^{-u}}{1+\rho u}\exp\left[-\frac{vu}{1+\rho u}\right].} \tag{12.1}
$$

Check the limits before plotting.

$$
\begin{aligned}
\rho = 0: \quad f &= \frac{1}{1+v} - \int_0^\infty \mathrm{d}u\,\mathrm{e}^{-(1+v)u} = 0,\\
\partial_\rho\left[(1+\rho u)^{-1}\mathrm{e}^{-vu/(1+\rho u)}\right]\Big|_{\rho=0} &= \mathrm{e}^{-vu}(-u+vu^2),\\
f(\rho,v) &= \rho\left[\frac{1}{(1+v)^2} - \frac{2v}{(1+v)^3}\right] + \mathcal{O}(\rho^2) = f_{\mathrm{dilute}} + \mathcal{O}(\rho^2),\\
\rho\to\infty \text{ at fixed } v: \quad f &\to \frac{1}{1+v}, \qquad N(r>0)\to 1.
\end{aligned}
$$

For small ρ, subtracting two nearly equal numbers loses precision. Evaluate instead

$$
f = \int_0^\infty \mathrm{d}u\,\mathrm{e}^{-(1+v)u}\left[1-\exp\left(\frac{v\rho u^2}{1+\rho u} - \ln(1+\rho u)\right)\right],
$$

using `expm1` where appropriate, or a truncated small-ρ asymptotic expansion. After the r integral the opacity series need not converge: at $v=0$, $f = \rho - 2\rho^2 + 6\rho^3 - 24\rho^4 + \cdots$. The supplied quadrature uses Eq. (12.1) in its validated parameter range.

Student check. Recover the dilute result by differentiating under the integral. Derive the large-ρ limit using dominated convergence.

Keep. The full and dilute answers differ only by a precisely specified target approximation.

Check. Does summing all powers of target opacity mean that all perturbative QCD corrections have been included?

Next. *Can the first correction predict which feature of the spectrum moves?*

13 Calculate the first correction and the shifted zero

Purpose. Explain analytically what the numerical comparison will show.

Before looking at a plot, predict how its zero must move. Near a cancellation, a small correction to the absolute amplitude need not be a small relative correction; this is why a single ratio evaluated exactly at the dilute zero would be misleading.

$$
\begin{aligned}
N(r) &= ar^2 - \frac{a^2 r^4}{2} + \mathcal{O}(a^3 r^6), \\
I_{\text{next}} &= -a\nabla_p^2 F_0 - \frac{a^2}{2}(\nabla_p^2)^2 F_0, \\
(\nabla_p^2)^2 F_0 &= 128\pi \frac{p^4 - 4p^2\epsilon^2 + \epsilon^4}{(p^2+\epsilon^2)^5}, \\
f_{\text{next}} &= \rho \frac{1-v}{(1+v)^3} - 2\rho^2 \frac{1-4v+v^2}{(1+v)^5}.
\end{aligned}
$$

Solve for the zero as a series.

$$
\begin{aligned}
v_0 &= 1 + c\rho + \mathcal{O}(\rho^2), \\
f_{\text{next}}(\rho, v_0) &= \rho\left[-\frac{c\rho}{8} + \mathcal{O}(\rho^2)\right] - 2\rho^2 \frac{-2 + \mathcal{O}(\rho)}{32 + \mathcal{O}(\rho)} = \frac{\rho^2}{8}(1-c) + \mathcal{O}(\rho^3), \\
c &= 1, \qquad v_0 = 1 + \rho + \mathcal{O}(\rho^2), \qquad\qquad f(\rho, 1) = \rho^2/8 + \mathcal{O}(\rho^3).
\end{aligned}
$$

There is no uniform relative-error estimate near a zero.

$$
\begin{aligned}
\frac{f_{\text{next}} - f_{\text{dilute}}}{f_{\text{dilute}}} &= -2\rho \frac{1-4v+v^2}{(1+v)^2(1-v)}, \\
\left|\frac{\delta f}{f}\right| \ll 1 &\Longrightarrow \frac{|f+\delta f|^2 - |f|^2}{|f|^2} = 2\,\mathrm{Re}(\delta f/f) + \mathcal{O}(|\delta f/f|^2).
\end{aligned}
$$

The expansion can be useful in absolute error while its relative error diverges at $v = 1$.

Student check. Reproduce the second Laplacian. Find a range of v where the nominal $\rho = 0.1$ correction is enhanced.

Keep. Saturation sensitivity is a property of the process and bin, not of x alone.

Check. Why can $\rho \ll 1$ coexist with a large relative effect near $v = 1$?

Next. *What finite-bin statement survives the zero and can be quoted as an approximation error?*

14 Quantify the effect and distinguish three approximations

Purpose. Turn the analytic result into a defensible statement about QCD.

Here is the first completed answer to our opening question. At one fixed hard scale and one fixed density parameter, repeated scattering suppresses one momentum bin but enhances a bin spanning the interference zero; sensitivity belongs to the specified observable, not to a universal value of Bjorken x.

A reproducible numerical question. Fix $m_c = 1.4$ GeV, $Q^2 = 8$ GeV^2, $z = 1/2$, $Q_s^2 = 0.396$ GeV^2, $R = 3$ GeV^{-1}. Then

$$\epsilon^2 = 3.96 \text{ GeV}^2, \qquad \rho = 0.1, \qquad A_\perp = 9\pi \text{ GeV}^{-2},$$
$$p^2_{0,\text{dilute}} = 3.96 \text{ GeV}^2, \quad p^2_{0,\text{full}} = 3.96(1.099349979\ldots) = 4.3534259\ldots \text{ GeV}^2.$$

For $\alpha_{\text{em}} = 1/137$, the common forward prefactor in Eq. (10.3) is $0.01256761376\ldots$ GeV^{-6}; $\mathcal{D}_L(p^2 = 0) = 8.94529245\times10^{-5}$ GeV^{-6}. The inputs illustrate a calculation, not a fit or a claimed experimental onset scale.

$$R_{[v_a,v_b]}(\rho) \equiv \frac{\int_{v_a}^{v_b} \mathrm{d}v\, f^2_{\text{full}}(\rho, v)}{\int_{v_a}^{v_b} \mathrm{d}v\, f^2_{\text{dilute}}(\rho, v)}. \tag{14.1}$$

The Jacobian $\mathrm{d}p^2 = \epsilon^2\,\mathrm{d}v$ and common prefactor cancel. For fixed $z, Q^2, x_{\mathbb{P}}$:

Bin $[v_a, v_b]$	$[0, 0.5]$	$[0.5, 1.5]$	$[1.5, 3]$	$[0, 4]$
$R_{[v_a,v_b]}$	0.825941	1.332619	0.873674	0.837738

The bin crossing the dilute zero is enhanced even though $N_{\text{full}}(r) < N_{\text{dilute}}(r)$ pointwise. Fourier amplitudes, not positive probabilities, are being modified.

Which kind of factorisation has been tested?

$$\begin{aligned}
\text{Inclusive or diffractive DIS:} &\quad \mathrm{d}\sigma = \sum_i \mathrm{d}\hat{\sigma}_i(Q, m_c, \mu) \otimes f_i \text{ or } f_i^D + \text{power corrections},\\
\text{Hard exclusive meson production:} &\quad \mathcal{M} = \sum_i C_i \otimes F_i^{\text{GPD}} \otimes \phi_M + \text{power corrections},\\
\text{Small-}x\text{ dipole separation:} &\quad \mathcal{A} = \psi_\gamma(\eta) \otimes [1 - S_\eta] \otimes \psi_f^*,\\
\text{Present test:} &\quad N = 1 - \mathrm{e}^{-\tau} \quad \longrightarrow \quad \tau.
\end{aligned}$$

$\otimes$ means integration over shared momentum fractions/coordinates. The first line concerns a cross section with PDFs f_i or diffractive PDFs f_i^D; the second concerns an exclusive amplitude with GPDs and a meson projection ϕ_M, under the process's hard-scale and polarization conditions. These are different factorisation statements. ξ is longitudinal momentum transfer, μ the ultraviolet factorisation scale, and η a rapidity separator.[9, 10]

Expansion	Large quantity / issue	Required response
Fixed-order coefficient	$\alpha_s \ln(1/x) \sim 1$	Resum rapidity logarithms; this can be necessary while $N \ll 1$.
Dilute scattering	$r^2 Q_s^2$ or observable correction appreciable	Retain higher powers of the same Wilson-line interaction.
Diffractive factorisation	Spectator exchanges fail to cancel in hadron–hadron diffraction	Additional soft physics; replacing a PDF with N is not a proof of factorisation.

Hard diffractive DIS has a leading-power collinear factorisation theorem; that statement does not transfer unchanged to diffractive hadron–hadron collisions.[9] Our ratio does not compare a fitted collinear GPD prediction with a fitted dipole prediction. It isolates one density expansion in a common model.

Student check. For the stated inputs, reproduce one bin ratio and the absolute forward spectrum. State exactly what a 20% departure would establish—and what additional matching a GPD comparison would need.

Keep. A quantitative approximation error is meaningful only after fixing the observable, reference calculation and target input.

Check. Have these numbers disproved collinear factorisation or located a universal experimental onset of saturation?

Checkpoint III: the opening quantitative question is answered.

$$\rho = 0.1: \qquad R_{[0,0.5]} = 0.825941, \qquad R_{[0.5,1.5]} = 1.332619,$$
$$p^2_{0,\text{dilute}} = 3.9600\ \text{GeV}^2, \qquad p^2_{0,\text{full}} = 4.3534\ \text{GeV}^2.$$

The first target term underestimates this bin and overestimates that one. There is no paradox: we changed an amplitude before its Fourier cancellations and its absolute square. We have measured the accuracy of the dilute approximation in this model; energy evolution and exchange quantum numbers are the next questions. **Check.** Which quantity was expanded, and which quantity was compared?

Numerical illustration of the computed spectrum

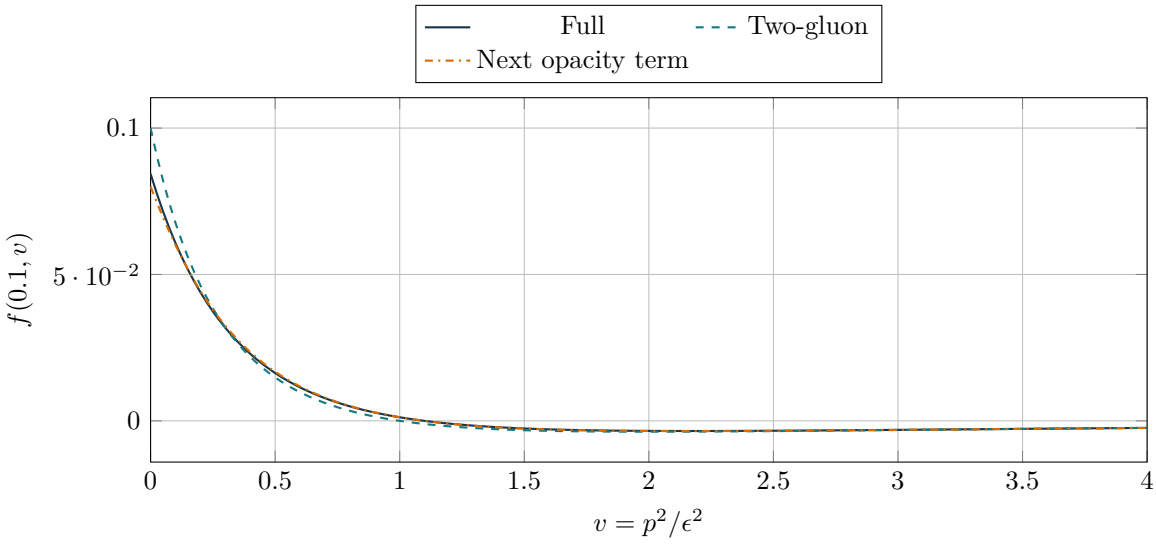


Figure 6: Signed longitudinal amplitude at fixed target rapidity. The full result is an integral; the two comparison curves are derived analytically. The common cross-section prefactor is given in Eq. (10.3).

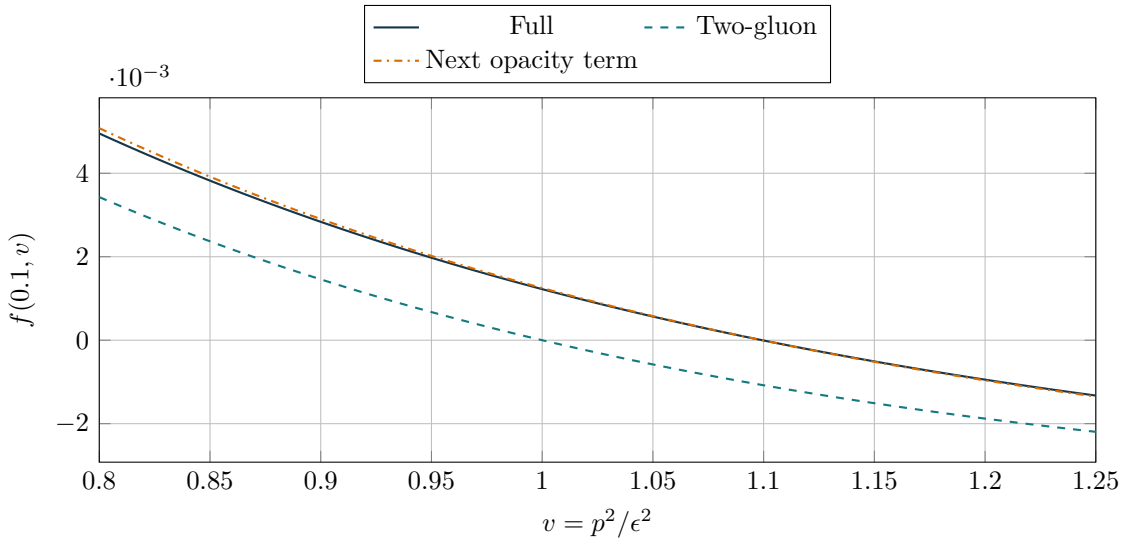


Figure 7: The zero moves from $v = 1$ to $v = 1.099349979\ldots$. A ratio to the dilute differential spectrum is undefined at its zero; the finite-bin ratios remain well defined.

Next. *So far density was an input at one energy. How does the target change when the energy increases?*

15 Add one softer gluon

Purpose. Derive the probability kernel which changes the target rapidity scale.

Increasing energy opens phase space for softer gluons. The same field-theory emission rule gives a probability proportional to a rapidity interval; emissions from the two ends interfere, and their field difference produces the dipole splitting kernel.

Why there is a logarithm. Successive emissions strongly ordered in plus momentum give

$$q^+ \gg k_1^+ \gg k_2^+ \gg \cdots, \qquad \int_{k_{\min}^+}^{q^+} \frac{\mathrm{d}k^+}{k^+} = Y,$$
$$\int_0^Y \mathrm{d}Y_1 \int_0^{Y_1} \mathrm{d}Y_2 \cdots \int_0^{Y_{n-1}} \mathrm{d}Y_n = \frac{Y^n}{n!}, \qquad Y \equiv \ln(x_0/x).$$

Even for small α_s, all $(\alpha_s Y)^n$ matter when $\alpha_s Y \sim 1$. This resummation is distinct from the fixed-energy expansion in target fields.

Soft emission from a fast charge. For a gluon of momentum k and transverse polarization ε, the external-line propagator and vertex reduce to an eikonal current:

$$\frac{\not{p} - \not{k} + m_c}{(p-k)^2 - m_c^2 + \mathrm{i}0}\gamma^\mu u(p) \simeq -\frac{(\not{p} + m_c)\gamma^\mu u(p)}{2p\cdot k} = -\frac{p^\mu}{p\cdot k}u(p),$$
$$p\cdot k \simeq p^+\frac{k_\perp^2}{2k^+}, \qquad \varepsilon^+ = 0, \quad \varepsilon^- = \frac{\boldsymbol{\varepsilon}\cdot\boldsymbol{k}}{k^+},$$
$$gt^a\frac{p\cdot\varepsilon}{p\cdot k} \simeq 2gt^a\frac{\boldsymbol{\varepsilon}\cdot\boldsymbol{k}}{k_\perp^2}.$$

The spin is unchanged at this order. Combining the square with the one-gluon phase space gives

$$\mathrm{d}\mathcal{P}_q = 4g^2C_F\frac{1}{k_\perp^2}\frac{\mathrm{d}k^+\,\mathrm{d}^2k}{2k^+(2\pi)^3} = \frac{\alpha_s C_F}{\pi^2}\frac{\mathrm{d}k^+}{k^+}\frac{\mathrm{d}^2k}{k_\perp^2}.$$

The exact leading soft denominator also contains $m_c^2 k^+/(2p^+)$: neglecting it requires $k_\perp^2 \gg (k^+/p^+)^2 m_c^2$. Strong ordering ensures this at fixed nonzero transverse momentum, not uniformly in the collinear region. The mass is neglected only in this strongly ordered high-energy kernel, not in the photon impact factor. A finite-mass non-eikonal emission calculation requires additional terms.

The color-singlet pair emits the difference of two fields. Differentiate the two-dimensional Coulomb Green function, $\int \mathrm{d}^2k\, \mathrm{e}^{\mathrm{i}k\cdot R}/[(2\pi)^2 k^2] = -\ln(R/R_0)/(2\pi)$, whose additive infrared constant disappears. With the unitary transverse Fourier convention,

$$\begin{aligned}
\int \frac{\mathrm{d}^2k}{2\pi} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}} \frac{k_i}{k^2} &= \mathrm{i}\frac{R_i}{R^2}, \\
\boldsymbol{e}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{z}) &\equiv \frac{\boldsymbol{x}-\boldsymbol{z}}{|\boldsymbol{x}-\boldsymbol{z}|^2} - \frac{\boldsymbol{y}-\boldsymbol{z}}{|\boldsymbol{y}-\boldsymbol{z}|^2}, \\
|\boldsymbol{e}|^2 &= \frac{1}{X^2} + \frac{1}{Z^2} - \frac{2(\boldsymbol{x}-\boldsymbol{z})\cdot(\boldsymbol{y}-\boldsymbol{z})}{X^2 Z^2} = \frac{r^2}{X^2 Z^2}, \\
X &\equiv |\boldsymbol{x}-\boldsymbol{z}|, \qquad\qquad Z \equiv |\boldsymbol{z}-\boldsymbol{y}|, \qquad r \equiv |\boldsymbol{x}-\boldsymbol{y}|.
\end{aligned}$$

Here the bold $\boldsymbol{z}$ is the emitted gluon's coordinate; the unbold z elsewhere is the charm momentum fraction.

$$\boxed{\mathrm{d}\mathcal{P}_{q\bar{q}\to q\bar{q}g} = \frac{\alpha_s C_F}{\pi^2}\,\mathrm{d}Y\;\mathrm{d}^2\boldsymbol{z}\,\frac{r^2}{X^2 Z^2}.} \tag{15.1}$$

Student check. Check the phase-space coefficient and the last vector identity. Why does a coincident color singlet fail to emit this soft field?

Keep. One logarithmic rapidity interval adds a calculable daughter-gluon fluctuation.

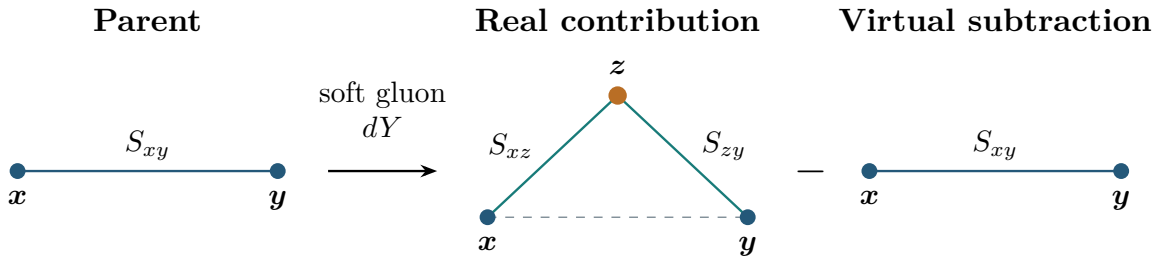
Check. Can rapidity logarithms require resummation even if every dipole interaction is still dilute?

Next. *How can real emission change the scattering while total fluctuation probability remains normalized?*

16 Real minus virtual: the hierarchy and BK

Purpose. Convert the emission probability into evolution of the same scattering operator.

A new soft-gluon state gains probability, so the original state loses the same probability. Combining those two contributions gives the real-minus-virtual equation; a further assumption about target correlations is what turns the hierarchy into the closed BK equation.



$$\frac{\partial S_{xy}}{\partial Y} = \frac{\alpha_s N_c}{2\pi^2} \int d^2\boldsymbol{z}\, \frac{(\boldsymbol{x}-\boldsymbol{y})^2}{(\boldsymbol{x}-\boldsymbol{z})^2(\boldsymbol{z}-\boldsymbol{y})^2} \left[S_{xz}S_{zy} - S_{xy}\right]$$

The product of averaged dipoles is the BK closure.
Before closure, the real term contains $\langle \widehat{S}_{xz}\widehat{S}_{zy}\rangle$.

Figure 8: An emission resolved over a rapidity step introduces the transverse gluon coordinate $\boldsymbol{z}$ (distinct from the longitudinal momentum fraction z). In the large-N_c dipole picture it generates two daughters. The virtual term subtracts the no-emission probability correction. This is a transverse-coordinate evolution diagram, not a spacetime trajectory.

Color algebra before taking a large-N_c limit. For fundamental generators,

$$t^a_{ij}t^a_{kl} = \frac{1}{2}\left(\delta_{il}\delta_{jk} - \frac{1}{N_c}\delta_{ij}\delta_{kl}\right), \qquad V^{ab}_{\boldsymbol{z}} = 2\,\mathrm{Tr}[t^a U_{\boldsymbol{z}} t^b U^\dagger_{\boldsymbol{z}}],$$

$$\begin{aligned}
\hat{S}_{q\bar{q}g} &= \frac{1}{N_c C_F}\,\mathrm{Tr}[t^a U_{\boldsymbol{x}} t^b U^\dagger_{\boldsymbol{y}}] V^{ab}_{\boldsymbol{z}} \\
&= \frac{1}{2N_c C_F}\left[\mathrm{Tr}(U_{\boldsymbol{x}}U^\dagger_{\boldsymbol{z}})\,\mathrm{Tr}(U_{\boldsymbol{z}}U^\dagger_{\boldsymbol{y}}) - \frac{1}{N_c}\,\mathrm{Tr}(U_{\boldsymbol{x}}U^\dagger_{\boldsymbol{y}})\right] \\
&= \frac{N_c^2}{N_c^2-1}\left[\hat{S}_{\boldsymbol{xz}}\hat{S}_{\boldsymbol{zy}} - \frac{1}{N_c^2}\hat{S}_{\boldsymbol{xy}}\right].
\end{aligned}$$

The initial singlet $q\bar{q}g$ color vector is $t^a_{ij}/\sqrt{N_c C_F}$; its norm is one. These equations also give $\hat{S}_{q\bar{q}g} = 1$ when all $U = 1$.

Wavefunction normalization supplies the virtual term. Use a common regulator $X, Z > a_\perp$ in real and virtual terms and take $\mathrm{d}Y$ small enough for first-order perturbation theory. Remove $a_\perp$ only after combining the terms; $\int \mathrm{d}\mathcal{P}$ alone is divergent.

$$\begin{aligned}
S_{xy}(Y+\mathrm{d}Y) &= \left(1 - \int \mathrm{d}\mathcal{P}\right) S_{xy}(Y) + \int \mathrm{d}\mathcal{P}\,\langle \hat{S}_{q\bar{q}g}\rangle_Y, \\
\hat{S}_{q\bar{q}g} - \hat{S}_{xy} &= \frac{N_c^2}{N_c^2-1}(\hat{S}_{xz}\hat{S}_{zy} - \hat{S}_{xy}), \\
\frac{\alpha_s C_F}{\pi^2}\frac{N_c^2}{N_c^2-1} &= \frac{\alpha_s N_c}{2\pi^2}.
\end{aligned}$$

Thus the leading-logarithmic first Balitsky equation is[3, 11]

$$\boxed{\partial_Y S_{xy} = \frac{\alpha_s N_c}{2\pi^2}\int \mathrm{d}^2\boldsymbol{z}\, \frac{|\boldsymbol{x}-\boldsymbol{y}|^2}{|\boldsymbol{x}-\boldsymbol{z}|^2|\boldsymbol{z}-\boldsymbol{y}|^2}\left[\langle \hat{S}_{xz}\hat{S}_{zy}\rangle_Y - S_{xy}\right].} \tag{16.1}$$

The two-dipole average is a new target correlator. Its evolution produces more correlators: a hierarchy, rather than an automatically closed equation.

$$\begin{aligned}
\langle \hat{S}_{xz}\hat{S}_{zy}\rangle_Y &\simeq S_{xz}S_{zy} \qquad \text{(mean-field closure)}, \\
S = 1 - N \quad (O = 0), &\qquad \bar{\alpha} \equiv \alpha_s N_c/\pi, \\
-\partial_Y N_{xy} &= \frac{\bar{\alpha}}{2\pi}\int \mathrm{d}^2 z\, \frac{r^2}{X^2 Z^2}[(1-N_{xz})(1-N_{zy}) - (1-N_{xy})].
\end{aligned}$$

$$\boxed{\partial_Y N_{xy} = \frac{\bar{\alpha}}{2\pi}\int \mathrm{d}^2 z\, \frac{r^2}{X^2 Z^2}[N_{xz} + N_{zy} - N_{xy} - N_{xz}N_{zy}].} \tag{16.2}$$

Large-N_c dipole dynamics motivates the closure; discarding target correlations is an additional assumption. The transverse positions carry impact parameter throughout.

Three checks and the meaning of resummation.

$$\begin{aligned} N \ll 1: &\quad \partial_Y N \simeq \mathcal{K} \otimes N \quad \text{(linear BFKL)},\\ N = 0 \text{ or all dipoles } N = 1: &\quad \partial_Y N = 0,\\ \boldsymbol{z} \to \boldsymbol{x}: &\quad N_{xz} \to 0,\ N_{zy} \to N_{xy}, \quad \text{real–virtual cancellation},\\ N(Y) = \mathrm{e}^{Y\mathcal{K}} N(0) &= \sum_{n=0}^{\infty} \frac{Y^n}{n!}\mathcal{K}^n N(0) \quad \text{in the linear limit.} \end{aligned}$$

The nonlinear product reduces growth in a dense target; it is not a correction to the electromagnetic splitting vertex. Define $Q_s(Y, b)$ by a convention, for example

$$N\big(Y, r = 2/Q_s(Y,b), b\big) = 1 - \mathrm{e}^{-1}. \tag{16.3}$$

Its numerical value depends on that convention and the fitted initial condition. LO fixed-coupling BK is the equation derived here. Running-coupling and collinearly resummed equations are improved approximations, not extra terms to append arbitrarily.[12, 13]

Student check. Use the Fierz identity to obtain $S_{q\bar{q}g}$. Recover all four terms of BK, and identify precisely the extra assumption between the hierarchy and BK.

Keep. The dipole framework separates a short-distance splitting from target operators whose energy dependence is itself calculable.

Check. Which step is an additional target assumption rather than the soft-emission calculation itself?

Checkpoint IV: keep the two resummations separate.

$$\begin{aligned} \text{Many target interactions at fixed } Y: &\quad 1 - \mathrm{e}^{-\tau} = \tau - \tau^2/2 + \cdots,\\ \text{Many ordered emissions as } Y \text{ increases}: &\quad (\alpha_s Y)^n, \qquad \partial_Y N = \mathcal{K} \otimes (N_1 + N_2 - N - N_1 N_2). \end{aligned}$$

The exponential was a target model at a specified rapidity. BK tells us how a target operator evolves, after a closure assumption. Both concern high-density QCD, but one is not a derivation of the other. **Check.** Is our Gaussian disk automatically an exact BK solution?

Next. *What changes if the measured charm pair binds into a particular charmonium state?*

17 Project the same produced pair onto J/ψ or h_c

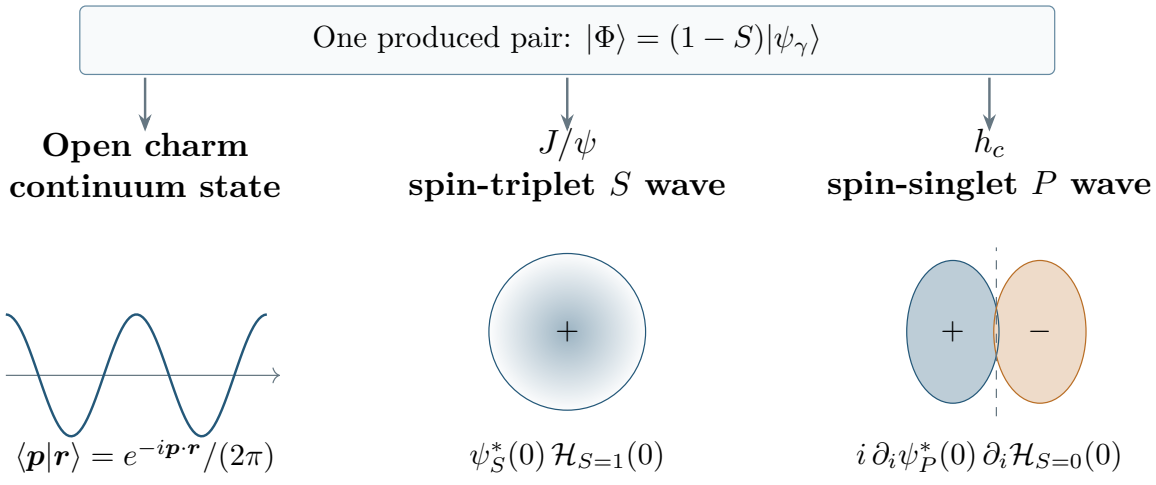
Purpose. Derive the difference between an S-wave value and a P-wave derivative, and identify one reliable dilute matching result.

Open charm asks for a final momentum state; charmonium asks for a final bound-state wavefunction. The target operator survives this change, but its weight is different: an S wave samples the spin-triplet amplitude at zero relative momentum, while the h_c requires a derivative of the spin-singlet amplitude.

First identify the state; then calculate its production amplitude. For a neutral $c\bar{c}$ state,

$$\begin{aligned} \boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}, \qquad & P = (-1)^{L+1}, \qquad & C = (-1)^{L+S}, \\ J/\psi: \quad {}^3S_1, \qquad & (L,S) = (0,1), \qquad & J^{PC} = 1^{--}, \\ h_c: \quad {}^1P_1, \qquad & (L,S) = (1,0), \qquad & J^{PC} = 1^{+-}. \end{aligned}$$

The minus sign in parity comes from the opposite intrinsic parities of a fermion and antifermion. Under charge conjugation, interchange contributes orbital parity $(-1)^L$, the spin interchange factor $(-1)^{S+1}$, and a minus sign from reordering fermionic operators; their product is $(-1)^{L+S}$. In h_c, the total angular momentum is orbital; its quark spins form a singlet. It is not an excited vector S wave.



Wave patterns show phase and orbital parity only, not fitted wavefunctions.
The P wave has an angular sign change; a radial node is not implied.

Figure 9: Different final states project the same pair-production amplitude. The first sketch is the real part of a plane wave; its full complex phase is retained in the formula. The bound-state sketches indicate the S- and P-wave orbital symmetries in the pair rest frame. The short-distance factors shown omit common external-state normalization. Spin projection is performed before a P-wave derivative.

Quantum-mechanical input, with its normalisation visible.

$$\left[-\frac{1}{m_c}\frac{\mathrm{d}^2}{\mathrm{d}R^2} + \frac{L(L+1)}{m_c R^2} + V(R)\right] u_L(R) = E_L u_L(R),$$

$$R_L(R) = u_L(R)/R, \qquad \int_0^\infty \mathrm{d}R\, R^2 |R_L(R)|^2 = 1. \tag{17.1}$$

The reduced mass is $m_c/2$, explaining $1/m_c$ in the kinetic energy. For a regular solution, $u_L(R) \sim$

R^{L+1} at the origin. Thus

$$\psi_S(\boldsymbol{R}) = \frac{R_S(R)}{\sqrt{4\pi}}, \qquad \psi_S(0) = \frac{R_S(0)}{\sqrt{4\pi}},$$
$$\psi_{P,m}(\boldsymbol{R}) = R_P(R)Y_{1m}(\widehat{\boldsymbol{R}}) = \sqrt{\frac{3}{4\pi}}R'_P(0)\,\boldsymbol{\varepsilon}_m\cdot\boldsymbol{R} + O(R^2). \tag{17.2}$$

Choose the phases of Y_{1m} and $\boldsymbol{\varepsilon}_m$ consistently. A potential, mass, numerical boundary and grid spacing must be specified before claiming numerical values for $R_S(0)$ or $R'_P(0)$.

Why the P wave differentiates the production amplitude. Use the Fourier convention $\psi(\boldsymbol{R}) = \int \mathrm{d}^3\ell\, \mathrm{e}^{\mathrm{i}\boldsymbol{\ell}\cdot\boldsymbol{R}}\widetilde{\psi}(\boldsymbol{\ell})/(2\pi)^3$ and expand the hard pair amplitude:

$$\mathcal{H}(\boldsymbol{\ell}) = \mathcal{H}(0) + \ell_i\partial_i\mathcal{H}(0) + O(\ell^2),$$
$$\int \frac{\mathrm{d}^3\ell}{(2\pi)^3}\widetilde{\psi}^*_S(\boldsymbol{\ell}) = \psi^*_S(0),$$
$$\int \frac{\mathrm{d}^3\ell}{(2\pi)^3}\widetilde{\psi}^*_{P,m}(\boldsymbol{\ell}) = 0,$$
$$\int \frac{\mathrm{d}^3\ell}{(2\pi)^3}\ell_i\widetilde{\psi}^*_{P,m}(\boldsymbol{\ell}) = \mathrm{i}\partial_i\psi^*_{P,m}(0) = \mathrm{i}\sqrt{\frac{3}{4\pi}}R'^*_P(0)\varepsilon^*_{m,i}.$$

Therefore, apart from the common chosen external-state normalisation,

$$\boxed{\mathcal{A}_S \propto \frac{R^*_S(0)}{\sqrt{4\pi}}\mathcal{H}_{S=1}(0), \qquad \mathcal{A}_P \propto \mathrm{i}\sqrt{\frac{3}{4\pi}}R'^*_P(0)\varepsilon^*_{h,i}\partial_i\mathcal{H}_{S=0}(0).} \tag{17.3}$$

The factor of i follows from the Fourier convention; a meson-state phase can change it. The relative momentum is an internal variable, not the measured proton recoil.

Supporting derivation. The covariant projector algebra below makes the momentum-moment product rule explicit and supplies a matching check.

Perform the spin projection before taking the derivative. Let $K = p_c + p_{\bar{c}}$ be the pair momentum, $p_c = K/2 + \ell$, $p_{\bar{c}} = K/2 - \ell$, and $K\cdot\ell = 0$. In the pair rest frame $\ell = (0, \boldsymbol{\ell})$ to the order required. Let $\mathcal{H}$ now be the open-pair Dirac and colour matrix with external quark spinors removed. A standard nonrelativistic projector convention, accurate through first order in ℓ, is [14]

$$C_1 = I_c/\sqrt{N_c},$$
$$\Pi_0(\ell) = \frac{(\not{K}/2 - \not{\ell} - m_c)\gamma_5(\not{K}/2 + \not{\ell} + m_c)}{\sqrt{8m_c^3}},$$
$$\Pi_1^\alpha(\ell) = \frac{(\not{K}/2 - \not{\ell} - m_c)\gamma^\alpha(\not{K}/2 + \not{\ell} + m_c)}{\sqrt{8m_c^3}},$$
$$H_J = \varepsilon^*_{J,\alpha}\,\mathrm{Tr}[C_1\Pi_1^\alpha\mathcal{H}]_{\ell=0},$$
$$H_h = \varepsilon^*_{h,\beta}\,\frac{\partial}{\partial\ell_\beta}\,\mathrm{Tr}[C_1\Pi_0(\ell)\mathcal{H}(\ell)]\bigg|_{\ell=0}. \tag{17.4}$$

Here $K^2 = 4m_c^2 + O(\ell^2)$ in the short-distance expansion, Tr includes Dirac and colour indices, and $\gamma_5 = \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$. Factors relating these short-distance projectors to a physical meson state are fixed by the radial convention and the matching coefficient; Eq. (17.8) below gives one fully specified published normalisation.

The product rule is essential:

$$\begin{aligned}
\partial_\beta \operatorname{Tr}[\Pi_0 \mathcal{H}] &= \operatorname{Tr}[(\partial_\beta \Pi_0)\mathcal{H} + \Pi_0(\partial_\beta \mathcal{H})], \\
\partial_\beta \Pi_0|_0 &= \frac{-\gamma^\beta \gamma_5 (\not{K}/2 + m_c) + (\not{K}/2 - m_c)\gamma_5 \gamma^\beta}{\sqrt{8 m_c^3}}, \\
G(L) &= \frac{\not{L} + m_c}{L^2 - m_c^2 + \mathrm{i}0}, \qquad L = K/2 + \ell - K_{\mathrm{sub}}, \\
\partial_\beta G(L) &= \frac{\gamma^\beta}{L^2 - m_c^2 + \mathrm{i}0} - \frac{2L^\beta(\not{L} + m_c)}{(L^2 - m_c^2 + \mathrm{i}0)^2}.
\end{aligned}$$

K_{sub} denotes a fixed sum of momenta already emitted or absorbed on that propagator. Dropping the derivative of the spin projector discards a term of the same order as the derivative of the propagator. Squared denominators in a P-wave coefficient are a direct consequence of this differentiation.

Check a zero before adding a target. For the vacuum vector current $\mathcal{H} = \gamma^\mu$, the singlet trace contains fewer than four ordinary gamma matrices next to γ_5 and vanishes. With the all-velocity projector containing an additional $(\not{K} + 2E)$ factor, the only possible term is proportional to

$$\epsilon^{\mu\alpha\beta\gamma} p_{\bar{c},\alpha} K_\beta p_{c,\gamma} = 0, \qquad p_c, p_{\bar{c}} \in \operatorname{span}\{K, \ell\}.$$

Its ℓ derivative vanishes too: a single vector current does not create h_c from the vacuum. The target inserts additional gluon vertices and momentum directions, so this zero does not forbid $\gamma p \to h_c p$.

Student check. Derive the two momentum integrals leading to Eq. (17.3). Then differentiate one propagator and the singlet projector explicitly. Why does changing R_S to R_P without changing the spin projector give the wrong state?

The same dipole interaction, a different final-state projection. Let $\Psi^{(1)}_{V,h\bar{h}}$ be a colour-singlet meson light-front wavefunction with $\int \mathrm{d}z\, \mathrm{d}^2 r \sum_{h\bar{h}} |\Psi^{(1)}_V|^2 = 1$ in the valence approximation. Define

$$\begin{aligned}
W_{V\lambda}(z, \boldsymbol{r}) &= \sum_{h\bar{h}} \Psi^{(1)*}_{V,h\bar{h}}(z, \boldsymbol{r}) \psi^{(1)}_{\gamma,\lambda,h\bar{h}}(z, \boldsymbol{r}), \\
\mathcal{A}_{V\lambda}(\boldsymbol{\Delta}) &= 2\mathrm{i} \int_0^1 \mathrm{d}z \int \mathrm{d}^2 r\, \mathrm{d}^2 b\, \mathrm{e}^{\mathrm{i}\boldsymbol{\Delta}\cdot[\boldsymbol{b} + (z-1/2)\boldsymbol{r}]} W_{V\lambda}(z, \boldsymbol{r})[1 - S(\boldsymbol{r}, \boldsymbol{b})], \\
\left.\frac{\mathrm{d}\sigma_{V\lambda}}{\mathrm{d}t}\right|_{t=0} &= \frac{|\mathcal{A}_{V\lambda}(0)|^2}{16\pi}.
\end{aligned} \tag{17.5}$$

The recoil is the target recoil, $\boldsymbol{\Delta} = \boldsymbol{p}'_p - \boldsymbol{p}_p$, and $\boldsymbol{b}$ is the dipole midpoint. The phase follows because the meson transverse centre of plus momentum is $\boldsymbol{b} + (z - 1/2)\boldsymbol{r}$.

For a simple leading nonrelativistic J/ψ_L spin wavefunction, choose $\Psi^{(1)}_{J,L,h\bar{h}} = \phi_S(z, r)\delta_{h,-\bar{h}}/\sqrt{2}$, with the overall meson phase fixed here and $\int \mathrm{d}z\, \mathrm{d}^2 r |\phi_S|^2 = 1$. Using the photon amplitude already derived,

$$\begin{aligned}
W_{JL}(z, r) &= -\frac{\sqrt{2N_c\alpha_{\mathrm{em}}}e_c}{\pi} Q z(1-z) K_0(\epsilon r)\phi^*_S(z, r), \\
W^{\mathrm{NR}}_{JT}(z, r) &= -\frac{e_c\sqrt{N_c\alpha_{\mathrm{em}}}}{\sqrt{2}\pi} m_c K_0(\epsilon r)\phi^*_S(z, r)(\boldsymbol{\varepsilon}^*_J \cdot \boldsymbol{\varepsilon}_\gamma).
\end{aligned} \tag{17.6}$$

The transverse vertex and this coefficient are derived explicitly in the derivation sheet on p. 44. The second expression keeps the leading spin-triplet mass term; derivative and spin-rotation contributions must be restored for a precision transverse calculation. It establishes the photoproduction limit: $W_{JL} \to 0$ when $Q \to 0$, while W_{JT} remains finite. In the symmetric static limit, the amplitude ratio for matched polarisations is $\mathcal{A}_L/\mathcal{A}_T \simeq Q/(2m_c)$. This is a limiting check, not an exact prediction for the measured longitudinal-to-transverse ratio.

The rest-frame and light-front wavefunctions are connected approximately, not by identifying them:

$$M_0^2 = \frac{m_c^2 + k_\perp^2}{z(1-z)}, \qquad k_z = (z - \tfrac{1}{2})M_0, \quad \frac{\partial k_z}{\partial z} = \frac{M_0}{4z(1-z)},$$
$$\phi_S(z, k_\perp) = \sqrt{\frac{1}{2\pi}\frac{\partial k_z}{\partial z}}\,\widetilde{\psi}_S^{\mathrm{RF}}\left(\sqrt{k_z^2 + k_\perp^2}\right), \quad \int \mathrm{d}z \frac{\mathrm{d}^2 k_\perp}{(2\pi)^2}|\phi_S|^2 = 1.$$

M_0 is the internal free-pair invariant mass, not the physical meson mass M_V. A consistent Melosh spin rotation is required beyond the static spin approximation. Ref. [15] develops the vector-meson construction, and Ref. [16] investigates its uncertainties. Define the vector decay constant by $\langle 0|\bar{c}\gamma^\mu c|J/\psi(K,\varepsilon)\rangle = f_J M_J \varepsilon^\mu$. A separate normalisation check is

$$\Gamma(J/\psi \to e^+e^-) = \frac{4\pi\alpha_{\mathrm{em}}^2 e_c^2 f_J^2}{3M_J}, \qquad f_J^2 \simeq \frac{4N_c}{M_J}|\psi_S^{\mathrm{RF}}(0)|^2. \tag{17.7}$$

The second equality is leading nonrelativistic matching; short-distance QCD and velocity corrections are additional. A measured f_J must not receive the same correction a second time.

h_c: a nonzero result that fixes the dilute matching target. Cui, Hu and Ma compute forward exclusive *photoproduction* at leading order in the nonrelativistic expansion and at twist two [17]. For this quoted collinear expression choose the proton to move in the + direction. Let $\bar{p} = (p+p')/2$, $\Delta = p' - p$, $\xi = -\Delta^+/(2\bar{p}^+)$, $t = \Delta^2$, $g_\perp^{ij} = -\delta^{ij}$ and $\epsilon_\perp^{12} = +1$. The following is the unpolarised-gluon *sector* of their amplitude; it is not the complete unpolarised cross section:

$$\mathcal{T}_{T,F_g}^{h_c} = 16 g_s^2 A_P\, \varepsilon_{h,\nu}^* \varepsilon_\gamma^\rho \epsilon_{\perp\mu\rho} g_\perp^{\mu\nu} \int_{-1}^{1} \mathrm{d}x\, \frac{\xi\, \mathcal{F}_g(x,\xi,t)}{(\xi - x - \mathrm{i}0)^2(\xi + x - \mathrm{i}0)},$$
$$A_P = (ee_c)\sqrt{\frac{3}{4\pi N_c}}\, R_P'^*(0)(2m_c)^{-5/2}. \tag{17.8}$$

Here $g_s = g$ is the QCD coupling already used on the preceding sections. The electrical charge is written explicitly as ee_c; R_P has the normalisation in Eq. (17.1). The invariant amplitude and proton spinors use relativistic external-state normalisation. Before comparing with the dipole amplitude, convert its invariant normalization:

$$\frac{\mathrm{d}\sigma(\gamma p \to h_c p)}{\mathrm{d}t} = \frac{\overline{\sum}|\mathcal{T}_{\mathrm{inv}}|^2}{16\pi(W^2 - m_p^2)^2}, \qquad \mathcal{A}_{\mathrm{reduced}} = \frac{\mathcal{T}_{\mathrm{inv}}}{W^2 - m_p^2}.$$

The bar denotes the specified initial-spin average and final-spin sum; use identical helicity conventions in both descriptions. $\mathcal{T}_{\mathrm{inv}}$ is dimensionless; $\mathcal{A}_{\mathrm{reduced}}$ has units GeV^{-2}. The corresponding source definition of the gluon matrix element is

$$\mathcal{F}_g(x,\xi,t) = \frac{1}{\bar{p}^+}\int \frac{\mathrm{d}\lambda}{2\pi} \mathrm{e}^{\mathrm{i}x\bar{p}^+\lambda}\langle p'|g_{\perp\mu\nu}F_a^{+\mu}(-\lambda n/2)\mathcal{U}_{ab}F_b^{+\nu}(\lambda n/2)|p\rangle$$
$$= -\frac{1}{2\bar{p}^+}\bar{u}(p')\left[\gamma^+ H_g(x,\xi,t) + \frac{\mathrm{i}\sigma^{+\alpha}\Delta_\alpha}{2m_p}E_g(x,\xi,t)\right]u(p). \tag{17.9}$$

Here n is lightlike with $p\cdot n = p^+$, $\mathcal{U}$ is the straight adjoint gauge link between the field strengths, colour indices are summed, and $\sigma^{\mu\nu} = \mathrm{i}[\gamma^\mu, \gamma^\nu]/2$. The link becomes unity in the source's light-cone-gauge treatment with its boundary prescription. The full result also contains helicity and gluon-transversity GPDs; keeping only $\mathcal{F}_g$ is an explicit restriction, not a consequence of averaging the proton spin.

Three checks can be made without a numerical GPD fit:

1. The transverse tensor is proportional, up to the stated metric convention, to $\boldsymbol{\varepsilon}_h^* \cdot (\boldsymbol{n} \times \boldsymbol{\varepsilon}_\gamma)$. A transverse h_c can therefore be produced at zero transverse recoil. Spin-singlet quantum numbers do not force this amplitude to vanish.

2. The coefficient is proportional to $R'_P(0)$, and a double pole occurs exactly as the derivative rule predicts.

3. $Q^2 = 0$ is assumed in this matching result. A longitudinal *produced meson* in the source's twist-three result is not a longitudinal incoming photon. Electroproduction requires inserting $q^2 = -Q^2$ and the appropriate ε_γ in the hard trace before redoing the expansion.

Equation (17.5) defines the small-x bound-state projection for either photon polarisation. For h_c, a properly constructed spin-singlet P-wave light-front state is required; its spin-orbit factor can be obtained from the singlet projector or the covariant light-front construction [18]. No h_c saturation curve is claimed here. The concrete continuation is to derive that impact factor in one convention, expand the Wilson lines to two gluons, and check the overlapping high-energy limit against Eq. (17.8). The finite-skewness collinear result is a matching anchor, not an all-density dipole formula.

A distinct failure mode: an endpoint, not an opacity. The published twist-three h_c amplitude contains terms of the form

$$I_3 = \int_{-\xi}^{\xi} \mathrm{d}x \, \frac{F(x,\xi)}{(x+\xi-\mathrm{i}0)^3}, \qquad F(x,\xi) \sim (\xi-x)^2(\xi+x)^2,$$
$$x = -\xi + u: \qquad I_3 \sim 4\xi^2 \int_0 \frac{\mathrm{d}u}{u}.$$

This exposes an endpoint sensitivity when one exchanged gluon becomes soft. It obstructs the stated power-suppressed collinear calculation; it does not invalidate the finite leading-twist h_c result. The multiple-scattering comparison performed earlier tests a different approximation and does not automatically cure this endpoint problem.

Keep. J/ψ samples a spin-triplet S-wave projection; h_c samples a derivative of a spin-singlet amplitude. Charge parity alone never replaces that calculation.

Check. Why is replacing the J/ψ radial wavefunction by a P-wave radial function insufficient to calculate h_c?

Next. *Which charge-conjugation sector of the same target operator can each meson select?*

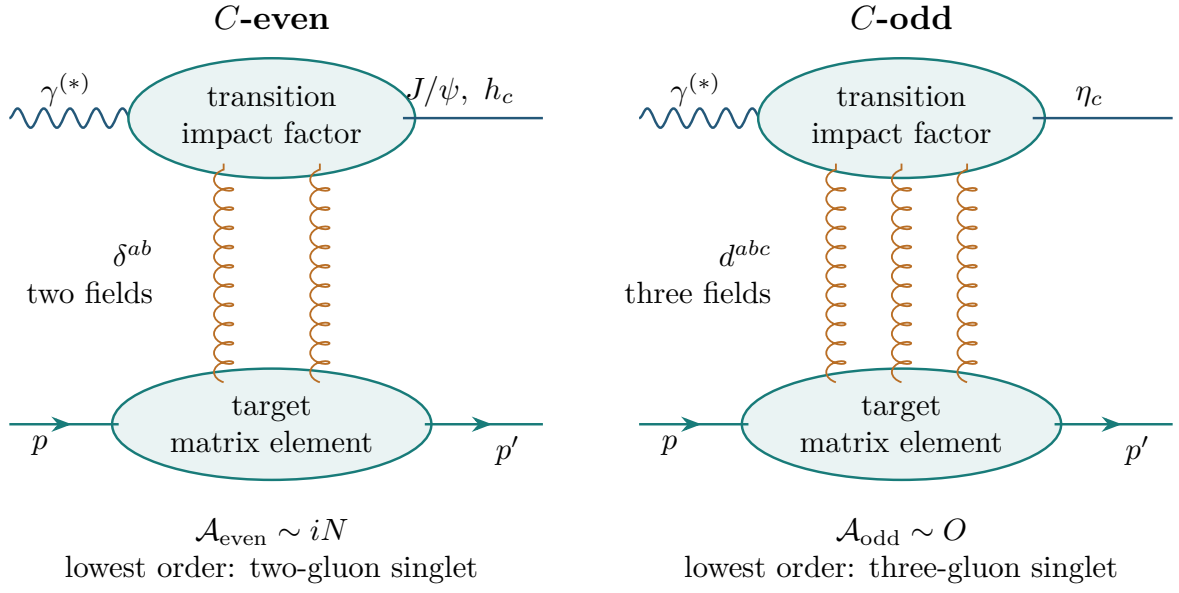
18 Which part of the target is Pomeron-like or Odderon-like?

Purpose. Use charge conjugation on the Wilson operator and apply it to the three charm final states.

Pomeron and Odderon name the even and odd exchange sectors that have already been present in our Wilson-line amplitude. We shall connect those operator symmetries to gluon exchange, energy evolution and possible measurements, while separating what a channel allows from what its measured rate can actually establish.

$$S_{xy} = \frac{1}{N_c}\langle \mathrm{Tr}(U_x U_y^\dagger)\rangle = 1 - N_{xy} + \mathrm{i}O_{xy},$$
$$S_{yx} = S_{xy}^*, \qquad N_{yx} = N_{xy}, \qquad O_{yx} = -O_{xy}. \tag{18.1}$$

Charge conjugation reverses the colour dipole. Its even part is associated with the Pomeron sector, its odd part with the dipole Odderon. These are operator sectors; an exact QCD Pomeron is not defined as just two noninteracting gluons.



The upper transition selects $C_{\text{exchange}} = C_\gamma C_M$.
Evolution and multiple scattering dress the displayed lowest-order exchanges.

Figure 10: Dilute gluonic exchange sectors. The ellipses are impact factors or target matrix elements, not extra particles, and already include the necessary attachment sums. J/ψ and h_c both have $C = -1$, whereas η_c has $C = +1$. The three-gluon Odderon uses the symmetric d^{abc} singlet. The indicated phases follow $S = 1 - N + iO$ and refer to the leading high-energy convention; corrections and other exchanges can alter physical phases.

Read the dilute content from the trace. For one weak shock with commuting classical field components, put $\alpha_x^a = \int \mathrm{d}x^+ A_a^-(x^+, \boldsymbol{x})$ and $\delta\alpha^a = \alpha_x^a - \alpha_y^a$. The symmetric colour structures follow from

$$\mathrm{Tr}\, t^a = 0, \qquad \mathrm{Tr}(t^a t^b) = \frac{\delta^{ab}}{2}, \qquad \mathrm{Tr}(t^a t^b t^c) = \frac{1}{4}(d^{abc} + \mathrm{i}f^{abc}).$$

Expand the Wilson lines and use the symmetric product of three classical fields:

$$S_{xy} = 1 - \frac{g_s^2}{4N_c}\langle \delta\alpha^a \delta\alpha^a \rangle - \frac{\mathrm{i}g_s^3}{24N_c} d^{abc}\langle \delta\alpha^a \delta\alpha^b \delta\alpha^c \rangle + \cdots ,$$
$$N_{xy}^{(2)} = \frac{g_s^2}{4N_c}\langle (\delta\alpha^a)^2 \rangle, \qquad O_{xy}^{(3)} = -\frac{g_s^3}{24N_c} d^{abc}\langle \delta\alpha^a \delta\alpha^b \delta\alpha^c \rangle. \quad (18.2)$$

The f^{abc} contraction vanishes against the symmetric product. The minus sign in $O^{(3)}$ follows from $U = \mathcal{P}\exp(+\mathrm{i}g_s \int A^-)$ and $\mathrm{i}^3 = -\mathrm{i}$. General path ordering must be retained beyond this weak-field identification. Repeated scattering and small-x evolution dress these first two- and three-field terms [19].

Saturation also changes the odd sector's evolution. Insert $S = 1 - N + \mathrm{i}O$ into the closed equation $\partial_Y S = \mathcal{K} \otimes (S_1 S_2 - S)$:

$$S_1 S_2 = (1 - N_1)(1 - N_2) - O_1 O_2 + \mathrm{i}[(1 - N_1)O_2 + (1 - N_2)O_1],$$
$$\partial_Y N = \mathcal{K} \otimes [N_1 + N_2 - N - N_1 N_2 + O_1 O_2],$$
$$\partial_Y O = \mathcal{K} \otimes [O_1 + O_2 - O - N_1 O_2 - O_1 N_2].$$

Here $1 = (x, z)$, $2 = (z, y)$ and $\mathcal{K}\otimes$ includes $\bar{\alpha}\ \mathrm{d}^2 z\, r^2/(2\pi X^2 Z^2)$. No odd term is generated from an exactly even initial condition in this approximation. Dense even fields modify odd evolution through the two NO terms. This connects the exchange sectors to the same saturation dynamics, not to a new target model.

Separate a model symmetry from a physical prohibition.

$$W[\alpha] = W[-\alpha] \implies \langle (\delta\alpha)^3 \rangle = 0,$$
$$O(\boldsymbol{r}) = O(r^2),\quad O(-\boldsymbol{r}) = -O(\boldsymbol{r}) \implies O = 0,$$
$$O(\boldsymbol{r}, \boldsymbol{b}) \sim (\boldsymbol{r}\cdot\boldsymbol{b})\, f(r^2, b^2, (\boldsymbol{r}\cdot\boldsymbol{b})^2) \quad \text{is symmetry-allowed.}$$

A symmetric Gaussian target and a translationally invariant isotropic ansatz remove the Odderon by construction. An unpolarised proton with impact-parameter dependence is not thereby forbidden to have an odd correlator. An angular or recoil integration can remove sensitivity even when the operator is nonzero.

Apply the selection to the photon transition.

$$C_{\text{exchange}} = C_\gamma C_M, \qquad C_\gamma = -1,$$

M	C_M	exchange selected by $\gamma^{(*)} \to M$
$J/\psi\ (^3S_1)$	-1	C-even
$h_c\ (^1P_1)$	-1	C-even
$\eta_c\ (^1S_0)$	$+1$	C-odd

(18.3)

The proton is not assumed to be a charge-conjugation eigenstate. The selection is at the upper photon-to-meson impact factor. Both J/ψ and h_c therefore probe the even sector, with different spin and orbital weights. The h_c is not an Odderon filter.

For comparison, the leading eikonal pseudoscalar trace has the form

$$H_\lambda^{\eta_c} \propto m_c \epsilon^{+-ij} \varepsilon_{\gamma,\perp i}(\ell_\perp - \ell'_\perp + zK_{\eta,\perp})_j, \qquad K_{\eta,\perp} = -\Delta_\perp.$$

Here ℓ and ℓ' are quark momenta before and after the shock and K_η is the produced meson momentum. The transverse antisymmetric tensor arises from the γ_5 trace; a longitudinal photon decouples in this leading eikonal pseudoscalar calculation [20]. Electromagnetic exchange can also produce η_c and must be separated from a gluonic Odderon signal. This is a comparison channel, not permission to reuse its radial overlap for h_c.

Student check. Derive P and C for all three states. Recover the coefficient of d^{abc} in Eq. (18.2). Give one target ansatz that sets $O = 0$ and one allowed geometric structure that does not. Explain why C-even exchange permits h_c production but does not determine its rate.

Keep. The same Wilson-line operator separates even and odd exchange; the final-state projector decides which sector and which momentum/spin dependence a measurement retains.

18A. From a colour-singlet operator to an energy-dependent exchange. Live route for this block. Operator sectors → phases and elastic evidence → the charm charge asymmetry → bound-state and electromagnetic filters → the distinction between density, symmetry and factorisation. The detailed signature and profile algebra remain available as supporting derivations.

Our Wilson-line calculation identifies what the target exchanges. We now ask two measurable questions: how does that exchange change with energy, and how can an experiment distinguish its even and odd parts?

Keep three levels separate:

$$\begin{aligned} \text{Born:}\quad & N^{(2)} \sim g_s^2 \langle \alpha\alpha \rangle , && O^{(3)} \sim g_s^3 d^{abc} \left\langle \alpha^a \alpha^b \alpha^c \right\rangle , \\ \text{Linear:}\quad & G_P(Y_R) = \mathrm{e}^{Y_R K_2} G_P(0), && G_O(Y_R) = \mathrm{e}^{Y_R K_3} G_O(0). \end{aligned}$$

$$\text{Dense target:}\quad N = \mathcal{O}(1), \qquad \partial_Y(N, O) \text{ contains the nonlinear terms just derived.}$$

Here $Y_R = \ln(s/s_0)$, with fixed energy-squared scale s_0; it differs from $Y = \ln(x_0/x)$ by a constant only when the relevant hard scales are fixed in the high-energy limit. K_2, K_3 evolve colour-singlet states of two or three *reggeized* gluons. A reggeized gluon is an effective exchanged degree of freedom containing virtual high-energy corrections, not a new elementary particle. BFKL denotes Balitsky–Fadin–Kuraev–Lipatov; BKP denotes Bartels–Kwieciński–Praszałowicz. The dilute dipole equation selects the components to which a dipole couples. Its expansion makes the resummation explicit:

$$\mathrm{e}^{Y_R K} = 1 + Y_R K + \frac{Y_R^2 K^2}{2!} + \cdots , \qquad K = \mathcal{O}(\alpha_s), \qquad (\alpha_s Y_R)^n \text{ need not be small.}$$

Thus "two versus three gluons" identifies the first nonzero term; it does not exhaust either exchange. As a *quoted leading-logarithmic, fixed-coupling spectral result*, the forward, azimuthally symmetric BFKL eigenmodes have [21]

$$\begin{aligned} G_\nu(Y_R) &\sim \mathrm{e}^{\omega(\nu) Y_R}, \qquad \omega(\nu) = \bar{\alpha}\,[2\psi(1) - \psi(\tfrac{1}{2} + \mathrm{i}\nu) - \psi(\tfrac{1}{2} - \mathrm{i}\nu)], \\ \psi(z) &= \frac{\mathrm{d}}{\mathrm{d}z} \ln \Gamma(z), \qquad \omega(0) = 4 \ln 2\, \bar{\alpha}, \qquad \alpha_P^{\mathrm{LL}} - 1 = 4 \ln 2\, \bar{\alpha}. \end{aligned}$$

ν is real; the leading saddle is at $\nu = 0$. This is a maximum along the $1/2 + \mathrm{i}\nu$ contour, not along real anomalous dimension. A perturbative Odderon solution has $\omega_O = 0$, hence $\alpha_O^{\mathrm{LL}} = 1$, and couples to $\gamma^{(*)} \to \eta_c$ [22]. We quote the spectral result rather than claim to solve BKP here; running coupling, higher logarithmic orders, density and infrared physics alter the interpretation of these intercepts.

Supporting derivation. The signature algebra below explains why nonzero odd exchange need not yield interference. The pole ansatz illustrates the phase structure; it is not an extra target model.

Define the Regge approximation rather than assuming it. For one spin amplitude, write the invariant amplitude

$$F_\varsigma(s,t) = \beta_\varsigma(t) \left(\frac{s}{s_0} \right)^{\alpha_\varsigma(t)} \eta_\varsigma(\alpha_\varsigma), \qquad \eta_\varsigma(\alpha) = -\frac{1 + \varsigma \mathrm{e}^{-\mathrm{i}\pi\alpha}}{\sin \pi\alpha}, \quad \varsigma = \pm 1. \tag{18.4}$$

β_ς is a real residue in this simple-pole illustration; $\alpha_\varsigma(t)$ is the trajectory and $\alpha_\varsigma(0)$ its intercept. The overall minus sign fixes a convenient positive-imaginary even amplitude for positive residue. Analytic continuation supplies $(-s - \mathrm{i}0)^\alpha = s^\alpha \mathrm{e}^{-\mathrm{i}\pi\alpha}$; combining direct and crossed terms gives

$1 + \varsigma e^{-i\pi\alpha}$. The pole denominator is the Regge-pole ansatz. Neither this ansatz nor real residues follow from the dipole calculation alone.

$$1 + e^{-i\pi\alpha} = 2e^{-i\pi\alpha/2}\cos(\pi\alpha/2), \qquad \eta_+ = -\cot(\pi\alpha/2) + i,$$
$$1 - e^{-i\pi\alpha} = 2ie^{-i\pi\alpha/2}\sin(\pi\alpha/2), \qquad \eta_- = -\tan(\pi\alpha/2) - i.$$

The signature is the parity under crossing $s \leftrightarrow u$. In the neutral vacuum-quantum-number exchange sectors considered here, even/odd signature goes with the Pomeron/Odderon C assignment. Crossing parity is not the operation of interchanging two measured charm jets.

Check. Why can a small coupling still require a resummation?

18B. The phase is part of the physics: what elastic data measure. For $\alpha_\pm = 1 + \delta_\pm$, the preceding identities give

$$\eta_+ = \tan\frac{\pi\delta_+}{2} + i, \qquad \eta_- = \cot\frac{\pi\delta_-}{2} - i.$$

An even exchange near intercept one is mostly imaginary; an odd exchange can be mostly real. The odd simple-pole factor has a pole at $\delta_- = 0$: a finite limit requires the residue and the full analytic structure to be treated together. Do not substitute zero into that factor alone. Regge cuts and logarithmic amplitudes need not have the phase of a single pole.

For a transparent finite-residue example, absorb the real factors into real functions $B_\pm(s,t)$ and write, away from signature singularities,

$$\begin{aligned} F_+ &= B_+(i + a), & a &= \tan(\pi\delta_+/2), \\ F_- &= B_-(1 - ib), & b &= \tan(\pi\delta_-/2), \\ \Re(F_+F_-^*) &= B_+B_-\Re[(i + a)(1 + ib)] = B_+B_-(a - b). \end{aligned} \tag{18.5}$$

With real impact factors, equal trajectories give a relative phase of $\pi/2$ and zero interference in this example. A nonzero odd amplitude does not automatically produce a nonzero interference observable.

The useful experimental question is not only whether an odd amplitude exists. We must choose an observable that retains its phase and its interference with the even amplitude.

Define crossing combinations in a fixed helicity and phase convention:

$$F^{pp} = F_+ + F_-, \qquad F^{p\bar{p}} = F_+ - F_-,$$
$$\frac{d\sigma_{el}}{dt} = \frac{|F|^2}{16\pi s^2}, \qquad \sigma_{tot} = \frac{\Im F(s,0)}{s} \qquad (s \gg m_p^2),$$
$$\sigma_{tot}^{pp} - \sigma_{tot}^{p\bar{p}} = \frac{2\Im F_-(s,0)}{s},$$
$$\frac{d\sigma^{pp}}{dt} - \frac{d\sigma^{p\bar{p}}}{dt} = \frac{\Re(F_+F_-^*)}{4\pi s^2}, \qquad \rho_{el} = \frac{\Re F(s,0)}{\Im F(s,0)}.$$

Spin sums and averages must be restored for an unpolarised measurement. ρ_{el} is an elastic phase ratio; it is unrelated to our opacity variable $\rho = Q_s^2/\epsilon^2$. A nearly real odd amplitude can be weak in a total-cross-section difference and still matter for ρ_{el} or near a dip of the even amplitude.

What established evidence does, and does not, establish. D0 and TOTEM compared $p\bar{p}$ data at 1.96 TeV with pp data extrapolated to that energy, finding a 3.4σ difference in the dip–bump region; combined with forward-scattering information, their analysis reported significance above 5σ for a C-odd contribution [23]. A complementary scaling analysis including 8 TeV data, coauthored by Pasechnik, reports its own significance and explicitly tests the scaling region [24]. These are distinct analyses with stated energy-comparison assumptions. They constrain a hadronic

crossing-odd amplitude; they do not by themselves determine the small-dipole $O(r, b, x)$, its proton coupling, or an exclusive η_c rate.

Check. If F_- is real, must it change $\sigma_{\rm tot}$?

18C. Return to our measured charm pair: construct the charge-odd observable. Let $\chi = (z, \boldsymbol{p}, \boldsymbol{\Delta}; h, \bar{h})$ denote the full pair configuration, with

$$\mathsf{C}\chi = (1 - z, -\boldsymbol{p}, \boldsymbol{\Delta}; \bar{h}, h).$$

Use the same canonically ordered pair basis as in the QFT construction. Choose quark charge-conjugation phases consistently, so

$$|\chi\rangle = b_h^\dagger(p_1) d_{\bar{h}}^\dagger(p_2)|0\rangle, \qquad C|\chi\rangle = d_h^\dagger(p_1) b_{\bar{h}}^\dagger(p_2)|0\rangle = -|\mathsf{C}\chi\rangle.$$

The reordering minus sign and $C_\gamma = -1$ cancel. For $C\widehat{T}_{P,O}C^{-1} = \pm\widehat{T}_{P,O}$, the numerical amplitudes therefore obey

$$\mathcal{A}_P(\mathsf{C}\chi) = +\mathcal{A}_P(\chi), \qquad \mathcal{A}_O(\mathsf{C}\chi) = -\mathcal{A}_O(\chi).$$

An even coefficient under this ordered relabelling is compatible with an intrinsically C-odd pair because the state itself supplies the minus sign. Other consistent external-state phases can move a common sign; the following observable is invariant. For charge-symmetric phase-space and acceptance factors, $\mathcal{J}(\mathsf{C}\chi) = \mathcal{J}(\chi)$, write $\mathcal{J} = \mathcal{J}(\chi)$:

$$\begin{aligned} \mathrm{d}\sigma(\chi) &= \mathcal{J}|\mathcal{A}_P + \mathcal{A}_O|^2, \\ \mathrm{d}\sigma(\mathsf{C}\chi) &= \mathcal{J}|\mathcal{A}_P - \mathcal{A}_O|^2, \\ \mathrm{d}\sigma(\chi) - \mathrm{d}\sigma(\mathsf{C}\chi) &= 4\mathcal{J}\Re(\mathcal{A}_P\mathcal{A}_O^*), \\ \mathrm{d}\sigma(\chi) + \mathrm{d}\sigma(\mathsf{C}\chi) &= 2\mathcal{J}(|\mathcal{A}_P|^2 + |\mathcal{A}_O|^2). \end{aligned}$$

Thus

$$\boxed{A_C(\chi) = \frac{2\Re(\mathcal{A}_P\mathcal{A}_O^*)}{|\mathcal{A}_P|^2 + |\mathcal{A}_O|^2} = \frac{2r_O\cos(\phi_P - \phi_O)}{1 + r_O^2}, \quad r_O = \frac{|\mathcal{A}_O|}{|\mathcal{A}_P|}.} \tag{18.6}$$

For spin sums, replace every product by the sum of products before taking the ratio. The phase formula is for one amplitude or mutually aligned spin amplitudes.

Derive the cancellation that dictates the measurement. For a domain D invariant under $\chi \mapsto \mathsf{C}\chi$, put $I(\chi) = 2\Re(\mathcal{A}_P\mathcal{A}_O^*)$. Then

$$\begin{aligned} I(\mathsf{C}\chi) &= -I(\chi), & \int_D \mathrm{d}\chi\, I(\chi) &= 0, \\ w(\mathsf{C}\chi) &= -w(\chi), & \int_D \mathrm{d}\chi\, w(\chi) I(\chi) &\text{ need not vanish.} \end{aligned}$$

For example $w = 2z - 1$ selects an energy-sharing asymmetry after a symmetric angular integration; angular weights can retain recoil-dependent structures that the energy asymmetry misses. A concrete moment is

$$\langle w\rangle_C = \frac{\int_D \mathrm{d}\chi\, w(\chi)\, \mathrm{d}\sigma/\,\mathrm{d}\chi}{\int_D \mathrm{d}\chi\, \mathrm{d}\sigma/\,\mathrm{d}\chi} = \frac{\int_D \mathrm{d}\chi\, \mathcal{J}(\chi) w(\chi)\, 2\Re(\mathcal{A}_P\mathcal{A}_O^*)}{\int_D \mathrm{d}\chi\, \mathcal{J}(\chi)(|\mathcal{A}_P|^2 + |\mathcal{A}_O|^2)}.$$

The factor $\mathcal{J}$ cancels in the pointwise asymmetry, but must remain inside the integrated moment unless it is constant. Charge identification, symmetric cuts and an estimate of electromagnetic odd exchange are essential. The original diffractive charm proposal makes precisely this interference strategy explicit [25]. A toy value $r_O = 0.10$ gives $A_C = 0$ for a 90° relative phase but $A_C \simeq 0.099$ for 60°; this illustrates phase sensitivity, not a prediction.

Result to keep. Our Gaussian, isotropic $O = 0$ benchmark predicts $A_C = 0$. Its shifted momentum-space zero is a multiple-scattering effect; it is not an Odderon signal. To predict A_C, specify an odd target correlator and phases, then check which integrations preserve the interference.

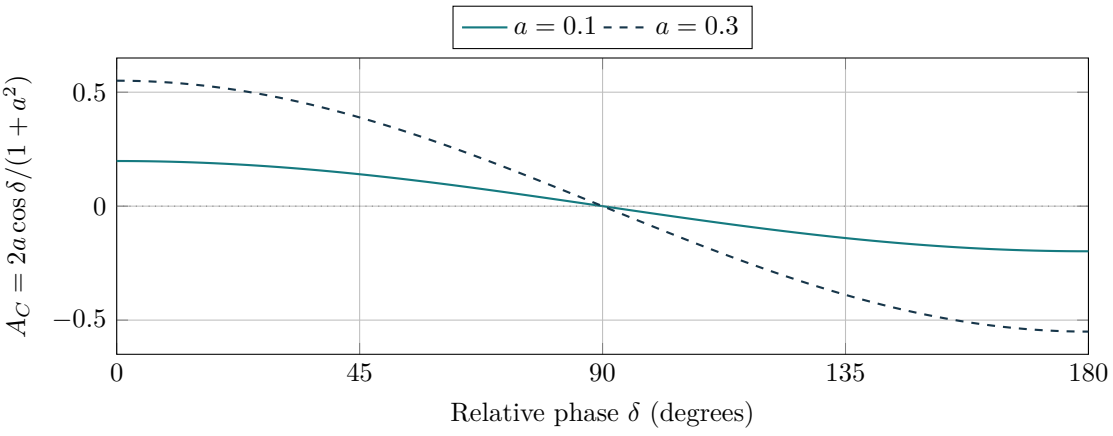


Figure 11: Illustrative phase dependence of a charge asymmetry. Here $a = r_O = |\mathcal{A}_O|/|\mathcal{A}_P|$ at one configuration; the curves use the derived ratio and are not cross-section predictions.

18D. Bound-state filters: which exchange can each experiment select?

Binding the same pair gives us a quantum-number filter. Open charm can retain interference between sectors; a state of definite charge parity selects one sector at its photon impact factor.

The photon-to-meson matrix element of a C-eigenoperator $\widehat{T}_\varsigma$ obeys

$$\begin{aligned}\langle M|\widehat{T}_\varsigma|\gamma\rangle &= \langle M|C^{-1}(C\widehat{T}_\varsigma C^{-1})C|\gamma\rangle \\ &= C_M C_\gamma \varsigma\, \langle M|\widehat{T}_\varsigma|\gamma\rangle, \\ \langle M|\widehat{T}_\varsigma|\gamma\rangle \neq 0 \quad &\Longrightarrow \quad \varsigma = C_M C_\gamma.\end{aligned}$$

With $C = (-1)^{L+S}$, this reproduces and explains the previous table:

$$\begin{aligned}\gamma^{(*)}p \to J/\psi p: &\quad C_{J/\psi} = -1, &&\mathcal{A} \sim \Psi^*_{^3S_1} \otimes N \otimes \psi_\gamma, \\ \gamma^{(*)}p \to h_c p: &\quad C_{h_c} = -1, &&\mathcal{A} \sim \Psi^*_{^1P_1} \otimes N \otimes \psi_\gamma, \\ \gamma^{(*)}p \to \eta_c p: &\quad C_{\eta_c} = +1, &&\mathcal{A}_{\text{strong}} \sim \Psi^*_{^1S_0} \otimes O \otimes \psi_\gamma.\end{aligned}$$

The symbols $\otimes$ include the already defined spin projection, transverse Fourier transform and integrations. This selection does not fix a rate. J/ψ needs the triplet S wave; h_c needs the differentiated singlet P wave. A radial-wavefunction substitution cannot change one into the other. Photoproduction uses the transverse photon; our longitudinal open-charm result vanishes at $Q^2 = 0$.

An Odderon filter still has a calculable electromagnetic competitor. For η_c, the exchanged photon also has $C = -1$. Use the invariant amplitude $\mathcal{T}$ and the reduced amplitude $\mathcal{A}$ in one convention:

$$\begin{aligned}\mathcal{A}(\gamma p \to \eta_c p) &= \mathcal{A}_O + \mathcal{A}_{\text{Prim}}, \\ \mathcal{A}_{\text{Prim}} &= \frac{\mathcal{T}_{\text{Prim}}}{W^2 - m_p^2}, \\ \mathcal{T}_{\text{Prim}} &= \frac{e}{t}\, J_p^\mu\, \Gamma^{\gamma^*\gamma\eta_c}_{\mu\nu}\varepsilon^\nu_\gamma, \\ \Gamma_{\mu\nu} &= F_{\eta\gamma\gamma}(t)\epsilon_{\mu\nu\alpha\beta}q^\alpha\Delta^\beta, \\ J_p^\mu &= \bar{u}(P')\left[F_1(t)\gamma^\mu + F_2(t)\frac{\mathrm{i}\sigma^{\mu\nu}\Delta_\nu}{2m_p}\right]u(P), \\ \frac{\mathrm{d}\sigma}{\mathrm{d}t} &= \frac{1}{16\pi}\overline{\sum}(|\mathcal{A}_O|^2 + |\mathcal{A}_{\text{Prim}}|^2 + 2\Re[\mathcal{A}_O\mathcal{A}^*_{\text{Prim}}]).\end{aligned}$$

F_1, F_2 are the proton Dirac and Pauli electromagnetic form factors, counted once in J_p. The transition form factor $F_{\eta\gamma\gamma}$ has dimension GeV^{-1} and here includes both photon couplings at the meson vertex; its phase fixes our definition of Γ and must be matched consistently to $\mathcal{A}_O$. Thus $\mathcal{T}$ is dimensionless and $\mathcal{A}$ has dimension GeV^{-2}. The bar denotes the physical initial-helicity average and final-helicity sum. A numerical prediction needs this transition form factor and the strong amplitude in the same phase convention. The photon propagator enhances small $|t|$, but the antisymmetric vertex also contains recoil momentum: infer the small-t power only after contracting the full amplitude. Published small-x calculations compare the two mechanisms and their nuclear diffraction patterns [20]; their numerical crossover is model- and kinematics-dependent.

Why hadroproduction changes the selection rule at the central vertex. In $pp \to p{+}J/\psi{+}p$, a central state can couple to two exchanges:

$$C_{J/\psi} = C_1 C_2 = -1, \qquad PO \to J/\psi \text{ and } \gamma P \to J/\psi \text{ are allowed.}$$

Hence J/ψ photoproduction selects even strong exchange, while exclusive hadroproduction can involve Pomeron–Odderon fusion [26]. Two even exchanges cannot produce the isolated C-odd meson through a C-conserving central vertex. Additional spectator interactions can fill the rapidity gaps; a hadronic prediction needs an absorptive correction or survival calculation and its uncertainty. This does not follow by replacing the incoming photon in our dipole formula.

Check. Does observing h_c in photoproduction identify an Odderon?

Supporting derivation. The slope and variance derivations establish two conclusions: coherent recoil resolves the mean profile; diffractive target dissociation probes fluctuations under the stated completeness assumption.

18E. Turn "proton structure" into measurements. Our amplitude has two separate Fourier resolutions:

$$\mathcal{A}_{c\bar{c}} \sim \int \mathrm{d}^2 r\, \mathrm{d}^2 b\, \mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}+\mathrm{i}\boldsymbol{\Delta}\cdot\boldsymbol{b}} \psi_\gamma(z,r)[N(r,b) - \mathrm{i}O(r,b)], \qquad |t| \simeq \Delta^2.$$

The overall common factor 2i is suppressed on this line. $\boldsymbol{p}$ resolves the dipole size; $\boldsymbol{\Delta}$ resolves the target profile. A zero in p caused by the dipole transform is not the same object as a diffractive minimum in t caused by the target profile.

For a mostly imaginary even amplitude, write a normalized impact-parameter weight w_b for one fixed channel. Expanding its Fourier transform,

$$\frac{\mathcal{A}(\Delta)}{\mathcal{A}(0)} = \int \mathrm{d}^2 b\, w_b(b) \mathrm{e}^{\mathrm{i}\boldsymbol{\Delta}\cdot\boldsymbol{b}} = 1 - \frac{\Delta^2 \langle b^2 \rangle_w}{4} + O(\Delta^4),$$
$$\frac{\mathrm{d}\sigma}{\mathrm{d}|t|} \simeq \left.\frac{\mathrm{d}\sigma}{\mathrm{d}|t|}\right|_0 \mathrm{e}^{-B|t|}, \qquad B = \frac{\langle b^2 \rangle_w}{2}.$$

This assumes azimuthal symmetry, a common phase and a positive normalized weight; a sign-changing overlap requires the corresponding amplitude-weighted moment. For the disk $\langle b^2 \rangle = R^2/2$, hence $B = R^2/4$. For a meson, the exact phase also contains $(z - 1/2)\boldsymbol{r}$, as in Eq. (17.5); the approximation above neglects that displacement.

A second measurable quantity: fluctuations rather than only the mean. Resolve the target as a coherent superposition of scattering eigenconfigurations ω, with probabilities $P(\omega)$. For a fixed final projectile state let A_ω be its amplitude. With $|p\rangle = \sum_\omega c_\omega |\omega\rangle$, $P(\omega) = |c_\omega|^2$ and $\widehat{T}|\omega\rangle = A_\omega |\omega\rangle$, completeness gives [8]

$$\langle p|\widehat{T}|p\rangle = \sum_\omega P(\omega) A_\omega, \qquad \sum_Y |\langle Y|\widehat{T}|p\rangle|^2 = \langle p|\widehat{T}^\dagger \widehat{T}|p\rangle = \sum_\omega P(\omega)|A_\omega|^2.$$

Therefore

$$\frac{\mathrm{d}\sigma_{\mathrm{coh}}}{\mathrm{d}t} = \frac{|\langle A_\omega\rangle_\omega|^2}{16\pi}, \qquad \frac{\mathrm{d}\sigma_{\mathrm{diff,all}}}{\mathrm{d}t} = \frac{\langle |A_\omega|^2\rangle_\omega}{16\pi},$$
$$\frac{\mathrm{d}\sigma_{\mathrm{incoh}}}{\mathrm{d}t} = \frac{\langle |A_\omega|^2\rangle_\omega - |\langle A_\omega\rangle_\omega|^2}{16\pi} \geq 0.$$

The coherent target remains intact; the displayed incoherent variance uses a complete sum over diffractive target excitations in this frozen-configuration approximation. A restricted dissociation mass or a target-fragment tag requires the corresponding final-state projector and acceptance, rather than an unqualified use of completeness. The smooth disk has no such fluctuations. Studies of coherent and incoherent J/ψ production use this distinction to constrain target geometry [27].

A controlled scan, with one purpose for each variable.

Change	Question answered
Q^2 or p^2	Does selecting smaller dipoles restore the dilute result? Use $\epsilon^2 = m_c^2 + z(1-z)Q^2$, not Q^2 alone.
$x_{\mathbb{P}}$ at fixed hard scales	Can linear energy evolution suffice, or do density corrections matter? Hold the bin definition fixed.
t and relative azimuth	Which transverse profiles, nodes and even/odd harmonics survive the transform?
Proton versus nucleus	Does increased thickness amplify multiple scattering with a consistently specified nuclear geometry?
Coherent versus incoherent	Are changes due to the mean profile or its configuration variance?
J/ψ, h_c, η_c, open charm	Can one target operator describe distinct spin/orbital/charge-parity projections within matching accuracy?

18F. The conceptual conclusion: three questions with different answers.

We can now make a precise statement about what saturation, the Pomeron and the Odderon mean in our calculation. They are connected, but they answer different questions. Confusing them would turn our calculable benchmark into an unjustified claim about QCD.

$$\underbrace{N_{xy} = N_{yx}, \quad O_{xy} = -O_{yx}}_{\text{Which charge-parity sector is exchanged?}} \implies \text{Pomeron-like versus Odderon-like operator,}$$
$$\underbrace{1 - S = \tau - \frac{\tau^2}{2} + \cdots}_{\text{Is one dilute interaction sufficient?}} \implies \text{our finite-bin multiple-scattering test,}$$
$$\underbrace{\mathcal{A} = \text{impact factor} \otimes \text{target matrix element}}_{\text{Can long and short distances be separated?}} \implies \text{a factorisation statement with conditions.}$$

Here $\tau = Q_s^2 r^2/4$ only in our Gaussian benchmark. Saturation can occur entirely in the even sector. An odd exchange can exist in a dilute target. A nonlinear correction does not automatically invalidate every form of factorisation: the high-energy Wilson-line separation itself is a factorised description, with a rapidity scale and an accuracy to specify.

Result to keep. The completed charm calculation measures the failure of the first term of a specified target-scattering expansion. The charge-parity analysis tells us how to design a separate measurement of the odd sector. Neither result alone proves the breakdown of the collinear factorisation theorem for a different observable.

What remains fundamental and unresolved? State the next calculable step.

1. **Soft and hard exchange.** How does a perturbatively defined small dipole connect to the large-distance crossing amplitudes of elastic hadron scattering? Next step: specify one common impact-parameter-dependent input and test where it can be matched, rather than equating fitted intercepts.

2. **Odd structure of the target.** Which non-Gaussian colour correlations and spatial structures give $O \neq 0$? Next step: propose an odd initial correlator that obeys $O_{yx} = -O_{xy}$, evolve it with the coupled equations, and construct an interference moment that survives angular integration.

3. **Universal target versus process-specific projection.** Can the same evolved S describe open charm and J/ψ with controlled mass, scale and wavefunction uncertainties? Next step: compare the already derived dilute limit before comparing fitted rates. For h_c, first complete the spin-singlet P-wave impact factor and its dilute matching test.

4. **Density corrections versus alternative explanations.** Does the finite-bin change persist after radiation, profile and scale effects? Next step: add these one at a time to our signed-amplitude benchmark. A fit improvement alone does not identify the physical mechanism.

Check. Give a process, an operator and an observable that test the odd sector.

Our end product is now more than a list of QCD concepts: it is a reproducible even-sector calculation, its error test, and a derivation of what a distinct odd-sector measurement would have to retain. This is the foundation for choosing the next research calculation.

Check. Would a vanishing Odderon in our symmetric Gaussian disk mean QCD forbids a proton Odderon?

Next. *Which results are now established, and what specific calculations remain before drawing experimental conclusions?*

19 The result we have earned

Purpose. Close the calculation: return to the first question, read the numerical answer, and state its physical scope.

The completed calculation

$$\mathcal{D}_L^{\text{model}} = \left.\frac{\mathrm{d}\sigma_L}{\mathrm{d}z\ \mathrm{d}p^2\ \mathrm{d}|t|}\right|_{t\simeq 0} = \frac{N_c \alpha_{\text{em}} e_c^2 Q^2 z^2 (1-z)^2 A_\perp^2}{2\pi^2 \epsilon^4} |f(\rho, v)|^2,$$

$$f_{\text{dilute}}(\rho, v) = \rho \frac{1-v}{(1+v)^3},$$

$$f_{\text{full}}(\rho, v) = \frac{1}{1+v} - \int_0^\infty \frac{\mathrm{d}u\, \mathrm{e}^{-u-vu/(1+\rho u)}}{1+\rho u},$$

$$\epsilon^2 = m_c^2 + z(1-z)Q^2, \qquad \rho = Q_s^2/\epsilon^2, \qquad v = p^2/\epsilon^2.$$

This is a complete leading-order longitudinal open-pair result within the declared high-energy approximation and finite-disk target model. The absolute normalization was fixed by the state conventions and checked by completeness; it was not adjusted after the comparison.

From the formula to a reproducible number.

$$m_c = 1.4 \text{ GeV}, \qquad Q^2 = 8 \text{ GeV}^2, \qquad z = \tfrac{1}{2},$$
$$Q_s^2 = 0.396 \text{ GeV}^2, \qquad R = 3 \text{ GeV}^{-1}, \qquad \rho = 0.1.$$

At fixed $x_{\mathbb{P}}, Q^2, z$, with bin boundaries expressed in v,

Result	Dilute versus full	Physical reading
First bin $[0, 0.5]$	$R = 0.825941$	17.4% suppression
Bin across the zero $[0.5, 1.5]$	$R = 1.332619$	33.3% enhancement
Open-pair amplitude zero	$3.9600 \to 4.3534$ GeV^2	The cancellation shifts

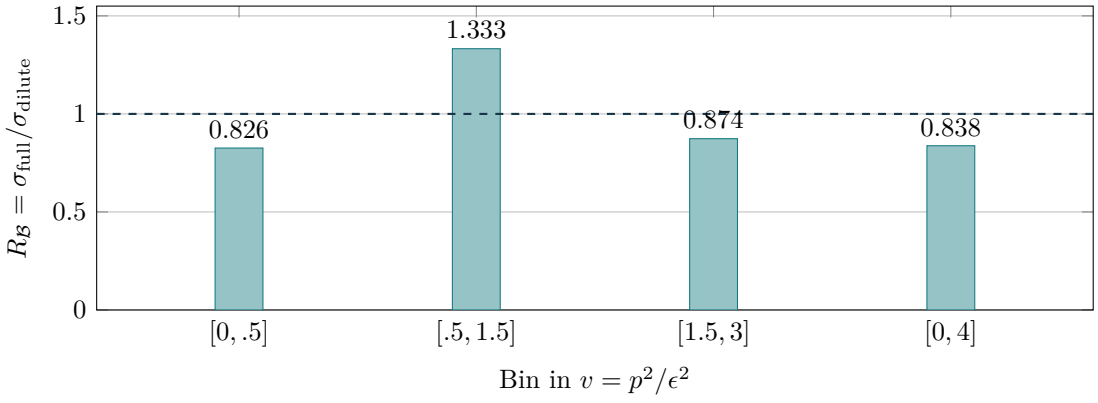


Figure 12: Finite-bin ratios computed from the same target model and normalized spectrum. The dashed line is equality; the bin across the shifted zero is enhanced even though the first bin is suppressed.

Here is the answer we earned. A hard probe can still show appreciable sensitivity to repeated target interactions, and the size and even the sign of the change depend on the measured momentum bin. We proved that with a calculation; we did not assume it from a small value of x.

From one spectrum to the physical framework.

$$\begin{aligned}
\text{Known photon fluctuation} &\longrightarrow \text{same defined target operator } S_{xy},\\
\text{Increasing rapidity} &\longrightarrow \text{Balitsky hierarchy; BK after closure,}\\
\text{Changing the final state} &\longrightarrow \text{continuum, } {}^3S_1\ (J/\psi), {}^1P_1\ (h_c),\\
\text{Charge conjugation} &\longrightarrow S = 1 - N + \mathrm{i}O, \quad N \text{ even}, \quad O \text{ odd.}
\end{aligned}$$

The Pomeron and Odderon discussion tells us which part of this target amplitude a channel can select. Both J/ψ and h_c select even exchange, with different spin and orbital weights; η_c supplies an odd-exchange comparison. Selection rules identify a mechanism, while impact factors, target structure, phases and competing exchanges decide whether an observable can reveal it.

What is established; what remains a research task.

Established in this course	Requires further input or calculation
A normalized open-charm spectrum and a controlled same-model comparison.	A data-constrained proton profile, realistic experimental bins and higher-order corrections.
The dipole's field-theory origin and its leading-log evolution with an explicit closure.	A precision evolved target and a matched comparison to a collinear GPD calculation.
The S-wave/P-wave projection rule and the even/odd exchange selection.	A complete matched h_c electroproduction impact factor and quantitative control of competing contributions in an Odderon measurement.

In particular, the worked ratio supplies neither a universal saturation threshold nor a counterexample to the hard-diffractive DIS factorisation theorem. A vanishing Odderon in our symmetric target model is a property of that model, not a statement that QCD has no odd exchange.

Questions for self-study.

1. Explain in one sentence why a hard scale and a dilute target are different conditions.
2. Identify the exact approximation changed between the two plotted spectra, and explain the finite-bin enhancement.
3. Explain why h_c is not an Odderon filter, and name one extra ingredient needed to turn an allowed odd exchange into an observable.

Further study. The calculation gives an observable whose approximation error can be computed. Further study can ask whether this sensitivity survives realistic target information, improved accuracy and the final-state projection that an experiment actually measures.

A Derivation sheet: the transverse photon and the J/ψ mass term

Purpose: obtain the transverse overlap on Section 17 from the same normalized spinors as the longitudinal calculation. No new photon normalization is introduced.

Let $\varepsilon_T^\mu = (0, 0, \boldsymbol{e})$, $\boldsymbol{e}^* \cdot \boldsymbol{e} = 1$, and $\not{\varepsilon}_T = -\boldsymbol{\gamma}_\perp \cdot \boldsymbol{e}$. Define $\mathsf{W} = (w_+, w_-)$, the 4×2 matrix made from Section 4's spinors. A 2×2 pair matrix has row h and column $\bar{h}$ in the order $(+, -)$; therefore its antiquark good component uses $\mathsf{W}\sigma^1$.

$$\mathsf{W}^\dagger \mathsf{W} = I_2, \qquad \mathsf{W}^\dagger \gamma^1 \mathsf{W} = -\mathrm{i}\sigma^2, \qquad \mathsf{W}^\dagger \gamma^2 \mathsf{W} = \mathrm{i}\sigma^1,$$
$$\mathsf{W}^\dagger \Sigma^3 \mathsf{W} = \sigma^3, \qquad \Sigma^3 = \begin{pmatrix} \sigma^3 & 0 \\ 0 & \sigma^3 \end{pmatrix}, \qquad \alpha^i \alpha^j = \delta^{ij} + \mathrm{i}\epsilon^{ij3}\Sigma^3 \quad (i, j = 1, 2).$$

Insert u_- and v_- from Section 4, using $\bar{\boldsymbol{k}} = -\boldsymbol{k}$:

$$\mathsf{D}_k \equiv \boldsymbol{\alpha}_\perp \cdot \boldsymbol{k} + \beta m_c, \qquad \mathsf{A}_e \equiv \boldsymbol{\alpha}_\perp \cdot \boldsymbol{e}.$$

$$\begin{aligned}
[\bar{u} \not{\varepsilon}_T v]_{h\bar{h}} &= -u_+^\dagger \left[-\frac{\mathsf{A}_e \mathsf{D}_k}{\sqrt{2}\bar{k}^+} + \frac{\mathsf{D}_k \mathsf{A}_e}{\sqrt{2}k^+} \right] v_+ \\
&= \sqrt{k^+ \bar{k}^+}\, \mathsf{W}^\dagger \left[\frac{\mathsf{A}_e \mathsf{D}_k}{\bar{k}^+} - \frac{\mathsf{D}_k \mathsf{A}_e}{k^+} \right] \mathsf{W}\sigma^1 \\
&= \frac{m_c \mathsf{M}(\boldsymbol{e}) + (2z-1)(\boldsymbol{e}\cdot\boldsymbol{k})\sigma^1 - (\boldsymbol{e}\times\boldsymbol{k})_3 \sigma^2}{\sqrt{z(1-z)}}.
\end{aligned}$$

$$\mathsf{M}(\boldsymbol{e}) \equiv e_x \sigma^3 - \mathrm{i} e_y I_2, \qquad (\boldsymbol{e}\times\boldsymbol{k})_3 \equiv e_x k_y - e_y k_x.$$

The three terms respectively come from the fermion mass, unequal longitudinal momentum sharing, and transverse spin algebra.

$$\begin{aligned}
\widetilde{\boldsymbol{\psi}}_T^{(1)} &= -\frac{e e_c \sqrt{N_c z(1-z)}}{2\sqrt{\pi}} \frac{\bar{u} \not{\varepsilon}_T v}{\boldsymbol{k}^2 + \epsilon^2} \\
&= -\frac{e e_c \sqrt{N_c}}{2\sqrt{\pi}} \frac{m_c \mathsf{M}(\boldsymbol{e}) + (2z-1)(\boldsymbol{e}\cdot\boldsymbol{k})\sigma^1 - (\boldsymbol{e}\times\boldsymbol{k})_3 \sigma^2}{\boldsymbol{k}^2 + \epsilon^2}.
\end{aligned} \tag{A.1}$$

The bold $\boldsymbol{\psi}$ denotes a helicity matrix, not a spatial vector. Its prefactor follows from $F = V/E_D$ and Eq. (3.2), followed by the color-singlet projection.

$$\int \frac{\mathrm{d}^2\boldsymbol{k}}{(2\pi)^2} \frac{k_i \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}}{\boldsymbol{k}^2 + \epsilon^2} = -\mathrm{i}\partial_i \frac{K_0(\epsilon r)}{2\pi}, \qquad \partial_i K_0(\epsilon r) = -\epsilon K_1(\epsilon r) \frac{r_i}{r}.$$

$$\begin{aligned}
\boldsymbol{\psi}_T^{(1)}(z, \boldsymbol{r}) = -\frac{e_c \sqrt{N_c \alpha_{\mathrm{em}}}}{2\pi} \big[& m_c \mathsf{M}(\boldsymbol{e}) K_0(\epsilon r) - \mathrm{i}(2z-1)\sigma^1 (\boldsymbol{e}\cdot\boldsymbol{\nabla}) K_0(\epsilon r) \\
& + \mathrm{i}\sigma^2 (\boldsymbol{e}\times\boldsymbol{\nabla})_3 K_0(\epsilon r) \big].
\end{aligned} \tag{A.2}$$

As a normalization check, average over the two orthogonal photon transverse polarizations (equivalently use either circular polarization):

$$\frac{1}{2} \sum_{\lambda=1,2} \sum_{h,\bar{h}} |\psi^{(1)}_{T\lambda,h\bar{h}}|^2 = \frac{N_c \alpha_{\mathrm{em}} e_c^2}{2\pi^2} \left[m_c^2 K_0^2(\epsilon r) + (z^2 + (1-z)^2) \epsilon^2 K_1^2(\epsilon r) \right]. \tag{A.3}$$

Use $\mathrm{Tr}(\sigma^i \sigma^j) = 2\delta^{ij}$, $\mathrm{Tr}\, \mathsf{M}^\dagger \mathsf{M} = 2$, and $[(2z-1)^2 + 1]/2 = z^2 + (1-z)^2$.

Project onto a leading nonrelativistic spin-triplet state.

$$\boldsymbol{\Psi}^{(1)}_{J,L} = \phi_S \frac{\sigma^1}{\sqrt{2}}, \qquad \boldsymbol{\Psi}^{(1)}_{J,T} = \phi_S \frac{\mathsf{M}(\boldsymbol{e}_J)}{\sqrt{2}}, \qquad \int \mathrm{d}z\ \mathrm{d}^2 r\, |\phi_S|^2 = 1.$$

$$\mathrm{Tr}\left[\frac{\mathsf{M}^\dagger(\boldsymbol{e}_J)}{\sqrt{2}}\mathsf{M}(\boldsymbol{e}_\gamma)\right] = \sqrt{2}\boldsymbol{e}_J^* \cdot \boldsymbol{e}_\gamma, \qquad \mathrm{Tr}[\mathsf{M}^\dagger \sigma^{1,2}] = 0,$$

$$\begin{aligned} W^{\mathrm{NR}}_{JT} &\equiv \mathrm{Tr}[(\boldsymbol{\Psi}^{(1)}_{J,T})^\dagger \boldsymbol{\psi}^{(1)}_T] \\ &= -\frac{e_c\sqrt{N_c\alpha_{\mathrm{em}}}}{\sqrt{2}\pi} m_c \phi_S^* K_0(\epsilon r)(\boldsymbol{e}_J^* \cdot \boldsymbol{e}_\gamma), \\ W^{\mathrm{NR}}_{JL} &= -\frac{e_c\sqrt{2N_c\alpha_{\mathrm{em}}}}{\pi} Q z(1-z)\phi_S^* K_0(\epsilon r). \end{aligned}$$

At $z = 1/2$, with the same static spatial wavefunction and matched polarizations,

$$\frac{W^{\mathrm{NR}}_{JL}}{W^{\mathrm{NR}}_{JT}} = \frac{Q}{2m_c}, \qquad Q \to 0: \quad \epsilon \to m_c, \quad W_{JL} \to 0, \quad W_{JT} \text{ remains finite.}$$

The transverse photon remains physical in photoproduction. The constant spin matrices above are the leading static approximation; momentum-dependent meson spin rotations generate additional transverse overlap terms beyond it.

B Supporting calculations and the student report

These calculations extend the main derivation and support independent study and the written report. **Universality means the same operator with the same definition.** The same dipole matrix element occurs in the leading high-energy approximations

$$\begin{aligned}\sigma_{\mathrm{tot}}^{\gamma^* p} &= 2\int \mathrm{d}z\,\mathrm{d}^2r\,\mathrm{d}^2b\,|\psi_\gamma|^2 N,\\ \mathcal{A}^{\gamma^* p\to c\bar{c}p} &= \langle \boldsymbol{p}|\,2\mathrm{i}(1-S)\,|\psi_\gamma\rangle,\\ \mathcal{A}^{\gamma^* p\to Vp} &= \langle \Psi_V|\,2\mathrm{i}(1-S)\,|\psi_\gamma\rangle.\end{aligned}$$

Momentum-transfer transforms and spin sums are understood as already derived. Consistent rapidity scale, target state, impact parameter, and perturbative accuracy are required. More inclusive multijet observables can involve quadrupoles such as

$$Q_{xyuv} = \frac{1}{N_c}\left\langle \mathrm{Tr}(U_x U_y^\dagger U_u U_v^\dagger)\right\rangle, \qquad Q_{xyuv} \neq S_{xy}S_{uv} \text{ in general.}$$

Different Wilson-link structures define different unintegrated gluon distributions. Neither the forward PDF identity of Section 8 nor a fitted dipole ansatz equates all of them.[28, 4]

Angular correlations are another projection of the same target. For a parity-even target, retain one geometric harmonic:

$$\begin{aligned}N(\boldsymbol{r},\boldsymbol{b}) &= N_0(r,b) + 2N_2(r,b)\cos 2(\varphi_r-\varphi_b)+\cdots,\\ \int_0^{2\pi}\mathrm{d}\varphi\,\mathrm{e}^{-\mathrm{i}kr\cos(\varphi-\varphi_k)}\mathrm{e}^{\mathrm{i}n\varphi} &= 2\pi(-\mathrm{i})^n J_n(kr)\mathrm{e}^{\mathrm{i}n\varphi_k},\\ F(\boldsymbol{p},\boldsymbol{\Delta}) &\equiv \int \mathrm{d}^2r\,\mathrm{d}^2b\,\mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}+\mathrm{i}\boldsymbol{\Delta}\cdot\boldsymbol{b}}K_0(\epsilon r)N(\boldsymbol{r},\boldsymbol{b})\\ &= (2\pi)^2\int r\,\mathrm{d}r\,b\,\mathrm{d}b\,K_0(\epsilon r)\Big[J_0(pr)J_0(\Delta b)N_0\\ &\qquad + 2J_2(pr)J_2(\Delta b)N_2\cos 2(\varphi_p-\varphi_\Delta)\Big].\end{aligned}$$

The two second-harmonic Fourier phases multiply to $(-1)(-1) = +1$. Since $J_2(0) = 0$, the anisotropic term disappears at strictly zero recoil in this smooth model. The disk benchmark has $N_2 = 0$ by construction.[7]

Which target input would change the quantitative answer?

Input class	Definition or distinguishing ingredient	What is supplied externally
GBW-type	$N = 1-\exp[-r^2Q_s^2(x)/4]$	$Q_s(x)$, area/profile; no automatic DGLAP or BK solution.
MV-type	Gaussian color sources; small-r exponent includes a Coulomb logarithm	Source density and infrared prescription; an initial condition, not an evolution equation.
IP-Sat / BGK	Exponent tied to a scale-dependent gluon PDF	Initial PDF and scale prescription; IP-Sat also a profile $T_p(b)$.
BK evolved	Specify $N(Y_0,r,b)$ and solve an evolution equation	Initial condition, coupling/resummation scheme and impact-parameter infrared treatment.

For example, with $\int \mathrm{d}^2b\,T_p(b) = 1$,

$$N_{\text{IP-Sat}} = 1-\exp\left[-\frac{\pi^2}{2N_c}r^2\alpha_s(\mu^2)xg(x,\mu^2)T_p(b)\right], \qquad \mu^2 = \mu_0^2 + C_\mu/r^2.$$

Expanding and integrating $2N$ recovers Section 8's coefficient. A model comparison should refit or consistently hold common data constraints, rather than interpret every difference as saturation.[5, 29, 6, 30]

One numerical step of BK, not an unvalidated evolution solver. For a translationally invariant initial condition $N_0(r) = 1 - \exp(-r^2/4)$ with $Q_s = 1$, choose $\boldsymbol{x} = 0$, $\boldsymbol{y} = (r, 0)$, $\boldsymbol{z} = u(\cos\theta, \sin\theta)$:

$$
\begin{aligned}
w^2 &= r^2 + u^2 - 2ru\cos\theta, \qquad u = r\mathrm{e}^s, \\
\frac{\partial_Y N(r)}{\bar{\alpha}} &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}s\int_0^{2\pi}\mathrm{d}\theta\,\frac{r^2}{w^2}[N(u) + N(w) - N(r) - N(u)N(w)], \\
B_0(r,u,w) &\equiv N_0(u) + N_0(w) - N_0(r) - N_0(u)N_0(w), \\
B_0(r,u,w) &= \mathrm{e}^{-r^2/4}\left[1 - \mathrm{e}^{-(u^2+w^2-r^2)/4}\right], \\
\left.\frac{\partial_Y N(1)}{\bar{\alpha}}\right|_{\mathrm{BK}} &= 0.2743732584\ldots, \\
\left.\frac{\partial_Y N(1)}{\bar{\alpha}}\right|_{\mathrm{linear}} &= 0.4417974474\ldots.
\end{aligned}
$$

Use the combined bracket near either daughter endpoint; integrating individually divergent terms is incorrect. Refine both the radial and angular quadrature and vary the finite s boundaries. The initial condition is a numerical test, not a fitted solution of QCD.

A report with testable deliverables.

1. **Normalization and kinematics (analytic).** Reproduce Sections 1, 3 and 5: $M_{c\bar{c}}$, the pair measure and w_L. Derive the Parseval check. Report dimensions and the $Q \to 0$ limit.

2. **The main result (analytic and numerical).** Derive f_{dilute}, f_{next} and f_{full}. Reproduce the shifted zero, one finite-bin ratio and one absolute spectrum value. Check quadrature convergence independently; plot signed amplitudes before squared ones.

3. **Energy evolution (analytic and numerical).** Derive the Fierz reduction and real–virtual subtraction. Reproduce the one-step BK value at $rQ_s = 1$, then compare with the linear value. State the target-correlation assumption.

4. **Charmonium (analytic).** Derive the P-wave momentum moment and the full product-rule derivative. Verify the vacuum h_c zero and the C table. Explain why the nonzero target-assisted h_c result is compatible with that zero.

Choose two substantial derivations and one numerical check for the written report; the remaining prompts support self-study. Include assumptions, equations, a reproducibility table, and one unresolved question. These are reproduction tasks, not a requirement to complete a new research project during the workshop.

Three questions for further investigation.

1. Does the open-charm node sensitivity survive a smooth fitted proton profile, realistic bins, scale variation and higher-order radiation? Start with the same finite-bin observable so each correction has a defined baseline.

2. Can one target operator and controlled wavefunction uncertainties describe both open charm and J/ψ? Reproduce the vector-meson spin-rotation and uncertainty analyses before adding new effects.[15, 16, 31]

3. Can a covariant h_c impact factor be derived, matched to the known two-gluon result, and then used with an evolved dipole? First pass the vacuum-current, C-parity and dilute-limit checks. This is an open calculation in this workshop, not an asserted new effect.

Hadroproduction adds colored projectile spectators and different hard processes. It requires a separate factorisation analysis, even when a heavy-quark dipole ingredient is shared.[32, 33]

Keep. The resulting framework comprises one reproduced observable, its controlled limits, and the tools to formulate the next calculation precisely.

Sources and reading route

For the main calculation, read the tree-level photon construction in Beuf and the coherent-dijet normalization and analytic limits in Salazar–Schenke. For evolution, read Balitsky and Kovchegov after reproducing the real–virtual step. For charmonium, use the spin-rotation and uncertainty papers, Petrelli's projectors and Cui–Hu–Ma's h_c coefficient: Eq. (43) in the arXiv version, Eq. (44) in the published version. For the exchange discussion, follow Hatta's operator evolution, Brodsky–Rathsman–Merino's charm interference proposal, the D0/TOTEM and scaling analyses for hadronic evidence, and Benić and collaborators for the electromagnetic competitor. Pasechnik–Taševský's review supplies the broader context. Quoted spectral and matching results are explicitly identified in the text.

References

[1] G. Beuf, *Dipole factorization for DIS at NLO: I. Loop correction to the photon to quark-antiquark light-front wave-functions*, Phys. Rev. D **94** (2016) 054016. arXiv:1606.00777.

[2] G. Beuf, T. Lappi and R. Paatelainen, *Massive quarks in NLO dipole factorization for DIS: Longitudinal photon*, Phys. Rev. D **104** (2021) 056032. arXiv:2103.14549.

[3] I. Balitsky, *Operator expansion for high-energy scattering*, Nucl. Phys. B **463** (1996) 99–160. arXiv:hep-ph/9509348.

[4] R. Pasechnik and M. Taševský, *Multi-dimensional hadron structure through the lens of gluon Wigner distribution*, Physics Reports **1084** (2024) 3–57. arXiv:2310.10793.

[5] K. Golec-Biernat and M. Wüsthoff, *Saturation Effects in Deep Inelastic Scattering at low Q^2 and its Implications on Diffraction*, Phys. Rev. D **59** (1999) 014017. arXiv:hep-ph/9807513.

[6] H. Kowalski and D. Teaney, *An Impact Parameter Dipole Saturation Model*, Phys. Rev. D **68** (2003) 114005. arXiv:hep-ph/0304189.

[7] F. Salazar and B. Schenke, *Diffractive dijet production in impact parameter dependent saturation models*, Phys. Rev. D **100** (2019) 034007. arXiv:1905.03763.

[8] M. L. Good and W. D. Walker, *Diffraction Dissociation of Beam Particles*, Phys. Rev. **120** (1960) 1857. doi:10.1103/PhysRev.120.1857.

[9] J. C. Collins, *Proof of Factorization for Diffractive Hard Scattering*, Phys. Rev. D **57** (1998) 3051; erratum **61** (2000) 019902. arXiv:hep-ph/9709499.

[10] J. C. Collins, L. Frankfurt and M. Strikman, *Factorization for hard exclusive electroproduction of mesons in QCD*, Phys. Rev. D **56** (1997) 2982. arXiv:hep-ph/9611433.

[11] Y. V. Kovchegov, *Small-x F_2 structure function of a nucleus including multiple pomeron exchanges*, Phys. Rev. D **60** (1999) 034008. arXiv:hep-ph/9901281.

[12] I. Balitsky, *Quark contribution to the small-x evolution of color dipole*, Phys. Rev. D **75** (2007) 014001. arXiv:hep-ph/0609105.

[13] E. Iancu, J. D. Madrigal, A. H. Mueller, G. Soyez and D. N. Triantafyllopoulos, *Resumming double logarithms in the QCD evolution of color dipoles*, Phys. Lett. B **744** (2015) 293. arXiv:1502.05642.

[14] A. Petrelli, M. Cacciari, M. Greco, F. Maltoni and M. L. Mangano, *NLO Production and Decay of Quarkonium*, Nucl. Phys. B **514** (1998) 245–309. arXiv:hep-ph/9707223.

[15] M. Krelina, J. Nemchik, R. Pasechnik and J. Cepila, *Spin rotation effects in diffractive electroproduction of heavy quarkonia*, Eur. Phys. J. C **79** (2019) 154. arXiv:1812.03001.

[16] J. Cepila, J. Nemchik, M. Krelina and R. Pasechnik, *Theoretical uncertainties in exclusive electroproduction S-wave heavy quarkonia*, Eur. Phys. J. C **79** (2019) 495. arXiv:1901.02664.

[17] Z. L. Cui, M. C. Hu and J. P. Ma, *Gluon GPDs and Exclusive Photoproduction of a Quarkonium in Forward Region*, Eur. Phys. J. C **79** (2019) 812. arXiv:1804.05293; published article.

[18] C.-W. Hwang, *Leading-twist light cone distribution amplitudes for p-wave heavy quarkonium states*, JHEP **10** (2009) 074. arXiv:0906.4412.

[19] Y. Hatta, E. Iancu, K. Itakura and L. McLerran, *Odderon in the Color Glass Condensate*, Nucl. Phys. A **760** (2005) 172–207. arXiv:hep-ph/0501171.

[20] S. Benić, D. Horvatić, A. Kaushik and E. Vivoda, *Exclusive η_c production from small-x evolved Odderon at an electron-ion collider*, Phys. Rev. D **108** (2023) 074005. arXiv:2306.10626.

[21] V. S. Fadin and L. N. Lipatov, *BFKL pomeron in the next-to-leading approximation*, Phys. Lett. B **429** (1998) 127–134. arXiv:hep-ph/9802290. The leading-order characteristic function quoted in the text is the leading term of the kernel eigenvalue.

[22] J. Bartels, L. N. Lipatov and G. P. Vacca, *A New Odderon Solution in Perturbative QCD*, Phys. Lett. B **477** (2000) 178–186. arXiv:hep-ph/9912423.

[23] V. M. Abazov et al. (D0 and TOTEM Collaborations), *Odderon Exchange from Elastic Scattering Differences between pp and $p\bar{p}$ Data at 1.96 TeV and from pp Forward Scattering Measurements*, Phys. Rev. Lett. **127** (2021) 062003. arXiv:2012.03981; published article.

[24] T. Csörgő, T. Novák, R. Pasechnik, A. Ster and I. Szanyi, *Model-independent Odderon results based on TOTEM data on elastic proton-proton scattering at 8 TeV*, arXiv:2302.04930 (2023).

[25] S. J. Brodsky, J. Rathsman and C. Merino, *Odderon-Pomeron Interference*, Phys. Lett. B **461** (1999) 114–122. arXiv:hep-ph/9904280.

[26] A. Bzdak, L. Motyka, L. Szymanowski and J.-R. Cudell, *Exclusive J/ψ and Υ hadroproduction and the QCD odderon*, Phys. Rev. D **75** (2007) 094023. arXiv:hep-ph/0702134.

[27] H. Mäntysaari and B. Schenke, *Evidence of strong proton shape fluctuations from incoherent diffraction*, Phys. Rev. Lett. **117** (2016) 052301. arXiv:1603.04349.

[28] F. Dominguez, C. Marquet, B.-W. Xiao and F. Yuan, *Universality of Unintegrated Gluon Distributions at small x*, Phys. Rev. D **83** (2011) 105005. arXiv:1101.0715.

[29] L. McLerran and R. Venugopalan, *Computing Quark and Gluon Distribution Functions for Very Large Nuclei*, Phys. Rev. D **49** (1994) 2233. arXiv:hep-ph/9309289.

[30] J. Bartels, K. Golec-Biernat and H. Kowalski, *A modification of the saturation model: DGLAP evolution*, Phys. Rev. D **66** (2002) 014001. arXiv:hep-ph/0203258.

[31] C. Henkels, E. G. de Oliveira, R. Pasechnik and H. Trebien, *Exclusive photoproduction of excited quarkonia in ultraperipheral collisions*, Phys. Rev. D **102** (2020) 014024. arXiv:2004.00607.

[32] V. P. Gonçalves, B. Z. Kopeliovich, J. Nemchik, R. Pasechnik and I. K. Potashnikova, *Heavy flavor production in high-energy pp collisions: color dipole description*, Phys. Rev. D **96** (2017) 014010. arXiv:1704.04699.

[33] B. Z. Kopeliovich, R. Pasechnik and I. K. Potashnikova, *Diffractive dijet production: breakdown of factorization*, Phys. Rev. D **98** (2018) 114021. arXiv:1807.05548.

The Dirac sea in a strong field

The Dirac sea in a strong field

Pair creation and the information in a divergent field expansion

The question, the model, and the result

One question, two calculational routes

How can a filled lower band develop excitations in a maintained electric field, and how can a divergent weak-field expansion encode an exponentially small transition?

We reproduce a known, solvable benchmark. It teaches tools for nonperturbative physics and the limits of a material analogy; it does not claim a new HgMnTe prediction.

The material motivates the model; it does not equal the model

A restricted, symmetric two-band approximation has

$$h_{\mathrm{eff}}(p) = \begin{pmatrix} \Delta & vp \\ vp & -\Delta \end{pmatrix}, \qquad E_{\pm}(p) = \pm\sqrt{\Delta^2 + v^2 p^2}, \qquad \Delta = E_g/2 > 0.$$

Section 1.1 derives this dispersion and defines its two components. Its resemblance to a massive Dirac dispersion motivates the workshop.

Retained in our benchmark	Needed for actual HgMnTe structures
A gapped avoided crossing; ideal isotropy and electron–hole symmetry; independent modes; a specified degeneracy	Additional Kane bands, Mn exchange, strain and geometry, occupations, defects, relaxation and the actual field profile

Bulk Kane fermions include heavy-hole states and need not have the Dirac magnetic spectrum. HgMnTe wells require multiband modelling with Mn exchange. Normal band ordering and numerical parameters are assumptions here, not measured properties of a particular experimental sample. See Teppe et al. (2016), Novik et al. (2005), and Linder et al. (2018) in the references.

Model contract. One ideal Dirac species in three spatial dimensions (two physical spin states), initially filled negative-energy states and empty positive-energy states. The classical external field is prescribed. We use a long, spatially uniform constant-field plateau, neglect switching transients, boundaries, finite-density blocking, interactions between excitations and their backreaction on the field. The source maintaining the field supplies the energy.

Meaning of exact. Exact dependence on the background field *within this model*: equivalently, one matter loop. This is distinct from all-loop interacting QED. A spatially bounded static field additionally needs sufficient potential drop to span the gap; a finite pulse has its own duration scale.

Charge, units, and the two observables

The charge q is signed and $e \equiv |q| > 0$. With $\hbar = v = 1$,

$$H_q = \boldsymbol{\alpha} \cdot (\mathbf{p} - q\mathbf{A}) + \beta m + q\phi, \qquad m \equiv \Delta, \qquad F \equiv |qE| > 0.$$

For the displayed algebra choose the representative $q = +e$ and fields $E > 0$ or $B > 0$. The electron has $q = -e$: force directions and magnetic spin labels reverse, but the probabilities and spin-summed rates below are unchanged. Thus $F = eE$ in our calculation. Restoring SI units means $m \mapsto \Delta$ and putting back $\hbar$ and v.

$$w_{\text{pairs}} \equiv \frac{\langle N_{\text{pairs}} \rangle}{VT} \qquad \text{mean number produced per volume and time,}$$
$$\Gamma_{\text{vac}} \equiv -\frac{\ln P_{\text{vac}}}{VT} \qquad \text{decay of the zero-pair probability.}$$

Here V is the quantisation volume and T the long field duration. These are different observables. In both cases the natural-unit dimension is energy^4.

The endpoint: derive these, then interpret their limitations

For the long homogeneous field,

$$w_{\text{pairs}} = \frac{F^2}{4\pi^3} e^{-\pi m^2/F},$$
$$\Gamma_{\text{vac}} = \frac{F^2}{4\pi^3} \sum_{k=1}^{\infty} \frac{e^{-k\pi m^2/F}}{k^2} = 2\,\mathrm{Im}\,\mathcal{L}(E).$$

Route I: solve/count independent two-state transitions and take a many-body probability product.
Route II: sum magnetic energy levels, specify subtractions, then continue the effective action to an electric field.

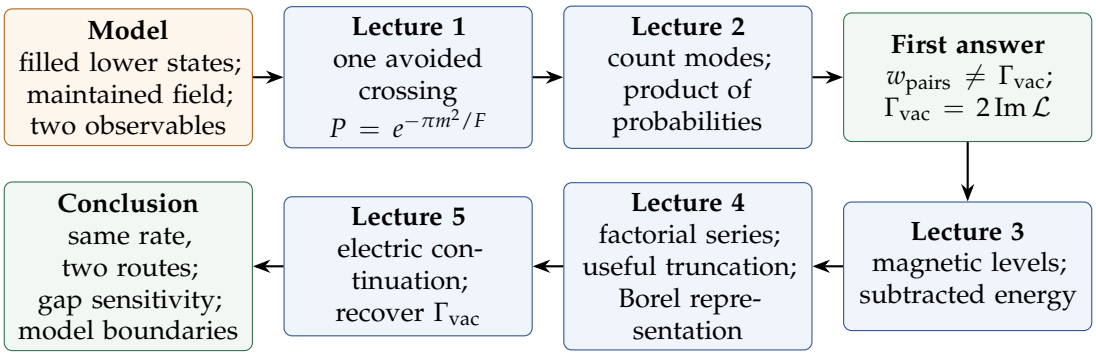
The weak-field coefficients grow factorially. Their dispersion integral reproduces the electric decay information, with a physical boundary prescription. The final material exercise quantifies sensitivity to the gap, without treating these rates as measured device currents.

Scale check (illustrative parameters).

	QED electron	ideal narrow-gap model
Rest energy / half gap Δ	511 keV	$E_g/2 = 50$ meV
Velocity v	$c = 3.0 \times 10^8$ m/s	10^6 m/s
$E_c = \Delta^2/(e\hbar v)$	1.3×10^{18} V/m	3.8×10^6 V/m

The latter is 38 kV/cm. This is a characteristic exponential scale, not a sharp threshold: the zero-transverse-momentum probability at $E = E_c$ is $e^{-\pi} = 0.0432$. It is nonzero below E_c in the infinite homogeneous model.

Roadmap of the calculation



Core calculation and supporting derivations

Core: follow the sequence of calculations and explain each change of description, not merely the endpoint.
Supplied mathematical inputs: exact Landau–Zener connection formula; the sinh product identity; the dispersion relation with its stated analytic assumptions. We derive their consequences explicitly.
Supporting derivations: full Clifford verification; the LZ component/Weber reduction; running-coupling outlook; the keyhole-contour proof; scalar-QED extensions. These provide additional detail for independent study.

PART I · FROM QUANTUM MECHANICS TO MANY MODES

one occupied lower state, then the filled sea

Lecture 1 One avoided crossing in a field

Section 1.1 Two amplitudes, two bands, and a hole

Goal: derive a gapped two-state spectrum before introducing the Dirac spin structure.

Choose two orthonormal band-edge basis states $|c\rangle, |v\rangle$. At fixed one-dimensional momentum p,

$$|\psi(p)\rangle = a(p)|c\rangle + b(p)|v\rangle, \qquad \psi(p) = \begin{pmatrix} a(p) \\ b(p) \end{pmatrix}, \qquad |a|^2 + |b|^2 = 1. \tag{1.1}$$

These are two amplitudes for *one* particle, not two particles. Choose the mean band-edge energy as zero. Retain a gap 2Δ and the leading interband coupling linear in p:

$$h_{\text{eff}}(p) = \begin{pmatrix} \Delta & vp \\ vp & -\Delta \end{pmatrix} = vp\,\sigma_x + \Delta\,\sigma_z, \tag{1.2}$$

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Here v is a velocity, not the band label v in $|v\rangle$. In a restricted $\mathbf{k}\cdot\mathbf{p}$ reduction along x, with envelope momentum $p = \hbar k_x$,

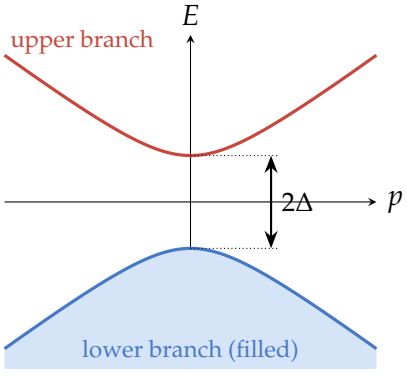
$$\langle c|\delta H|v\rangle = p\langle c|\hat{v}_x|v\rangle \equiv vp.$$

Choose the relative phase of the basis states so the interband velocity matrix element is real. For the simplest microscopic kinetic Hamiltonian $\hat{v}_x = \hat{p}_x/m_0$, with m_0 the bare electron mass; a more general microscopic model defines its own velocity operator. We have omitted diagonal p^2 terms and other bands: this is a *model choice*.

$$\begin{aligned} 0 &= \det(h_{\text{eff}} - E\mathbb{1}) = (\Delta - E)(-\Delta - E) - v^2p^2 \\ &= E^2 - \Delta^2 - v^2p^2 \quad \Longrightarrow \quad E_\pm(p) = \pm\varepsilon(p), \quad \varepsilon(p) = \sqrt{\Delta^2 + v^2p^2}, \end{aligned} \tag{1.3}$$

$$E_+(p) = \Delta + \frac{v^2p^2}{2\Delta} + \cdots = \Delta + \frac{p^2}{2m_{\text{eff}}} + \cdots, \qquad m_{\text{eff}} = \frac{\Delta}{v^2}.$$

The Pauli matrices here act on the **band pseudospin** (basis amplitudes); they are not yet the physical spin of an electron.



Occupation, not negative physical particle energy. Initially every lower-energy one-particle state in our model is occupied and every upper state empty. An excitation at momentum p removes one lower-state particle and puts one in the upper state:

$$\begin{aligned} (n_-, n_+) &: (1,0) \longrightarrow (0,1), \\ \Delta E &= E_+(p) - E_-(p) = 2\varepsilon(p) \geq 2\Delta. \end{aligned}$$

The missing lower-state particle is a *hole*, with positive excitation energy. This is the useful filled-sea language; all measured excitation energies are relative to the filled reference state.

Two basis amplitudes give a gapped spectrum; occupations turn a transition into a particle–hole pair. From now on $\hbar = v = 1$ and $m = \Delta$.

The three-dimensional Dirac benchmark and its two copies. For later transverse motion and magnetic fields, use

$$H = \boldsymbol{\alpha} \cdot \mathbf{p} + \beta m, \tag{1.4}$$

$$\alpha_i = \tau_x \otimes \sigma_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}, \qquad \beta = \tau_z \otimes \mathbb{1} = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}. \tag{1.5}$$

The 2×2 identity is understood in each block. Here σ_i act on physical spin and τ_i on upper/lower Dirac components. Thus this σ has a *different role* from the band pseudospin in Eq. (1.2). For $\mathbf{p} = (p, 0, 0)$, $\mathbb{1} \otimes \sigma_x$ commutes with H. In its eigenbasis $s = \pm 1$,

$$H\big|_s = sp\,\tau_x + m\tau_z, \qquad \tau_z(-p\tau_x + m\tau_z)\tau_z = p\tau_x + m\tau_z.$$

The two spin sectors therefore have equivalent dynamics. Renaming the two-component matrices in either sector gives

$$h(p) = p\sigma_x + m\sigma_z, \qquad h(p)^2 = (p^2 + m^2)\mathbb{1}. \tag{1.6}$$

This supplies the factor of two in the later state count. It does not supply the actual magnetic degeneracies of multiband HgMnTe.

Supporting algebra. The matrix products obey

$$\alpha_i\alpha_j + \alpha_j\alpha_i = 2\delta_{ij}\mathbb{1}, \qquad \alpha_i\beta + \beta\alpha_i = 0, \qquad \beta^2 = \mathbb{1}. \tag{1.7}$$

$$H^2 = \sum_{ij} \alpha_i\alpha_j p_i p_j + m\sum_i (\alpha_i\beta + \beta\alpha_i)p_i + \beta^2 m^2 = (p^2 + m^2)\mathbb{1},$$

$$E = \pm\varepsilon(\mathbf{p}), \qquad \varepsilon(\mathbf{p}) = \sqrt{p^2 + m^2}. \tag{1.8}$$

The first equality follows by multiplying out H. Since $p_i p_j = p_j p_i$ for the free particle, its antisymmetric matrix part cancels; the mixed term vanishes by anticommutation.

Section 1.2 Tunnelling in space: tilted bands

Goal: derive the exponent and explain why probability has twice the amplitude exponent.

For our representative charge $q = +e$, take a scalar potential $\phi_s(x) = -Ex$ and $\mathbf{A}_s = 0$. Then

$$\mathbf{E} = -\nabla\phi_s = E\hat{x}, \qquad U = q\phi_s = -Fx, \qquad F = eE > 0. \tag{1.9}$$

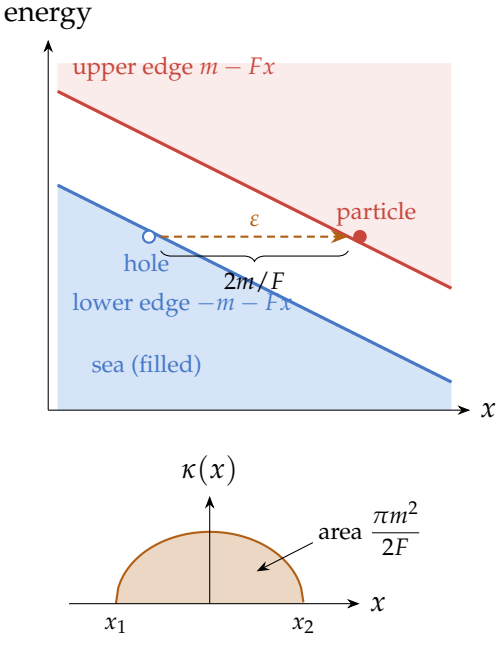
A physical electron accelerates oppositely; the following probability depends on $|qE|$.

At conserved total energy ε,

$$\begin{aligned} \varepsilon &= -Fx \pm \sqrt{m^2 + p(x)^2} \\ p(x)^2 &= (\varepsilon + Fx)^2 - m^2. \end{aligned} \tag{1.10}$$

The band edges are $\pm m - Fx$. The forbidden interval is

$$\begin{aligned} x_1 &= \frac{-m-\varepsilon}{F}, \quad x_2 = \frac{m-\varepsilon}{F}, \\ x_2 - x_1 &= \frac{2m}{F}, \qquad p = i\kappa, \\ \kappa(x) &= \sqrt{m^2 - (\varepsilon + Fx)^2}. \end{aligned} \tag{1.11}$$



The leading WKB wave amplitude follows the local momentum:

$$\begin{aligned} \psi(x) &\sim \exp\left(\frac{i}{\hbar}\int^x p(x')\,dx'\right) \xrightarrow{p=i\kappa} \exp\left(-\frac{1}{\hbar}\int^x \kappa(x')\,dx'\right), \\ P &\sim |\text{transmission amplitude}|^2 \simeq \exp\left(-2\int_{x_1}^{x_2}\kappa(x)\,dx\right) \qquad (\hbar = 1). \end{aligned} \tag{1.12}$$

WKB fixes the leading exponential; matching wave amplitudes is needed for a general prefactor. Set $y = \varepsilon + Fx$ and then $y = m\sin\theta$:

$$\begin{aligned} 2\int_{x_1}^{x_2}\kappa\,dx &= \frac{2}{F}\int_{-m}^{m}\sqrt{m^2-y^2}\,dy, \\ &= \frac{2m^2}{F}\int_{-\pi/2}^{\pi/2}\cos^2\theta\,d\theta, \\ &= \frac{2m^2}{F}\left[\frac{\theta}{2} + \frac{\sin 2\theta}{4}\right]_{-\pi/2}^{\pi/2} = \frac{\pi m^2}{F}. \end{aligned} \tag{1.13}$$

$P \simeq e^{-\pi m^2/F} = e^{-\pi E_c/E}$; the exponent is independent of the conserved ε.

Where the field must act. In SI units the forbidden width is $\ell_t = 2\Delta/(eE)$. A bounded *static* field region needs a potential drop spanning the gap, $e|\Delta\phi| \geq 2\Delta$, to connect occupied and empty asymptotic bands. A nearly homogeneous treatment further needs the turning points to lie in that region. This restriction is absent for the infinite uniform field.

Self-study task T1

Reproduce the integral in Eq. (1.13) with $y = m\sin\theta$, including the transformed integration limits.
Checkpoint. Why is the exponent $2\int \kappa\,\mathrm{d}x$, rather than $\int \kappa\,\mathrm{d}x$? Locate the two turning points on the band diagram.

Section 1.3 Tunnelling in time: gauge change and Landau–Zener

Goal: map the same field to a standard two-level crossing and identify the supplied exact result.

Gauge change, including the wavefunction. For a signed charge q and a real function $\chi(x,t)$,

$$\mathbf{A}' = \mathbf{A} + \nabla\chi, \qquad \phi' = \phi - \partial_t\chi, \qquad \psi' = e^{iq\chi/\hbar}\psi. \tag{1.14}$$

These transformations preserve the Schrödinger equation because

$$(-i\hbar\nabla - q\mathbf{A}')\psi' = e^{iq\chi/\hbar}(-i\hbar\nabla - q\mathbf{A})\psi,$$
$$i\hbar\partial_t\psi' = e^{iq\chi/\hbar}(i\hbar\partial_t - q\partial_t\chi)\psi.$$

Starting with $\phi_s = -Ex, \mathbf{A}_s = 0$, choose $\chi = -Ext$:

$$\phi_t = -Ex - (-Ex) = 0, \qquad A_{t,x} = -Et, \qquad \psi_t = e^{-iqExt/\hbar}\psi_s,$$
$$E_x = -\partial_x\phi_t - \partial_t A_{t,x} = E.$$

Probability densities are unchanged because the multiplying phase has unit modulus. For $q = +e$ the Hamiltonian in this gauge is

$$H(t) = \alpha_x(p - eA_x) + \beta m = \alpha_x(p + Ft) + \beta m, \tag{1.15}$$
$$\pi(t) = p + Ft, \qquad \dot{\pi} = F. \tag{1.16}$$

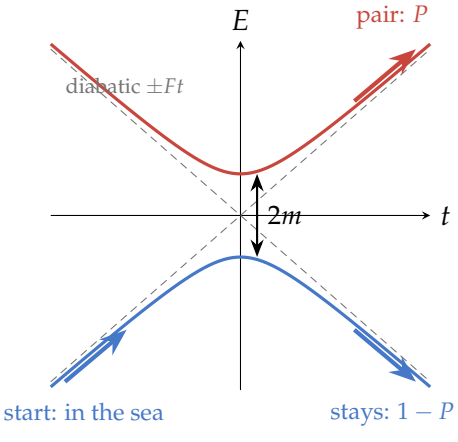
p is the conserved *canonical* momentum (no x in H); π is the changing *kinetic* momentum. Set $\tau = t + p/F$ so the crossing occurs at $\tau = 0$, then rename τ as t in each mode:

$$h(t) = Ft\sigma_x + m\sigma_z. \tag{1.17}$$

A fixed basis change makes the crossing explicit.

$$R = \frac{\sigma_x + \sigma_z}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad R^\dagger R = \mathbb{1},$$
$$R\sigma_x R^\dagger = \sigma_z, \quad R\sigma_z R^\dagger = \sigma_x, \qquad \widetilde{\psi} = R\psi,$$
$$i\partial_t\widetilde{\psi} = \widetilde{h}(t)\widetilde{\psi}, \qquad \widetilde{h} = RhR^\dagger = \begin{pmatrix} Ft & m \\ m & -Ft \end{pmatrix}. \tag{1.18}$$

R is constant, so it generates no extra time-derivative term. The *diabatic* basis is this fixed two-state basis: its uncoupled diagonal energies $\pm Ft$ cross. The *adiabatic* basis consists of the instantaneous eigenvectors of the full $\widetilde{h}$; its energies $\pm\sqrt{F^2t^2 + m^2}$ remain separated.



Supplied theorem: the exact Landau–Zener connection formula

For $\widetilde{h} = \left(\begin{smallmatrix} at & g \\ g & -at \end{smallmatrix}\right)$, $a > 0$, start in the lower instantaneous-energy state at $t \to -\infty$. The probability to end in the *upper* instantaneous-energy state at $t \to +\infty$ is

$$P = \exp\left[-\frac{2\pi g^2}{\hbar\,|\mathrm{d}(at-(-at))/\mathrm{d}t|}\right] \xrightarrow{g=m,\ a=F,\ \hbar=1} \exp\left[-\frac{2\pi m^2}{2F}\right] = e^{-\pi m^2/F}. \tag{1.19}$$

This is an exact infinite-time result for a linear sweep, not a general WKB prefactor rule. Its special-function connection formula is supplied; supporting notes reduce the coupled equations to the Weber equation. The numerical task checks it independently with finite-time convergence.

At large negative time the lower state approaches the first fixed basis vector. At large positive time that same fixed vector is the upper state. Hence a nonadiabatic transition between energy branches is survival in the corresponding diabatic label; confusing these labels would replace P by $1-P$. For $m = F = 1$, the theorem gives $P = e^{-\pi} = 0.0432139183$.

Section 1.4 Beyond all weak-field orders: the checkpoint

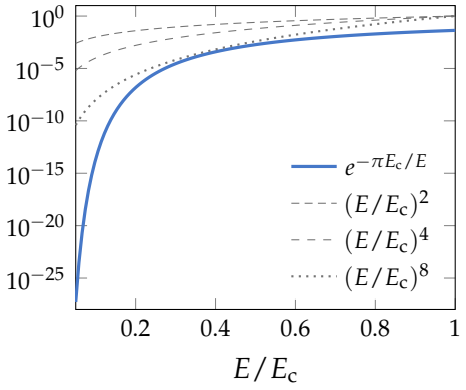
Goal: distinguish field expansion from loop expansion.

For positive field magnitude E let $a = \pi E_c > 0$:

$$P(E) = e^{-a/E}, \tag{1.20}$$

$$\frac{\mathrm{d}}{\mathrm{d}E}e^{-a/E} = \frac{a}{E^2}e^{-a/E},$$

$$\frac{\mathrm{d}^n P}{\mathrm{d}E^n} = \mathrm{poly}(1/E)e^{-a/E} \to 0 \quad (E \to 0^+).$$



For example, with $y = a/E$,

$$E^{-N}e^{-a/E} = a^{-N}y^N e^{-y} \to 0.$$

Extending by $P(0) = 0$, every right derivative vanishes, so its formal weak-field series is

$$0 + 0E + 0E^2 + \cdots. \tag{1.21}$$

The physical probability for reversed field uses $|E|$; it does not blow up. The complex function $e^{-a/z}$ has an essential singularity at $z = 0$, and no convergent power series there represents it.

Checkpoint before counting the whole sea

A finite weak-field power expansion misses $e^{-\pi E_c/E}$. This does *not* mean that the effect is absent at every loop order: our later effective action contains it already at one loop, summed to all orders in the classical background. Nonanalyticity alone also does not prove that another asymptotic series diverges; Lecture 4 will exhibit its factorial coefficients.

Lecture 2 The whole Dirac sea

Section 2.1 Motion in all directions

Goal: reduce a three-dimensional mode to the two-level problem already solved.

The representative charge is $q = +e$, with $eE = F > 0$. For $\mathbf{p} = (p_x, \mathbf{p}_\perp)$ the field changes $\pi_x = p_x + Ft$ but conserves $\mathbf{p}_\perp$:

$$H(t) = \alpha_x \pi_x(t) + \boldsymbol{\alpha}_\perp \cdot \mathbf{p}_\perp + \beta m = \alpha_x \pi_x(t) + m_\perp \beta', \tag{2.1}$$

$$m_\perp^2 = m^2 + p_\perp^2, \qquad \beta' = \frac{\boldsymbol{\alpha}_\perp \cdot \mathbf{p}_\perp + \beta m}{m_\perp}. \tag{2.2}$$

The Clifford algebra gives, with all cross terms cancelling,

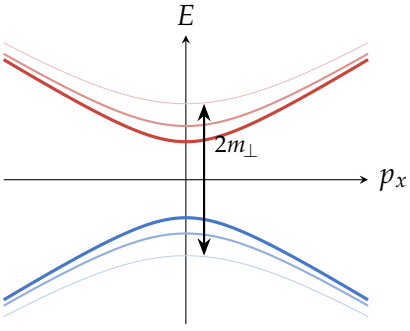
$$(\boldsymbol{\alpha}_\perp \cdot \mathbf{p}_\perp + \beta m)^2 = p_\perp^2 + m^2,$$
$$\beta'^2 = 1, \qquad \alpha_x \beta' + \beta' \alpha_x = 0. \tag{2.3}$$

Why this is again two identical two-level systems. Choose two orthonormal eigenvectors $|+, a\rangle$ of β' with eigenvalue $+1$, $a = 1, 2$. Then $|-, a\rangle = \alpha_x |+, a\rangle$ obeys

$$\beta' |-, a\rangle = -|-, a\rangle, \qquad \alpha_x |-, a\rangle = |+, a\rangle.$$

In each ordered basis $(|+, a\rangle, |-, a\rangle)$, $\beta' = \sigma_z$ and $\alpha_x = \sigma_x$. After shifting the crossing time,

$$h_a(t) = Ft\,\sigma_x + m_\perp \sigma_z \quad \Longrightarrow \quad \boxed{P(p_\perp) = \exp\left[-\frac{\pi(m^2 + p_\perp^2)}{F}\right]}. \tag{2.4}$$



At $m = 1$, $p_\perp = 0, 1, \sqrt{3}$ from dark to light: transverse motion increases the gap.

Checkpoint: what has changed?

Does transverse momentum create a new tunnelling mechanism, or change the gap in the old one?

Section 2.2 Occupations and the mean number of pairs

Goal: explain what a pair is and compute its mean production rate.

One pair sector j. The label j specifies one momentum channel and one of its two spin channels. Use the positive/negative-energy modes appropriate to the incoming or outgoing field-free regions. Each fermion mode has occupation $n = 0$ or 1 (Pauli exclusion). Relative to a filled lower level:

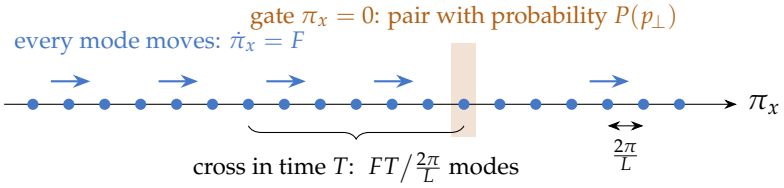
	n_+	n_-	particle number $n_e = n_+$	hole number $n_h = 1 - n_-$
no pair	0	1	0	0
one pair	1	0	1	1

Promoting the lower-level electron conserves electron number in the two-level description and creates an *particle–hole excitation* relative to the sea. The hole's charge is $-q$. For the physical electron field $q = -e$, it is the positron; our representative $q = +e$ convention interchanges the charge labels and leaves the probabilities unchanged. The outgoing state of this sector has the form

$$|\psi_j\rangle = a_j|0 \text{ pairs}\rangle_j + b_j|1 \text{ pair}\rangle_j, \qquad |a_j|^2 + |b_j|^2 = 1, \quad |b_j|^2 = P_j. \tag{2.5}$$

Thus $\langle n_j\rangle = 0|a_j|^2 + 1|b_j|^2 = P_j$. The mean total number is $\langle N\rangle = \sum_j P_j$; a probability is not yet a rate.

Which modes cross during a time T? Use a large periodic box $V = L^3$, so $p_x = 2\pi n/L$. Since $\pi_x = p_x + Ft$, the crossing time is $t_0 = -p_x/F$. Crossings in $-T/2 < t_0 < T/2$ occupy a canonical-momentum interval of width FT.



$$\#p_x = \frac{FT}{2\pi/L} = \frac{FTL}{2\pi}, \qquad \#\mathbf{p}_\perp = L^2\int\frac{\mathrm{d}^2p_\perp}{(2\pi)^2}, \qquad g = 2, \tag{2.6}$$

$$\langle N\rangle = 2\frac{FTL}{2\pi}L^2\int\frac{\mathrm{d}^2p_\perp}{(2\pi)^2}e^{-\pi(m^2+p_\perp^2)/F}. \tag{2.7}$$

This is a *long-plateau rate*: the crossing must lie far from switching boundaries. End effects vanish in $\langle N\rangle/(VT)$ as $T \to \infty$; finite pulses require solving their actual time dependence.

$$\int\mathrm{d}^2p_\perp\, e^{-\pi p_\perp^2/F} = 2\pi\int_0^\infty p\,\mathrm{d}p\, e^{-\pi p^2/F} = F, \tag{2.8}$$

$$w_{\text{pairs}} \equiv \lim_{V,T\to\infty}\frac{\langle N\rangle}{VT} = 2\frac{F}{2\pi}\frac{F}{(2\pi)^2}e^{-\pi m^2/F} = \boxed{\frac{F^2}{4\pi^3}e^{-\pi m^2/F}}. \tag{2.9}$$

For the isotropic two-spin Dirac benchmark, units restored:

$$w_{\text{pairs}} = \frac{(eE)^2}{4\pi^3\hbar^2 v}\exp\left[-\frac{\pi\Delta^2}{eE\hbar v}\right].$$

A crystal calculation must check its actual dispersions, occupation factors and degeneracies before using this prefactor.

Self-study task T2

Do the Gaussian (2.8) in polar coordinates; track both factors of 2π in the transverse state density.

Section 2.3 Vacuum survival is a different observable

Goal: derive the zero-pair probability and distinguish it from the mean number.

Why a product? In the prescribed uniform background, momenta are not mixed and dynamical-photon interactions are absent. The Hamiltonian is a sum of independent pair-sector Hamiltonians. The initially filled sea evolves into a product over these sectors:

$$|\Psi_{\text{out}}\rangle = \bigotimes_j (a_j|0 \text{ pairs}\rangle_j + b_j|1 \text{ pair}\rangle_j),$$

$$P_{\text{vac}} = \left|\prod_j a_j\right|^2 = \prod_j (1 - P_j), \tag{2.10}$$

$$\ln P_{\text{vac}} = \sum_j \ln(1 - P_j) = -\sum_j \sum_{k=1}^{\infty} \frac{P_j^k}{k}. \tag{2.11}$$

The tensor factors here are the even-fermion pair sectors; no exchange-sign assumption changes their probabilities.

Two-mode checkpoint: probability versus counting

If two independent sectors each have $P = 1/2$, then $P(N = 0) = 1/4$, $P(N = 1) = 1/2$, $P(N = 2) = 1/4$. Compute $\langle N \rangle$ and $-\ln P(N = 0)$. Are they equal?

For our modes $P_j^k = e^{-k\pi(m^2+p_\perp^2)/F}$, and

$$\int \mathrm{d}^2 p_\perp \, e^{-k\pi p_\perp^2/F} = F/k, \tag{2.12}$$

$$-\frac{\ln P_{\text{vac}}}{VT} = 2\frac{F}{2\pi} \sum_{k=1}^{\infty} \frac{1}{k} \frac{F}{4\pi^2 k} e^{-k\pi m^2/F}$$

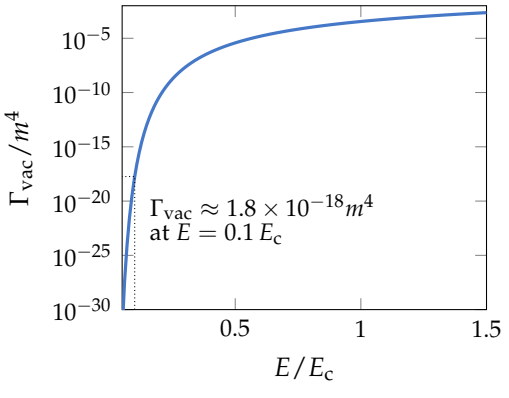
$$\equiv \Gamma_{\text{vac}} = \boxed{\frac{F^2}{4\pi^3} \sum_{k=1}^{\infty} \frac{e^{-k\pi m^2/F}}{k^2}} = \frac{e^2E^2}{4\pi^3} \sum_{k=1}^{\infty} \frac{e^{-k\pi E_c/E}}{k^2}. \tag{2.13}$$

One $1/k$ comes from $-\ln(1 - P)$; the other comes from the two-dimensional Gaussian. The integer k is an index in this logarithmic expansion, *not* the number of pairs in one fermion mode.

The two distinct rates

$$\boxed{\langle N \rangle = VTw_{\text{pairs}}}, \qquad \boxed{P(N = 0) = e^{-VT\Gamma_{\text{vac}}}}, \qquad \Gamma_{\text{vac}} \geq w_{\text{pairs}}.$$

Only when all relevant $P_j \ll 1$ does $-\ln(1 - P_j) \simeq P_j$, so that the two rates nearly coincide. The $k = 1$ term of Γ_{vac} equals the exact mean rate in this constant-field model.



Vacuum-decay density. The first three terms are shown; omitted terms are below 0.02% on this range.

Energy source. The excitation energy comes from work done by the external source:

$$\frac{\mathrm{d}U_{\text{matter}}}{\mathrm{d}t} = \int \mathrm{d}^3x\, \mathbf{j}\cdot\mathbf{E}. \tag{2.14}$$

Maintaining the prescribed field supplies this energy. If the field is dynamical, its depletion and the produced current must also be solved for (backreaction); neither is included in the benchmark rate.

Optional comparison: bosons. For a charged scalar the same constant-field mode solution gives a *mean occupation* $P = e^{-\pi m_\perp^2/F}$. A bosonic pair sector can contain $0, 1, 2, \ldots$ pairs and has a geometric distribution with zero-pair probability $(1+P)^{-1}$. Hence $-\ln(1+P)$ replaces $\ln(1-P)$ in $\ln P_{\text{vac}}$; there is no factor of two for spin. Report R2 supplies the further steps.

Section 2.4 One quantity for the phase and the loss of vacuum

Goal: define the effective Lagrangian without presupposing a QFT course.

In and out. Prepare the empty excitation vacuum before the field is applied, and compare with the empty excitation vacuum after it is removed. Their overlap is the amplitude for no pairs. Divide out the field-free vacuum amplitude, thereby fixing an irrelevant reference phase. For the large uniform plateau, define

$$\begin{aligned} Z[A] &\equiv \frac{\langle 0_{\text{out}}|U_A|0_{\text{in}}\rangle}{\langle 0_{\text{out}}|U_0|0_{\text{in}}\rangle} = e^{i\Gamma^{(1)}[A]}, \\ \Gamma^{(1)}[A] &= VT\mathcal{L} + \text{boundary terms}, \qquad \mathcal{L} = \operatorname{Re}\mathcal{L} + i\operatorname{Im}\mathcal{L}. \end{aligned} \tag{2.15}$$

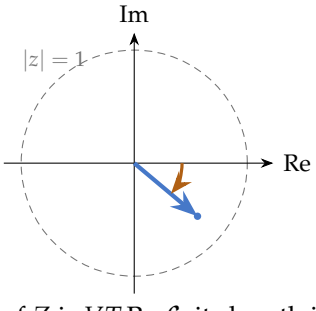
U_A is time evolution in the applied field; $\Gamma^{(1)}$ is its one-loop matter effective action. Here $\mathcal{L}$ denotes the matter correction, not the classical Maxwell term. Boundary terms do not contribute to its long-time bulk density.

$$\begin{aligned} Z[A] &= e^{iVT\operatorname{Re}\mathcal{L}} e^{-VT\operatorname{Im}\mathcal{L}}, \\ |Z[A]|^2 &= e^{-2VT\operatorname{Im}\mathcal{L}} = P_{\text{vac}} = e^{-VT\Gamma_{\text{vac}}} \quad\Longrightarrow\quad \boxed{2\operatorname{Im}\mathcal{L} = \Gamma_{\text{vac}}}. \end{aligned} \tag{2.16}$$

For a *stable static magnetic background*, ordinary quantum mechanics gives

$$Z[B] = e^{-i\Delta E_{\text{vac}}(B)T}, \qquad \boxed{\operatorname{Re}\mathcal{L}(B) = -\frac{\Delta E_{\text{vac}}(B)}{V} \equiv -\Delta\varepsilon_{\text{vac}}(B)}. \tag{2.17}$$

The electric background is unstable; its complex amplitude is not the energy of a stationary ground state.



The angle of Z is $VT\operatorname{Re}\mathcal{L}$; its length is $e^{-VT\operatorname{Im}\mathcal{L}}$.

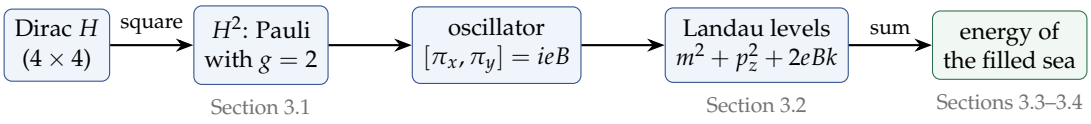
Answer to Q1; the precise bridge to Q2

The ideal constant, spatially extended electric field has a nonzero vacuum-decay density for every $E > 0$. We have computed both the mean pair rate and the vacuum survival probability. The exponentially small decay is invisible in a finite power series in this static field. Next we calculate the real vacuum response in B, expand it, and ask how the resulting coefficients encode the decay in E.

PART II · QUANTUM FIELD THEORY

a stable vacuum response whose analytic continuation reproduces the decay

Lecture 3 The energy of the Dirac sea in a magnetic field



Section 3.1 Squaring the Dirac Hamiltonian

Goal: derive the magnetic spectrum of the ideal isotropic Dirac benchmark.

For the representative charge $q=+e$ and $B>0$,

$$H=\boldsymbol{\alpha}\cdot\boldsymbol{\pi}+\beta m,\qquad \boldsymbol{\pi}=\mathbf{p}-e\mathbf{A},\qquad \mathbf{B}=B\hat{z},\qquad \mathbf{p}=-i\boldsymbol{\nabla}.$$

This benchmark is not a complete HgMnTe Landau spectrum. For a physical electron $q=-e$, the spin labels reverse; the spectrum and degeneracy depend on $|qB|$.

$$H=\begin{pmatrix} m & \boldsymbol{\sigma}\cdot\boldsymbol{\pi} \\ \boldsymbol{\sigma}\cdot\boldsymbol{\pi} & -m \end{pmatrix},$$
$$H^2=\begin{pmatrix} (\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2+m^2 & 0 \\ 0 & (\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2+m^2 \end{pmatrix}. \tag{3.1}$$

The off-diagonal block vanishes because m is constant: $m\boldsymbol{\sigma}\cdot\boldsymbol{\pi}-\boldsymbol{\sigma}\cdot\boldsymbol{\pi}\,m=0$. Now use $\sigma_i\sigma_j=\delta_{ij}\mathbb{1}+i\epsilon_{ijk}\sigma_k$, keeping operator order:

$$\begin{aligned}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2&=\sum_{ij}(\delta_{ij}+i\epsilon_{ijk}\sigma_k)\pi_i\pi_j\\ &=\boldsymbol{\pi}^2+i\sigma_z(\pi_x\pi_y-\pi_y\pi_x) &(3.2)\\ [\pi_x,\pi_y]&=-e[p_x,A_y]-e[A_x,p_y]\\ &=ie(\partial_xA_y-\partial_yA_x)=ieB. &(3.3)\end{aligned}$$

Thus

$$\boxed{H^2=m^2+p_z^2+\pi_x^2+\pi_y^2-eB\Sigma_z},\qquad \Sigma_z=\begin{pmatrix}\sigma_z & 0\\ 0 & \sigma_z\end{pmatrix}. \tag{3.4}$$

The $-eB\Sigma_z$ term is the Dirac spin coupling ($g=2$), produced by the noncommuting kinetic momenta. It was not inserted as an independent phenomenological Zeeman term.

Self-study task T3

Check (3.3) in the gauge $\mathbf{A}=(0,Bx,0)$, by acting on a test wavefunction.

Section 3.2 Relativistic Landau levels and state counting

Goal: obtain both the eigenvalues and their correct multiplicities.

Define the oscillator ladder operators

$$a = \frac{\pi_x + i\pi_y}{\sqrt{2eB}}, \qquad a^\dagger = \frac{\pi_x - i\pi_y}{\sqrt{2eB}},$$

$$[a, a^\dagger] = \frac{-2i[\pi_x, \pi_y]}{2eB} = 1, \tag{3.5}$$

$$a^\dagger a = \frac{\pi_x^2 + \pi_y^2 + i[\pi_x, \pi_y]}{2eB} = \frac{\pi_x^2 + \pi_y^2 - eB}{2eB},$$

$$\pi_x^2 + \pi_y^2 = eB(2a^\dagger a + 1). \tag{3.6}$$

On an oscillator state $n = 0, 1, \ldots$ and a spin state $s_z = \pm 1$,

$$E^2 = m^2 + p_z^2 + eB(2n + 1 - s_z) = \boxed{m^2 + p_z^2 + 2eBk}, \qquad k = 0, 1, \ldots. \tag{3.7}$$

k	(n, s_z) first possibility	second possibility	multiplicity g_k per energy sign
0	$(0, +1)$	none	1
$k \geq 1$	$(k, +1)$	$(k-1, -1)$	2

Why the upper/lower Dirac blocks do not give another factor of two. H^2 contains two copies of the Pauli operator, but diagonalising H splits them between positive and negative energies. For $k \geq 1$, each sign has two states. In the $k = 0$ spin-up oscillator subspace,

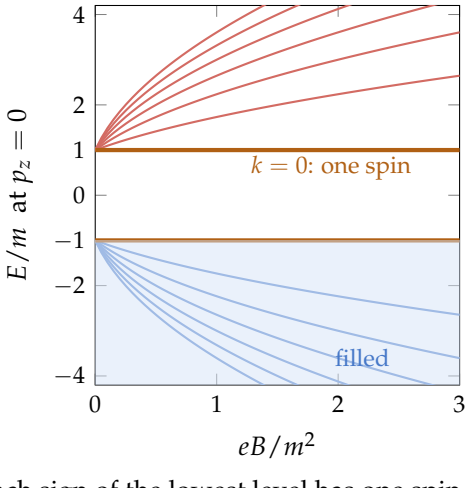
$$H|_{k=0} = \begin{pmatrix} m & p_z \\ p_z & -m \end{pmatrix},$$

with one eigenstate at $+\sqrt{m^2 + p_z^2}$ and one at $-\sqrt{m^2 + p_z^2}$. This explains $g_0 = 1$ without double counting the Dirac components.

The separate orbital degeneracy. In $A_y = Bx$, $\pi_y = p_y - eBx$ and the oscillator centre is $x_0 = p_y/(eB)$. Periodic y gives $p_y = 2\pi n/L_y$; centres within the sample satisfy

$$0 \leq x_0 < L_x \quad \Longrightarrow \quad N_{\text{centres}} = \frac{eBL_xL_y}{2\pi}, \qquad \frac{N_{\text{centres}}}{L_xL_y} = \frac{eB}{2\pi}. \tag{3.8}$$

Boundary corrections vanish in the large-area limit.



Dirac benchmark, $p_z = 0$. Each sign of the lowest level has one spin state; higher levels have two.

Checkpoint: three independent counts

For a negative-energy level, what factors count (i) orbit centres, (ii) longitudinal momenta, and (iii) spin/oscillator multiplicity?

Section 3.3 A regulated energy difference: proper time

Goal: sum the filled levels while keeping the energy reference and regulator explicit.

The formal sea energy per volume is

$$\varepsilon(B) = -\frac{eB}{2\pi}\sum_{k\geq 0} g_k \int \frac{\mathrm{d}p_z}{2\pi}\sqrt{m^2+p_z^2+2eBk}. \tag{3.9}$$

It diverges. We do not assign a measured value to this absolute sum. The relevant quantity is the change from zero field, regulated in the *same* way in both backgrounds.

$$-\sum_{\text{states}}\sqrt{A},\ A=m^2+p_z^2+2eBk \xrightarrow{\text{(i)}} \int\frac{\mathrm{d}s}{s^{3/2}}\sum_{\text{states}} e^{-sA} \xrightarrow{\text{(ii)}} \frac{e^{-m^2 s}}{2\sqrt{\pi s}}\sum_k g_k e^{-2eBks} \xrightarrow{\text{(iii)}} \coth(eBs)$$

Step 1: turn the square root into an exponential. For $A>0$, let $I(A)=\int_0^\infty \mathrm{d}s\, s^{-3/2}(e^{-sA}-1)$. Integration by parts gives

$$\begin{aligned} I(A) &= \left[-\frac{2(e^{-sA}-1)}{\sqrt{s}}\right]_0^\infty - 2A\int_0^\infty \frac{\mathrm{d}s}{\sqrt{s}}e^{-sA} \\ &= -2A\sqrt{\frac{\pi}{A}} = -2\sqrt{\pi}\sqrt{A}, \\ \sqrt{A} &= -\frac{1}{2\sqrt{\pi}}\int_0^\infty \frac{\mathrm{d}s}{s^{3/2}}(e^{-sA}-1). \end{aligned} \tag{3.10}$$

At $s\to 0$, $e^{-sA}-1\simeq -As$; at $s\to\infty$, $(e^{-sA}-1)/\sqrt{s}\to 0$. These are the two boundary terms. The auxiliary variable s has units of energy^{-2}; it is not laboratory time.

Step 2: sum the exponentials, including every degeneracy. Define the heat-kernel density of one energy-sign branch,

$$K_B(s) = \frac{eB}{2\pi}\sum_{k\geq 0} g_k \int \frac{\mathrm{d}p_z}{2\pi} e^{-s(m^2+p_z^2+2eBk)}. \tag{3.11}$$

Equivalently $K_B = (2V)^{-1}\,\mathrm{Tr}\, e^{-sH_B^2}$; the factor $1/2$ removes the duplicated energy signs. With $q_s = e^{-2eBs}$ (a summation variable, not the charge),

$$\begin{aligned} \int\frac{\mathrm{d}p_z}{2\pi}e^{-sp_z^2} &= \frac{1}{2\sqrt{\pi s}}, \\ \sum_{k\geq 0} g_k e^{-2eBks} &= 1+2\frac{q_s}{1-q_s} = \frac{1+q_s}{1-q_s} = \coth(eBs), \end{aligned} \tag{3.12}$$

$$\begin{aligned} K_B(s) &= \frac{e^{-m^2 s}}{4\pi^{3/2}s^{3/2}}(eBs)\coth(eBs), \\ K_0(s) &= 2\int\frac{\mathrm{d}^3p}{(2\pi)^3}e^{-s(m^2+p^2)} = \frac{e^{-m^2 s}}{4\pi^{3/2}s^{3/2}}. \end{aligned} \tag{3.13}$$

Step 3: subtract first, then regulate the remaining short-time integral. The identity operator in the square-root formula cancels in the difference of traces over the same Hilbert space. Introduce a common, gauge-respecting proper-time cutoff $s_0 = 1/\Lambda^2$:

$$\begin{aligned} \Delta\varepsilon_{s_0}(B) &\equiv \frac{1}{2\sqrt{\pi}}\int_{s_0}^\infty \frac{\mathrm{d}s}{s^{3/2}}[K_B(s)-K_0(s)] \\ &= \boxed{\frac{1}{8\pi^2}\int_{s_0}^\infty \frac{\mathrm{d}s}{s^3}e^{-m^2 s}[(eBs)\coth(eBs)-1]}. \end{aligned} \tag{3.14}$$

This is a *definition of the regulated difference*, not an equality between two unregulated infinities. Large energies are weighted at small s, so Λ is an ultraviolet regulator. $\Delta\varepsilon_{s_0}(0) = 0$ holds before the cutoff is removed.

Self-study task T4

Verify $1 + 2\sum_{k\geq 1} q_s^k = (1+q_s)/(1-q_s) = \coth(eBs)$.

Section 3.4 Fix the measured quadratic response; predict the nonlinear remainder

Goal: understand what renormalisation measures and what the calculation predicts.

At small proper time,

$$(eBs)\coth(eBs) = 1 + \frac{(eBs)^2}{3} - \frac{(eBs)^4}{45} + O(s^6). \tag{3.15}$$

The field-independent 1 is already subtracted. The remaining divergence is purely quadratic:

$$\begin{aligned}\Delta\varepsilon_{s_0}(B) &= \frac{e^2B^2}{24\pi^2} I_\Lambda + \Delta\varepsilon_{\rm nl}(B),\\ I_\Lambda &= \int_{1/\Lambda^2}^{\infty} \frac{\mathrm{d}s}{s} e^{-m^2 s} = \ln\frac{\Lambda^2}{m^2} - \gamma_E + O(m^2/\Lambda^2).\end{aligned} \tag{3.16}$$

The equality in the first line uses the same cutoff in the nonlinear remainder until its convergent $\Lambda \to \infty$ limit is taken. Here γ_E is Euler's constant.

The measured field and charge. Temporarily denote the bare parameters by B_0, e_0, so the total magnetic energy is

$$\frac{B_0^2}{2} + \Delta\varepsilon_{s_0}(B_0) = \frac{B_0^2}{2}\left(1 + \frac{e_0^2}{12\pi^2} I_\Lambda\right) + \Delta\varepsilon_{\rm nl}(e_0 B_0). \tag{3.17}$$

The measured coefficient of the quadratic response defines the renormalized field normalization. The field-renormalization constant Z_3 is unrelated to the vacuum amplitude $Z[A]$. Choose

$$\begin{aligned} Z_3^{-1} &= 1 + \frac{e_0^2}{12\pi^2} I_\Lambda, \qquad B_R = Z_3^{-1/2} B_0, \qquad e_R = Z_3^{1/2} e_0,\\ e_R B_R &= e_0 B_0, \qquad \frac{B_0^2}{2} + \Delta\varepsilon_{s_0} = \frac{B_R^2}{2} + \Delta\varepsilon_{\rm nl}(e_R B_R).\end{aligned} \tag{3.18}$$

Thus the Landau-level combination eB is preserved. Once the measured quadratic coefficient (equivalently, in QED, the low-energy charge) is fixed, the remaining nonlinear response is a prediction. This is not a prescription for making the crystal's measured linear susceptibility vanish: its finite-band background and other degrees of freedom must be matched separately.

The renormalized one-loop matter Lagrangian

Drop the subscript R after fixing the measured parameters:

$$\boxed{\mathcal{L}(B) = -\Delta\varepsilon_{\rm nl}(B) = -\frac{1}{8\pi^2}\int_0^\infty \frac{\mathrm{d}s}{s^3} e^{-m^2 s}\left[(eBs)\coth(eBs) - 1 - \frac{(eBs)^2}{3}\right]}. \tag{3.19}$$

The integrand behaves as s near $s = 0$ and falls exponentially at large s for $m > 0$: the cutoff can now be removed. Exact in *B at one loop for the constant-field Dirac benchmark*; dynamical-photon radiative corrections are omitted.

Checkpoint: which parts were measured and which were predicted?

The constant energy reference and the quadratic electromagnetic response were fixed. The $B^4, B^6, \ldots$ response remains. Why could we not also subtract each of those terms while claiming the same QED theory?

Supporting note: running charge

At one loop, $e_R^2 = e_0^2/(1 + e_0^2 I_\Lambda/(12\pi^2))$ implies

$$\frac{1}{e_R^2} = \frac{1}{e_0^2} + \frac{I_\Lambda}{12\pi^2}.$$

The corresponding one-loop QED running is $\mu\,\mathrm{d}e/\mathrm{d}\mu = e^3/(12\pi^2)$. Extrapolating this formula for a theory containing only the electron gives the formal Landau scale $\Lambda_{\mathrm{L}} \sim m\exp[3\pi/(2\alpha)]$, of order 10^{286} eV for $\alpha \simeq 1/137$. That extrapolation is not a reliable prediction at the pole and ignores all other charged species. It diagnoses the limitations of perturbation theory; it is neither a computed physical singularity of a crystal nor a proof of the ultraviolet fate of complete QED.

Lecture 4 The perturbative series, and why it diverges

Section 4.1 Expanding in the field

Goal: expand the renormalised *one-loop* answer; locate the factorial growth.

Dimensionless variables: $t = m^2 s$, $x \equiv eB/m^2 = B/B_c > 0$, and $\ell(x) \equiv \mathcal{L}(B)/m^4$:

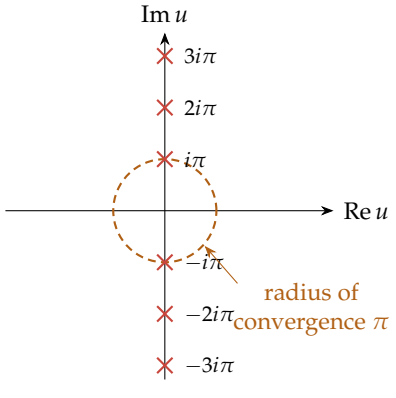
$$\ell(x) = -\frac{1}{8\pi^2}\int_0^\infty \frac{\mathrm{d}t}{t^3}\, e^{-t}\, G(xt), \qquad G(u) \equiv u\coth u - 1 - \frac{u^2}{3}. \tag{4.1}$$

Supplied mathematical identity. We use the normalised product for sinh below; its full proof is not needed here. Its zeros alone would not determine the function.

$$\sinh u = u\prod_{k\geq 1}\left(1 + \frac{u^2}{k^2\pi^2}\right) \tag{4.2}$$

$$\xrightarrow{\frac{\mathrm{d}}{\mathrm{d}u}\ln} \coth u = \frac{1}{u} + \sum_{k\geq 1}\frac{2u}{u^2 + k^2\pi^2}$$

$$\Longrightarrow\ u\coth u = 1 + 2\sum_{k\geq 1}\frac{u^2}{u^2 + k^2\pi^2}. \tag{4.3}$$



Poles of $u\coth u$: the expansion in u converges only inside the dashed circle.

Geometric series, $\frac{u^2}{u^2+k^2\pi^2} = \sum_{n\geq 1}(-1)^{n+1}\left(\frac{u}{k\pi}\right)^{2n}$ for $|u| < k\pi$:

$$u\coth u = 1 + 2\sum_{n\geq 1}(-1)^{n+1}\zeta(2n)\left(\frac{u}{\pi}\right)^{2n}, \qquad \zeta(2n) = \sum_{k\geq 1}\frac{1}{k^{2n}}. \tag{4.4}$$

Use $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$, $\zeta(6) = \pi^6/945$:

$$\begin{aligned} u\coth u &= 1 + \frac{u^2}{3} - \frac{u^4}{45} + \frac{2u^6}{945} + \cdots, \\ G(u) &= -\frac{u^4}{45} + \frac{2u^6}{945} + \cdots, \\ \ell(x) &\sim \frac{x^4}{360\pi^2}\underbrace{\int_0^\infty te^{-t}\mathrm{d}t}_{1!} - \frac{x^6}{3780\pi^2}\underbrace{\int_0^\infty t^3e^{-t}\mathrm{d}t}_{3!} + \cdots \\ &= \frac{x^4}{360\pi^2} - \frac{x^6}{630\pi^2} + \cdots. \end{aligned}$$

For every fixed integer $n \geq 2$, repeated integration by parts gives

$$\int_0^\infty \mathrm{d}t\; t^{2n-3}e^{-t} = (2n-3)!. \tag{4.5}$$

$$\boxed{\mathcal{L} \sim m^4 \sum_{n\geq 2} d_n\, x^{2n}, \qquad d_n = \frac{(-1)^n\,\zeta(2n)\,(2n-3)!}{4\pi^{2n+2}}} \tag{4.6}$$

Why $\sim$, not $=$? At fixed t the expansion needs $|xt| < \pi$, whereas the integral extends to $t = \infty$. Termwise expansion gives an *asymptotic* series as $x \to 0^+$, not a convergent infinite sum. In particular, after the x^{2N} term the remainder is of higher order for fixed N as $x \to 0^+$; sending $N \to \infty$ at fixed x is a different limit.

Section 4.2 What perturbation theory computes

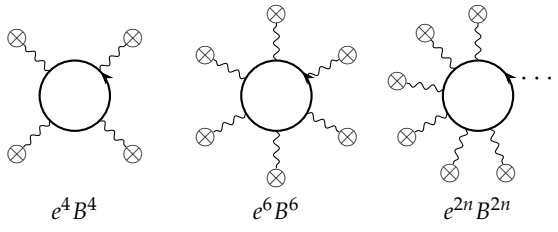
Goal: the meaning and the size of the coefficients.

First term, $n = 2$:

$$\begin{aligned} d_2 &= \frac{\zeta(4)\cdot 1!}{4\pi^6} = \frac{1}{360\pi^2} \\ \Rightarrow\ \mathcal{L} &= \frac{e^4B^4}{360\pi^2m^4} + \cdots = \frac{2\alpha^2}{45}\frac{B^4}{m^4} + \ldots \end{aligned} \tag{4.7}$$

Here $\alpha = e^2/(4\pi)$ is the QED fine-structure constant. The B^4 term is the constant-field limit of the four-photon electron-loop response; its general electromagnetic version describes low-energy light-by-light scattering.

Each x^{2n} contains $e^{2n}B^{2n}$: **one loop,** $2n$ **external-field insertions**. A higher weak-field order is not a higher loop order. Dynamical-photon corrections are absent throughout this calculation.



e^4B^4 e^6B^6 $e^{2n}B^{2n}$

$\otimes$ = an insertion of the prescribed field; the directed closed line is a fermion loop.
All three sketches are at one loop; the wavy lines here are not emitted real photons.

n	2	3	4	5	6	7
d_n	2.81×10^{-4}	-1.61×10^{-4}	3.22×10^{-4}	-1.36×10^{-3}	9.95×10^{-3}	-1.11×10^{-1}

$$\frac{d_{n+1}}{d_n} = -\frac{\zeta(2n+2)}{\zeta(2n)}\,\frac{(2n-1)(2n-2)}{\pi^2} \xrightarrow[n\to\infty]{} -\infty. \tag{4.8}$$

The ratio test gives zero radius of convergence for this *one-loop weak-field series*. We have explicitly proved its divergence, not the divergence of the full QED loop expansion.

Self-study task T5

Check $d_2 = 1/(360\pi^2)$ from (4.6), and $\frac{e^4}{360\pi^2} = \frac{2\alpha^2}{45}$.

Section 4.3 Divergent but useful: optimal truncation

Goal: how accurate can a divergent series be?

$$|d_n|\,x^{2n} \approx \frac{(2n-3)!}{4\pi^2}\left(\frac{x}{\pi}\right)^{2n} \qquad [\zeta \to 1] \tag{4.9}$$

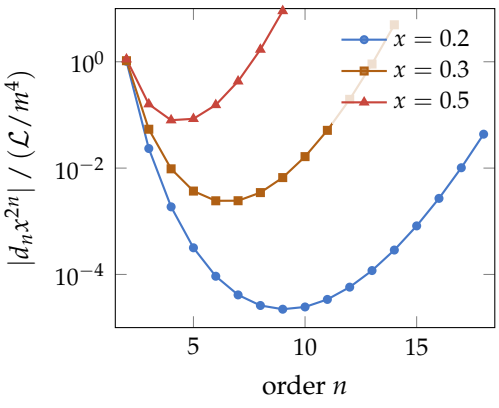
$$\frac{|d_{n+1}|x^{2n+2}}{|d_n|x^{2n}} \approx \frac{(2n-1)(2n-2)\,x^2}{\pi^2} \approx \left(\frac{(2n-\frac{3}{2})\,x}{\pi}\right)^2$$

Terms decrease while this ratio is < 1, then grow:

$$n_* \approx \frac{\pi}{2x} + \frac{3}{4}. \tag{4.10}$$

Stirling, $N! \approx \sqrt{2\pi N}(N/e)^N$:

$$\frac{\text{smallest term}}{\text{first term}} \approx \frac{90\sqrt{2}}{\pi^2}\,x^{-3/2}\,e^{-\pi/x} \;\propto\; e^{-\pi m^2/eB}. \tag{4.11}$$



Size of an individual term, divided by the exact one-loop answer. This is not the partial-sum error.

N (last term kept)	2	3	4	5	6	7	8	9	exact
$x = 0.3$: $10^6\mathcal{L}/m^4$	2.2797	2.1625	2.1836	2.1755	**2.1808**	2.1755	2.1831	2.1686	2.1784
$x = 0.5$: $10^5\mathcal{L}/m^4$	1.7590	1.5078	**1.6334**	1.5001	1.7430	1.0664	3.7398	−10.48	1.5775

A *plain partial sum*, stopped near its smallest term, typically stops improving at the scale $e^{-\pi B_c/B}$ up to powers of B/B_c. This is not a bound on what *resummation of all coefficients* can recover. The same exponential scale will reappear in electric-field pair creation.

Section 4.4 Borel summation: the proper time was the answer

Goal: replace a divergent sum by a well-defined integral, within the one-loop model.

Euler's example.

$$f(g) = \int_0^\infty \mathrm{d}t\, \frac{e^{-t}}{1+gt}\,; \qquad \frac{1}{1+gt} = \sum_n (-gt)^n, \quad \int_0^\infty t^n e^{-t}\mathrm{d}t = n! \tag{4.12}$$

$$\Longrightarrow \quad f(g) \sim \sum_n (-1)^n\, n!\, g^n = 1 - g + 2g^2 - 6g^3 + \ldots \qquad \text{[divergent; } f(g) \text{ is finite]}$$

At $g = 0.1$: partial sums 1, 0.9, 0.92, 0.914, 0.9164, 0.9152, 0.91592, ... around $f(0.1) = 0.915633$; best near $n_* \approx 1/g$.

Borel's rule. Divide by $n!$, sum, put the $n!$ back as an integral:

$$\sum_n a_n g^n \;\longrightarrow\; \mathcal{B}(t) \equiv \sum_n \frac{a_n}{n!} t^n \;\longrightarrow\; \int_0^\infty \mathrm{d}t\, e^{-t}\, \mathcal{B}(gt). \tag{4.13}$$

Euler: $\mathcal{B}(t) = \sum(-t)^n = 1/(1+t)$, which gives (4.12) back.

Our series needs a shifted convention. The factorial is $(2n-3)!$, not $(2n)!$. Define the *generalised Borel function* explicitly:

$$b(u) \equiv \sum_{n\geq 2} \frac{d_n}{(2n-3)!} u^{2n-3} = -\frac{G(u)}{8\pi^2 u^3}, \tag{4.14}$$

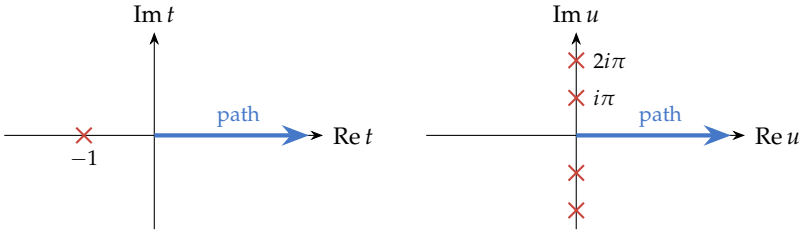
$$\ell(x) = x^2 \int_0^\infty \mathrm{d}u\, e^{-u/x} b(u), \qquad x > 0. \tag{4.15}$$

The apparent singularity at $u = 0$ is removable: $G(u) = -u^4/45 + O(u^6)$, so $b(u) = u/(360\pi^2) + O(u^3)$. Check each coefficient with $u = xt$:

$$x^2 \int_0^\infty \mathrm{d}u\, e^{-u/x} \frac{u^{2n-3}}{(2n-3)!} = \frac{x^{2n}}{(2n-3)!} \int_0^\infty \mathrm{d}t\, e^{-t} t^{2n-3} = x^{2n},$$

$$x^2 \int_0^\infty \mathrm{d}u\, e^{-u/x} \left[-\frac{G(u)}{8\pi^2 u^3}\right] = -\frac{1}{8\pi^2} \int_0^\infty \frac{\mathrm{d}t}{t^3} e^{-t} G(xt).$$

Thus the proper-time integral is exactly this shifted Borel–Laplace reconstruction. G by itself is *not* the ordinary Borel transform defined in (4.13). The analytically continued rational/hyperbolic function supplies the continuation of the locally convergent Borel series.



(a) Euler: $\mathcal{B}(t) = \frac{1}{1+t}$ (b) magnetic field: $b(u)$, poles $\pm ik\pi$

Borel singularities: $u = \pm ik\pi$ (poles of coth) ; path: $u = xt \in [0, \infty)$ real. (4.16)

For real $x > 0$ no pole obstructs this contour, and the integral converges. Its value agrees with the renormalised constant-magnetic-field effective action at one loop. The weak-field coefficients alone do not permit arbitrary extra exponentially small terms to be excluded without a function class and boundary conditions; here the proper-time construction supplies that information.

- Finite radius (π) in the Borel plane $\Leftrightarrow$ factorial growth of the series.
- Distance π of the nearest singularity $\Leftrightarrow$ the limit $e^{-\pi/x}$ of optimal truncation (4.11).

Checkpoint: more terms or more information?

Why can a divergent series be accurate? Why does Borel reconstruction do more than add more terms at fixed x?

Lecture 5 The electric field: one function, two faces

Section 5.1 From B to E

Goal: continue the one-loop answer and identify the extra physical prescription.

For constant electromagnetic fields in relativistic QED, the effective Lagrangian depends on the Lorentz invariants $B^2 - E^2$ and $\mathbf{E}\cdot\mathbf{B}$ (in vacuum units $c = 1$). For pure fields the second invariant vanishes:

$$\text{pure } B \longrightarrow \text{pure } E: \qquad B^2 \longrightarrow -E^2, \quad eB \longrightarrow ieE. \tag{5.1}$$

This is **analytic continuation of a function**, not a real Lorentz boost: a boost preserves $B^2 - E^2$ and cannot reverse its sign. The result here belongs to the continuum Dirac model; a crystal does not acquire exact electromagnetic Lorentz invariance by replacing c with a band velocity.

Use $\cosh(iy) = \cos y$, $\sinh(iy) = i\sin y$:

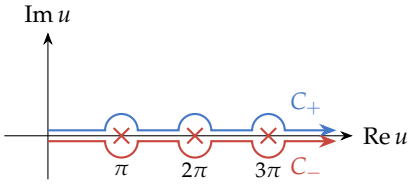
$$(iy)\coth(iy) = (iy)\frac{\cos y}{i\sin y} = y\cot y,$$
$$-\frac{(eBs)^2}{3} \longrightarrow +\frac{(eEs)^2}{3}.$$

The continued one-loop expression is therefore

$$\mathcal{L}(E) = -\frac{1}{8\pi^2}\int_C \frac{\mathrm{d}s}{s^3}\, e^{-m^2 s}\left[(eEs)\cot(eEs) - 1 + \frac{(eEs)^2}{3}\right]. \tag{5.2}$$

At $\sin(eEs) = 0$ the poles lie on the positive proper-time axis:

$$s_k = \frac{k\pi}{eE}, \qquad k = 1, 2, \ldots. \tag{5.3}$$



(c) electric field: Borel poles on the path

The scaled variable is $u = eEs$. C_+ passes above every positive pole; C_- passes below.

A boundary condition is needed

The contour choices are complex conjugates. Their difference is fixed by the pole residues. The physical in/out prescription must give a probability $P_{\text{vac}} = \exp[-2VT\,\mathrm{Im}\,\mathcal{L}] \le 1$. With our overall sign in (5.2), this selects C_+ and $\mathrm{Im}\,\mathcal{L} \ge 0$. The boundary prescription is physical input; coefficients alone do not select it.

Section 5.2 The poles are the pairs

Goal: derive the imaginary part, then compare it with the independently counted vacuum-decay density.

First learn the pole rule. For a real variable a and $\epsilon > 0$,

$$\frac{1}{a+i\epsilon} = \frac{a}{a^2+\epsilon^2} - i\frac{\epsilon}{a^2+\epsilon^2},$$
$$\int_{-\infty}^{\infty} \frac{\epsilon\,\mathrm{d}a}{a^2+\epsilon^2} = \int_{-\infty}^{\infty} \frac{\mathrm{d}v}{1+v^2} = \pi \qquad (a = \epsilon v).$$

As $\epsilon \to 0^+$ the imaginary kernel becomes infinitely narrow with area $-\pi$. More precisely, integrated against a smooth test function f it tends to $-\pi f(0)$. The real part tends to the principal value, defined by symmetric exclusion around $a = 0$. Thus

$$\frac{1}{s - s_k + i0} = \mathcal{P}\frac{1}{s-s_k} - i\pi\delta(s-s_k). \tag{5.4}$$

Here the pole is at $s_k - i0$, below the integration path, exactly the C_+ prescription.

Now compute its weight. Set $eEs = k\pi + \delta$:

$$\sin(k\pi+\delta) \simeq (-1)^k\delta, \qquad \cos(k\pi+\delta) \simeq (-1)^k,$$
$$\cot(eEs) \simeq \frac{1}{eE(s-s_k)}, \qquad (eEs)\cot(eEs) \simeq \frac{s_k}{s-s_k}. \tag{5.5}$$

The subtraction terms have no poles. The residue of the *whole* integrand, including its negative prefactor, is

$$-\frac{e^{-m^2 s_k}}{8\pi^2 s_k^2}.$$

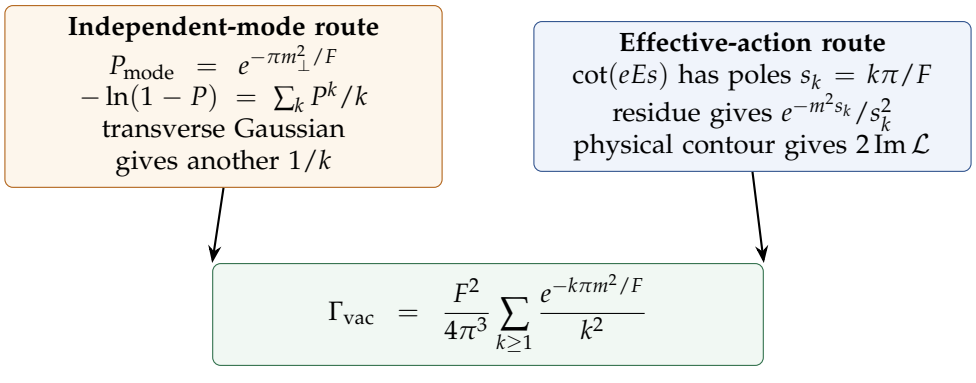
Multiplying that real negative weight by $-i\pi$ gives a positive imaginary part:

$$\operatorname{Im}\mathcal{L}(E) = \frac{1}{8\pi}\sum_{k\geq 1}\frac{e^{-m^2 s_k}}{s_k^2} = \frac{e^2E^2}{8\pi^3}\sum_{k\geq 1}\frac{e^{-k\pi m^2/(eE)}}{k^2}. \tag{5.6}$$

The independent calculations meet

$$2\operatorname{Im}\mathcal{L}(E) = \frac{e^2E^2}{4\pi^3}\sum_{k\geq 1}\frac{e^{-k\pi E_c/E}}{k^2} = \Gamma_{\mathrm{vac}}. \tag{5.7}$$

This equals (2.13), the loss of the zero-pair outcome per space-time volume. The *mean pair rate* w is its $k = 1$ term, (2.9); it is not the full sum.



Section 5.3 What perturbation theory knows: the dispersion relation

Goal: recover each one-loop weak-field coefficient from the decay density.

Use a scalar field variable $\beta \equiv (eB/m^2)^2 = x^2$ and define $\widehat{\ell}(\beta) \equiv \mathcal{L}(B)/m^4 = \ell(\sqrt{\beta})$. This β is an expansion variable, not the Dirac matrix used earlier.

$$\beta > 0 : \text{magnetic field, } \widehat{\ell} \text{ real,}$$
$$\beta = -y - i0, \quad y = (eE/m^2)^2 > 0 : \text{physical electric boundary, } \operatorname{Im}\widehat{\ell} > 0.$$

Write $\rho(y)$ for this boundary imaginary part, *not* a density of material states:

$$\rho(y) \equiv \operatorname{Im}\widehat{\ell}_{\text{phys}}(y) = \frac{y}{8\pi^3}\sum_{k\geq 1}\frac{1}{k^2}e^{-k\pi/\sqrt{y}}. \tag{5.8}$$

Supplied dispersion relation; proof on the supporting sheet. For this renormalised one-loop function: it is analytic off the negative real cut, obeys complex conjugation across that cut, has its constant and linear terms subtracted, and grows no faster than $\beta \ln \beta$ at large $|\beta|$. These conditions give

$$\widehat{\ell}(\beta) = \frac{\beta^2}{\pi}\int_0^\infty \frac{\rho(y)\,\mathrm{d}y}{y^2(y+\beta)}, \qquad \beta > 0. \tag{5.9}$$

For a fixed finite order, expand the kernel geometrically. Its exact remainder is controlled on the supporting sheet; we do not assume convergence of the final series. The coefficient of β^n is

$$d_n = \frac{(-1)^n}{\pi}\int_0^\infty \frac{\rho(y)}{y^{n+1}}\,\mathrm{d}y, \qquad n \geq 2. \tag{5.10}$$

Do the moment integral explicitly. Put $a = k\pi > 0$ and $u = a/\sqrt{y}$:

$$y = \frac{a^2}{u^2}, \qquad \mathrm{d}y = -2a^2u^{-3}\mathrm{d}u, \qquad (y: 0 \to \infty) \Longleftrightarrow (u: \infty \to 0),$$
$$y^{-n}\mathrm{d}y = -2a^{2-2n}u^{2n-3}\mathrm{d}u.$$

Hence

$$\begin{aligned}\int_0^\infty \mathrm{d}y\, y^{-n}e^{-k\pi/\sqrt{y}} &= 2(k\pi)^{2-2n}\int_0^\infty \mathrm{d}u\, u^{2n-3}e^{-u} \\ &= 2(k\pi)^{2-2n}(2n-3)!.\end{aligned} \tag{5.11}$$

Insert this into (5.10):

$$\begin{aligned}d_n &= \frac{(-1)^n}{8\pi^4}\sum_{k\geq 1}\frac{2(k\pi)^{2-2n}(2n-3)!}{k^2} \\ &= \frac{(-1)^n(2n-3)!}{4\pi^{2n+2}}\sum_{k\geq 1}k^{-2n}.\end{aligned}$$

$$\boxed{d_n = \frac{(-1)^n\zeta(2n)(2n-3)!}{4\pi^{2n+2}}} \tag{5.12}$$

The quantitative answer to Q2

We recovered *every coefficient of the one-loop constant-field weak-field series*, (4.6). The electric tunnelling exponential produces the factorial moment; the sum of pole contributions produces $\zeta(2n)$. This explicit factorial growth proves zero radius of convergence. Nonanalyticity alone would not prove that: $e^{-1/\beta}$ can have an identically zero asymptotic power series as $\beta \to 0^+$.

Checkpoint: what extra information entered?

The weak-field coefficients are real. Why does the calculation nevertheless recover a nonzero Γ_{vac}? What selects its physical sign?

Section 5.4 Conclusions: what we calculated, and why it matters

Goal: close both questions, then translate one controlled consequence back to a narrow-gap model.

The final three equations

For prescribed uniform fields, one isotropic Dirac species with two spin states, and the long-duration, empty-particle initial state:

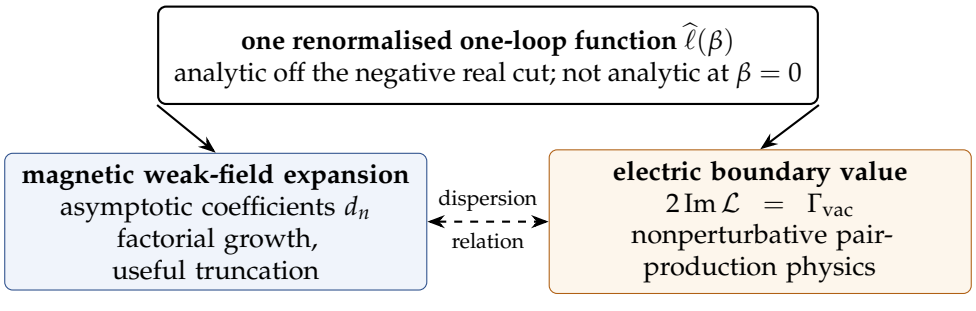
$$\text{mean pair production:} \quad w = \frac{\langle N_{\text{pairs}} \rangle}{VT} = \frac{F^2}{4\pi^3} e^{-\pi m^2/F},$$

$$\text{vacuum persistence:} \quad \Gamma_{\text{vac}} = -\frac{\ln P_{\text{vac}}}{VT} = 2\,\text{Im}\,\mathcal{L}(E) = \frac{F^2}{4\pi^3} \sum_{k\geq 1} \frac{e^{-k\pi m^2/F}}{k^2},$$

$$\text{weak-field coefficients:} \quad d_n = \frac{(-1)^n}{\pi} \int_0^\infty \frac{\rho(y)}{y^{n+1}} \mathrm{d}y = \frac{(-1)^n \zeta(2n)(2n-3)!}{4\pi^{2n+2}}.$$

Q1. In this ideal constant-field limit the initially filled lower band loses vacuum probability for every $F > 0$. The mean number created and the vacuum-persistence exponent are distinct; they agree approximately when each mode's transition probability is small. This says neither that E_c is a sharp threshold nor that every finite sample/pulse has the same rate.

Q2. Every finite weak-field polynomial has zero imaginary part, yet the all-order coefficients grow factorially. The generalised Borel reconstruction plus the physical continuation prescription reproduces the nonperturbative vacuum-decay density. The agreement of two routes is our calculation's central check.



Narrow-gap materials: a controlled sensitivity test. Restore units for the ideal symmetric two-band model, with half gap Δ, band velocity v, and $e = |q|$:

$$w(\Delta) = \frac{(eE)^2}{4\pi^3 \hbar^2 v} \exp\left[-\frac{\pi \Delta^2}{eE\hbar v}\right], \qquad E_c(\Delta) = \frac{\Delta^2}{e\hbar v}.$$

For two gaps at the *same* E and v, the prefactor cancels:

$$\ln \frac{w(\Delta_2)}{w(\Delta_1)} = -\frac{\pi(\Delta_2^2 - \Delta_1^2)}{eE\hbar v}, \qquad \left.\frac{\partial \ln w}{\partial \ln \Delta}\right|_{E,v} = -2\pi \frac{E_c}{E}.$$

Example at fixed $E = 0.3E_c(\Delta_1)$: let $r = \Delta_2/\Delta_1 = E_{g,2}/E_{g,1}$. Then

$$\frac{w(\Delta_2)}{w(\Delta_1)} = \exp\left[-\frac{\pi}{0.3}(r^2 - 1)\right] = \begin{cases} 0.341852, & r = 1.05, \\ 2.77602, & r = 0.95. \end{cases}$$

A 5% gap variation changes the ideal rate by a factor of several. This explains why a field-driven process can be sensitive to gap uniformity; it is *not* a prediction of a real HgMnTe sample's leakage current. Band structure, occupations, temperature, scattering, exchange splitting, contacts, pulse duration and the spatial field profile must be specified for that prediction. In particular the ideal spatial tunnelling length $2\Delta/(eE)$ must fit inside the approximately uniform region.

What remains fundamental and open? The calculation is an exactly solvable *one-loop benchmark* for a broader question: what nonperturbative information can the divergent expansions of an interacting quantum field theory determine, and which boundary/data inputs remain independent? Adding dynamical photons, backreaction or inhomogeneous fields requires new calculations; this equality is not an all-loop reconstruction theorem.

Dyson: related logic, different argument. Dyson's original argument reverses the sign of e^2 in *interacting QED*: like charges then attract, suggesting vacuum instability and obstructing analyticity of the full coupling expansion. It is a physical heuristic. Here we varied the *background invariant* $B^2 \to -E^2$ at one loop and proved factorial growth directly. Our example illustrates the instability/analyticity connection; it does not derive the full QED loop series from Dyson's argument.

Exit checkpoint

Name the calculated observable, one approximation controlling its applicability, and one way the result would change when moving from the ideal Dirac model to a fabricated semiconductor.

Report: reproduce, test, explain

Suggested scope: 4–6 pages plus figures/code; legible handwritten derivations are welcome. Required core: R1(a,c), R3 and R5. R1(b), R2 and R4 are optional extensions, not prerequisites for completing this workshop.

R1: a single avoided crossing — **required (a,c)**

(a) Write the two component equations for $h(t) = Ft\sigma_x + m\sigma_z$. Explain the initial negative-energy state and final positive-energy projection. Using the midpoint time-evolution algorithm given in the numerical guide, take $m = 1$ and $F = 0.25, 0.5, 1, 2, 4$; plot $\ln P$ against $1/F$ and compare with $-\pi/F$. Here the integration parameter T is the half-window $[-T, T]$. At $F = 1$, separately halve the time step at fixed T, then double T at fixed step. Record the norm error. Explain why norm conservation alone does not prove the transition probability is accurate. Include either the WKB semicircle calculation or the transverse state-counting derivation.

(b) **Optional:** evolve both negative-energy states of $H(t) = Ft\alpha_x + p_y\alpha_y + p_z\alpha_z + m\beta$. Project onto the *whole* positive-energy subspace, not one arbitrary eigenvector. Verify Eq. (2.4) and its two identical spin copies. Repeat the convergence checks.

(c) Show how $\chi = -Ext$ transforms $(\phi, \mathbf{A}) = (-Ex, 0)$ to $(0, -Et\,\hat{\mathbf{x}})$. Include the wavefunction phase, and explain why the pair probability is gauge invariant.

R3: a divergent series that gives useful numbers — **required**

For $x = eB/m^2 = 0.2, 0.3, 0.5$, compare partial sums through $N = 2, \ldots, 12$ with the proper-time integral. Plot the absolute error and identify the best displayed N. Compare with Eq. (4.10), remembering that the least *term* and best last-kept *partial sum* need not have the same index. Check quadrature order and large-t cutoff independently. Explain what Borel summation adds beyond simply keeping more terms.

R5: the physical conclusion — **required, about one page**

(a) For two independent modes with $P_1 = 0.2$, $P_2 = 0.3$, compute $\langle N \rangle$, P_{vac} and $-\ln P_{\text{vac}}$. Explain which gives the mean pair rate and which gives vacuum decay after division by VT. Where does the energy of produced pairs come from?

(b) Explain the distinction between the one-loop weak-*field* expansion and the loop expansion of interacting QED. What result did we reproduce twice? What physical information selects the electric contour?

(c) Restore $\hbar, v$ in the exponent, then derive $\mathrm{d}\ln P/\mathrm{d}\ln E_g$ at fixed E, v. For $E = 0.3E_c$ at the nominal gap, compute the response to $\pm 5\%$ changes in E_g. List three ingredients that must be restored before interpreting this as a prediction for an actual HgMnTe sample.

Completion criterion

The argument and approximations must be traceable; numerical comparisons must include convergence checks; the two rates must be distinguished; and the ideal material analogy must not be presented as a calibrated device prediction. A complete calculation of all optional topics is not expected.

Optional extensions

R2: charged scalar field, with the needed input supplied

A bosonic mode has a geometric pair-number distribution $\Pr(N = j) = \bar{n}^j/(1+\bar{n})^{j+1}$, $j = 0, 1, \ldots$, where $\bar{n} = e^{-\pi m_\perp^2/F}$. Check its normalization and mean. There is no spin factor.

(a) Use $P_0 = 1/(1+\bar{n})$ and the same state counting to obtain

$$\Gamma_{\mathrm{vac}}^{(0)} = \frac{F^2}{8\pi^3} \sum_{k\geq 1} \frac{(-1)^{k+1}}{k^2} e^{-k\pi m^2/F}.$$

(b) The scalar Landau levels have $E_n^2 = m^2 + p_z^2 + eB(2n+1)$, and particle plus antiparticle zero-point energies sum to $+E_n$ per orbital. With the same common proper-time regulator, derive the *energy difference*

$$\varepsilon^{(0)}(B) - \varepsilon^{(0)}(0) = -\frac{1}{16\pi^2} \int_{s_0}^{\infty} \frac{\mathrm{d}s}{s^3} e^{-m^2 s} \left[\frac{eBs}{\sinh(eBs)} - 1 \right].$$

Use $\sum_{n\geq 0} e^{-(2n+1)a} = 1/(2\sinh a)$; identify what the reference subtraction removes.

(c) Fix the measured quadratic response as in Lecture 3. Show that

$$\mathcal{L}^{(0)} = \frac{1}{16\pi^2} \int_0^{\infty} \frac{\mathrm{d}s}{s^3} e^{-m^2 s} \left[\frac{eBs}{\sinh(eBs)} - 1 + \frac{(eBs)^2}{6} \right] = \frac{7e^4B^4}{5760\pi^2 m^4} + \cdots.$$

(d) Continue to E with the physical prescription. Explain why alternating pole residues yield (a).

R4: reconstruct the scalar coefficients

Use $\mathrm{Im}\,\mathcal{L}^{(0)} = \Gamma_{\mathrm{vac}}^{(0)}/2$ in the dispersion moment. Derive

$$d_n^{(0)} = \frac{(-1)^n (2n-3)!}{8\pi^{2n+2}} (1 - 2^{1-2n}) \zeta(2n).$$

Check $n = 2$ against R2(c). The statistics change the sum over pole number; the same exponential moment produces the factorial.

Supporting derivation: the Landau–Zener equation

Purpose. Explain the differential equation underlying the supplied exact theorem. The reduction below is elementary; the special-function connection formula is a separate mathematical input.

Use the fixed basis of Eq. (1.18), $\hbar = 1$, and write its components as $a(t), b(t)$:

$$\begin{aligned} i\dot{a} &= Ft\,a + mb, \qquad i\dot{b} = ma - Ft\,b, \\ i\ddot{a} &= Fa + Ft\dot{a} + m\dot{b} \\ &= Fa + Ft\dot{a} - im^2 a + iFt\,(mb) \\ &= Fa + Ft\dot{a} - im^2 a + iFt\,(i\dot{a} - Fta) \\ &= Fa - i(m^2 + F^2t^2)a, \\ \therefore \quad \ddot{a} + (F^2t^2 + m^2 + iF)a &= 0. \end{aligned}$$

Eliminating a gives the companion equation

$$\ddot{b} + (F^2t^2 + m^2 - iF)b = 0.$$

The first-order pair fixes the relation between the two components; independent arbitrary solutions of both second-order equations would overcount the initial data.

Choose the complex variable and dimensionless adiabaticity parameter

$$z = e^{i\pi/4}\sqrt{2F}\,t, \qquad \delta = \frac{m^2}{2F}, \qquad \frac{\mathrm{d}^2}{\mathrm{d}t^2} = 2iF\frac{\mathrm{d}^2}{\mathrm{d}z^2}, \qquad F^2t^2 = -\frac{iF}{2}z^2.$$

Substituting and dividing by $2iF$ gives

$$\begin{aligned} 0 &= 2iFa_{zz} + \left(-\frac{iF}{2}z^2 + m^2 + iF\right)a, \\ 0 &= a_{zz} + \left(\frac{1}{2} - i\delta - \frac{z^2}{4}\right)a. \end{aligned}$$

This is Weber's equation,

$$\frac{\mathrm{d}^2 D_\nu}{\mathrm{d}z^2} + \left(\nu + \frac{1}{2} - \frac{z^2}{4}\right)D_\nu = 0, \qquad \nu = -i\delta.$$

For $m > 0$ a general solution can be written

$$a(t) = A\,D_{-i\delta}(z) + B\,D_{-i\delta}(-z), \qquad b(t) = \frac{i\dot{a} - Fta}{m}.$$

The constants A, B must be fixed by a normalized incoming lower-energy state; the outgoing coefficients must be read in the instantaneous-energy basis. These boundary conditions are essential. Merely writing down D_ν does not calculate a transition probability.

The supplied special-function matching, stated precisely

Matching the asymptotic Weber solutions on the rays traversed as real t runs from $-\infty$ to $+\infty$ gives a diabatic-survival amplitude of modulus $e^{-\pi\delta}$. Therefore

$$P = |A_{\text{diabatic survival}}|^2 = e^{-2\pi\delta} = e^{-\pi m^2/F}.$$

The complementary probability is $1 - P$ by norm conservation. Phases are required for repeated passages and interference, but not for this single-passage probability. We supply this connection formula as the Landau–Zener theorem, rather than claiming to prove all special-function asymptotics here.

Direct numerical check avoids special functions. Let $u_-(t)$ and $u_+(t)$ be normalized instantaneous eigenvectors of $h(t)$, with eigenvalues $-\sqrt{F^2t^2+m^2}$ and $+\sqrt{F^2t^2+m^2}$. Solve

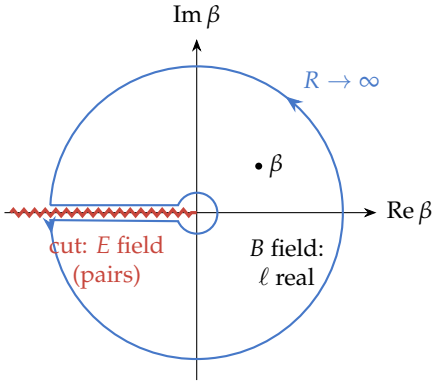
$$i\dot{\psi} = h(t)\psi, \qquad \psi(-T) = u_-(-T), \qquad P_T = \left|u_+^\dagger(T)\psi(T)\right|^2.$$

Here T is the numerical half-window, not the full plateau duration used in VT. Increase T and reduce the integration step independently. Both tests are required: exact norm preservation alone does not bound the error in a small transition probability. The $m = 0$ limit is immediate from the uncoupled diabatic equations and gives $P = 1$; the reconstruction formula for b above assumes $m \neq 0$.

Supporting derivation: the subtracted dispersion relation

Purpose. Supply the contour proof deferred from Section 5.3. It requires the Cauchy integral formula. The core lecture uses the stated relation and performs its moment integral explicitly.

Use $\widehat{\ell}(\beta) = \mathcal{L}(B)/m^4$, $\beta = (eB/m^2)^2$. In this constant-field one-loop problem, the only cut is on the negative real β axis. Reflection relates the two banks. The renormalisation subtractions remove constant and linear terms; near the origin, away from the cut, $\widehat{\ell}(\beta)/\beta^2 \to d_2$. At large argument $\widehat{\ell} = O(\beta \ln \beta)$, so two subtractions suffice. These analytic/growth assumptions are properties of the proper-time function, not consequences of knowing an arbitrary finite set of coefficients.



The contour surrounds the observation point β while excluding the negative-axis cut.

Cauchy's formula on the cut plane gives

$$\frac{\widehat{\ell}(\beta)}{\beta^2} = \frac{1}{2\pi i}\oint_C \mathrm{d}\beta' \, \frac{\widehat{\ell}(\beta')}{\beta'^2(\beta'-\beta)}. \tag{S.1}$$

The large-circle integral is $O(\ln R/R)$ and vanishes as $R \to \infty$. The small-circle integral vanishes as its radius tends to zero. The upper bank runs from $-\infty$ towards 0 and the lower bank returns. Define

$$\mathrm{Disc}\,\widehat{\ell}(\beta') \equiv \widehat{\ell}(\beta'+i0) - \widehat{\ell}(\beta'-i0).$$

For the physical prescription, $eE \to eE + i0$ and hence $\beta = -y - i0$. Thus the lower bank has positive imaginary part $\rho(y)$, while the upper bank has its conjugate:

$$\begin{aligned}
\mathrm{Disc}\,\widehat{\ell}(-y) &= -2i\rho(y), \\
\frac{\widehat{\ell}(\beta)}{\beta^2} &= \frac{1}{2\pi i}\int_{-\infty}^{0} \mathrm{d}\beta' \, \frac{\mathrm{Disc}\,\widehat{\ell}(\beta')}{\beta'^2(\beta'-\beta)} \\
&= -\frac{1}{\pi}\int_0^\infty \mathrm{d}y \, \frac{\rho(y)}{y^2(-y-\beta)} = \frac{1}{\pi}\int_0^\infty \mathrm{d}y \, \frac{\rho(y)}{y^2(y+\beta)}.
\end{aligned}$$

This proves (5.9) with a positive magnetic answer.

Why the coefficient extraction is asymptotic. For integer $N \geq 2$ the finite geometric identity is exact:

$$\frac{1}{y+\beta} = \sum_{j=0}^{N-2} \frac{(-\beta)^j}{y^{j+1}} + \frac{(-\beta)^{N-1}}{y^{N-1}(y+\beta)}.$$

Substitute it into (5.9):

$$\widehat{\ell}(\beta) = \sum_{n=2}^{N} \frac{(-1)^n \beta^n}{\pi} \int_0^\infty \mathrm{d}y \, \frac{\rho(y)}{y^{n+1}} + R_N(\beta),$$
$$R_N(\beta) = \frac{(-1)^{N-1}\beta^{N+1}}{\pi} \int_0^\infty \mathrm{d}y \, \frac{\rho(y)}{y^{N+1}(y+\beta)}.$$

For real $\beta > 0$, $\rho(y) > 0$ and $y + \beta \geq y$. Therefore

$$|R_N(\beta)| \leq \frac{\beta^{N+1}}{\pi} \int_0^\infty \mathrm{d}y \, \frac{\rho(y)}{y^{N+2}} = |d_{N+1}|\beta^{N+1}.$$

Each moment is finite: at $y \to 0^+$ the tunnelling exponential beats any inverse power, and at $y \to \infty$, $\rho(y) = O(y)$ so moments with $n \geq 2$ converge. This proves a controlled asymptotic expansion at every fixed order, while the factorial growth of d_n prevents convergence as $N \to \infty$ at fixed β. It also explains why the next omitted term controls the magnetic partial-sum error.

Numerical guide: algorithms and checks

The algorithms below specify how to reproduce the numerical illustrations and test their convergence. They can be implemented with any numerical language supporting matrix arithmetic and quadrature. Record the parameters, approximation errors and convergence checks alongside each result.

The numerical time window is not the phase-space duration

In the time-evolution calculation, T denotes the *half-window*: integrate from $-T$ to $+T$. In the rate calculation, T denotes the full duration of the long field plateau in VT. The ODE calculation concerns one mode and does not insert its endpoint parameter into the state-counting prefactor.

1. Time evolution without a black-box ODE solver. At a time midpoint t_j, $h_j = Ft_j\sigma_x + m\sigma_z$ and

$$h_j^2 = \epsilon_j^2 I, \quad \epsilon_j = \sqrt{F^2 t_j^2 + m^2}, \qquad U_j = \cos(\epsilon_j \delta t) I - i\frac{\sin(\epsilon_j \delta t)}{\epsilon_j} h_j.$$

Thus $\psi_{j+1} = U_j \psi_j$ is exactly unitary for the frozen midpoint Hamiltonian. Time ordering is still approximated: at fixed endpoints the global method is second order in δt.

1. Initialize the normalized negative-energy eigenvector at $-T$.
2. Apply the midpoint matrices in time order.
3. Project onto the positive-energy eigenvector at $+T$ and square its modulus.
4. Halve δt at fixed T; then increase T independently. Record the norm as a separate check.

For the four-component extension use the whole positive-energy projector. A choice of basis within its degenerate subspace cannot change a probability.

F ($m = 1$)	finite-window midpoint P	$e^{-\pi/F}$	relative difference
0.25	$3.48691946 \times 10^{-6}$	$3.48734236 \times 10^{-6}$	-1.21×10^{-4}
0.5	0.00186743535	0.00186744273	-3.95×10^{-6}
1	0.0432138984	0.0432139183	-4.60×10^{-7}
2	0.207879940	0.207879576	1.75×10^{-6}
4	0.455937892	0.455938128	-5.18×10^{-7}

These finite-endpoint results use $m = 1$, $T = 80 \max(1, F^{-1/2}) F^{-1/2}$ and a time step no larger than 0.002 in these units; they are not exact values. The small difference at $F = 0.25$ requires an endpoint check before interpreting its last digits.

2. The magnetic integral. Gauss–Legendre quadrature integrates $e^{-t} G(xt)/t^3$ on $0 < t < t_{\max}$. Near zero, evaluating $u \coth u - 1 - u^2/3$ by subtraction loses precision; use its convergent small-u series there. Double the quadrature order and increase $t_{\max}$ independently. At $x = 0.2, 0.3, 0.5$ the checked integrals are $4.40736251308 \times 10^{-7}$, $2.17839981750 \times 10^{-6}$, $1.57747326488 \times 10^{-5}$, respectively. The best displayed partial sums end at $N = 9, 6, 4$.

3. The dispersion integral. Integrate the sum of decay exponentials after $u = \pi/\sqrt{y}$, keeping this numerical integral independent of the analytic gamma formula. Coefficients from this route are compared to a separate exact-rational Bernoulli construction of d_n. The finite sum over pole number has an integrated tail bounded by $K^{1-2n}/(2n-1)$ in the associated zeta sum. Quadrature resolution and the tail are distinct error sources. Also change the angle of the deformed electric proper-time contour and check its imaginary part against the residue sum.

4. Landau levels in a finite basis. A truncated oscillator cannot obey $[a, a^\dagger] = I$ at its highest state. It creates a spurious low-energy boundary state. Therefore identify cutoff support within degenerate eigenspaces, check invariance under a basis rotation there, and compare only the physical low levels to the analytic spectrum. This is a useful lesson about verifying a discretization rather than merely diagonalizing a matrix.

Reading guide

Shortest route: Cohen–McGady for the two rates; Dunne for the constant-field calculation and weak-field asymptotics; Linder–Lorke–Schützhold for the semiconductor analogy and its conditions. Teppe and Novik prevent over-identifying the ideal Dirac model with actual Hg-based materials.

References

[1] T. D. Cohen and D. A. McGady, *The Schwinger mechanism revisited*, Phys. Rev. D **78**, 036008 (2008). arXiv:0807.1117. Mean number versus vacuum persistence; long constant-field limit (Lecture 2).

[2] G. V. Dunne, *Heisenberg–Euler effective Lagrangians: Basics and extensions* (2004). arXiv:hep-th/0406216. Spinor/scalar proper-time integrals, renormalization, weak-field asymptotics and higher-loop scope (Lectures 3–5).

[3] J. Schwinger, *On gauge invariance and vacuum polarization*, Phys. Rev. **82**, 664 (1951). doi:10.1103/PhysRev.82.664. Primary constant-field effective-action calculation.

[4] C. Zener, *Non-adiabatic crossing of energy levels*, Proc. R. Soc. A **137**, 696 (1932). doi:10.1098/rspa.1932.0165. Exact connection formula used in Lecture 1 and the Weber supporting sheet.

[5] F. J. Dyson, *Divergence of perturbation theory in quantum electrodynamics*, Phys. Rev. **85**, 631 (1952). doi:10.1103/PhysRev.85.631. Original coupling-sign argument, distinguished from our one-loop field expansion.

[6] F. Teppe et al., *Temperature-driven massless Kane fermions in HgCdTe crystals*, Nat. Commun. **7**, 12576 (2016). arXiv:1602.05999. Multiband Kane spectrum and its distinction from genuine Dirac fermions.

[7] E. G. Novik et al., *Band structure of semimagnetic $Hg_{1-y}Mn_yTe$ quantum wells*, Phys. Rev. B **72**, 035321 (2005). arXiv:cond-mat/0409392. Eight-band model and Mn exchange.

[8] M. F. Linder, A. Lorke and R. Schützhold, *Analog Sauter–Schwinger effect in semiconductors for spacetime-dependent fields*, Phys. Rev. B **97**, 035203 (2018). arXiv:1503.07108. Conditions and limits of semiconductor analogies.

[9] W. Zawadzki, *Semirelativity in semiconductors: a review*, J. Phys.: Condens. Matter **29**, 373004 (2017). arXiv:1701.07067. Broader background on the band/relativistic correspondence.

[10] M. V. Berry, *Histories of adiabatic quantum transitions*, Proc. R. Soc. A **429**, 61 (1990). doi:10.1098/rspa.1990.0051. Optional adiabatic/superadiabatic context; not needed for the core derivation.

Quantum fluctuations and physical memory

Quantum fluctuations and physical memory

Stochastic quantization and recurrent dynamics

Contents

Motivation and scope 2

Notation and the three different times 3

1 Quantum equilibrium as a measure on paths 4

2 The first interaction correction and independent checks 7

3 Stochastic quantization on a finite lattice 10

4 Recurrent updates: stability and bias 14

5 Eliminating a quantum environment 16

6 Quantum noise and fluctuation–dissipation 18

7 An exact memory kernel and its recurrent representation 20

8 One bath, two representations 24

References 29

Motivation and scope

A recurrent system can have the correct stationary distribution and still reproduce the wrong temporal correlations. For a physical model, stability, equilibrium fluctuations and response to an external force are separate requirements. What constraints must a recurrence satisfy to represent quantum fluctuations and environmental memory?

The quantum oscillator provides two controlled answers. First, a weak quartic interaction changes its ground-state fluctuations. The correction is derived from a Euclidean path integral and independently recovered from a stochastic evolution of paths. Second, eliminating a harmonic environment produces a causal memory equation. For an exponential friction kernel, its response has an exact representation using one additional memory variable. Integrating out the same bath in equilibrium connects these two descriptions.

For research on nonlinear recurrent neural networks, these calculations provide analytic examples of nonlinear drift, invariant measures, discretization bias, stability and hidden-state memory. The recurrent parameters here follow from the Hamiltonian; no learning algorithm is assumed. The unresolved extension is how compact representations of interacting quantum dynamics can preserve equilibrium statistics, causal response and quantum fluctuation–dissipation together.

The two solved limits are distinct: an isolated weakly anharmonic oscillator, and a bath-coupled harmonic oscillator. These are established benchmarks, not a solution of generic nonlinear quantum dynamics.

Conventions. The system mass and Planck's constant are set to $M = \hbar = 1$. Energy has inverse-time units; $[q] = E^{-1/2}$, $[\lambda] = E^3$, and $[s] = E^{-2}$. Physical time t and imaginary time τ have units E^{-1}. Vacuum denotes the ground state. The subscript 0 denotes a free quantity in G_0, but the ground state in E_0; initial-time values are written $q(0)$.

Prerequisites. Basic calculus, linear algebra and the idea of a quantum observable as an operator. The path-integral, perturbative and stochastic constructions are developed explicitly; no previous quantum-field-theory course is assumed.

Notation and the three different times

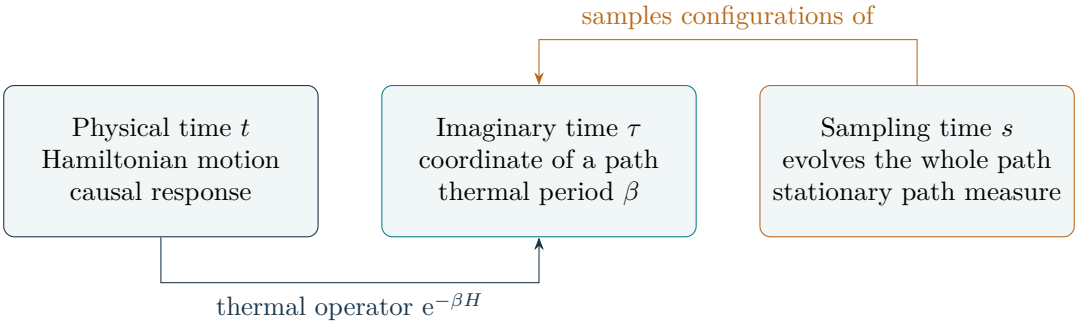


Figure. The sampling clock does not become physical time at equilibrium. A decay in s measures convergence of the sampler; a decay in t measures physical relaxation.

Symbol	Meaning
$q, p;\ x(\tau)$	Quantum operators; ordinary real integration variable specifying a Euclidean path.
$H, \lvert n\rangle, E_n$	Hamiltonian, energy eigenstate, energy eigenvalue: $H\lvert n\rangle = E_n\lvert n\rangle$.
$\rho, \mathrm{Tr}, \langle O\rangle$	Density operator, trace, quantum expectation $\mathrm{Tr}(\rho O)$. A probability *density* for the sampler will instead be written $\rho(X, s)$ and identified explicitly.
β, T, k_{B}	Inverse temperature $\beta = 1/(k_{\mathrm{B}}T)$, temperature, Boltzmann constant. Ground-state results use $\beta \to \infty$.
ω_0, λ	Positive oscillator frequency; nonnegative quartic coupling. Perturbative parameter $\lambda/\omega_0^3 \ll 1$.
$S_E,\ S_a,\ X_\ell$	Continuum Euclidean action; finite-lattice action; rescaled path values $X_\ell = \sqrt{a}\, x(\ell a)$.
$a, N;\ s, \epsilon, n$	Imaginary-time mesh $a = \beta/N$ and number of sites; auxiliary time, its step size, and step index.
$A, \kappa_j;\ J$	Positive free lattice matrix and its eigenvalues; a source for a Gaussian integral. $J(\omega)$ with a frequency argument later denotes the *bath spectral density*, not this source.
G_0, G	Free and interacting Euclidean two-point functions; inverse free quadratic kernel and its interacting correction.
ν, ν_n	Continuous Euclidean frequency (zero temperature); periodic frequencies $\nu_n = 2\pi n/\beta$ at finite temperature.
W_i, η, ξ_n	Wiener processes, continuum white noise, and independent standard Gaussian increments. These belong to the auxiliary sampler.
$Q_j, P_j, m_j, \omega_j, c_j$	Bath coordinates, momenta, masses, frequencies, and real linear couplings. Bath index j is local to that calculation.
K, F, C_F	Retarded friction kernel, operator bath force, and its symmetrized quantum covariance.
f, v, h	Applied ordinary force; velocity $\dot{q}$; hidden variable storing exponentially weighted past velocity.
γ, Ω	Drude damping strength and inverse memory time; both positive.
$z, \widetilde{K}(z), \chi(z)$	Laplace variable with $\mathrm{Re}\, z > 0$, transformed memory kernel, and response.
$\omega, \chi_R(\omega)$	Real physical Fourier frequency and causal susceptibility.
$Q, h_R, R(t)$	Force-induced position and memory responses; causal response kernel. The subscripted Q_j still denotes a bath coordinate.
$Y, \mathsf{M}, \mathsf{b}; W, U$	Real response-state vector, its drift matrix and input vector; exact sampled transition and input weights.
$\mathbb{E}, \delta, \delta_{ij}$	Average over classical noise, Dirac delta distribution, Kronecker delta.

1 Quantum equilibrium as a measure on paths

The equilibrium quantum problem can be represented as a positive probability measure on coordinate paths.

1.1 The operator problem and the meaning of the three times

The coordinate and momentum operators are q and p, with

$$[q,p]=\mathrm{i}, \qquad H_A \equiv H=\frac{p^2}{2}+V(q), \qquad V(q)=\frac{\omega_0^2 q^2}{2}+\frac{\lambda q^4}{24}. \tag{1.1}$$

Here $[A,B]=AB-BA$ is the commutator and $\mathrm{i}^2=-1$. We set the oscillator mass and $\hbar$ equal to one. The frequency ω_0 is positive; $\lambda \geq 0$ makes the potential confining. For our perturbative calculation the dimensionless quantity $g=\lambda/\omega_0^3$ must be small. An operator acts on a quantum state; a lower-case x below will instead denote a possible value of the coordinate, or a coordinate path.

Let $H|n\rangle = E_n|n\rangle$, with normalized energy eigenstates and E_0 the lowest energy. At temperature T, define

$$\beta=\frac{1}{k_B T}, \qquad Z(\beta)=\mathrm{Tr}\ \mathrm{e}^{-\beta H}=\sum_n \mathrm{e}^{-\beta E_n}, \qquad \langle O\rangle_\beta=\frac{\mathrm{Tr}(\mathrm{e}^{-\beta H}O)}{Z(\beta)}. \tag{1.2}$$

Here k_B is Boltzmann's constant, Tr sums diagonal matrix elements over a complete basis, and O is an observable. If the ground state is unique,

$$\lim_{\beta\to\infty}\langle O\rangle_\beta=\langle 0|O|0\rangle, \qquad E_0=-\lim_{\beta\to\infty}\frac{\log Z(\beta)}{\beta}. \tag{1.3}$$

The factor $\mathrm{e}^{-\beta(E_n-E_0)}$ suppresses every excited state: this is why an equilibrium trace determines vacuum properties.

Symbol	Meaning	Physical or mathematical role
t	Physical time	Actual response after a force or a disturbance
τ	Euclidean time	Coordinate along a quantum path
s	Stochastic time	Auxiliary relaxation used to sample paths

Euclidean evolution uses $\mathrm{e}^{-\tau H}$, whereas physical evolution uses $\mathrm{e}^{-\mathrm{i}tH}$. They are different evolutions. The positive operator $\mathrm{e}^{-\beta H}$ is already enough to derive the Euclidean construction; no unproved change of a real-time integration contour is needed here.

1.2 Derive the path integral by inserting ordinary coordinate states

Coordinate states obey $q|x\rangle = x|x\rangle$, $\langle x|x'\rangle=\delta(x-x')$, and $\int \mathrm{d}x\,|x\rangle\langle x| = 1$. Divide the Euclidean interval into N pieces of length $a=\beta/N$. The product formula gives

$$\mathrm{e}^{-\beta H}=\lim_{N\to\infty}\left(\mathrm{e}^{-ap^2/2}\mathrm{e}^{-aV(q)}\right)^N. \tag{1.4}$$

At finite a this product is an approximation; the limit is the quantum operator we started from. Insert N coordinate resolutions of the identity in the trace. The trace closes the path: $x_N=x_0$. Each kinetic matrix element is a Gaussian integral,

$$\begin{aligned}\langle x_{\ell+1}|\mathrm{e}^{-ap^2/2}|x_\ell\rangle &= \int_{-\infty}^{\infty}\frac{\mathrm{d}p}{2\pi}\exp\left[-\frac{ap^2}{2}+\mathrm{i}p(x_{\ell+1}-x_\ell)\right]\\ &= \frac{1}{\sqrt{2\pi a}}\exp\left[-\frac{(x_{\ell+1}-x_\ell)^2}{2a}\right].\end{aligned} \tag{1.5}$$

The momentum integral follows by completing the square; the Fourier normalization is $\langle x|p\rangle = \mathrm{e}^{\mathrm{i}px}/\sqrt{2\pi}$. The potential is diagonal in this basis. Consequently

$$Z(\beta) = \lim_{N\to\infty} (2\pi a)^{-N/2} \int_{\mathbb{R}^N} \prod_{\ell=0}^{N-1} \mathrm{d}x_\ell \, \mathrm{e}^{-S_a(x)}, \tag{1.6}$$

$$S_a(x) = \sum_{\ell=0}^{N-1} \left[\frac{(x_{\ell+1} - x_\ell)^2}{2a} + aV(x_\ell) \right]. \tag{1.7}$$

An insertion of q^2 in the trace multiplies the integrand by x_ℓ^2. For two insertions with $0 \le \tau \le \beta$, the operator expression is

$$G_\beta(\tau) = \frac{1}{Z} \operatorname{Tr}\left[\mathrm{e}^{-(\beta-\tau)H} q \, \mathrm{e}^{-\tau H} q\right] = \langle q_E(\tau) q \rangle_\beta, \qquad q_E(\tau) = \mathrm{e}^{\tau H} q \, \mathrm{e}^{-\tau H}. \tag{1.8}$$

Inserting coordinate states in the two evolution intervals gives the path factors $x(\tau)x(0)$. For either ordering of the times, the corresponding operator product is ordered in Euclidean time. In the vacuum limit, insertion of energy eigenstates between the two q operators gives the useful spectral interpretation

$$G(\tau) = \sum_n |\langle 0|q|n\rangle|^2 \mathrm{e}^{-(E_n - E_0)|\tau|}. \tag{1.9}$$

The decays measure quantum excitation energies; they do not describe physical friction. No classical trajectory has been selected: all paths are integrated over, with the weight specified by the quantum Hamiltonian.

The continuum shorthand for this limit is

$$Z = \int_{x(\beta)=x(0)} \mathcal{D}x \, \mathrm{e}^{-S_E[x]}, \qquad S_E[x] = \int_0^\beta \mathrm{d}\tau \left[\frac{1}{2} \left(\frac{\mathrm{d}x}{\mathrm{d}\tau} \right)^2 + \frac{\omega_0^2 x^2}{2} + \frac{\lambda x^4}{24} \right]. \tag{1.10}$$

The symbol $\mathcal{D}x$ includes the limiting measure in (1.6). It is not an ordinary product of independent uniform probabilities. The finite lattice defines the formulas; typical paths in the continuum measure need not have an ordinary derivative. The derivative in the action is understood through its lattice limit.

Key result. Quantum equilibrium has become a positive measure on entire paths, $\mathcal{P}[x] \propto \mathrm{e}^{-S_E[x]}$. It has not become the one-variable classical distribution $\mathrm{e}^{-\beta V(x)}$. The kinetic term couples different Euclidean times and is essential for quantum fluctuations.

1.3 Invert the free quadratic form: the propagator

For $\lambda = 0$, integrate the kinetic term by parts, using periodic boundary conditions. Then

$$S_0[x] = \frac{1}{2} \int_0^\beta \mathrm{d}\tau \, x(\tau) \mathcal{K}_0 x(\tau), \qquad \mathcal{K}_0 = -\partial_\tau^2 + \omega_0^2. \tag{1.11}$$

A propagator here means the two-point correlation, not an additional particle:

$$G_0(\tau - \tau') = \langle x(\tau) x(\tau') \rangle_0. \tag{1.12}$$

The subscript 0 denotes the normalized free Gaussian measure. Why is its covariance $\mathcal{K}_0^{-1}$? First do an ordinary d-dimensional integral with a real symmetric positive matrix A, a vector X, and a source vector b:

$$-\frac{1}{2} X^{\mathsf{T}} A X + b^{\mathsf{T}} X = -\frac{1}{2} (X - A^{-1}b)^{\mathsf{T}} A (X - A^{-1}b) + \frac{1}{2} b^{\mathsf{T}} A^{-1} b, \tag{1.13}$$

$$\frac{\int \mathrm{d}^d X \, \mathrm{e}^{-X^{\mathsf{T}} A X/2 + b^{\mathsf{T}} X}}{\int \mathrm{d}^d X \, \mathrm{e}^{-X^{\mathsf{T}} A X/2}} = \mathrm{e}^{b^{\mathsf{T}} A^{-1} b/2}, \tag{1.14}$$

$$\langle X_i X_j \rangle_0 = \frac{\partial^2}{\partial b_i \partial b_j} \mathrm{e}^{b^{\mathsf{T}} A^{-1} b/2} \bigg|_{b=0} = (A^{-1})_{ij}. \tag{1.15}$$

The continuum version is the same inverse problem:

$$\mathcal{K}_0 G_0(\tau) = \delta(\tau). \tag{1.16}$$

At zero temperature, $\tau \in \mathbb{R}$, and we demand decay as $|\tau| \to \infty$. For $\tau \neq 0$, the solution is $C\mathrm{e}^{-\omega_0|\tau|}$. Integrating (1.16) across $\tau = 0$ determines its normalization:

$$-[G_0'(0^+) - G_0'(0^-)] = 1 \quad \Longrightarrow \quad 2\omega_0 C = 1. \tag{1.17}$$

Thus

$$\boxed{G_0(\tau) = \frac{\mathrm{e}^{-\omega_0|\tau|}}{2\omega_0}, \qquad \langle q^2 \rangle_{\mathrm{vac},0} = G_0(0) = \frac{1}{2\omega_0}.} \tag{1.18}$$

Our Fourier convention and the corresponding inverse are

$$G_0(\tau) = \int_{-\infty}^{\infty} \frac{\mathrm{d}\nu}{2\pi} \mathrm{e}^{\mathrm{i}\nu\tau} \widetilde{G}_0(\nu), \qquad \widetilde{G}_0(\nu) = \frac{1}{\nu^2 + \omega_0^2}. \tag{1.19}$$

The variable ν is a Euclidean frequency. It is not yet the real frequency of an externally applied physical force.

The periodic finite-temperature inverse follows by summing images:

$$G_{0,\beta}(\tau) = \sum_{n \in \mathbb{Z}} \frac{\mathrm{e}^{-\omega_0|\tau + n\beta|}}{2\omega_0}. \tag{1.20}$$

Each term solves the inverse problem with its source translated by $n\beta$; the sum is periodic and therefore has the required periodic delta source. At coincident times, the geometric series gives

$$G_{0,\beta}(0) = \frac{1}{2\omega_0} \left(1 + 2 \sum_{n=1}^{\infty} \mathrm{e}^{-n\beta\omega_0} \right) = \frac{1}{2\omega_0} \coth \frac{\beta\omega_0}{2}. \tag{1.21}$$

The limits identify the physics: $G_{0,\beta}(0) \to 1/(2\omega_0)$ as $T \to 0$, whereas $G_{0,\beta}(0) \sim k_B T/\omega_0^2$ when $k_B T \gg \omega_0$. Quantum vacuum fluctuations survive when thermal fluctuations vanish.

Self-study question 1. Why would a one-coordinate Langevin process with stationary density proportional to $\mathrm{e}^{-\beta V(x)}$ fail to reproduce the zero-temperature variance?

2 The first interaction correction and independent checks

A first-order calculation of the interacting variance provides a quantitative benchmark for the stochastic construction.

2.1 Wick's rule follows from the Gaussian generating function

Introduce a real source function $J(\tau)$, coupled linearly to the path. The same square completion as above gives

$$\frac{Z_0[J]}{Z_0[0]} = \exp\left[\frac{1}{2}\int \mathrm{d}\tau\ \mathrm{d}\tau'\, J(\tau)G_0(\tau-\tau')J(\tau')\right]. \tag{2.1}$$

Differentiation with respect to $J(\tau_i)$ inserts $x(\tau_i)$. Every surviving term at $J=0$ must pair all source derivatives into factors of G_0. This is Wick's rule. Odd moments vanish; for four variables,

$$\langle x_1x_2x_3x_4\rangle_0 = G_{12}G_{34} + G_{13}G_{24} + G_{14}G_{23}, \qquad G_{ij} = G_0(\tau_i-\tau_j). \tag{2.2}$$

In particular, $\langle x(u)^4\rangle_0 = 3G_0(0)^2$. Here u is an integration variable for Euclidean time.

2.2 Expand the numerator and the normalization together

Write $S_E = S_0 + \lambda I$, where

$$I[x] = \frac{1}{24}\int_0^\beta \mathrm{d}u\, x(u)^4. \tag{2.3}$$

The exact normalized two-point function is

$$\begin{aligned} G_\beta(\tau) &= \frac{\langle x(\tau)x(0)\mathrm{e}^{-\lambda I}\rangle_0}{\langle \mathrm{e}^{-\lambda I}\rangle_0} \\ &= G_{0,\beta}(\tau) - \lambda\left[\langle x(\tau)x(0)I\rangle_0 - G_{0,\beta}(\tau)\langle I\rangle_0\right] + O(\lambda^2). \end{aligned} \tag{2.4}$$

The subtraction comes from $1/(1-\lambda\langle I\rangle_0) = 1+\lambda\langle I\rangle_0 + O(\lambda^2)$. It is required even in ordinary probability theory. In the next formulas we take the vacuum limit, so all G_0 are the zero-temperature functions and the u integral runs over the real line.

There are $5\cdot 3\cdot 1 = 15$ pairings of six Gaussian variables. For $x(\tau)x(0)x(u)^4$ they split into two classes:

$$\langle x(\tau)x(0)x(u)^4\rangle_0 = 3G_0(\tau)G_0(0)^2 + 12G_0(\tau-u)G_0(u)G_0(0). \tag{2.5}$$

In the first class the two external variables pair with one another; the four remaining variables have three pairings. In the second class there are four choices for the partner of $x(\tau)$ and three for that of $x(0)$; the remaining two variables at u pair together. These $3+12$ possibilities exhaust the 15 pairings.

The first class cancels against the normalization term in (2.4). Therefore

$$\boxed{G(\tau) = G_0(\tau) - \frac{\lambda}{2}G_0(0)\int_{-\infty}^{\infty}\mathrm{d}u\, G_0(\tau-u)G_0(u) + O(\lambda^2).} \tag{2.6}$$

This derives the diagram and its factor $12/24 = 1/2$; no diagrammatic rule has been assumed.

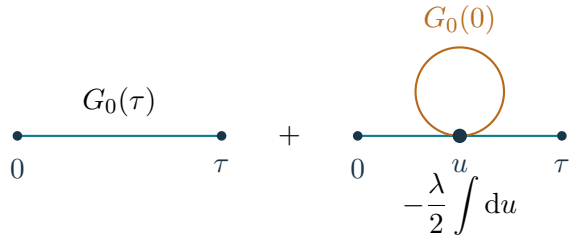


Figure. The line means a free correlation. The loop pairs the two unused fields at the same vertex. This "tadpole" is a compact picture of the second term in (2.6), not a literal orbit of a particle. The displayed plus sign adds the correction whose coefficient is negative.

2.3 Evaluate the correction, rather than stopping at a diagram

A convolution becomes a product under the Fourier convention (1.19). Since $G_0(0) = 1/(2\omega_0)$,

$$\boxed{\widetilde{G}(\nu) = \frac{1}{\nu^2+\omega_0^2} - \frac{\lambda}{4\omega_0}\frac{1}{(\nu^2+\omega_0^2)^2} + O(\lambda^2).} \tag{2.7}$$

The combination $\Sigma = \lambda/(4\omega_0)$ is called the first-order self-energy: expanding $1/(\nu^2+\omega_0^2+\Sigma)$ to first order in Σ reproduces this expression. This rewriting is an order-by-order identity, not an exact solution of the interacting oscillator.

At coincident times we can do the integral without contour integration:

$$\int_{-\infty}^{\infty} \mathrm{d}u\, G_0(u)^2 = \frac{2}{4\omega_0^2}\int_0^{\infty} \mathrm{d}u\, \mathrm{e}^{-2\omega_0 u} = \frac{1}{4\omega_0^3}, \tag{2.8}$$

$$\boxed{\langle q^2\rangle_{\mathrm{vac}} = G(0) = \frac{1}{2\omega_0} - \frac{\lambda}{16\omega_0^4} + O(\lambda^2).} \tag{2.9}$$

The sign has a physical explanation: a positive quartic potential penalizes large displacements and decreases the vacuum width.

The full first-order function also follows by differentiating the free one:

$$\int \frac{\mathrm{d}\nu}{2\pi}\frac{\mathrm{e}^{\mathrm{i}\nu\tau}}{(\nu^2+\omega_0^2)^2} = -\frac{\partial G_0(\tau)}{\partial(\omega_0^2)} = \frac{\mathrm{e}^{-\omega_0|\tau|}}{4\omega_0^3}(1+\omega_0|\tau|), \tag{2.10}$$

$$G(\tau) = \frac{\mathrm{e}^{-\omega_0|\tau|}}{2\omega_0}\left[1 - \frac{\lambda}{8\omega_0^3}(1+\omega_0|\tau|)\right] + O(\lambda^2). \tag{2.11}$$

The expansion is at fixed τ. Its term proportional to $\lambda|\tau|$ should not be extrapolated to arbitrarily long times at fixed λ: a small change of the decay exponent accumulates with time. To this order the shifted decay frequency is $\omega_0 + \lambda/(8\omega_0^2)$.

2.4 Check the result through energy, with a different calculation

Expand the partition function itself at finite β:

$$\log Z = \log Z_0 + \log\langle \mathrm{e}^{-\lambda I}\rangle_0 = \log Z_0 - \lambda\langle I\rangle_0 + O(\lambda^2), \tag{2.12}$$

$$\langle I\rangle_0 = \frac{\beta}{24}\, 3G_{0,\beta}(0)^2. \tag{2.13}$$

The free ground-state energy can be checked directly in the coordinate representation $p = -\mathrm{i}\,\partial_x$:

$$\psi_0(x) = \left(\frac{\omega_0}{\pi}\right)^{1/4}\mathrm{e}^{-\omega_0 x^2/2}, \qquad \psi_0'' = (\omega_0^2 x^2 - \omega_0)\psi_0, \tag{2.14}$$

$$\left(-\frac{1}{2}\partial_x^2 + \frac{\omega_0^2 x^2}{2}\right)\psi_0 = \frac{\omega_0}{2}\psi_0. \tag{2.15}$$

It is the lowest state: writing $a_0 = (\omega_0 q + \mathrm{i}p)/\sqrt{2\omega_0}$ gives $H_0 = \omega_0(a_0^\dagger a_0 + 1/2) \geq \omega_0/2$, and $a_0\psi_0 = 0$. The dagger denotes the Hermitian adjoint. Thus $E_0^{(0)} = \omega_0/2$. Equations (1.3) and (1.21) now give

$$\boxed{E_0 = \frac{\omega_0}{2} + \frac{\lambda}{32\omega_0^2} + O(\lambda^2).} \tag{2.16}$$

For a normalized eigenstate, differentiation of $H|0\rangle = E_0|0\rangle$ gives the Hellmann–Feynman identity. The derivatives of the state cancel because $\partial\langle 0|0\rangle = 0$:

$$\frac{\partial E_0}{\partial(\omega_0^2)} = \langle 0|\frac{\partial H}{\partial(\omega_0^2)}|0\rangle = \frac{1}{2}\langle q^2\rangle_{\mathrm{vac}}, \tag{2.17}$$

$$2\frac{\partial}{\partial(\omega_0^2)}\left(\frac{\omega_0}{2} + \frac{\lambda}{32\omega_0^2}\right) = \frac{1}{2\omega_0} - \frac{\lambda}{16\omega_0^4}. \tag{2.18}$$

The derivative is taken with λ fixed. It reproduces the two-point calculation. In these formulas $O(\lambda^2)$ means an expansion at fixed $\omega_0 > 0$; the actual small parameter is λ/ω_0^3.

Self-study question 2. Where did the vacuum bubble go, and why does removing it matter?

Key result. We have a definite target: an interacting quantum variance and its first-order propagator, checked against the ground-state energy. Next we construct a stochastic dynamics whose equilibrium measure is precisely the measure used in this calculation.

3 Stochastic quantization on a finite lattice

A finite Euclidean lattice turns path sampling into an ordinary diffusion with a specified invariant measure.

The construction below is the finite-lattice version of the Parisi–Wu prescription [1]; the interacting scalar-field extension uses the same drift and perturbative logic [2].

3.1 Keep a regulator so every probability density is ordinary

Return to a periodic path lattice with $N \geq 3$ sites, $a = \beta/N$, and indices understood modulo N. Rescale its coordinates as $X_\ell = \sqrt{a}\, x_\ell$. Equation (1.7) becomes

$$S_a(X) = \frac{1}{2} X^{\mathsf{T}} A X + \frac{\lambda}{24a} \sum_{\ell=0}^{N-1} X_\ell^4, \tag{3.1}$$

$$A_{\ell m} = \left(\frac{2}{a^2} + \omega_0^2 \right) \delta_{\ell m} - \frac{1}{a^2} \left(\delta_{\ell+1,m} + \delta_{\ell-1,m} \right). \tag{3.2}$$

Here X is a real N-component vector and $\delta_{\ell m}$ is the periodic Kronecker delta. A constant Jacobian from the rescaling cancels in normalized expectations. Define the finite normalizing integral

$$Z_a = \int_{\mathbb{R}^N} \mathrm{d}^N X\, \mathrm{e}^{-S_a(X)}, \qquad \rho_\star(X) = Z_a^{-1} \mathrm{e}^{-S_a(X)}. \tag{3.3}$$

This Z_a is the normalization of the rescaled coordinate density; recovering the absolute quantum partition function also requires the known measure prefactors in (1.6). We use Z_a only in normalized correlations.

The matrix spectrum is easy to check by applying A to $\mathrm{e}^{2\pi \mathrm{i} j \ell / N}$:

$$\kappa_j = \omega_0^2 + \frac{2 - \mathrm{e}^{2\pi \mathrm{i} j/N} - \mathrm{e}^{-2\pi \mathrm{i} j/N}}{a^2} = \omega_0^2 + \frac{4}{a^2} \sin^2 \frac{\pi j}{N}, \quad j = 0, \ldots, N-1. \tag{3.4}$$

All eigenvalues are positive. Complex exponentials were used only to calculate them; the stochastic normal coordinates below can be chosen real and orthonormal. In particular $X^{\mathsf{T}} A X \geq \omega_0^2 |X|^2$, so Z_a is finite.

3.2 Specify the random increments and derive their evolution equation

Let $W_\ell(s)$ be independent standard Wiener processes, meaning Gaussian increments with

$$\mathbb{E}[\mathrm{d}W_\ell] = 0, \qquad \mathbb{E}[\mathrm{d}W_\ell\, \mathrm{d}W_m] = \delta_{\ell m}\, \mathrm{d}s, \tag{3.5}$$

and independent increments over disjoint intervals. The symbol $\mathbb{E}$ denotes averaging over these random histories. Define the Itô stochastic differential equation

$$\boxed{\mathrm{d}X_\ell = -\frac{\partial S_a}{\partial X_\ell}\, \mathrm{d}s + \sqrt{2}\, \mathrm{d}W_\ell, \qquad \frac{\partial S_a}{\partial X_\ell} = (AX)_\ell + \frac{\lambda}{6a} X_\ell^3.} \tag{3.6}$$

The drift moves toward smaller action and the noise prevents collapse onto its minimum. The normalization $\sqrt{2}$ fixes the intended weight to e^{-S_a}. The parameter s is independent of τ; with our rescaling its units are inverse frequency squared.

To derive the equation for the probability density $\rho(X, s)$, take any smooth test function $f(X)$ that decays sufficiently rapidly. Over a small interval Δs, the conditional increment moments are

$$\mathbb{E}[\Delta X_i | X] = -\partial_i S_a\, \Delta s + o(\Delta s), \qquad \mathbb{E}[\Delta X_i \Delta X_j | X] = 2\delta_{ij} \Delta s + o(\Delta s), \tag{3.7}$$

where $\partial_i = \partial/\partial X_i$. Taylor expansion includes the second derivative because a noise increment is of order $\sqrt{\Delta s}$:

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}[f(X)] = \mathbb{E}\left[-\sum_i (\partial_i S_a)(\partial_i f) + \sum_i \partial_i^2 f\right] \tag{3.8}$$

$$= \int \mathrm{d}^N X\, f(X) \sum_i \partial_i \left[(\partial_i S_a)\rho + \partial_i \rho\right]. \tag{3.9}$$

The last line integrates by parts once for the drift and twice for the noise term. Boundary terms vanish for the decaying densities under consideration. Since f is arbitrary,

$$\boxed{\partial_s \rho = \sum_i \partial_i \left[(\partial_i S_a)\rho + \partial_i \rho\right] = -\sum_i \partial_i \mathcal{J}_i, \quad \mathcal{J}_i = -(\partial_i S_a)\rho - \partial_i \rho.} \tag{3.10}$$

This is the Fokker–Planck equation. The current $\mathcal{J}_i$ is probability flow in configuration space, not a physical electric current.

3.3 Check equilibrium and explain why it is reached in this model

For the proposed target density,

$$\partial_i \rho_\star = -(\partial_i S_a)\rho_\star \quad \Longrightarrow \quad \mathcal{J}_i[\rho_\star] = 0 \quad \Longrightarrow \quad \partial_s \rho_\star = 0. \tag{3.11}$$

Zero current is stronger than stationarity: drift and diffusion balance separately in each coordinate. Finding a stationary density alone would not prove that a process starting elsewhere approaches it. Here confinement and convexity give a direct check.

For two solutions $X(s)$ and $Y(s)$ driven by the same Wiener increments, the noise cancels in their difference $D = X - Y$:

$$\frac{\mathrm{d}}{\mathrm{d}s}|D|^2 = -2D \cdot [\nabla S_a(X) - \nabla S_a(Y)] \tag{3.12}$$

$$= -2D^{\mathsf{T}} A D - \frac{\lambda}{3a}\sum_i (X_i - Y_i)(X_i^3 - Y_i^3) \le -2\omega_0^2 |D|^2. \tag{3.13}$$

The inequality uses $(X_i - Y_i)(X_i^3 - Y_i^3) = (X_i - Y_i)^2(X_i^2 + X_i Y_i + Y_i^2) \ge 0$. Multiplying by $\mathrm{e}^{2\omega_0^2 s}$ and integrating gives

$$|X(s) - Y(s)|^2 \le \mathrm{e}^{-2\omega_0^2 s}|X(0) - Y(0)|^2. \tag{3.14}$$

Choosing $Y(0)$ from $\rho_\star$ makes $Y(s)$ stationary. The bound therefore shows convergence toward that stationary process in mean-square coupling for initial distributions with finite second moments. It also rules out two distinct invariant distributions with finite second moments. The underlying diffusion is well defined globally: applying the same increment calculation to $|X|^2$ gives

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}|X|^2 = -2\mathbb{E}[X \cdot \nabla S_a(X)] + 2N \le -2\omega_0^2 \mathbb{E}|X|^2 + 2N. \tag{3.15}$$

The confining polynomial drift thus supplies a moment bound rather than a finite-time escape. These claims concern the continuous stochastic equation. A finite-step explicit recurrence requires its own stability analysis.

3.4 Solve the free stochastic process and recover the propagator

Set $\lambda = 0$ and choose a real orthogonal matrix O satisfying $O^{\mathsf{T}} A O = \mathrm{diag}(\kappa_0, \ldots, \kappa_{N-1})$. The real coordinates $y = O^{\mathsf{T}} X$ have independent transformed Wiener noises because an orthogonal transformation preserves their covariance. Each mode obeys

$$\mathrm{d}y_j = -\kappa_j y_j\, \mathrm{d}s + \sqrt{2}\, \mathrm{d}\widehat{W}_j. \tag{3.16}$$

Multiplication by $\mathrm{e}^{\kappa_j s}$ and integration yields

$$y_j(s) = \mathrm{e}^{-\kappa_j s} y_j(0) + \sqrt{2} \int_0^s \mathrm{e}^{-\kappa_j (s-u)} \, \mathrm{d}\widehat{W}_j(u). \tag{3.17}$$

This linear Gaussian diffusion is the Ornstein–Uhlenbeck process. For a deterministic initial value its mean is $\mathrm{e}^{-\kappa_j s} y_j(0)$ and its variance is

$$\operatorname{Var} y_j(s) = 2 \int_0^s \mathrm{d}u \, \mathrm{e}^{-2\kappa_j (s-u)} = \frac{1 - \mathrm{e}^{-2\kappa_j s}}{\kappa_j}. \tag{3.18}$$

The first equality follows directly from independent noise increments: in a discretized stochastic integral, cross terms have zero average and the diagonal term is the squared coefficient times the interval length. At large s, the covariance is therefore

$$\mathbb{E}_\star[X_\ell X_m] = (A^{-1})_{\ell m}, \qquad \mathbb{E}_\star[x_\ell x_m] = \frac{1}{a} (A^{-1})_{\ell m}. \tag{3.19}$$

This is exactly the inverse quadratic form derived from the quantum Gaussian integral, rather than a fitted agreement.

Using the Fourier eigenvectors to display the inverse gives

$$\frac{1}{a}(A^{-1})_{\ell m} = \frac{1}{aN} \sum_{j=0}^{N-1} \frac{\mathrm{e}^{2\pi \mathrm{i} j(\ell - m)/N}}{\kappa_j} \xrightarrow[a \to 0]{\beta \text{ fixed}} \frac{1}{\beta} \sum_{n \in \mathbb{Z}} \frac{\mathrm{e}^{\mathrm{i}\nu_n (\tau - \tau')}}{\nu_n^2 + \omega_0^2}, \quad \nu_n = \frac{2\pi n}{\beta}, \tag{3.20}$$

$$\xrightarrow[\beta \to \infty]{} \quad \int \frac{\mathrm{d}\nu}{2\pi} \frac{\mathrm{e}^{\mathrm{i}\nu(\tau - \tau')}}{\nu^2 + \omega_0^2} = \frac{\mathrm{e}^{-\omega_0 |\tau - \tau'|}}{2\omega_0}. \tag{3.21}$$

The index n in the continuum sum runs over positive and negative Euclidean frequencies; the finite periodic lattice labels near $j = N$ supply the negative ones. We have explicitly taken stochastic equilibrium first, removed the Euclidean lattice spacing second, and taken the vacuum limit third.

3.5 Recover the interaction correction directly from the noise equation

There is a useful check stronger than comparing the free solutions. We can derive the first interacting correction from the stochastic equation without expanding its stationary density. Let $C_{ij} = \mathbb{E}_\star[X_i X_j]$ denote its stationary covariance. The mean vanishes because the action is unchanged by $X \mapsto -X$. The product increment contains a noise term of order $\mathrm{d}s$:

$$\mathrm{d}(X_i X_j) = X_i \, \mathrm{d}X_j + X_j \, \mathrm{d}X_i + \mathrm{d}X_i \, \mathrm{d}X_j, \qquad \mathbb{E}[\mathrm{d}X_i \, \mathrm{d}X_j] = 2\delta_{ij} \, \mathrm{d}s. \tag{3.22}$$

Substitute (3.6), average, and set the time derivative to zero in equilibrium. Since A is symmetric,

$$(AC + CA)_{ij} + \frac{\lambda}{6a} \left[\mathbb{E}_\star(X_i^3 X_j) + \mathbb{E}_\star(X_i X_j^3) \right] = 2\delta_{ij}. \tag{3.23}$$

This is an exact moment equation. At general coupling it is not closed: the second moment depends on fourth moments. Perturbation theory closes it to the accuracy we need.

Write $C = C_0 + \delta C + O(\lambda^2)$, with $C_0 = A^{-1}$ and δC of first order in λ. Inside the term already multiplied by λ, use the free Gaussian average. Translation invariance makes the free diagonal $d = (C_0)_{ii}$ independent of i. Wick's rule gives

$$\mathbb{E}_0(X_i^3 X_j) = 3d(C_0)_{ij}, \qquad \mathbb{E}_0(X_i X_j^3) = 3d(C_0)_{ij}. \tag{3.24}$$

The zeroth-order terms in (3.23) cancel because $AC_0 + C_0 A = 2I$, where I is the identity matrix. What remains is

$$A\delta C + \delta C A = -\frac{\lambda d}{a} A^{-1} \quad \Longrightarrow \quad \boxed{\delta C = -\frac{\lambda d}{2a} A^{-2}.} \tag{3.25}$$

Multiplying the proposed solution by A from both sides checks it immediately. It is unique: in the eigenbasis of A, each matrix element of the left side is $(\kappa_i + \kappa_j)\delta C_{ij}$, and all these coefficients are positive.

Convert back to path coordinates. The free equal-time variance is d/a, and the correction to the path correlator is $\delta C_{ij}/a$. Taking the same limits as in (3.21),

$$\frac{d}{a} \longrightarrow G_0(0), \qquad \frac{1}{a}(A^{-2})_{ij} \longrightarrow \int \frac{\mathrm{d}\nu}{2\pi} \frac{\mathrm{e}^{\mathrm{i}\nu(\tau-\tau')}}{(\nu^2+\omega_0^2)^2}. \tag{3.26}$$

Consequently the stochastic moment equation yields

$$\delta\widetilde{G}(\nu) = -\frac{\lambda}{2} G_0(0)\widetilde{G}_0(\nu)^2 = -\frac{\lambda}{4\omega_0}\frac{1}{(\nu^2+\omega_0^2)^2}. \tag{3.27}$$

This is precisely (2.7). The quartic quantum correction has now been obtained from a normalized path integral and independently from stationary stochastic moment equations. Both use Gaussian contractions at first order; neither requires coding or training a network.

3.6 Only now write the continuum Langevin equation

Varying the action under $x \mapsto x + \delta x$, and integrating its kinetic term by parts, gives

$$\delta S_E = \int \mathrm{d}\tau \left[-\partial_\tau^2 x + \omega_0^2 x + \frac{\lambda}{6}x^3\right]\delta x, \qquad \frac{\delta S_E}{\delta x(\tau)} = -\partial_\tau^2 x + \omega_0^2 x + \frac{\lambda}{6}x^3. \tag{3.28}$$

The functional derivative is defined by the first equality: it is the coefficient of an arbitrary variation at each time. Equation (3.6) becomes, in continuum shorthand,

$$\boxed{\partial_s x(\tau,s) = \partial_\tau^2 x - \omega_0^2 x - \frac{\lambda}{6}x^3 + \eta(\tau,s), \quad \mathbb{E}[\eta(\tau,s)\eta(\tau',s')] = 2\delta(\tau-\tau')\delta(s-s').} \tag{3.29}$$

The noise has zero mean. Its continuum normalization follows from dividing the lattice equation by $\sqrt{a}$: the site noise covariance becomes $2\delta_{\ell m}/a$, which represents $2\delta(\tau-\tau')$. White noise is a generalized random function defined through its integrals, not a differentiable force history. Equation (3.29) summarizes the regulated construction; our probability proof was performed on the finite lattice.

Self-study question 3. Does the decay rate κ_j in stochastic time measure the oscillator's physical damping?

Key result. The noisy evolution reproduces quantum correlations because its equilibrium distribution is the quantum path measure. The equality is a statement about stationary averages, not an identification of stochastic trajectories with quantum real-time motion.

4 Recurrent updates: stability and bias

Time discretization connects this diffusion to a recurrent update, whose stability and invariant distribution require separate checks.

Same fluctuations, different temporal correlations

Let y_n be a real scalar state at step n, r a real recurrence coefficient, and $\sigma^2 > 0$ a prescribed variance. Let ξ_n be independent Gaussian numbers of mean zero and variance one, independent of the current state y_n. Consider

$$y_{n+1} = r\, y_n + \sqrt{(1-r^2)\sigma^2}\, \xi_n, \qquad |r| < 1.$$

If $\mathbb{E}[y_n^2] = \sigma^2$, then independence gives

$$\mathbb{E}[y_{n+1}^2] = r^2\sigma^2 + (1-r^2)\sigma^2 = \sigma^2.$$

Yet iterating the update gives $\mathbb{E}[y_{n+k}y_n] = \sigma^2 r^k$ in its zero-mean stationary state. Every allowed r has the same Gaussian invariant distribution, but a different temporal correlation. Identical invariant distributions therefore do not imply identical temporal correlations.

4.1 A path is the recurrent state

Let n label stochastic updates, with step $\epsilon > 0$ in s. Let $\xi_{\ell,n}$ be independent normal random variables with mean zero and variance one: $\mathbb{E}[\xi_{\ell,n}\xi_{m,r}] = \delta_{\ell m}\delta_{nr}$. The formal explicit Euler update of (3.6) is

$$X_{n+1} = X_n - \epsilon \nabla S_a(X_n) + \sqrt{2\epsilon}\, \xi_n, \tag{4.1}$$

$$X_{\ell,n+1} = \left[1 - \epsilon\left(\frac{2}{a^2} + \omega_0^2\right)\right] X_{\ell,n} + \frac{\epsilon}{a^2}(X_{\ell+1,n} + X_{\ell-1,n}) - \frac{\epsilon\lambda}{6a} X_{\ell,n}^3 + \sqrt{2\epsilon}\, \xi_{\ell,n}. \tag{4.2}$$

The vector X_n is reused at the next step: this is recurrence. Its N components store a Euclidean path, not N different physical particles. Equation (4.2) has the structure of a noisy recurrent network with fixed weights and a local cubic nonlinearity. This structural correspondence introduces no learning rule and no claim about training performance.

4.2 Solve stability and discretization bias exactly in the free case

For $\lambda = 0$, a real normal mode with eigenvalue $\kappa > 0$ obeys

$$y_{n+1} = r\, y_n + \sqrt{2\epsilon}\, \xi_n, \qquad r = 1 - \epsilon\kappa. \tag{4.3}$$

Iterating gives

$$y_n = r^n y_0 + \sqrt{2\epsilon} \sum_{k=0}^{n-1} r^{n-1-k} \xi_k. \tag{4.4}$$

Initial conditions decay and the variance series converges precisely when

$$|r| < 1 \quad \Longleftrightarrow \quad 0 < \epsilon\kappa < 2. \tag{4.5}$$

For the full free lattice require this for every eigenvalue. The sufficient bound $\epsilon < 2/(\omega_0^2 + 4/a^2)$ follows from (3.4). Even when stable, the limiting variance is biased:

$$C_\epsilon = r^2 C_\epsilon + 2\epsilon, \tag{4.6}$$

$$\boxed{C_\epsilon = \frac{2\epsilon}{1-(1-\epsilon\kappa)^2} = \frac{1}{\kappa(1-\epsilon\kappa/2)} \neq \frac{1}{\kappa}.} \tag{4.7}$$

The correct variance is $1/\kappa$. The explicit recurrence inflates it by a relative amount $\epsilon\kappa/2 + O(\epsilon^2)$; stability does not imply the correct stationary distribution.

For a free mode we can remove this error completely by using the exact solution (3.17) over one finite interval:

$$\boxed{y_{n+1} = \mathrm{e}^{-\kappa\epsilon} y_n + \sqrt{\frac{1-\mathrm{e}^{-2\kappa\epsilon}}{\kappa}}\,\xi_n.} \tag{4.8}$$

Substituting $C = 1/\kappa$ into its variance recurrence proves that it preserves the correct Gaussian equilibrium at any positive ϵ. This exact formula is available because the free equation is linear.

4.3 Why the quartic case needs a separate warning

The continuous quartic diffusion is confining, as proved in (3.13)–(3.15). Explicit Euler need not inherit that property. For a sufficiently large component, its deterministic update is dominated by

$$X_{\ell,n+1} \simeq -\frac{\epsilon\lambda}{6a} X_{\ell,n}^3. \tag{4.9}$$

At fixed ϵ and $\lambda > 0$ the magnitude of this term eventually exceeds $|X_{\ell,n}|$: the update can overshoot the confining potential and drive the next magnitude larger. Gaussian noise can reach large values. The free-mode bound therefore supplies neither global stability nor an exact invariant measure for the interacting explicit scheme. Strong moment-convergence claims for this naive scheme would require additional justification and can fail for superlinear drifts.

We use (4.2) to expose the recurrent structure and analyze the free benchmark. We do not prescribe it as a validated finite-step sampler of the quartic theory. The equilibrium construction of Section 3 concerns the continuous stochastic equation. No numerical implementation is required for either conclusion.

Self-study question 4. Which two independent checks would a proposed stochastic recurrent sampler have to pass?

Key result. First endpoint: quantum perturbation theory fixes the observable; a derived stochastic dynamics reproduces its underlying measure; and the recurrent update exposes concrete stability and bias conditions. Next we introduce a physical bath. Its memory will live in physical time t, not in the sampling time s.

5 Eliminating a quantum environment

A Hamiltonian oscillator bath determines the memory force, quantum fluctuations and initial-condition term together.

5.1 The system, the bath and the initial preparation

Keep the system mass and $\hbar$ equal to one. The system coordinate and momentum are operators q, p, with $[q, p] = \mathrm{i}$. Here $[A, B] = AB - BA$ denotes a commutator. Introduce bath coordinates Q_j, momenta P_j, masses $m_j > 0$, frequencies $\omega_j > 0$ and real couplings c_j. At the initial time,

$$[Q_j, P_k] = \mathrm{i}\delta_{jk}, \qquad [Q_j, Q_k] = [P_j, P_k] = 0, \qquad [q, Q_j] = [q, P_j] = [p, Q_j] = [p, P_j] = 0. \tag{5.1}$$

The same equal-time canonical relations hold after unitary evolution of the complete system. Consider

$$H = \frac{p^2}{2} + V(q) + \sum_j \left\{ \frac{P_j^2}{2m_j} + \frac{m_j\omega_j^2}{2}\left(Q_j - \frac{c_j q}{m_j\omega_j^2}\right)^2 \right\} - f(t)q. \tag{5.2}$$

The real prescribed function $f(t)$ is an external force used to measure response, not a fluctuating bath operator. The elimination below allows differentiable V. For the explicitly solved response in Section 7, however, we set

$$V(q) = \frac{\omega_0^2 q^2}{2}, \qquad \omega_0 > 0, \qquad \boxed{\lambda = 0 \text{ in the solved bath problem.}} \tag{5.3}$$

The isolated anharmonic oscillator and the open harmonic oscillator are two controlled problems, not two exact descriptions of the same dynamics.

Expanding one square identifies the coupling and counterterm:

$$\frac{m_j\omega_j^2 Q_j^2}{2} - c_j q Q_j + \underbrace{\frac{c_j^2 q^2}{2m_j\omega_j^2}}_{\text{static counterterm}}. \tag{5.4}$$

At a fixed classical value of q, minimizing over Q_j gives $Q_j = c_j q/(m_j\omega_j^2)$ and the whole square is zero. Thus the bath does not replace the specified static potential $V(q)$ by $V(q) - \sum_j c_j^2 q^2/(2m_j\omega_j^2)$. It can still change dynamical response and the quantum reduced equilibrium state: the counterterm does not remove these effects.

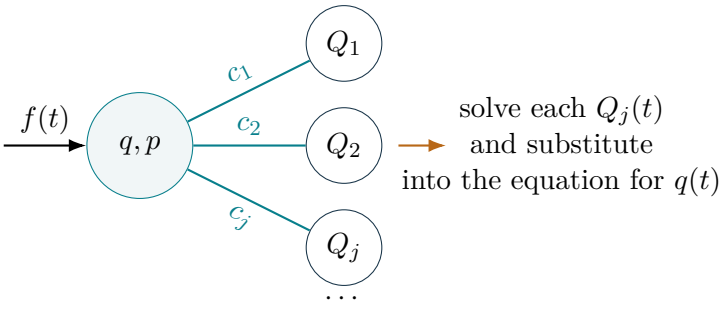


Figure. The complete oscillator system evolves locally in time. Retaining q and eliminating the bath coordinates leaves their history-dependent force.

To make the fluctuation calculation precise, specify the preparation:

$$\begin{aligned} \rho_{\mathrm{tot}}(0) &= \rho_S(0) \otimes \rho_B, \qquad \rho_B = \frac{\mathrm{e}^{-\beta H_B}}{Z_B}, \\ H_B &= \sum_j \left(\frac{P_j^2}{2m_j} + \frac{m_j\omega_j^2 Q_j^2}{2} \right), \qquad Z_B = \mathrm{Tr}_B\, \mathrm{e}^{-\beta H_B}. \end{aligned} \tag{5.5}$$

The density operator $\rho_S(0)$ specifies the initial system state; ρ_B is the thermal state of the *bare* bath, with $\beta = (k_B T)^{-1}$. A density operator has nonnegative eigenvalues and unit trace; $\langle A \rangle = \mathrm{Tr}(\rho A)$. This product preparation has no initial system–bath correlations. It is generally not the thermal state of the coupled Hamiltonian.

5.2 Heisenberg equations and the forced-oscillator solution

In the Heisenberg picture the state is fixed and operators evolve. Without explicit operator time dependence, $\dot{A} = \mathrm{i}[H, A]$. For a polynomial potential, $[q^n, p] = \mathrm{i}nq^{n-1}$ proves $[V(q), p] = \mathrm{i}V'(q)$; suitable functions follow by extension. Applying these identities gives

$$\dot{q} = p, \qquad \dot{p} = -V'(q) + \sum_j c_j Q_j - \sum_j \frac{c_j^2}{m_j \omega_j^2} q + f(t), \tag{5.6}$$

$$\dot{Q}_j = P_j/m_j, \qquad \dot{P}_j = -m_j \omega_j^2 Q_j + c_j q. \tag{5.7}$$

Thus $\ddot{Q}_j + \omega_j^2 Q_j = c_j q(t)/m_j$, with solution

$$Q_j(t) = Q_j(0)\cos\omega_j t + \frac{P_j(0)}{m_j\omega_j}\sin\omega_j t + \frac{c_j}{m_j\omega_j}\int_0^t \sin[\omega_j(t-u)] q(u)\, \mathrm{d}u. \tag{5.8}$$

Verify the last term by differentiation. Its first derivative has no upper-limit term because $\sin 0 = 0$. Its second derivative gives $c_j q(t)/m_j$ from the upper limit and $-\omega_j^2$ times the original integral from the integrand. The integral and its first derivative vanish at $t = 0$, so both initial data agree. The argument remains valid for operator-valued $q(u)$: the bath equation is linear and no system operators were reordered.

5.3 The integration by parts that produces friction memory

Define the real kernel K and the operator force F:

$$K(t) = \sum_j \frac{c_j^2}{m_j\omega_j^2}\cos\omega_j t, \qquad F(t) = \sum_j c_j \left[Q_j(0)\cos\omega_j t + \frac{P_j(0)}{m_j\omega_j}\sin\omega_j t \right]. \tag{5.9}$$

Use K at nonnegative arguments in the causal convolution, and its even extension when discussing correlations. The crucial integration by parts has two boundary terms:

$$\begin{aligned} \int_0^t \sin[\omega_j(t-u)] q(u)\, \mathrm{d}u &= \frac{1}{\omega_j}\int_0^t \frac{\partial}{\partial u}\cos[\omega_j(t-u)] q(u)\, \mathrm{d}u \\ &= \frac{q(t) - q(0)\cos\omega_j t}{\omega_j} - \frac{1}{\omega_j}\int_0^t \cos[\omega_j(t-u)]\dot{q}(u)\, \mathrm{d}u. \end{aligned} \tag{5.10}$$

Multiply by $c_j^2/(m_j\omega_j)$ and sum:

$$\sum_j c_j Q_j(t) = F(t) + K(0)q(t) - K(t)q(0) - \int_0^t K(t-u)\dot{q}(u)\, \mathrm{d}u. \tag{5.11}$$

Substitution in (5.6) cancels $K(0)q(t)$ against the counterterm contribution and yields

$$\boxed{\ddot{q}(t) + V'(q(t)) + \int_0^t K(t-u)\dot{q}(u)\, \mathrm{d}u = F(t) - K(t)q(0) + f(t).} \tag{5.12}$$

This generalized quantum Langevin equation is an exact operator identity for the specified oscillator bath, before weak-coupling, Markov or classical approximations [3, 4].

The integral is causal: only earlier velocities enter. $F(t)$ contains bath initial data. The *initial-slip term* $-K(t)q(0)$ retains the mismatch between our initially unshifted bath and the coupled system coordinate. One can absorb it into a redefined force, but that force is then correlated with $q(0)$ and no longer has the bare-bath statistics derived next.

Self-study question 5. At $t = 0$ the memory integral vanishes. What equation must remain?

Key result. Solving hidden variables produces a memory equation. Friction, fluctuations and the preparation term arise together; none is an arbitrary addition.

For finitely many bath oscillators, $K(t)$ is a finite sum of cosines. The closed harmonic system has normal modes and finite-bath recurrences; it does not provide irreversible exponential damping for all times. A smooth decaying kernel requires a continuum bath limit or an approximation appropriate before recurrences.

6 Quantum noise and fluctuation–dissipation

The same bath spectrum fixes dissipation and force correlations.

6.1 One thermal oscillator supplies the calculation

Introduce annihilation and creation operators $b_j, b_j^\dagger$:

$$Q_j = \frac{b_j + b_j^\dagger}{\sqrt{2m_j\omega_j}}, \qquad P_j = -\mathrm{i}\sqrt{\frac{m_j\omega_j}{2}}(b_j - b_j^\dagger), \qquad [b_j, b_k^\dagger] = \delta_{jk}. \tag{6.1}$$

Then $H_{B,j} = \omega_j(b_j^\dagger b_j + 1/2)$. Occupation-number states satisfy $b_j^\dagger b_j|n\rangle = n|n\rangle$, $n = 0, 1, \ldots$. With $r = \mathrm{e}^{-\beta\omega_j}$ their thermal probabilities are $(1-r)r^n$, and

$$\bar{n}_j = (1-r)\sum_{n=0}^{\infty} nr^n = (1-r)r\frac{\mathrm{d}}{\mathrm{d}r}\frac{1}{1-r} = \frac{1}{\mathrm{e}^{\beta\omega_j} - 1}. \tag{6.2}$$

An operator b_j changes the occupation number, so it has no diagonal matrix elements in this ensemble. Independence of the bare bath modes gives

$$\langle b_j\rangle_B = 0, \quad \langle b_j b_k\rangle_B = 0, \quad \langle b_j^\dagger b_k\rangle_B = \delta_{jk}\bar{n}_j, \quad \langle b_j b_k^\dagger\rangle_B = \delta_{jk}(\bar{n}_j + 1). \tag{6.3}$$

Here $\langle A\rangle_B = \mathrm{Tr}_B(\rho_B A)$, which for bath-only operators equals the expectation in the product state (5.5). Rewriting the force gives

$$F(t) = \sum_j \frac{c_j}{\sqrt{2m_j\omega_j}}\left(b_j\mathrm{e}^{-\mathrm{i}\omega_j t} + b_j^\dagger \mathrm{e}^{\mathrm{i}\omega_j t}\right), \qquad \langle F(t)\rangle_B = 0. \tag{6.4}$$

Multiply two brackets and use (6.3). For $\Delta = t - u$,

$$\begin{aligned}\langle F(t)F(u)\rangle_B &= \sum_j \frac{c_j^2}{2m_j\omega_j}\left[(\bar{n}_j + 1)\mathrm{e}^{-\mathrm{i}\omega_j\Delta} + \bar{n}_j\mathrm{e}^{\mathrm{i}\omega_j\Delta}\right] \\ &= \sum_j \frac{c_j^2}{2m_j\omega_j}\left[(2\bar{n}_j + 1)\cos\omega_j\Delta - \mathrm{i}\sin\omega_j\Delta\right].\end{aligned} \tag{6.5}$$

These force correlations depend only on a time difference. This stationarity follows from the thermal bare bath; it does not imply that the arbitrarily prepared system coordinate $q(t)$ is stationary.

6.2 Symmetric fluctuations and the quantum commutator

The anticommutator is $\{A, B\} = AB + BA$. Symmetrizing (6.5) gives the real correlation

$$C_F(\Delta) \equiv \frac{1}{2}\langle\{F(t), F(u)\}\rangle_B = \sum_j \frac{c_j^2}{2m_j\omega_j} \coth\left(\frac{\beta\omega_j}{2}\right) \cos\omega_j\Delta, \tag{6.6}$$

where $2\bar{n}_j + 1 = (\mathrm{e}^{\beta\omega_j} + 1)/(\mathrm{e}^{\beta\omega_j} - 1) = \coth(\beta\omega_j/2)$. Antisymmetrizing instead, or directly using $[b_j, b_k^\dagger] = \delta_{jk}$, gives

$$[F(t), F(u)] = -\mathrm{i}\sum_j \frac{c_j^2}{m_j\omega_j} \sin\omega_j(t-u) = \mathrm{i}K'(t-u). \tag{6.7}$$

The prime is the derivative of the even extension of K. The commutator is a multiple of the identity, not a classical covariance. Ordinary real random variables commute. Therefore reproducing C_F with classical random noise does not reproduce the complete quantum force.

Self-study question 6. Does the force vanish as $T \to 0$?

6.3 Spectral density and fluctuation–dissipation

Package the bath frequencies and couplings into

$$J(\omega) = \frac{\pi}{2}\sum_j \frac{c_j^2}{m_j\omega_j}\delta(\omega - \omega_j), \qquad \omega > 0. \tag{6.8}$$

The Dirac delta obeys $\int \delta(\omega - \omega_j)g(\omega)\,\mathrm{d}\omega = g(\omega_j)$. A continuum bath can have smooth $J \geq 0$. Substitution gives the two linked formulas

$$K(t) = \frac{2}{\pi}\int_0^\infty \frac{J(\omega)}{\omega}\cos\omega t\,\mathrm{d}\omega, \tag{6.9}$$

$$C_F(t) = \frac{1}{\pi}\int_0^\infty J(\omega)\coth\left(\frac{\beta\omega}{2}\right)\cos\omega t\,\mathrm{d}\omega. \tag{6.10}$$

This is the fluctuation–dissipation relation for this bath: the same J fixes friction memory and thermal quantum fluctuations. Its assumptions include the specified thermal preparation [5, 4].

Define the two-sided symmetric noise spectrum by

$$S_F(\omega) = \int_{-\infty}^\infty \mathrm{e}^{\mathrm{i}\omega t}C_F(t)\,\mathrm{d}t. \tag{6.11}$$

Since $\int \mathrm{e}^{\mathrm{i}\omega t}\cos\omega' t\,\mathrm{d}t = \pi[\delta(\omega - \omega') + \delta(\omega + \omega')]$,

$$S_F(\omega) = J(|\omega|)\coth\left(\frac{\beta|\omega|}{2}\right), \qquad \omega \neq 0. \tag{6.12}$$

The value at zero is understood by a limit if it exists. The spectrum is even because we symmetrized; the unsymmetrized quantum spectrum need not be even.

If the bath has a finite frequency cutoff and $\beta\omega \ll 1$ throughout its contributing spectrum,

$$\coth(\beta\omega/2) = \frac{2}{\beta\omega} + O(\beta\omega), \qquad C_F(t) \simeq k_BT\,K(t). \tag{6.13}$$

For a classical thermal oscillator bath the latter relation is exact: equipartition gives $\langle Q_j^2\rangle = k_BT/(m_j\omega_j^2)$ and $\langle P_j^2\rangle = m_jk_BT$, which substituted in F yield k_BTK. For an unbounded quantum spectrum the expansion can instead be used in the low-frequency spectral range where $\beta|\omega| \ll 1$; it is not uniform at arbitrarily large frequency. An ultraviolet limit and a thermal limit may fail to commute.

Key result. A physical quantum memory model must connect its damping to its fluctuations, and distinguish symmetric fluctuations from quantum commutators.

7 An exact memory kernel and its recurrent representation

An exponential friction kernel admits an analytic causal response and an exact local representation of its memory.

7.1 A bath with one friction-memory time

Choose the Drude spectral density

$$J_D(\omega) = \gamma \frac{\Omega^2 \omega}{\omega^2 + \Omega^2}, \qquad \gamma > 0, \quad \Omega > 0. \tag{7.1}$$

γ is the low-frequency damping scale and Ω^{-1} the friction-memory time; both γ and Ω have dimensions of inverse time. Equation (6.9) gives

$$K(t) = \frac{2\gamma\Omega^2}{\pi} \int_0^\infty \frac{\cos \omega t}{\omega^2 + \Omega^2} \, \mathrm{d}\omega = \gamma\Omega \mathrm{e}^{-\Omega t}, \qquad t \geq 0. \tag{7.2}$$

To evaluate the integral, extend it to half the integral of $\mathrm{e}^{\mathrm{i}\omega t}/(\omega^2 + \Omega^2)$ over the real line. For $t > 0$, close the contour above: the pole at $\omega = \mathrm{i}\Omega$ has residue $\mathrm{e}^{-\Omega t}/(2\mathrm{i}\Omega)$, giving the full integral $\pi \mathrm{e}^{-\Omega t}/\Omega$. At $t = 0$, the same result follows by an arctangent integral.

The friction kernel is finite at zero, but the unregulated quantum force variance is not. At high frequency $J_D(\omega) \sim \gamma\Omega^2/\omega$ and $\coth(\beta\omega/2) \to 1$, so $C_F(0)$ diverges logarithmically. Use the spectrum, separated-time correlations with a regulated interpretation, or an additional ultraviolet cutoff. Such a cutoff changes the exact single-exponential kernel. We never use a finite unregulated $\langle F^2 \rangle$ for this model.

7.2 The retarded response has specified initial data

Compare two harmonic-system evolutions with identical initial state and bath initial operators, differing only in the prescribed force. Their difference is a scalar response times the identity. Write $Q(t) = \delta\langle q(t) \rangle$, with $Q(0) = \dot{Q}(0) = 0$. The noise and slip cancel in this difference:

$$\ddot{Q}(t) + \omega_0^2 Q(t) + \int_0^t K(t-u)\dot{Q}(u) \, \mathrm{d}u = f(t). \tag{7.3}$$

This removes preparation ambiguities by subtraction, not by deleting a term from the original operator equation.

For a causal function g, define $\widetilde{g}(z) = \int_0^\infty \mathrm{e}^{-zt} g(t) \, \mathrm{d}t$, initially at $\Re z > 0$. Integration by parts gives $\mathcal{L}[\dot{g}] = z\widetilde{g} - g(0)$; interchanging integrations in a convolution gives $\mathcal{L}[K * g] = \widetilde{K}\widetilde{g}$. Therefore

$$[z^2 + \omega_0^2 + z\widetilde{K}(z)]\widetilde{Q}(z) = \widetilde{f}(z), \qquad \widetilde{K}(z) = \frac{\gamma\Omega}{z + \Omega}. \tag{7.4}$$

The susceptibility and its pole polynomial are

$$\boxed{\chi(z) \equiv \frac{\widetilde{Q}(z)}{\widetilde{f}(z)} = \frac{1}{z^2 + \omega_0^2 + z\gamma\Omega/(z+\Omega)} = \frac{z+\Omega}{P(z)},} \tag{7.5}$$

$$P(z) = z^3 + \Omega z^2 + (\omega_0^2 + \gamma\Omega)z + \Omega\omega_0^2. \tag{7.6}$$

The frequency response is $\chi_R(\omega) = \chi(0^+ - \mathrm{i}\omega)$. A sinusoidal force $f(t) = \Re(f_0 \mathrm{e}^{-\mathrm{i}\omega t})$ produces a late-time displacement $\Re[\chi_R(\omega) f_0 \mathrm{e}^{-\mathrm{i}\omega t}]$ if transients decay.

Define the causal time-domain response by $Q(t) = \int_0^t R(t-u)f(u)\,\mathrm{d}u$. For three distinct roots r_α of P, partial fractions give

$$R(t) = \sum_{\alpha=1}^{3} \frac{r_\alpha + \Omega}{P'(r_\alpha)} \mathrm{e}^{r_\alpha t}, \qquad t \geq 0. \tag{7.7}$$

Repeated roots are obtained by a limit, producing $t\mathrm{e}^{rt}$ terms. Since $\chi(z) = z^{-2} + O(z^{-4})$ at large z, $R(0^+) = 0$ and $\dot{R}(0^+) = 1$: a unit impulse changes velocity by one, not position.

7.3 One hidden state replaces the friction integral

For a trajectory with velocity $v = \dot{q}$ define

$$h(t) = \Omega \int_0^t \mathrm{e}^{-\Omega(t-u)} v(u)\,\mathrm{d}u. \tag{7.8}$$

Differentiation immediately gives

$$\dot{h} = \Omega v - \Omega h, \qquad h(0) = 0, \qquad \int_0^t K(t-u)v(u)\,\mathrm{d}u = \gamma h(t). \tag{7.9}$$

The original operator equation can thus be written

$$\begin{aligned} \dot{q} &= v, \qquad \dot{h} = \Omega v - \Omega h, \\ \dot{v} &= -\omega_0^2 q - \gamma h + F(t) - K(t)q(0) + f(t). \end{aligned} \tag{7.10}$$

Only friction history has been compressed. $F(t)$ is still the quantum force of Section 6.

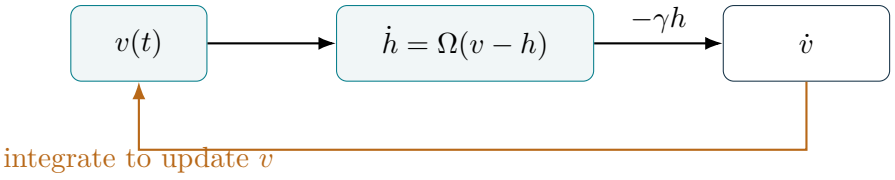


Figure. The hidden state is an exponentially weighted velocity history. Its feedback supplies retarded friction.

For response define the corresponding real memory coordinate

$$h_R(t) = \delta\langle h(t)\rangle = \Omega \int_0^t \mathrm{e}^{-\Omega(t-u)} \dot{Q}(u)\,\mathrm{d}u. \tag{7.11}$$

It obeys $\dot{h}_R = \Omega\dot{Q} - \Omega h_R$. For the homogeneous response, stability does not require solving a cubic. Define

$$\mathcal{E} = \frac{\dot{Q}^2 + \omega_0^2 Q^2}{2} + \frac{\gamma}{2\Omega} h_R^2. \tag{7.12}$$

This positive function contains real response coordinates, not products of quantum operators. Their homogeneous equations give

$$\dot{\mathcal{E}} = \dot{Q}(-\omega_0^2 Q - \gamma h_R) + \omega_0^2 Q\dot{Q} + \frac{\gamma}{\Omega} h_R(\Omega\dot{Q} - \Omega h_R) = -\gamma h_R^2 \leq 0. \tag{7.13}$$

An exponentially growing mode contradicts this positive nonincreasing function. For a nondecaying periodic mode, integrating $\dot{\mathcal{E}}$ over its period requires $h_R = 0$ identically. Then $\dot{h}_R = \Omega\dot{Q}$ implies $\dot{Q} = 0$, and $\ddot{Q} = -\omega_0^2 Q$ implies $Q = 0$. The same argument excludes a nonzero constant mode. Thus all modes of this finite linear response system decay when $\omega_0, \gamma, \Omega > 0$. The equivalent cubic check is positivity of its coefficients and $\Omega(\omega_0^2 + \gamma\Omega) - \Omega\omega_0^2 = \gamma\Omega^2 > 0$.

7.4 What the short-memory approximation discards

Integrating the definition of h by parts yields

$$h(t) = v(t) - \mathrm{e}^{-\Omega t}v(0) - \int_0^t \mathrm{e}^{-\Omega(t-u)}\dot{v}(u)\,\mathrm{d}u. \tag{7.14}$$

After the initial layer $t \sim \Omega^{-1}$, if v changes slowly on that scale, $h \simeq v$. Equivalently require $|z|/\Omega \ll 1$ at the frequencies being resolved; $\Omega \gg \omega_0, \gamma$ is a useful sufficient scale separation. Then

$$\begin{gathered}\ddot{Q} + \gamma\dot{Q} + \omega_0^2 Q \simeq f, \qquad \chi_{\mathrm{M}}(z) = \frac{1}{z^2 + \gamma z + \omega_0^2}, \\ r_\pm = -\frac{\gamma}{2} \pm \sqrt{\frac{\gamma^2}{4} - \omega_0^2}.\end{gathered} \tag{7.15}$$

M labels the Markov response approximation. We discarded the resolved memory pole and the fast preparation transient. This does not itself make the quantum noise white: its spectrum still contains $\coth(\beta|\omega|/2)$.

Self-study question 7. Is short friction memory sufficient for a classical white-noise bath?

7.5 The precise recurrent-network connection

For the response problem let $Y = (Q, \dot{Q}, h_R)^T$, a column vector; the superscript T denotes transpose. Then

$$\dot{Y} = \mathsf{M}Y + \mathsf{b}f(t), \qquad \mathsf{M} = \begin{pmatrix} 0 & 1 & 0 \\ -\omega_0^2 & 0 & -\gamma \\ 0 & \Omega & -\Omega \end{pmatrix}, \qquad \mathsf{b} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \tag{7.16}$$

Multiply by $\mathrm{e}^{-\mathsf{M}t}$ and integrate. At $t_n = n\Delta t$, if the imposed force equals f_n throughout each interval, the result is the exact update

$$\boxed{Y_{n+1} = WY_n + Uf_n, \qquad W = \mathrm{e}^{\mathsf{M}\Delta t}, \qquad U = \int_0^{\Delta t} \mathrm{e}^{\mathsf{M}(\Delta t - u)}\mathsf{b}\,\mathrm{d}u.} \tag{7.17}$$

This is a linear recurrent state-space model: its hidden state has a physical definition and its weights are fixed by the bath. No training is required. It is a controlled benchmark for recurrent memory, not a result about nonlinear network chaos or a solution of general quantum dynamics.

The illustrative scalar memory-cell update

$$h_{n+1} = \mathrm{e}^{-\Omega\Delta t}h_n + (1 - \mathrm{e}^{-\Omega\Delta t})v_n \tag{7.18}$$

is exact if $v(t) = v_n$ throughout the interval. The coupled oscillator's velocity generally varies, so freezing it is an approximation. Equation (7.17) instead propagates the coupled response exactly for the specified piecewise-constant force.

For $K(t) = \sum_{a=1}^{M} k_a \mathrm{e}^{-\Omega_a t}$, one hidden variable per exponential suffices. A general continuum kernel need not have a finite representation: eliminating a finite linear state matrix produces rational functions of z, whereas a continuum response can be nonrational. Exponential approximations therefore have a domain of accuracy, not universal exactness [6].

Classical comparison. A classical thermal Drude force is Gaussian with

$$\mathbb{E}[F(t)F(u)] = k_B T\gamma\Omega\mathrm{e}^{-\Omega|t-u|}.$$

A stationary Ornstein–Uhlenbeck process reproduces it:

$$\mathrm{d}\eta_{\mathrm{cl}} = -\Omega\eta_{\mathrm{cl}}\,\mathrm{d}t + \sqrt{2k_BT\gamma\Omega^2}\,\mathrm{d}W_t.$$

Its stationary variance is $k_BT\gamma\Omega$ and its correlation decays as $\mathrm{e}^{-\Omega|t-u|}$. Here W_t is a Wiener process with $\mathbb{E}[(\mathrm{d}W_t)^2] = \mathrm{d}t$, and η_{cl} denotes a classical force, distinct from the auxiliary-time noise of Section 3. This matches the classical Gaussian force, not the quantum force: it lacks the quantum commutator and thermal spectrum. The initial-slip term still belongs to a product-prepared system equation.

Key result. We derived an exact retarded response and an exact recurrent representation of its exponential friction memory. Quantum noise, initial preparation and the Markov approximation remain explicit and distinct.

7.6 What could be measured: the absorption curve

For a positive real drive frequency, substitute $z = 0^+ - \mathrm{i}\omega$ in $\chi(z)$ and multiply the memory term by $(\Omega + \mathrm{i}\omega)/(\Omega + \mathrm{i}\omega)$:

$$\chi_R(\omega)^{-1} = \omega_0^2 - \omega^2 + \frac{\gamma\Omega\omega^2}{\Omega^2+\omega^2} - \mathrm{i}\,\frac{\gamma\Omega^2\omega}{\Omega^2+\omega^2} \equiv A_\omega - \mathrm{i}B_\omega, \tag{7.19}$$

$$\operatorname{Im}\chi_R(\omega) = \frac{B_\omega}{A_\omega^2 + B_\omega^2} = J_D(\omega)|\chi_R(\omega)|^2 \geq 0. \tag{7.20}$$

A_ω, B_ω are real scalar functions defined by the first line, not the lattice matrix. The last equality uses precisely the spectral-density normalization of Section 6.

The sign has a direct meaning. For $f(t) = f_0\cos\omega t$ with real amplitude f_0,

$$Q(t) = f_0\big[\operatorname{Re}\chi_R\cos\omega t + \operatorname{Im}\chi_R\sin\omega t\big], \tag{7.21}$$

$$\overline{f(t)\dot{Q}(t)} = \frac{f_0^2\omega}{2}\operatorname{Im}\chi_R(\omega) \geq 0. \tag{7.22}$$

The bar means an average over a drive period after transients. We used $\overline{\cos^2} = 1/2$ and $\overline{\sin\cos} = 0$. The oscillator and bath absorb nonnegative power. Thus the positive bath spectrum is a physical constraint on a possible recurrent response model, not merely a choice of stable parameters.

For this harmonic model the mean response has the same form in classical and quantum mechanics. Its poles alone do not demonstrate quantum behavior. The quantum content is explicit in the equilibrium fluctuations, the thermal factor and the noncommuting force operators derived earlier.

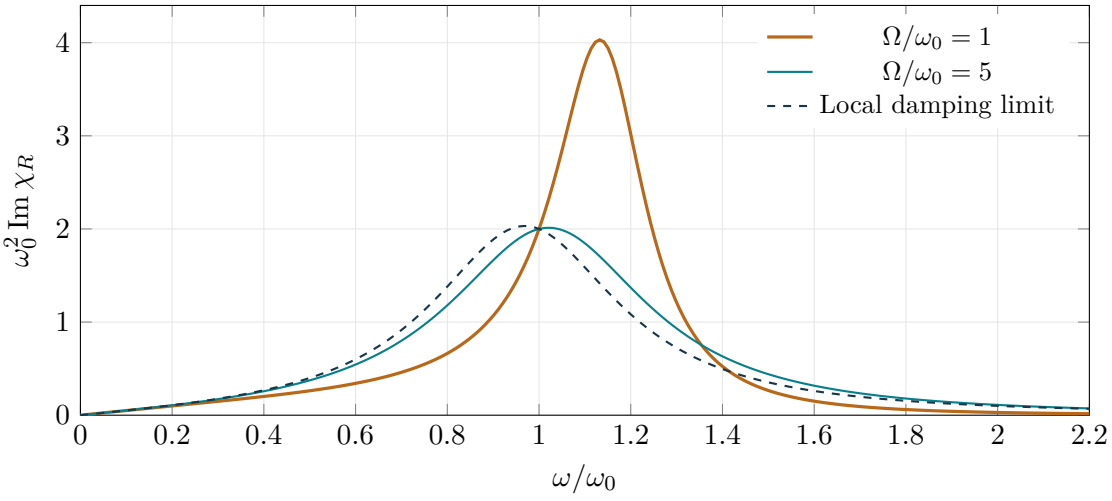


Figure. Analytic Drude response at fixed $\gamma/\omega_0 = 0.5$. Finite memory changes the resonance position and width. The dashed curve uses χ_{M}, corresponding to $\Omega \to \infty$ at fixed resolved frequency. All curves have the same static susceptibility; that agreement does not determine their frequency response. These are plots of the derived formulas, not simulation data or fits.

8 One bath, two representations

Gaussian bath elimination connects the equilibrium path measure to the physical-time response.

8.1 The coupled equilibrium ensemble

Sections 5–7 used a factorized initial state to define an initial-value problem. Here the question is equilibrium of the *coupled* Hamiltonian H_{tot}, with $\lambda = 0$ and no external force:

$$\rho_{\mathrm{eq}}^{\mathrm{tot}} = \frac{\mathrm{e}^{-\beta H_{\mathrm{tot}}}}{\operatorname{Tr} \mathrm{e}^{-\beta H_{\mathrm{tot}}}}, \qquad \rho_{\mathrm{eq}}^{S} = \operatorname{Tr}_B \rho_{\mathrm{eq}}^{\mathrm{tot}}. \tag{8.1}$$

Tr_B traces over all bath coordinates. In general $\rho_{\mathrm{eq}}^{S} \neq \mathrm{e}^{-\beta H_A} / \operatorname{Tr}_S \mathrm{e}^{-\beta H_A}$: the coupling changes the reduced equilibrium state. Nor is $\rho_{\mathrm{eq}}^{\mathrm{tot}}$ the factorized initial state of Section 5. A stable continuum bath can relax local observables toward equilibrium; a finite set of undamped bath oscillators can recur.

8.2 The Euclidean action before elimination

Use the position-space construction of Section 1 for the system and for every bath oscillator. Write $y_j(\tau)$ for a bath path, so it is not confused with the bath operator $Q_j(t)$. The action is

$$S_E[x, \{y_j\}] = \int_0^\beta \mathrm{d}\tau \left\{ \frac{\dot{x}_\tau^2}{2} + \frac{\omega_0^2 x^2}{2} + \sum_j \left[\frac{m_j \dot{y}_{j,\tau}^2}{2} + \frac{m_j \omega_j^2}{2} \left(y_j - \frac{c_j x}{m_j \omega_j^2} \right)^2 \right] \right\}. \tag{8.2}$$

Here $\dot{x}_\tau = \partial_\tau x$ and $\dot{y}_{j,\tau} = \partial_\tau y_j$, not physical-time velocities. All paths are periodic with period β. Introduce the Fourier series

$$x(\tau) = \frac{1}{\sqrt{\beta}} \sum_{n \in \mathbb{Z}} x_n \mathrm{e}^{\mathrm{i}\nu_n \tau}, \quad y_j(\tau) = \frac{1}{\sqrt{\beta}} \sum_n y_{j,n} \mathrm{e}^{\mathrm{i}\nu_n \tau}, \quad \nu_n = \frac{2\pi n}{\beta}. \tag{8.3}$$

Reality gives $x_{-n} = x_n^*$ and $y_{j,-n} = y_{j,n}^*$. The orthogonality integral $\int_0^\beta \mathrm{d}\tau\, \mathrm{e}^{\mathrm{i}(\nu_n-\nu_m)\tau} = \beta\delta_{nm}$ implies

$$\int_0^\beta x^2\,\mathrm{d}\tau = \sum_n |x_n|^2, \qquad \int_0^\beta (\partial_\tau x)^2\,\mathrm{d}\tau = \sum_n \nu_n^2 |x_n|^2. \tag{8.4}$$

One may equivalently work with independent real sine and cosine coefficients; the sum over positive and negative n below includes the required paired modes.

8.3 Complete the square, one bath mode at a time

Let $D_{j,n} = \nu_n^2 + \omega_j^2 > 0$. Substituting the Fourier series gives

$$S_E = \frac{1}{2}\sum_n \Bigg[(\nu_n^2+\omega_0^2)|x_n|^2 + \sum_j \left\{ m_j D_{j,n}|y_{j,n}|^2 - 2c_j\,\mathrm{Re}(x_n^* y_{j,n}) + \frac{c_j^2}{m_j\omega_j^2}|x_n|^2 \right\}\Bigg]. \tag{8.5}$$

The elementary identity used for each j, n is

$$m_j D_{j,n}|y|^2 - 2c_j\,\mathrm{Re}(x^* y) = m_j D_{j,n}\left|y - \frac{c_j x}{m_j D_{j,n}}\right|^2 - \frac{c_j^2}{m_j D_{j,n}}|x|^2. \tag{8.6}$$

Shift the bath integration variable by $c_j x/(m_j D_{j,n})$. Its Gaussian integral produces a determinant independent of x; this factor cancels between numerator and denominator of normalized system correlators. It would have to be retained when calculating the total partition function or free energy. The remaining action is

$$S_{\mathrm{eff}}[x] = \frac{1}{2}\sum_n \mathcal{A}_{\mathrm{eff}}(\nu_n)|x_n|^2, \tag{8.7}$$

$$\mathcal{A}_{\mathrm{eff}}(\nu) = \nu^2 + \omega_0^2 + \sum_j \frac{c_j^2}{m_j}\left[\frac{1}{\omega_j^2} - \frac{1}{\nu^2+\omega_j^2}\right] = \nu^2 + \omega_0^2 + \sum_j \frac{c_j^2}{m_j\omega_j^2}\frac{\nu^2}{\nu^2+\omega_j^2}. \tag{8.8}$$

The square bracket vanishes at $\nu = 0$: this is the static cancellation supplied by the counterterm. At nonzero ν it is positive and changes the fluctuations. The counterterm therefore does not make the dynamical effect of the bath disappear.

8.4 Recognize the same memory kernel

From Section 5,

$$K(t) = \sum_j \frac{c_j^2}{m_j\omega_j^2}\cos\omega_j t. \tag{8.9}$$

For real $z > 0$, integrate the cosine as the real part of an exponential:

$$\int_0^\infty \mathrm{d}t\,\mathrm{e}^{-zt}\cos\omega_j t = \mathrm{Re}\,\frac{1}{z-\mathrm{i}\omega_j} = \frac{z}{z^2+\omega_j^2}, \tag{8.10}$$

$$z\widetilde{K}(z) = \sum_j \frac{c_j^2}{m_j\omega_j^2}\frac{z^2}{z^2+\omega_j^2}. \tag{8.11}$$

Comparison with (8.8) proves

$$\boxed{\mathcal{A}_{\rm eff}(\nu) = \nu^2 + \omega_0^2 + |\nu|\widetilde{K}(|\nu|).} \tag{8.12}$$

At $\nu = 0$ the last term is defined by its continuous limit, zero for the bath under consideration. For the Drude kernel,

$$\boxed{\mathcal{A}_{\rm eff}(\nu) = \nu^2 + \omega_0^2 + \frac{\gamma\Omega|\nu|}{\Omega + |\nu|}, \qquad G_{\rm eq}(\nu_n) = \frac{1}{\mathcal{A}_{\rm eff}(\nu_n)}.} \tag{8.13}$$

$G_{\rm eq}$ is the Fourier coefficient of the coupled-equilibrium Euclidean covariance. Direct Gaussian differentiation, as in Section 2, gives

$$\langle x_n x_m\rangle_{\rm eq} = \frac{\delta_{n,-m}}{\mathcal{A}_{\rm eff}(\nu_n)}, \tag{8.14}$$

$$\langle q^2\rangle_{\rm eq} = \frac{1}{\beta}\sum_n \frac{1}{\mathcal{A}_{\rm eff}(\nu_n)} \xrightarrow{\beta\to\infty} \int_{-\infty}^{\infty}\frac{{\rm d}\nu}{2\pi}\frac{1}{\mathcal{A}_{\rm eff}(\nu)}. \tag{8.15}$$

This integral converges since $\mathcal{A}_{\rm eff}(\nu) \sim \nu^2$ at large $|\nu|$. It is a position variance; its finiteness does not imply a finite equal-time variance of the ideal Drude *force*.

The classical high-temperature limit of this *position variance* can be bounded directly, without a nonuniform expansion of the force spectrum. The $n = 0$ term is $1/(\beta\omega_0^2)$ and every nonzero-frequency term is positive with $\mathcal{A}_{\rm eff}(\nu_n) \geq \nu_n^2$. Therefore

$$0 \leq \langle q^2\rangle_{\rm eq} - \frac{1}{\beta\omega_0^2} \leq \frac{2}{\beta}\sum_{n=1}^{\infty}\frac{1}{(2\pi n/\beta)^2} = \frac{\beta}{2\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\beta}{12}. \tag{8.16}$$

The final step uses $\sum_{n\geq 1} n^{-2} = \pi^2/6$. At fixed model parameters the relative correction is thus at most $(\beta\omega_0)^2/12$ as $\beta \to 0$.

Key result. For each positive Euclidean frequency,

$$G_{\rm eq}(\nu) = \chi(z = \nu).$$

Both follow from the same eliminated bath, with the appropriate boundary conditions. This identity does not identify the auxiliary sampling time s with physical time t.

Self-study question 8. Why does the bath leave the static susceptibility $\chi(0) = 1/\omega_0^2$ unchanged but alter the quantum position variance?

8.5 What a recurrence must preserve

Discretizing $S_{\rm eff}$ would give another positive quadratic matrix. The argument of Section 3 then supplies a sampler with stationary measure $\mathrm{e}^{-S_{\rm eff}}$. Its drift is generally nonlocal along τ, because the bath couples different points of the Euclidean path. This is an equilibrium representation of the coupled oscillator, not its physical trajectory.

In physical time the single hidden variable h instead stores the exponential convolution exactly. More generally, if a friction kernel is a sum of exponentials,

$$K(t) = \sum_{r=1}^{R} g_r\Omega_r \mathrm{e}^{-\Omega_r t}, \qquad g_r > 0, \quad \Omega_r > 0, \tag{8.17}$$

define $h_r(t) = \Omega_r \int_0^t \mathrm{e}^{-\Omega_r(t-u)} v(u)\,\mathrm{d}u$. Differentiation gives

$$\dot{h}_r = -\Omega_r h_r + \Omega_r v, \qquad \int_0^t K(t-u)v(u)\,\mathrm{d}u = \sum_{r=1}^{R} g_r h_r. \tag{8.18}$$

For these positive coefficients the enlarged homogeneous linear system has the energy function

$$\mathcal{E} = \frac{v^2 + \omega_0^2 q^2}{2} + \sum_r \frac{g_r h_r^2}{2\Omega_r}, \qquad \dot{\mathcal{E}} = -\sum_r g_r h_r^2 \le 0. \tag{8.19}$$

This elementary proof explains why physically constrained memory parameters can be preferable to arbitrary recurrent weights. It concerns the deterministic response. Reproducing the full quantum process also requires the correct force covariance and force commutator; fitting the friction kernel alone does not supply them. Auxiliary-variable representations are widely used for classical colored-noise dynamics [6]; their full quantum interpretation needs the distinctions derived here.

	Equilibrium sampler	Physical memory representation
State updated	Entire discretized Euclidean path X	Position, velocity and stored history (q, v, h)
Clock	Auxiliary s, or iteration n	Physical t
Correctness test	Stationary density and quantum correlators	Causality, response and quantum noise identities
Exact statement	Continuous diffusion samples e^{-S_a}	One hidden variable realizes exponential friction
Controlled error	Mesh, finite β, step bias; perturbative truncation	Kernel approximation or Markov limit, if imposed
What is not implied	Real-time quantum trajectories	Complete quantum noise from one classical random variable

8.6 Main results and a research question

$$\text{Quantum fluctuation:} \quad \langle q^2 \rangle_{\mathrm{vac}} = \frac{1}{2\omega_0} - \frac{\lambda}{16\omega_0^4} + O(\lambda^2), \tag{8.20}$$

$$\text{Sampling principle:} \quad \mathrm{d}X = -\nabla S_a\,\mathrm{d}s + \sqrt{2}\,\mathrm{d}W \implies \rho_*(X) \propto \mathrm{e}^{-S_a(X)}, \tag{8.21}$$

$$\text{Physical memory:} \quad \chi(z)^{-1} = z^2 + \omega_0^2 + \frac{\gamma\Omega z}{z+\Omega}. \tag{8.22}$$

The open extension is a nonlinear quantum system with a nontrivial bath: can a compact recurrent representation retain the correct equilibrium measure, causal response and quantum fluctuation–dissipation relation simultaneously? The harmonic example is an exact benchmark, and the anharmonic result is a perturbative benchmark, against which such approximations can be tested. These controlled results provide tests of proposed approximations to interacting quantum dynamics.

8.7 Self-study and final-report tasks

The report can be a short sequence of calculations with a sentence explaining each result. No simulation is required.

1. **Quantum fluctuation.** Starting with the normalized path integral, count the Wick pairings and derive the $O(\lambda)$ correction to $\langle q^2 \rangle$. Check the sign by the energy derivative. Refer to Section 2.
2. **Sampling and recurrence.** Derive the Fokker–Planck current and its stationary density. For one free mode derive the finite-step variance and explain why stability is not the same as correct sampling. Use Sections 3–4.
3. **Memory from a Hamiltonian.** Solve one bath oscillator, perform the integration by parts including both boundaries, and obtain the initial-slip term. State the initial preparation. Use Section 5.
4. **What makes the noise quantum?** Derive the symmetrized force covariance and its commutator from $[Q_j, P_k] = \mathrm{i}\delta_{jk}$. State the approximation needed for the classical fluctuation–dissipation relation. Use Section 6.
5. **One hidden variable.** Eliminate h from the local equations and recover the exponential convolution. Derive $\chi(z)$ and explain when the memory can be approximated by local damping. Use Section 7.
6. **Connect the two halves.** Complete the bath square in (8.2) and derive (8.12). Explain why this uses a different initial ensemble from the factorized preparation. Refer to Section 8.

An optional short extension, still analytic, is to repeat the hidden-state energy argument for two exponentials and compare the resulting rational susceptibility with the single-exponential result. This tests the memory-representation idea without adding a new formalism.

References

The calculations above are derived explicitly in the conventions of these notes. The sources establish the methods and give wider context; reading them is not a prerequisite for following the derivations. In particular, an auxiliary-time Langevin equation in stochastic quantization and a physical-time quantum Langevin equation answer different questions.

[1] G. Parisi and Y.-S. Wu, *Perturbation Theory without Gauge Fixing*, Scientia Sinica **24**, 483–496 (1981). Author's publication record, entry 90. The original stochastic-quantization proposal; Sections 3–4 use its elementary scalar construction.

[2] G. Menezes and N. F. Svaiter, *Stochastic Quantization of $(\lambda\phi^4)_d$ Scalar Theory: Generalized Langevin Equation with Memory Kernel*, Physica A **374**, 617–630 (2007). arXiv:hep-th/0511224. Contains the scalar stochastic perturbative framework. Its auxiliary-time memory extension should not be confused with the physical oscillator bath in Sections 5–8.

[3] G. W. Ford, J. T. Lewis and R. F. O'Connell, *Quantum Langevin equation*, Physical Review A **37**, 4419–4428 (1988). doi:10.1103/PhysRevA.37.4419. Microscopic oscillator-bath construction, operator noise and physical restrictions on the memory function; Sections 5–7.

[4] A. Ghosh, M. Bandyopadhyay, S. Dattagupta and S. Gupta, *Independent-oscillator model and the quantum Langevin equation for an oscillator: A review*, Journal of Statistical Mechanics: Theory and Experiment **2024**, 074002. arXiv:2306.02665. Useful extended treatment of the bath Hamiltonian, noise and equilibrium. Initial conditions and counterterm conventions must be compared before borrowing a formula.

[5] P. Hänggi and G.-L. Ingold, *Fundamental Aspects of Quantum Brownian Motion*, Chaos **15**, 026105 (2005). arXiv:quant-ph/0412052. Quantum versus classical noise, fluctuation–dissipation and path-integral treatment of dissipation; Sections 6–8.

[6] M. Ceriotti, G. Bussi and M. Parrinello, *Langevin Equation with Colored Noise for Constant-Temperature Molecular Dynamics Simulations*, Physical Review Letters **102**, 020601 (2009). arXiv:0812.1168. Auxiliary variables for classical colored-noise dynamics; a connection to finite-state memory, not a proof of a finite classical representation of all quantum noise.

Beyond the Born–Oppenheimer approximation

Beyond the Born–Oppenheimer approximation

Effective Hamiltonians and isotope spectra

Study notes and report questions

Contents

The question and the calculation **2**

1 From the molecular Hamiltonian to a nuclear potential **4**

2 An exact spectrum against which to test the approximation **9**

3 Virtual excitations and the effective Hamiltonian **13**

4 From an effective Hamiltonian to isotope spectra **16**

5 From the calculation to an interpretable spectral test **22**

References and reading guide **24**

The question and the calculation

What do we lose when electrons are replaced by a potential acting on nuclei, and how can isotope spectra reveal the missing effects?

The spectroscopy project motivating these notes concerns the $B^2\Sigma^+ - X^2\Sigma^+$ emission spectrum of $^{13}C^{17}O^+$. Its molecular constants help test the mass dependence of the ground-state spectrum. Related work on $^{12}C^{17}O^+$ provides the immediate experimental setting [7]. The theoretical issue is how to connect a fitted spectral correction to the approximation used to describe the molecule.

We solve one closed, conservative quantum model exactly, approximate it by freezing its fast motion, and calculate what that approximation misses. We then construct an isotope diagnostic for molecular spectra. The model is a benchmark for the mechanism; it does not supply numerical CO^+ constants. Born–Oppenheimer breakdown corrections have already been included in CO^+ isotope analyses [6]; our goal is to understand and test their interpretation.

Three results to obtain.

1. Projecting the molecular Schrödinger equation gives derivative couplings and a positive diagonal correction to the nuclear potential.
2. Eliminating fast excitations also changes the *kinetic* term. An exactly solvable example gives an induced inertia $\delta M = ma^2$; correcting only the potential misses its effect on line spacings.
3. Reduced-mass scaling and combinations of measured lines give a test of a common electronic potential, provided ordinary BO rovibrational effects, spin structure, and beyond-clamped-nuclei recoil effects are distinguished consistently.

Prerequisites and scope. Complex amplitudes, derivatives, matrices, the Schrödinger equation, and elementary harmonic-oscillator quantum mechanics. Necessary oscillator identities are derived below. The effective-theory connection uses matrix elimination and second-order perturbation theory; no path integral, thermal environment, or stochastic formalism is assumed. All dynamics in the solvable benchmark is unitary.

Units and conventions. Keep $\hbar$ explicit; $h = 2\pi\hbar$. Angular frequencies such as Ω have units s^{-1}; spectroscopic wavenumbers $\widetilde{\nu} = \Delta E/(hc)$ have units cm^{-1} when c is expressed in $\mathrm{cm\,s}^{-1}$. Nuclear masses and the electron mass must be expressed in the same mass unit. A mass number such as 13 is an isotope label, not its exact nuclear mass.

Notation

Symbol	Meaning and where it is used
R, μ	Internuclear separation and nuclear reduced mass in the molecular problem. In the benchmark $R \in \mathbb{R}$ is a slow displacement, with mass M.
$\boldsymbol{r}, \epsilon_n, \phi_n$	All electronic coordinates, clamped-nuclei electronic energy, and normalized electronic state in channel n.
χ_n, d_{mn}, t_{mn}	Nuclear channel amplitude, first and second derivative couplings. Electronic inner products integrate only electronic coordinates.
Φ	Diagonal Born–Oppenheimer correction (DBOC) in the declared one-coordinate model.
x, m, K, κ, a	Fast displacement, fast mass, two positive stiffnesses, and dimensionless coupling in the benchmark.
$\omega_f, \Omega_{\mathrm{BO}}, \omega_\pm$	Frozen-coordinate fast frequency, BO slow frequency, and exact high/low normal-mode frequencies.
P, p	Momenta conjugate to the benchmark coordinates. Projection operators in Part 3 are $\mathcal{P}, \mathcal{Q}$, not momenta.
v, N, J	Vibrational quantum number; rotational angular momentum excluding spin; total electronic-spin-plus-rotation angular momentum.
B_e, B_v, D_v	Equilibrium rotational constant, vibrationally dependent rotational constant, and centrifugal-distortion constant; here in wavenumber units.
$Y_{kl}, U_{kl}, \Delta_{kl}^A$	Dunham coefficients, leading isotope-invariant coefficients, and dimensionless atom-specific breakdown parameters in the stated convention.
α, β	Isotopologue labels when superscripts; α_e is instead the vibration–rotation constant.

1 From the molecular Hamiltonian to a nuclear potential

Physical meaning. The measured spectrum consists of differences of molecular energy levels. Our first task is to identify which approximation allows us to calculate those levels using a potential for the nuclei.

1.1 The quantum-mechanical starting point

For a time-independent Hamiltonian H, a stationary state obeys

$$H\Psi = E\Psi, \qquad \Psi(t) = \mathrm{e}^{-\mathrm{i}Et/\hbar}\Psi, \qquad \hbar = \frac{h}{2\pi}. \tag{1.1}$$

Here E is the total energy, h is Planck's constant, and $\mathrm{i}^2 = -1$. In the coordinate representation, momentum is $-\mathrm{i}\hbar\nabla$; hence a particle of mass m has kinetic-energy operator $-\hbar^2\nabla^2/(2m)$. An inner product integrates over the coordinates on which the states depend. For example, $\langle\phi|\psi\rangle = \int \mathrm{d}\mathbf{r}\, \phi^*(\mathbf{r})\psi(\mathbf{r})$, with any discrete spin variables summed.

The ion CO^+ contains two nuclei with charge numbers $Z_\mathrm{C} = 6$, $Z_\mathrm{O} = 8$, and $N_e = 13$ electrons. Let $\mathbf{R}_A$, M_A be the position and *nuclear* mass of nucleus $A \in \{\mathrm{C}, \mathrm{O}\}$, and let $\mathbf{r}_i$ be the position of electron i, of mass m_e and charge $-e$, with $e > 0$. The isolated, nonrelativistic point-charge Coulomb Hamiltonian in laboratory coordinates is

$$H_\mathrm{lab} = -\sum_{A=\mathrm{C,O}} \frac{\hbar^2}{2M_A}\nabla^2_{\mathbf{R}_A} - \sum_{i=1}^{13} \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} + V_\mathrm{Coul}, \tag{1.2}$$

$$V_\mathrm{Coul} = \frac{e^2}{4\pi\varepsilon_0}\left[\frac{Z_\mathrm{C}Z_\mathrm{O}}{|\mathbf{R}_\mathrm{C} - \mathbf{R}_\mathrm{O}|} + \sum_{i<j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,A}\frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|}\right]. \tag{1.3}$$

ε_0 is the vacuum permittivity; the three potential terms are nuclear repulsion, electron repulsion, and electron–nucleus attraction. Electronic states obey fermionic antisymmetry. This Hamiltonian omits relativistic, spin-dependent, radiative, and finite-nuclear-size interactions. They are separate corrections when the target precision requires them.

Translational invariance permits an exact separation

$$H_\mathrm{lab} = \frac{\mathbf{P}^2_\mathrm{CM}}{2M_\mathrm{tot}} + H_\mathrm{int}, \qquad M_\mathrm{tot} = M_\mathrm{C} + M_\mathrm{O} + 13m_e, \tag{1.4}$$

where $\mathbf{P}_\mathrm{CM}$ is total momentum. The internal Hamiltonian H_int includes electron–nuclear recoil terms; in common internal coordinates these include mixed kinetic derivatives, also called mass-polarization terms. Consequently, simply deleting the center-of-mass kinetic energy does *not* establish the one-coordinate form used below.

1.2 A declared model for separating fast and slow motion

To derive the essential mechanism without rotational bookkeeping, we use the spin-independent, one-coordinate model

$$H = -\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial R^2} + H_\mathrm{el}(\mathbf{r}; R), \qquad R = |\mathbf{R}_\mathrm{C} - \mathbf{R}_\mathrm{O}|, \qquad \mu = \frac{M_\mathrm{C}M_\mathrm{O}}{M_\mathrm{C} + M_\mathrm{O}}. \tag{1.5}$$

Here $\mathbf{r}$ collectively denotes electronic variables in a chosen fixed electronic representation; H_el contains the clamped-nuclei electronic Hamiltonian *and the nuclear repulsion.* The coordinate $R > 0$ is the internuclear distance. The nuclear amplitude is a reduced radial amplitude, with

measure $\mathrm{d}R$, so its radial kinetic term is a second derivative. We initially suppress rotational motion and electron spin. Recoil and mass-polarization corrections of the complete internal molecular problem are not included in (1.5). Thus the following projection is exact for this stated model, while its application to precision CO^+ requires those additional physical terms.

At each fixed R, solve the electronic eigenproblem

$$H_{\mathrm{el}}(R)|n;R\rangle = \epsilon_n(R)|n;R\rangle, \qquad \langle m;R|n;R\rangle = \delta_{mn}, \qquad \phi_n(\mathbf{r};R) = \langle \mathbf{r}|n;R\rangle. \tag{1.6}$$

δ_{mn} equals one for $m = n$ and zero otherwise. The index n labels electronic channels; $\epsilon_n(R)$ is their potential energy curve, including nuclear repulsion. The inner product in (1.6) is over electronic variables only, holding R fixed. For point nuclei with fixed charges, the clamped-nuclei problem contains no nuclear masses. Different isotopes therefore share these curves at this level of approximation.

Checkpoint. A potential energy curve is an electronic eigenvalue evaluated at every nuclear separation. Its mass independence does not imply that the molecular energy levels are mass independent: the nuclear kinetic energy contains μ.

Expand the molecular state in the complete electronic basis:

$$\Psi(\mathbf{r},R) = \sum_n \chi_n(R)\phi_n(\mathbf{r};R), \qquad \sum_n \int_0^\infty \mathrm{d}R\,|\chi_n(R)|^2 = 1. \tag{1.7}$$

The coefficient χ_n is a nuclear amplitude in channel n. A complete sum includes electronic continuum integrals when needed. This is the Born–Huang expansion, a change of representation rather than an approximation [1]. Both factors depend on R, so the product rule gives

$$\partial_R(\chi_n\phi_n) = \chi_n'\phi_n + \chi_n\partial_R\phi_n, \tag{1.8}$$

$$\partial_R^2(\chi_n\phi_n) = \chi_n''\phi_n + 2\chi_n'\partial_R\phi_n + \chi_n\partial_R^2\phi_n. \tag{1.9}$$

Primes denote R derivatives. Multiplying $H\Psi = E\Psi$ by ϕ_m^* and integrating over electronic variables yields

$$-\frac{\hbar^2}{2\mu}\left[\chi_m'' + 2\sum_n d_{mn}\chi_n' + \sum_n t_{mn}\chi_n\right] + \epsilon_m\chi_m = E\chi_m, \tag{1.10}$$

where

$$d_{mn}(R) = \langle m;R|\partial_R|n;R\rangle, \qquad t_{mn}(R) = \langle m;R|\partial_R^2|n;R\rangle. \tag{1.11}$$

The derivative acts on the ket. The dimensions of d_{mn} and t_{mn} are inverse length and inverse length squared. These couplings arise because the electronic basis itself changes as the nuclei move.

Keeping only channel n and dropping all derivatives of its electronic state gives the elementary Born–Oppenheimer equation

$$\left[-\frac{\hbar^2}{2\mu}\frac{\mathrm{d}^2}{\mathrm{d}R^2} + \epsilon_n(R)\right]\chi_n^{\mathrm{BO}}(R) = E^{\mathrm{BO}}\chi_n^{\mathrm{BO}}(R). \tag{1.12}$$

This is the precise approximation to be tested: nuclear motion sees one electronic energy curve, and the kinetic operator is assumed not to probe the changing electronic state.

Physical meaning. The product rule has identified the omitted terms. We now determine their signs, physical meaning, and conditions for being small.

1.3 A positive diagonal correction

Differentiate orthonormality to obtain

$$0 = \partial_R \langle m; R | n; R \rangle = d^*_{nm} + d_{mn}. \tag{1.13}$$

Thus the matrix d is anti-Hermitian, and d_{nn} is purely imaginary. For a smooth isolated orbital channel in a field-free real representation, the electronic eigenfunction may locally be chosen real; normalization then gives $d_{nn} = 0$. Spin degeneracy is a spectator in this statement: we fix one spin component and ignore spin couplings.

In this local convention, differentiate $\langle n | \partial_R | n \rangle = 0$ once more:

$$0 = \langle \partial_R n | \partial_R n \rangle + \langle n | \partial_R^2 | n \rangle, \tag{1.14}$$

$$t_{nn} = -\langle \partial_R n | \partial_R n \rangle = -\sum_{m \neq n} |d_{mn}|^2. \tag{1.15}$$

The final equality inserts the electronic identity $\sum_m |m\rangle\langle m| = 1$ and uses $d_{nn} = 0$. Retaining this diagonal term, but still neglecting transitions to other channels, gives

$$\left[-\frac{\hbar^2}{2\mu} \frac{\mathrm{d}^2}{\mathrm{d}R^2} + \epsilon_n(R) + \Phi_n(R) \right] \chi_n = E \chi_n, \qquad \boxed{\Phi_n(R) = \frac{\hbar^2}{2\mu} \sum_{m \neq n} |d_{mn}(R)|^2 \geq 0.} \tag{1.16}$$

Φ_n is the radial diagonal Born–Oppenheimer correction for our model. It has dimensions of energy and an explicit inverse-mass dependence. A changing electronic cloud costs kinetic energy even if no real transition to another electronic channel occurs. The full molecular diagonal correction involves all relevant nuclear derivatives and the chosen internal-coordinate treatment.

The result can be stated without selecting a phase convention:

$$\Phi_n = \frac{\hbar^2}{2\mu} \left[\langle \partial_R n | \partial_R n \rangle - |\langle n | \partial_R | n \rangle|^2 \right] = \frac{\hbar^2}{2\mu} \langle \partial_R n | (1 - |n\rangle\langle n|) | \partial_R n \rangle. \tag{1.17}$$

A replacement $|n\rangle \mapsto \mathrm{e}^{\mathrm{i}\alpha(R)} |n\rangle$ changes only the arbitrary electronic phase. The component parallel to $|n\rangle$ cancels in (1.17); the correction measures the physical change orthogonal to that state. We use the real local convention throughout, so no additional phase formalism is needed.

Checkpoint. The diagonal correction is a positive addition to the channel potential. Its contribution to a transition frequency is a *difference* of level shifts and can have either sign. Including it does not include all nonadiabatic effects.

1.4 Why a small electronic gap matters

Differentiate the electronic eigenvalue equation:

$$H'_{\mathrm{el}} |n\rangle + H_{\mathrm{el}} |\partial_R n\rangle = \epsilon'_n |n\rangle + \epsilon_n |\partial_R n\rangle. \tag{1.18}$$

Projection onto $\langle m|$, using $\langle m | H_{\mathrm{el}} = \epsilon_m \langle m|$, gives

$$\langle m | H'_{\mathrm{el}} | n \rangle + \epsilon_m d_{mn} = \epsilon'_n \delta_{mn} + \epsilon_n d_{mn}. \tag{1.19}$$

Consequently,

$$\epsilon'_n = \langle n | H'_{\mathrm{el}} | n \rangle, \tag{1.20}$$

$$\boxed{d_{mn} = \frac{\langle m | H'_{\mathrm{el}} | n \rangle}{\epsilon_n - \epsilon_m}}, \qquad m \neq n, \quad \epsilon_m \neq \epsilon_n. \tag{1.21}$$

The first identity is the Hellmann–Feynman relation; the second identifies an energy-gap denominator. A small gap permits rapid variation of the electronic eigenvectors if the numerator is nonzero. It is therefore insufficient to argue that heavy nuclei alone guarantee a reliable single-channel description.

The off-diagonal operator omitted from channel m is

$$W_{mn} = -\frac{\hbar^2}{2\mu}\left(2d_{mn}\frac{\mathrm{d}}{\mathrm{d}R} + t_{mn}\right), \qquad m \neq n. \tag{1.22}$$

It can mix nuclear amplitudes on different electronic curves. In a locally semiclassical region, write a nuclear wave as $\chi \sim \mathrm{e}^{\mathrm{i}p_R R/\hbar}$, where p_R is its local momentum and $v_R = p_R/\mu$ its velocity. The first-derivative contribution then has energy scale $\hbar|v_R d_{mn}|$. One useful local diagnostic is

$$\eta_{mn} = \frac{\hbar|v_R d_{mn}|}{|\epsilon_m - \epsilon_n|} \ll 1. \tag{1.23}$$

It is a diagnostic for this first-derivative term, not a universal error bound. Near turning points a local plane wave is inadequate; the second-derivative term, nuclear wavefunction support, and possible near-degeneracy of *rovibronic* levels must also be examined in stationary spectroscopy.

1.5 An avoided crossing: an explicit two-state check

Let $|1\rangle, |2\rangle$ be an R-independent orthonormal electronic basis and consider

$$H_{\mathrm{el}}(R) = \begin{pmatrix} \Delta(R)/2 & g \\ g & -\Delta(R)/2 \end{pmatrix}, \qquad \Delta(R) = b(R - R_c), \qquad g > 0. \tag{1.24}$$

R_c is the crossing position of the uncoupled levels, b has dimensions of energy per length, and g is the mixing energy. A common multiple of the identity was omitted because it does not change the eigenvectors or their coupling. The characteristic polynomial is

$$\det(H_{\mathrm{el}} - \epsilon I) = \epsilon^2 - \frac{\Delta^2}{4} - g^2 = 0, \qquad \epsilon_\pm = \pm\frac{1}{2}\sqrt{\Delta^2 + 4g^2}. \tag{1.25}$$

Here I is the two-dimensional identity; the minimum gap is $2g$. Choose smooth eigenvectors

$$|+\rangle = \cos\theta|1\rangle + \sin\theta|2\rangle, \qquad |-\rangle = -\sin\theta|1\rangle + \cos\theta|2\rangle, \tag{1.26}$$

where

$$\cos 2\theta = \frac{\Delta}{\sqrt{\Delta^2 + 4g^2}}, \qquad \sin 2\theta = \frac{2g}{\sqrt{\Delta^2 + 4g^2}}. \tag{1.27}$$

These two equations fix a continuous branch; using $\tan 2\theta = 2g/\Delta$ alone would obscure it. Differentiation gives

$$\theta' = -\frac{bg}{\Delta^2 + 4g^2}, \qquad \partial_R|+\rangle = \theta'|-\rangle, \qquad \partial_R|-\rangle = -\theta'|+\rangle. \tag{1.28}$$

Thus

$$|d_{-+}| = |d_{+-}| = \frac{|bg|}{b^2(R - R_c)^2 + 4g^2}, \qquad |d_{-+}(R_c)| = \frac{|b|}{4g}, \tag{1.29}$$

and the diagonal correction in either channel is

$$\Phi_\pm(R) = \frac{\hbar^2}{2\mu}\frac{b^2 g^2}{[b^2(R - R_c)^2 + 4g^2]^2}. \tag{1.30}$$

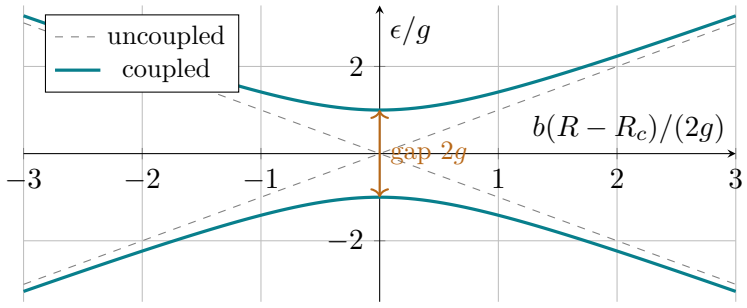


Figure. Avoided crossing in the declared two-state model. Nonzero mixing replaces the crossing by a gap; eigenvectors rotate most rapidly near its minimum. No specific crossing in CO^+ is assumed.

The electronic curves remain smooth for $g > 0$, but the eigenvectors change increasingly rapidly as the gap shrinks. A large $\Phi_\pm$ near a narrow crossing signals the danger of retaining only one channel; it is not evidence that adding this scalar term alone repairs the approximation.

Physical meaning. We have isolated two effects of electronic motion: a scalar energy correction and coupling to other channels. The next calculation will show, in an exactly solvable system, why eliminating the second effect can also modify nuclear inertia.

2 An exact spectrum against which to test the approximation

Physical meaning. The electronic problem of CO^+ has many interacting electrons. To test the approximation itself, first choose a quantum system for which both the full answer and every approximation can be calculated.

2.1 Define the model and hold the slow coordinate fixed

Consider real displacements $R, x \in \mathbb{R}$ and conjugate momenta

$$P = -\mathrm{i}\hbar\partial_R, \quad p = -\mathrm{i}\hbar\partial_x, \quad [R,P] = [x,p] = \mathrm{i}\hbar, \quad [R,x] = [R,p] = [P,p] = 0. \tag{2.1}$$

The square-integrable wavefunction is $\Psi(R,x)$ and the Hamiltonian is

$$H = \frac{P^2}{2M} + \frac{p^2}{2m} + \frac{KR^2}{2} + \frac{\kappa(x-aR)^2}{2}, \qquad M,m,K,\kappa > 0, \quad a \in \mathbb{R}. \tag{2.2}$$

The stiffnesses have dimensions energy/length^2; a is dimensionless. This is a confined system with a discrete spectrum. Here R is a displacement near equilibrium, not the positive molecular separation on its full domain. Both potential terms are nonnegative and vanish together only at $R = x = 0$.

For fixed R, put $y = x - aR$. The fast Hamiltonian is an oscillator,

$$h_f(R) = \frac{p^2}{2m} + \frac{\kappa(x-aR)^2}{2}, \qquad \omega_f = \sqrt{\kappa/m}, \tag{2.3}$$

$$\phi_0(x;R) = \left(\frac{m\omega_f}{\pi\hbar}\right)^{1/4} \exp\left[-\frac{m\omega_f(x-aR)^2}{2\hbar}\right], \qquad \epsilon_0 = \frac{\hbar\omega_f}{2}. \tag{2.4}$$

Normalization follows from $\int_{-\infty}^{\infty} \mathrm{e}^{-by^2}\,\mathrm{d}y = \sqrt{\pi/b}$ for $b > 0$. Translating the centre changes its wavefunction, but not its energy. BO therefore predicts

$$H_{\mathrm{BO}} = \frac{P^2}{2M} + \frac{KR^2}{2} + \frac{\hbar\omega_f}{2}, \qquad \Omega_{\mathrm{BO}} = \sqrt{K/M}, \tag{2.5}$$

$$E_v^{\mathrm{BO}} = \frac{\hbar\omega_f}{2} + \hbar\Omega_{\mathrm{BO}}(v+\tfrac{1}{2}), \qquad v = 0,1,2,\ldots. \tag{2.6}$$

These are the states with the fast oscillator in its ground channel.

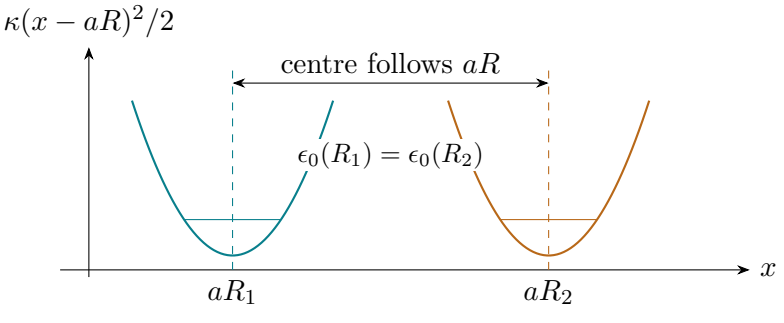


Figure. The fast ground-state energy is independent of the slow displacement, but the fast wavefunction moves with it. Vertical offsets and energy spacings in this sketch are schematic.

2.2 Solve the coupled equations before approximating

The Heisenberg equations follow from $\dot{O} = (\mathrm{i}/\hbar)[H,O]$. Since H is quadratic, they have the same linear form as classical equations:

$$\dot{R} = P/M, \qquad \dot{P} = -KR + a\kappa(x-aR), \tag{2.7}$$

$$\dot{x} = p/m, \qquad \dot{p} = -\kappa(x-aR). \tag{2.8}$$

Try a normal mode $(R, x) = (R_0, x_0)\mathrm{e}^{-\mathrm{i}\omega t}$. Its amplitudes obey

$$\begin{pmatrix} K + \kappa a^2 - M\omega^2 & -\kappa a \\ -\kappa a & \kappa - m\omega^2 \end{pmatrix} \begin{pmatrix} R_0 \\ x_0 \end{pmatrix} = 0. \tag{2.9}$$

A nonzero vector requires a vanishing determinant:

$$\begin{aligned} 0 &= (K + \kappa a^2 - M\omega^2)(\kappa - m\omega^2) - \kappa^2 a^2 \\ &= Mm\omega^4 - \left[M\kappa + mK + m\kappa a^2\right]\omega^2 + K\kappa. \end{aligned} \tag{2.10}$$

With $A = M\kappa + mK + m\kappa a^2$,

$$\omega_\pm^2 = \frac{A \pm \sqrt{A^2 - 4MmK\kappa}}{2Mm}, \qquad \omega_-^2 = \frac{2K\kappa}{A + \sqrt{A^2 - 4MmK\kappa}}. \tag{2.11}$$

The second expression avoids subtracting almost equal numbers. Both roots are positive: their product is $K\kappa/(Mm) > 0$, their sum is $A/(Mm) > 0$, and the mass-weighted stiffness matrix below is real symmetric and positive definite.

2.3 Why these frequencies give the exact quantum spectrum

Introduce mass-weighted coordinates and conjugate momenta

$$\boldsymbol{Q} = \begin{pmatrix} \sqrt{M}R \\ \sqrt{m}x \end{pmatrix}, \quad \boldsymbol{\Pi} = \begin{pmatrix} P/\sqrt{M} \\ p/\sqrt{m} \end{pmatrix}, \quad [Q_i, \Pi_j] = \mathrm{i}\hbar\delta_{ij}. \tag{2.12}$$

Then $H = \boldsymbol{\Pi}^T\boldsymbol{\Pi}/2 + \boldsymbol{Q}^T\mathsf{C}\boldsymbol{Q}/2$, where

$$\mathsf{C} = \begin{pmatrix} (K + \kappa a^2)/M & -\kappa a/\sqrt{Mm} \\ -\kappa a/\sqrt{Mm} & \kappa/m \end{pmatrix}. \tag{2.13}$$

A real orthogonal matrix O diagonalizes C. Applying the same transformation to coordinates and momenta, $\boldsymbol{q} = O^T\boldsymbol{Q}$ and $\boldsymbol{\pi} = O^T\boldsymbol{\Pi}$, preserves the commutators. Thus

$$H = \sum_{s=\pm} \left(\frac{\pi_s^2}{2} + \frac{\omega_s^2 q_s^2}{2} \right). \tag{2.14}$$

For each s, define

$$b_s = \sqrt{\frac{\omega_s}{2\hbar}}q_s + \frac{\mathrm{i}\pi_s}{\sqrt{2\hbar\omega_s}}, \quad [b_s, b_{s'}^\dagger] = \delta_{ss'}, \quad H_s = \hbar\omega_s(b_s^\dagger b_s + \tfrac{1}{2}). \tag{2.15}$$

The number operator has eigenvalues $n_s = 0, 1, \ldots$, with $b_s|n_s\rangle = \sqrt{n_s}|n_s - 1\rangle$. The exact energies are

$$\boxed{E_{n_-,n_+} = \hbar\omega_-(n_- + \tfrac{1}{2}) + \hbar\omega_+(n_+ + \tfrac{1}{2}).} \tag{2.16}$$

The low-energy spacing is $E_{v+1,0} - E_{v,0} = \hbar\omega_-$. Excitations of the fast mode cost $\hbar\omega_+$.

2.4 State the small parameters, then expand

Define

$$\varepsilon = \frac{m}{M}, \quad \rho = \frac{K}{\kappa}, \quad \frac{\Omega_{\mathrm{BO}}^2}{\omega_f^2} = \varepsilon\rho, \quad z_\pm = \frac{\omega_\pm^2}{\omega_f^2}. \tag{2.17}$$

Equation (2.10) becomes

$$z^2 - [1 + \varepsilon(\rho + a^2)]z + \varepsilon\rho = 0. \tag{2.18}$$

For $\varepsilon \to 0$ at fixed finite ρ, a, insert $z_- = \varepsilon\rho + c_2\varepsilon^2 + c_3\varepsilon^3 + \cdots$:

$$\varepsilon^2: \quad \rho^2 - c_2 - \rho(\rho + a^2) = 0 \quad \Rightarrow c_2 = -\rho a^2, \tag{2.19}$$

$$\varepsilon^3: \quad 2\rho c_2 - c_3 - (\rho + a^2)c_2 = 0 \quad \Rightarrow c_3 = \rho a^2(a^2 - \rho). \tag{2.20}$$

Using $\sqrt{1+u} = 1 + u/2 - u^2/8 + \cdots$,

$$\omega_- = \Omega_{\mathrm{BO}}\left[1 - \frac{a^2\varepsilon}{2} + \left(\frac{3a^4}{8} - \frac{a^2\rho}{2}\right)\varepsilon^2 + O(\varepsilon^3)\right], \tag{2.21}$$

$$\omega_+ = \omega_f\left[1 + \frac{a^2\varepsilon}{2} + O(\varepsilon^2)\right]. \tag{2.22}$$

For a coupled system, a sufficient regime for the controlled slow–fast expansion is separated frequencies and a small induced-mass fraction: $\varepsilon\rho \ll 1$ and $a^2\varepsilon \ll 1$. The mass ratio alone is insufficient if stiffnesses or couplings scale unfavourably.

2.5 The diagonal correction is real, but it is not enough

Differentiate the normalized fast ground state at fixed x:

$$\partial_R\phi_0 = \frac{m\omega_f a}{\hbar}(x - aR)\phi_0, \qquad \langle(x - aR)^2\rangle_0 = \frac{\hbar}{2m\omega_f}. \tag{2.23}$$

Its first derivative expectation vanishes by oddness, while

$$\langle\partial_R\phi_0|\partial_R\phi_0\rangle = \frac{m^2\omega_f^2 a^2}{\hbar^2}\frac{\hbar}{2m\omega_f} = \frac{m\omega_f a^2}{2\hbar}, \tag{2.24}$$

$$\Phi = \frac{\hbar^2}{2M}\langle\partial_R\phi_0|\partial_R\phi_0\rangle = \frac{\hbar m\omega_f a^2}{4M}. \tag{2.25}$$

This constant shifts every level in the retained channel equally. It reproduces the leading change $\hbar(\omega_+ - \omega_f)/2$ in the fast zero-point energy, but cancels from every slow line spacing.

At fixed v and fixed K, κ, m, a,

$$E_{v,0} = \frac{\hbar\omega_f}{2} + \Phi + \hbar\Omega_{\mathrm{BO}}(v + \tfrac{1}{2})\left(1 - \frac{ma^2}{2M}\right) + O(M^{-2}). \tag{2.26}$$

The powers differ: Φ is of order M^{-1}; the correction to a slow spacing is of order $M^{-3/2}$. The remainder refers to total energy as M increases with the other parameters fixed.

There is a useful variational check. For the normalized BO product ground state $\Psi_{\mathrm{trial}}(R, x) = \chi_0^{\mathrm{BO}}(R)\phi_0(x; R)$,

$$\langle\Psi_{\mathrm{trial}}|H|\Psi_{\mathrm{trial}}\rangle = E_0^{\mathrm{BO}} + \Phi \;\geq\; E_{0,0}. \tag{2.27}$$

The full expectation includes the derivative kinetic cost. The bare BO eigenvalue alone is not this variational expectation. A positive diagonal correction and a negative correction from further mixing are therefore compatible.

Checkpoint. The fast state changes with R even though its eigenvalue does not. The diagonal correction notices that change, but a constant potential correction cannot repair the wrong slow frequency.

2.6 A quantitative accuracy check

For illustration only, take $a = 1$ and $\rho = 1$. These dimensionless numbers are model results, not CO^+ predictions.

m/M	Exact $\omega_-/\Omega_{\mathrm{BO}}$	BO value	$1/\sqrt{1+m/M}$
0.001	0.999500	1	0.999500
0.010	0.994988	1	0.995037
0.100	0.949200	1	0.953463

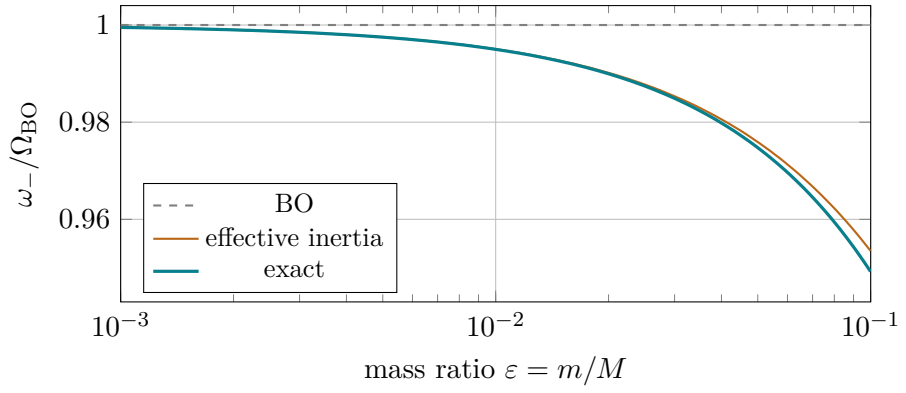


Figure. For $a = \rho = 1$, the exact low-frequency spacing approaches BO as the mass ratio tends to zero. The induced-inertia approximation captures the first correction. Both axes are dimensionless.

The last column retains induced inertia but neglects higher time derivatives, as derived in Section 3. For an isotope-like change, vary M holding m, K, κ, a fixed. BO predicts $M\omega_-^2 = K$; the first correction gives

$$M\omega_-^2 \simeq \frac{KM}{M + ma^2}, \qquad \frac{M_A\omega_-^2(M_A)}{M_B\omega_-^2(M_B)} - 1 \simeq ma^2\left(\frac{1}{M_B} - \frac{1}{M_A}\right). \tag{2.28}$$

The last equality is first order in $ma^2/M_{A,B}$. This is a mass-scaled residual with a known microscopic cause in the benchmark. Establishing the cause of a real molecular residual requires additional information.

3 Virtual excitations and the effective Hamiltonian

Physical meaning. The exact result tells us what is missing. We now recover it by eliminating fast motion, first through linear equations and then through quantum perturbation theory. These are elementary operations behind an effective field theory.

3.1 Eliminate one amplitude exactly

Add a weak classical force $f(t)$ through $H \mapsto H - f(t)R$. It is a probe of the closed system, not a fluctuating environment. For a Fourier component proportional to $\mathrm{e}^{-\mathrm{i}\omega t}$, the second row of (2.9) gives

$$x_0 = \frac{\kappa a}{\kappa - m\omega^2} R_0 = \frac{a}{1 - \omega^2/\omega_f^2} R_0. \tag{3.1}$$

Substitution in the first row gives $D(\omega)R_0 = f_0$, where

$$\begin{aligned} D(\omega) &= K + \kappa a^2 - M\omega^2 - \frac{\kappa^2 a^2}{\kappa - m\omega^2} \\ &= K - M\omega^2 - \frac{ma^2\omega^2}{1 - \omega^2/\omega_f^2}. \end{aligned} \tag{3.2}$$

For $a \neq 0$, the poles of $1/D$ are both coupled normal modes. At $a = 0$ the fast oscillator decouples and is invisible to a force acting on R alone. The elimination formula itself is singular at the uncoupled fast frequency; use the original 2×2 system there. A pole in an intermediate elimination formula is not automatically a normal-mode pole of the full system.

For $|\omega| < \omega_f$, the geometric series $1/(1-u) = 1 + u + u^2 + \cdots$ yields

$$D(\omega) = K - (M + ma^2)\omega^2 - \frac{ma^2}{\omega_f^2}\omega^4 + O(ma^2\omega^6/\omega_f^4). \tag{3.3}$$

Retaining the first two terms gives

$$M_{\text{eff}} = M + ma^2, \qquad \Omega_{\text{eff}} = \sqrt{\frac{K}{M + ma^2}}. \tag{3.4}$$

For slow motion $x \simeq aR$, the fast kinetic energy is already $m\dot{x}^2/2 \simeq ma^2\dot{R}^2/2$. This estimate predicts the sign and size of the result; exact elimination specifies its corrections and its range of validity.

The ω^4 term is a higher-derivative correction, used perturbatively at low frequency. Solving a truncated higher-derivative equation for additional high-frequency roots would introduce solutions outside its accuracy; the exact system has only the two modes in (2.11).

3.2 The same elimination in stationary quantum mechanics

Let $\mathcal{P}$ project onto the retained fast channel and $\mathcal{Q} = 1 - \mathcal{P}$ onto the other channels. Thus $\mathcal{P}^2 = \mathcal{P}$, $\mathcal{Q}^2 = \mathcal{Q}$, and $\mathcal{P}\mathcal{Q} = 0$. Set

$$|\psi\rangle = \mathcal{P}|\Psi\rangle, \qquad |\eta\rangle = \mathcal{Q}|\Psi\rangle. \tag{3.5}$$

Project $H|\Psi\rangle = E|\Psi\rangle$ on each subspace:

$$\mathcal{P}H\mathcal{P}|\psi\rangle + \mathcal{P}H\mathcal{Q}|\eta\rangle = E|\psi\rangle, \tag{3.6}$$

$$\mathcal{Q}H\mathcal{P}|\psi\rangle + \mathcal{Q}H\mathcal{Q}|\eta\rangle = E|\eta\rangle. \tag{3.7}$$

When $E - \mathcal{Q}H\mathcal{Q}$ is invertible in the eliminated subspace,

$$|\eta\rangle = (E - \mathcal{Q}H\mathcal{Q})^{-1}\mathcal{Q}H\mathcal{P}|\psi\rangle. \tag{3.8}$$

Substitute back to obtain the exact, energy-dependent operator

$$\boxed{H_{\text{eff}}(E) = \mathcal{P}H\mathcal{P} + \mathcal{P}H\mathcal{Q}\frac{1}{E - \mathcal{Q}H\mathcal{Q}}\mathcal{Q}H\mathcal{P}.} \tag{3.9}$$

This is block-matrix elimination, not a new postulate. The second term describes leaving the retained channel, propagating virtually in the other channels, and returning. Below the eliminated spectrum its denominators are negative; coupling to higher states lowers a low-lying level.

The inverse operator is called a *resolvent*. If $H_0|j\rangle = E_j^{(0)}|j\rangle$ and V weakly couples distinct levels, expanding (3.9) gives

$$\delta E_i^{(2)} = \sum_{j\neq i} \frac{|\langle j|V|i\rangle|^2}{E_i^{(0)} - E_j^{(0)}}. \tag{3.10}$$

The numerator has units energy squared and the denominator energy. A small denominator signals the need to retain the nearly degenerate state explicitly, as in Part 1.

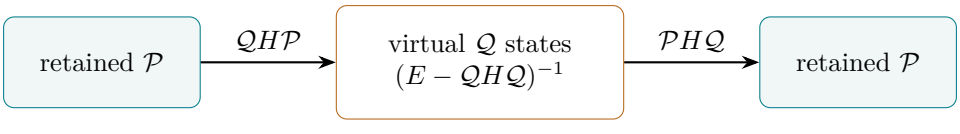


Figure. Second-order channel elimination. Read the picture from left to right as state transitions; the operator product in (3.9) acts from right to left. This is a virtual transition diagram, not an assertion that particles are created in the molecular calculation.

3.3 Recover the induced inertia from one virtual excitation

Change variables to $R, y = x - aR$. The Jacobian is one, preserving the inner product. The chain rule gives

$$\partial_R|_x = \partial_R|_y - a\partial_y, \qquad \partial_x = \partial_y. \tag{3.11}$$

With $\Pi = -\mathrm{i}\hbar\partial_R|_y$ and $p_y = -\mathrm{i}\hbar\partial_y$,

$$H = \underbrace{\frac{\Pi^2}{2M} + \frac{KR^2}{2} + \frac{p_y^2}{2m} + \frac{\kappa y^2}{2}}_{H_0} + \underbrace{\left(-\frac{a}{M}\Pi p_y\right)}_{V_1} + \underbrace{\frac{a^2 p_y^2}{2M}}_{V_2}. \tag{3.12}$$

Π and p_y commute because they act on different coordinates. For the fast oscillator,

$$p_y = \mathrm{i}\sqrt{\frac{m\hbar\omega_f}{2}}(b_f^\dagger - b_f), \qquad \langle 0|p_y|0\rangle = 0, \tag{3.13}$$

$$\langle 0|p_y^2|0\rangle = \frac{m\hbar\omega_f}{2}, \qquad |\langle 1|p_y|0\rangle|^2 = \frac{m\hbar\omega_f}{2}. \tag{3.14}$$

First-order projection of V_2 gives $\langle 0|V_2|0\rangle = \Phi$ in (2.25); V_1 has zero diagonal expectation. At second order V_1 connects the fast ground state only to $|1\rangle_f$. To leading order in slow energies divided by $\hbar\omega_f$, the denominator is $-\hbar\omega_f$, so

$$\begin{aligned}\delta H_{\text{kin}} &\simeq \frac{a^2}{M^2}\Pi^2\frac{|\langle 1|p_y|0\rangle|^2}{-\hbar\omega_f} \\ &= -\frac{ma^2}{2M^2}\Pi^2.\end{aligned} \tag{3.15}$$

Consequently

$$H_{\text{eff}} \simeq \frac{\Pi^2}{2M}\left(1-\frac{ma^2}{M}\right)+\frac{KR^2}{2}+\frac{\hbar\omega_f}{2}+\Phi. \tag{3.16}$$

Since $1/(M+ma^2) = M^{-1}(1-ma^2/M+\cdots)$, this agrees with the response calculation. The correction lowers a kinetic coefficient; expressed as an inertia it is positive.

3.4 Check the energy denominator rather than just assuming it

In an unperturbed slow state $|v\rangle$ with frequency $\Omega = \Omega_{\text{BO}}$,

$$|\langle v+1|\Pi|v\rangle|^2 = \frac{M\hbar\Omega}{2}(v+1), \quad |\langle v-1|\Pi|v\rangle|^2 = \frac{M\hbar\Omega}{2}v. \tag{3.17}$$

The intermediate states are $|v+1,1\rangle$ and $|v-1,1\rangle$. Their denominators are $-\hbar(\omega_f+\Omega)$ and $-\hbar(\omega_f-\Omega)$. Thus

$$\delta E^{(2)}_{v,0}[V_1] = -\frac{a^2 m\hbar\Omega\omega_f}{4M}\left[\frac{v+1}{\omega_f+\Omega}+\frac{v}{\omega_f-\Omega}\right]. \tag{3.18}$$

For $\Omega/\omega_f \ll 1$,

$$\delta E^{(2)}_{v,0}[V_1] = -\frac{ma^2}{2M}\hbar\Omega(v+\tfrac{1}{2}) + O\left(\frac{ma^2}{M}\frac{\hbar\Omega^2}{\omega_f}\right). \tag{3.19}$$

This verifies (2.26) at the retained order. Equation (3.18) checks the V_1^2 contribution, not the complete higher-order expansion: terms involving V_2 must also be included beyond the accuracy of (3.16). Near $\Omega = \omega_f$, nondegenerate perturbation theory cannot describe resonant mixing of the relevant excited states. For $v = 0$ the downward intermediate state is absent.

Checkpoint. Two distinct effects survive: a scalar energy shift Φ and a kinetic correction. The latter changes isotope-dependent line spacings. Both have been matched to an exactly solvable spectrum.

3.5 The precise connection to quantum field theory

In an effective theory, *matching* means choosing coefficients in a reduced description so it reproduces specified low-energy results of a more complete theory. Here we match the low-frequency response and low-lying spectrum. Expanding the resolvent or (3.2) in small energies generates local operators, starting with a potential and a kinetic term. More derivatives carry additional powers of slow-to-fast frequency ratios.

The projected Green operator illustrates the same statement:

$$\mathcal{P}(E-H)^{-1}\mathcal{P} = [E-H_{\text{eff}}(E)]^{-1}_{\mathcal{P}}. \tag{3.20}$$

This follows by solving the block equations with a source in $\mathcal{P}$, at energies where the inverses exist. In QFT the corresponding energy-dependent insertion is called a self-energy. Here it arises from a finite set of quantum coordinates and is explicitly calculable. No ultraviolet divergence or renormalization prescription is needed.

For a molecule, the retained motion is nuclear and the eliminated excitations are electronic. In some heavy-quark systems, slowly moving heavy quarks interact with rapidly responding gluonic or light-quark configurations. Born–Oppenheimer effective theories formalize both settings [2]. Their potentials, spin structures and expansion parameters differ; the common operation is to separate scales and retain virtual-excitation effects as operators. The calculation above establishes that operation without NRQED or QCD effective-theory notation.

4 From an effective Hamiltonian to isotope spectra

Physical meaning. The coupled model established a mechanism: eliminating fast motion can change both a potential and an inertia. We now ask which combinations of molecular line positions can test a common electronic potential. The oscillator model illustrates the mechanism; its parameters are not quantitative parameters of CO^+.

4.1 The reference problem: one potential, different nuclear masses

Choose one electronic state of a diatomic molecule, with a smooth BO potential $U(R)$ and internuclear distance $R > 0$. For nuclear masses M_C and M_O, the reduced mass is

$$\mu = \frac{M_C M_O}{M_C + M_O}. \tag{4.1}$$

Here and below a mass convention must be held fixed; we use nuclear masses in the derivation. We consider a spin-free Σ electronic state. At this stage we omit electronic spin and corrections to the one-surface BO Hamiltonian. Let $N = 0, 1, \ldots$ denote rotational angular momentum excluding electron spin, and M_N its projection quantum number. With normalized spherical harmonics Y_{NM_N}, write the nuclear wavefunction as

$$\chi_{vNM_N}(R, \theta, \varphi) = \frac{u_{vN}(R)}{R} Y_{NM_N}(\theta, \varphi), \qquad \int_0^\infty |u_{vN}(R)|^2 \, \mathrm{d}R = 1. \tag{4.2}$$

The factor $1/R$ removes the radial Jacobian from the normalization. In fact,

$$\nabla_R^2 = \frac{1}{R^2}\partial_R(R^2\partial_R) - \frac{\hat{\boldsymbol{N}}^2}{\hbar^2 R^2}, \qquad \frac{1}{R^2}\partial_R\left[R^2\partial_R\left(\frac{u}{R}\right)\right] = \frac{u''}{R}. \tag{4.3}$$

Using $\hat{\boldsymbol{N}}^2 Y_{NM_N} = \hbar^2 N(N+1) Y_{NM_N}$ therefore gives

$$\boxed{\left[-\frac{\hbar^2}{2\mu}\frac{\mathrm{d}^2}{\mathrm{d}R^2} + U(R) + \frac{\hbar^2 N(N+1)}{2\mu R^2}\right] u_{vN} = E_{vN} u_{vN}.} \tag{4.4}$$

Regularity imposes $u(0) = 0$; bound states decay at large R. The electronic potential is held fixed when changing isotope. Its eigenvalues nevertheless change because the kinetic and centrifugal terms contain μ.

Checkpoint. An isotope shift occurs already within the BO approximation. A test of BO corrections must first account for this ordinary mass dependence.

4.2 Harmonic vibration and the first mass-scaling law

Let R_e be a minimum of U, $q = R - R_e$, and $k_n = U^{(n)}(R_e)$ its nth derivative. Then

$$U(R) = U_e + \frac{k_2 q^2}{2} + \frac{k_3 q^3}{6} + \frac{k_4 q^4}{24} + \cdots, \quad U_e = U(R_e), \quad k_2 > 0. \tag{4.5}$$

For low vibrational states whose spatial extent is much smaller than R_e, we may extend the local coordinate from $q > -R_e$ to the whole real line. For $N = 0$ define

$$\Omega = \sqrt{\frac{k_2}{\mu}}, \qquad \ell = \sqrt{\frac{\hbar}{2\mu\Omega}}, \qquad \hat{b} = \frac{q}{2\ell} + \frac{\mathrm{i}\ell p}{\hbar}, \qquad p = -\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}q}. \tag{4.6}$$

Since $[q,p]=\mathrm{i}\hbar$, these definitions imply

$$[\hat{b},\hat{b}^\dagger]=1, \qquad q=\ell(\hat{b}+\hat{b}^\dagger), \qquad p=\frac{\hbar}{2\mathrm{i}\ell}(\hat{b}-\hat{b}^\dagger), \tag{4.7}$$

$$H_0-U_e=\hbar\Omega(\hat{b}^\dagger\hat{b}+\tfrac{1}{2}), \qquad \hat{b}|v\rangle=\sqrt{v}|v-1\rangle, \qquad \hat{b}^\dagger|v\rangle=\sqrt{v+1}|v+1\rangle. \tag{4.8}$$

The integer $v\geq 0$ labels vibration, so

$$E_v^{(0)}=U_e+\hbar\Omega(v+\tfrac{1}{2}), \qquad \omega_e=\frac{\Omega}{2\pi c}\propto\mu^{-1/2}. \tag{4.9}$$

Ω is an angular frequency; ω_e is the conventional spectroscopic *wavenumber*, despite its omega symbol. Energies divided by hc have the same units as ω_e. Use a consistent length unit for c and wavenumbers.

4.3 Anharmonicity is part of the BO baseline

The following algebra supplies the correction needed before interpreting isotope residuals. Treat $V_3=k_3q^3/6$ and $V_4=k_4q^4/24$ perturbatively, retaining first order in k_4 and second order in k_3. The matrix elements linking the relevant low states must be small compared with their harmonic energy separations. The local Taylor polynomial is not assumed to describe the potential at arbitrarily large $|q|$.

Put $A=\hat{b}+\hat{b}^\dagger$. Applying the ladder operators twice gives

$$A^2|v\rangle=\sqrt{v(v-1)}|v-2\rangle+(2v+1)|v\rangle+\sqrt{(v+1)(v+2)}|v+2\rangle. \tag{4.10}$$

Consequently, by orthonormality,

$$\langle v|A^4|v\rangle=\|A^2|v\rangle\|^2 \tag{4.11}$$

$$=v(v-1)+(2v+1)^2+(v+1)(v+2)=6v^2+6v+3, \tag{4.12}$$

$$E_{v,4}^{(1)}=\frac{k_4\ell^4}{8}(2v^2+2v+1). \tag{4.13}$$

Odd parity gives $\langle v|q^3|v\rangle=0$. Applying A once more instead yields

$$A^3|v\rangle=\sqrt{v(v-1)(v-2)}|v-3\rangle+3v\sqrt{v}|v-1\rangle + 3(v+1)\sqrt{v+1}|v+1\rangle+\sqrt{(v+1)(v+2)(v+3)}|v+3\rangle. \tag{4.14}$$

Terms with a negative occupation number are absent. Nondegenerate second-order perturbation theory now gives

$$E_{v,3}^{(2)}=\sum_{w\neq v}\frac{|\langle w|V_3|v\rangle|^2}{E_v^{(0)}-E_w^{(0)}} \tag{4.15}$$

$$=\frac{k_3^2\ell^6}{36\hbar\Omega}\left[\frac{v(v-1)(v-2)}{3}+9v^3-9(v+1)^3-\frac{(v+1)(v+2)(v+3)}{3}\right] \tag{4.16}$$

$$=-\frac{k_3^2\ell^6}{36\hbar\Omega}(30v^2+30v+11). \tag{4.17}$$

The negative denominators for intermediate states above v are essential. Thus the result through these orders is

$$E_v=U_e+\hbar\Omega(v+\tfrac{1}{2})+\frac{k_4\ell^4}{8}(2v^2+2v+1)-\frac{k_3^2\ell^6}{36\hbar\Omega}(30v^2+30v+11)+\cdots. \tag{4.18}$$

For an isotope-independent U, the derivatives k_n do not change, while

$$\Omega\propto\mu^{-1/2}, \quad \ell\propto\mu^{-1/4}, \quad k_4\ell^4\propto\mu^{-1}, \quad \frac{k_3^2\ell^6}{\hbar\Omega}\propto\mu^{-1}. \tag{4.19}$$

Anharmonic contributions therefore scale differently from the harmonic spacing, even though the potential is identical for every isotope.

4.4 Rotation, vibration–rotation coupling, and distortion

At fixed R_e, the rotational energy is $hcB_eN(N+1)$, where

$$B_e = \frac{\hbar^2}{2\mu R_e^2 hc} \propto \mu^{-1}. \tag{4.20}$$

At first order in rotation the vibrating molecule samples R^{-2}:

$$B_v = \frac{\hbar^2}{2\mu hc}\langle R^{-2}\rangle_v, \tag{4.21}$$

$$(R_e+q)^{-2} = R_e^{-2}\left(1 - \frac{2q}{R_e} + \frac{3q^2}{R_e^2} + \cdots\right). \tag{4.22}$$

To leading harmonic order, $\langle q^2\rangle_v = \ell^2(2v+1)$. Anharmonicity also shifts the mean bond length. In a stationary bound state $\langle \dot{p}\rangle = -\langle U'(q)\rangle = 0$, so to first order in k_3,

$$k_2\langle q\rangle_v + \frac{k_3}{2}\langle q^2\rangle_v = 0, \qquad \langle q\rangle_v = -\frac{k_3\ell^2}{2k_2}(2v+1) + \cdots . \tag{4.23}$$

Substitution gives the leading relation

$$B_v = B_e - \alpha_e(v+\tfrac{1}{2}) + \cdots , \tag{4.24}$$

$$\alpha_e = -2B_e\ell^2\left(\frac{k_3}{k_2R_e} + \frac{3}{R_e^2}\right) \propto \mu^{-3/2}. \tag{4.25}$$

Neither the sign nor the size of α_e follows from mass scaling alone. In particular, B_0 and B_e are different quantities.

Rotation also stretches the bond. Write $L = N(N+1)$ and $C = hcB_e$. Keeping the first displacement term in the centrifugal potential gives

$$\frac{k_2q^2}{2} + CL - \frac{2CL}{R_e}q = \frac{k_2}{2}\left(q - \frac{2CL}{k_2R_e}\right)^2 + CL - \frac{2C^2}{k_2R_e^2}L^2, \tag{4.26}$$

$$D_e = \frac{2(hcB_e)^2}{k_2R_e^2hc} = \frac{4B_e^3}{\omega_e^2} \propto \mu^{-2}. \tag{4.27}$$

This leading harmonic estimate explains the negative centrifugal-distortion term $-D_eL^2$. It requires a small rotational displacement relative to R_e; at higher N, additional terms and state-dependent constants are necessary.

4.5 What a Dunham expansion does and does not assume

Define the spin-independent term value $T_{vN} = E_{vN}/(hc)$. A local spectroscopic representation is

$$T_{vN} = \sum_{k,l\geq 0} Y_{kl}(v+\tfrac{1}{2})^k[N(N+1)]^l. \tag{4.28}$$

Its coefficients are not new energy levels: they organize the dependence of the levels on vibration and rotation. The calculations above give the leading identifications $Y_{10} \simeq \omega_e$, $Y_{01} \simeq B_e$, $Y_{11} \simeq -\alpha_e$ and $Y_{02} \simeq -D_e$. They illustrate the usual leading BO mass scaling

$$Y_{kl}^{(\alpha)} \simeq U_{kl}\mu_\alpha^{-(k/2+l)}, \tag{4.29}$$

where α labels an isotopologue and U_{kl} is independent of its masses at this order. This is a controlled leading-order statement, not a claim that every exact fitted coefficient has one pure mass power.

Higher-order Dunham corrections can modify the identification of Y_{01} with the equilibrium constant B_e, even for a common potential [3, 4].

A common first-order parameterization of isotope corrections is

$$Y_{kl}^{(\alpha)} = U_{kl}\mu_\alpha^{-p}\left[1 + \frac{m_e}{M_{\mathrm{C}}^{(\alpha)}}\Delta_{kl}^{\mathrm{C}} + \frac{m_e}{M_{\mathrm{O}}^{(\alpha)}}\Delta_{kl}^{\mathrm{O}} + \cdots\right], \qquad p = \frac{k}{2} + l. \tag{4.30}$$

Here m_e is the electron mass and the Δ_{kl} are dimensionless fitted parameters in the displayed sign convention. The units of U_{kl} depend on the common mass unit used. Published conventions may use atomic rather than nuclear masses and a corresponding redistribution of correction terms; quoted parameters must be converted consistently before comparison.

The coefficients in Eq. (4.30) do not identify a unique microscopic mechanism by themselves. For CO^+ rotational constants, electronic-structure analysis distinguishes the adiabatic change in equilibrium bond length, electronic contributions to rotational inertia, and Dunham corrections needed when connecting fitted coefficients to equilibrium quantities [5]. The last are not, by themselves, evidence of BO failure.

4.6 From CO^+ emission lines to a ground-state interval

The target emission system is the $B^2\Sigma^+ \to X^2\Sigma^+$ system of $^{13}C^{17}O^+$; related FT-UV work concerns $^{12}C^{17}O^+$ [7]. Here B and X name electronic states; they must not be confused with a rotational constant B_v. The superscript 2 means electron spin $S = 1/2$; Σ means zero projection of electronic orbital angular momentum on the molecular axis; $+$ specifies even reflection symmetry in a plane containing that axis.

Energy conservation for a photon gives

$$hc\widetilde{\nu}_{u\to l} = E_u - E_l, \qquad \widetilde{\nu} = \nu/c, \tag{4.31}$$

where ν is ordinary frequency. A line thus measures a *difference* between two electronic-state term values. A residual in a single line cannot automatically be assigned to the ground state. Two transitions sharing the *same resolved upper level* eliminate it exactly:

$$\boxed{\widetilde{\nu}_{u\to l_1} - \widetilde{\nu}_{u\to l_2} = \frac{E_{l_2} - E_{l_1}}{hc}.} \tag{4.32}$$

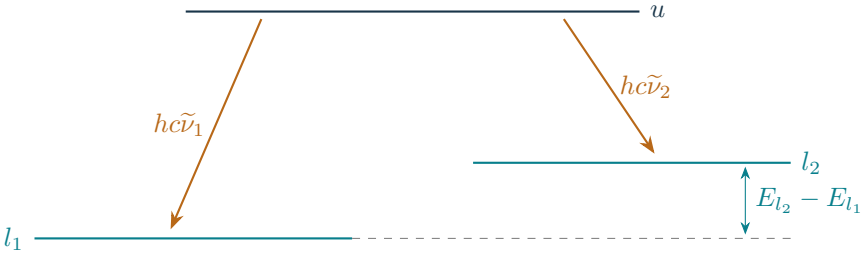


Figure. Two assigned emissions share the same upper level. Subtraction cancels its energy exactly, leaving a lower-state interval. The diagram is schematic, with $E_{l_2} > E_{l_1}$.

The subtraction is performed on assigned transitions, not on unrelated nearby features. In the spin-free electric-dipole Σ–Σ model, the rotational selection rule is $\Delta N = \pm 1$. A common upper N_u therefore connects lower levels $N_u - 1$ and $N_u + 1$, separated by $\Delta N = 2$.

Before using that simple model for CO^+, the doublet fine structure must be handled. Let $\boldsymbol{J} = \boldsymbol{N}+\boldsymbol{S}$. At leading order,

$$H_{\rm sr} = hc\gamma_v \frac{\boldsymbol{N}\cdot\boldsymbol{S}}{\hbar^2}, \quad \boldsymbol{N}\cdot\boldsymbol{S} = \tfrac{1}{2}(\boldsymbol{J}^2 - \boldsymbol{N}^2 - \boldsymbol{S}^2), \tag{4.33}$$

$$T_{vN,J} = T_{vN}^{\rm sf} + \frac{\gamma_v}{2}[J(J+1) - N(N+1) - \tfrac{3}{4}], \tag{4.34}$$

$$T_{F_1} = T_{vN}^{\rm sf} + \frac{\gamma_v N}{2}, \quad J = N + \tfrac{1}{2}, \tag{4.35}$$

$$T_{F_2} = T_{vN}^{\rm sf} - \frac{\gamma_v (N+1)}{2}, \quad J = N - \tfrac{1}{2} \quad (N \geq 1). \tag{4.36}$$

The superscript sf denotes the spin-free level and γ_v is a wavenumber. Each J level has degeneracy $2J+1$. Thus, in this model,

$$T_{vN}^{\rm sf} = \frac{(2N+2)T_{F_1} + 2NT_{F_2}}{4N+2} \quad (N \geq 1). \tag{4.37}$$

At $N = 0$ only $J = 1/2$ exists and its spin–rotation shift is zero. These are degeneracy weights for *energy levels*, not arbitrary intensity weights for unresolved emission lines. Nuclear hyperfine structure and additional perturbations must be fitted or otherwise accounted for if they matter at the measurement precision; they are not included in this minimal Hamiltonian. The following extraction applies to spin-free term values inferred from such an appropriate analysis.

4.7 A closed analytic diagnostic from the rotational intervals

At fixed v, retain

$$T_{vN}^{\rm sf} = G_v + B_v L - D_v L^2, \qquad L = N(N+1), \tag{4.38}$$

where G_v is independent of N. Set $L_+ = (N+2)(N+3)$. Since $L_+ - L = 4N+6$ and $L_+^2 - L^2 = (L_+ - L)(L_+ + L)$,

$$S_N = \frac{T_{v,N+2}^{\rm sf} - T_{vN}^{\rm sf}}{4N+6} = B_v - D_v f_N, \tag{4.39}$$

$$f_N = L_+ + L = 2N^2 + 6N + 6. \tag{4.40}$$

Two different intervals, at $N_a \neq N_b$, determine

$$D_v = \frac{S_{N_a} - S_{N_b}}{f_{N_b} - f_{N_a}}, \qquad B_v = \frac{f_{N_b} S_{N_a} - f_{N_a} S_{N_b}}{f_{N_b} - f_{N_a}}. \tag{4.41}$$

Additional intervals test the straight-line form in Eq. (4.40); curvature can reveal omitted higher rotational terms rather than BO breakdown. Data in $v = 0$ and $v = 1$ additionally give, at the linear order of Eq. (4.25),

$$\alpha_e = B_0 - B_1, \qquad B_e \simeq \frac{3B_0 - B_1}{2}. \tag{4.42}$$

Quadratic vibration–rotation terms would change this extrapolation. Its finite-order uncertainty must accompany any precision isotope test.

In the following test, $B_e^{(a,b)}$ means an inferred *effective equilibrium* rotational coefficient after vibrational extrapolation and the required Dunham corrections. It may contain adiabatic bond-length and electronic-inertia effects. It need not equal the bare BO expression in Eq. (4.20); equality is the hypothesis under test.

For isotopologues a, b, construct

$$\overline{B}_a = \mu_a B_e^{(a)}, \qquad \mathcal{R}_{ab} = \frac{\mu_a B_e^{(a)}}{\mu_b B_e^{(b)}} - 1. \tag{4.43}$$

The ideal fixed-R_e model predicts $\overline{B}_a = \hbar^2/(2hcR_e^2)$ and $\mathcal{R}_{ab} = 0$. This is a leading equilibrium test, not a claim that the directly measured μB_0 or an uncorrected fitted μY_{01} must be exactly isotope independent.

If mass uncertainties are negligible and the rotational constants have uncertainties σ_a, σ_b, linear error propagation gives

$$\mathrm{Var}(\mathcal{R}_{ab}) \simeq (1+\mathcal{R}_{ab})^2 \left[\frac{\sigma_a^2}{(B_e^{(a)})^2} + \frac{\sigma_b^2}{(B_e^{(b)})^2} - \frac{2\,\mathrm{Cov}(B_e^{(a)}, B_e^{(b)})}{B_e^{(a)} B_e^{(b)}} \right]. \tag{4.44}$$

Covariance matters when constants share lines, calibration, or fit parameters. Model truncation and unresolved spectral structure require separate assessment; a small statistical error alone does not establish the physical origin of a residual.

Finally, one isotope cannot determine independent carbon and oxygen corrections. At fixed k, l, Eq. (4.30) has the linear structure

$$Z_\alpha \equiv \mu_\alpha^p Y_{kl}^{(\alpha)} = U_{kl} + \frac{A_{kl}}{M_{\mathrm{C}}^{(\alpha)}} + \frac{C_{kl}}{M_{\mathrm{O}}^{(\alpha)}}, \quad A_{kl} = m_e U_{kl} \Delta_{kl}^{\mathrm{C}}, \quad C_{kl} = m_e U_{kl} \Delta_{kl}^{\mathrm{O}}. \tag{4.45}$$

For three unknown coefficients the isotope design matrix, whose rows are $(1, 1/M_{\mathrm{C}}^{(\alpha)}, 1/M_{\mathrm{O}}^{(\alpha)})$, must have rank three. Varying only carbon leaves the oxygen column proportional to the constant column. Independent isotope substitutions or external theoretical information are required to separate the contributions.

Checkpoint. The achievable result is a chain from an effective Hamiltonian to a testable isotope residual, with explicit baseline corrections and uncertainty. CO^+ fits already include BO-breakdown terms [6]; these notes derive what such terms can mean and what the spectra alone can determine.

5 From the calculation to an interpretable spectral test

5.1 The chain of inference

An isotope comparison joins several distinct calculations. Common-upper-level differences isolate lower-state intervals; a rotational fit determines B_v and D_v; vibrational data constrain an equilibrium coefficient; only then is a mass-scaled residual interpreted. Each step inherits the assignments and approximations of the previous one.

The report questions at the end return to the central derivations and their interpretation in molecular spectroscopy.

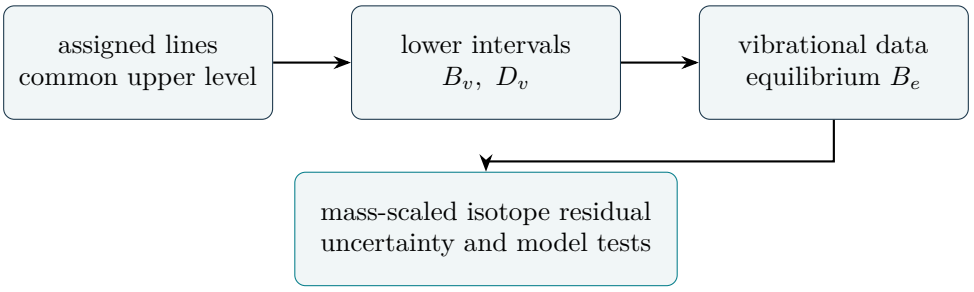


Figure. The endpoint is an observable comparison with an explicit baseline. Each arrow introduces assumptions that can be checked using additional lines or isotopologues.

5.2 What precision does the thesis target imply?

A relative line-position accuracy of 10^{-7} in the range 30,000–50,000 cm^{-1} corresponds to

$$\delta\widetilde{\nu} \sim 0.003\text{–}0.005\ \mathrm{cm}^{-1}. \tag{5.1}$$

This is a scale for individual line positions, not the fractional uncertainty of a small isotope correction. If two line errors have standard deviations σ_1, σ_2 and covariance C_{12}, their difference has

$$\mathrm{Var}(\widetilde{\nu}_1 - \widetilde{\nu}_2) = \sigma_1^2 + \sigma_2^2 - 2C_{12}. \tag{5.2}$$

Independent equal errors add in quadrature; a shared additive calibration error cancels, whereas a multiplicative scale error generally remains in proportion to the interval. Multiple lines can constrain constants more tightly, but systematic effects and model error do not disappear merely by increasing their number. The quoted accuracy need not itself be a one-standard-deviation uncertainty; that interpretation must come from the experimental analysis.

5.3 The three conclusions, with their limits

1. **A potential is the first term of an effective description.** The product rule shows exactly what is dropped in BO: derivatives of the fast eigenstates. Their diagonal part produces a positive scalar correction. Coupling through virtual fast excitations produces additional operators.

2. **We have calculated one missing operator and checked it.** The coupled model gives $\delta M = ma^2$ at leading low-frequency order. Normal modes, exact elimination, and second-order perturbation theory agree. A constant DBOC changes absolute energies but not slow line spacings in this model.

3. **Spectra test combinations of physical effects.** A common potential predicts leading isotope mass scaling. Anharmonicity, vibration–rotation coupling, spin structure and recoil

conventions must be distinguished from an inferred residual. Isotope comparisons constrain an effective Hamiltonian; a fitted residual alone does not identify a unique microscopic mechanism.

For an actual CO^+ analysis the remaining scientific task is quantitative: which correction functions and coupled states are required at the available precision, and which combinations can the chosen isotopologues determine? That requires molecular input and measured lines beyond the analytic models developed here. The model establishes a mechanism and an error-controlled method, not a numerical conclusion about BO validity for the particular $^{13}C^{17}O^+$ dataset.

5.4 Questions for the report

The following six questions define the report. Develop the arguments and re-derive the relevant results, making the assumptions and limits of each approximation explicit.

1. **What is discarded in the Born–Oppenheimer approximation?** Starting from the expansion in electronic channels, re-derive the coupled nuclear equations and the phase-independent diagonal correction. Explain why a large nuclear mass alone does not guarantee a reliable single-channel description.

2. **What does the exactly solvable model reveal?** Re-derive the coupled-oscillator spectrum and its slow-frequency expansion. Use the comparison with the BO spectrum to identify the controlling small parameters and explain why correcting the potential alone cannot recover the slow line spacings.

3. **How do virtual excitations change nuclear inertia?** Derive the leading kinetic correction by eliminating the fast channel, and compare it with the low-frequency expansion of the exact response. Explain in what sense this calculation is a matching calculation for an effective Hamiltonian.

4. **Which isotope effects occur within BO itself?** Re-derive the leading mass dependence of vibrational, anharmonic and rotational terms for a common electronic potential. Explain how these results constrain the interpretation of a departure from simple reduced-mass scaling.

5. **How can measured emission lines define an isotope test?** Reconstruct the chain from common-upper-level differences to rotational constants and a mass-scaled equilibrium residual. Derive its uncertainty propagation, including correlated line errors, and identify the spectroscopic assumptions needed to interpret a nonzero residual.

6. **What can isotope spectra determine about BO breakdown?** Analyze the identifiability of separate carbon and oxygen correction terms when one or both elements are isotopically substituted. Discuss what additional molecular information is needed to connect fitted coefficients to a microscopic mechanism in CO^+.

Checkpoint. A successful report connects a derived correction to an observable test while retaining its assumptions. The benchmark proves a mechanism; a molecular isotope residual constrains combinations of effects and requires a controlled spectroscopic baseline.

References

[1] M. Born and K. Huang, *Dynamical Theory of Crystal Lattices*, Oxford University Press (1954). Standard source for expansion in electronic channels and separation of nuclear and electronic motion. The needed one-coordinate derivation is given in Section 1.

[2] N. Brambilla, G. Krein, J. Tarrús Castellà and A. Vairo, *The Born–Oppenheimer approximation in an effective field theory language*, Physical Review D **97**, 016016 (2018). arXiv:1707.09647; doi:10.1103/PhysRevD.97.016016. The molecular-QED and heavy-quark connection; its specialist matching calculations are not prerequisites here.

[3] J. L. Dunham, *The Energy Levels of a Rotating Vibrator*, Physical Review **41**, 721–731 (1932). doi:10.1103/PhysRev.41.721. The rovibrational expansion and corrections to leading coefficient identifications.

[4] J. K. G. Watson, *The isotope dependence of diatomic Dunham coefficients*, Journal of Molecular Spectroscopy **80**, 411–421 (1980). doi:10.1016/0022-2852(80)90152-6. Isotope scaling and conventions for effective spectroscopic parameters.

[5] C. Puzzarini and J. Gauss, *Quantum-chemical determination of Born–Oppenheimer breakdown parameters for rotational constants: the open-shell species CN, CO^+ and BO*, Molecular Physics **111**, 2204–2210 (2013). doi:10.1080/00268976.2013.797614. Microscopic contributions and the connection between fitted constants and equilibrium quantities.

[6] S. Spezzano, S. Brünken, H. S. P. Müller, G. Klapper, F. Lewen, K. M. Menten and S. Schlemmer, *Accurate High-N Rest Frequencies for CO^+, an Ideal Tracer of Photon-Dominated Regions*, Journal of Physical Chemistry A **117**, 9814–9818 (2013). doi:10.1021/jp312616u. An isotopically invariant CO^+ analysis including BO-breakdown corrections.

[7] W. Szajna et al., *The $B^2\Sigma^+ - X^2\Sigma^+$, $0 - v''$ progression of $^{12}C^{17}O^+$ reinvestigated by FT-UV emission spectroscopy*, Journal of Quantitative Spectroscopy and Radiative Transfer **347**, 109633 (2025). doi:10.1016/j.jqsrt.2025.109633. Immediate experimental context; this isotopologue differs from the $^{13}C^{17}O^+$ target considered in these notes.

Why forbidden molecular transitions radiate

Why forbidden molecular transitions radiate

Spin–orbit mixing, photon emission, and the lifetime of CO

Study notes and report questions

Contents

The question and the calculation **2**

1 Why the transition is forbidden in the first approximation **4**

2 One mixing interaction, two observable consequences **9**

3 A mixed molecule emits a photon: deriving the radiative rate **13**

4 From spectral shifts to radiative lifetimes **19**

5 Conclusions and report questions **24**

References **26**

The question and the calculation

How can the same quantum mixing shift an energy level and enable a forbidden decay? What can line positions tell us about a radiative lifetime?

The molecular example is high-resolution spectroscopy of neutral $^{12}C^{16}O$, particularly the Cameron $a^3\Pi - X^1\Sigma^+$ system and deperturbation of $a^3\Pi(v = 0\text{–}6)$. A deperturbation model extracts interactions from displaced energy levels. Our goal is to derive how those interactions also enter a transition amplitude, and why energies alone need not determine it.

We solve a declared three-level model: a triplet level, a singlet level, and a common lower singlet level. Its excited-state mixing is exact; emission is treated to leading order in the coupling to the quantized radiation field. We then extend the amplitude to several intermediate states and compare the physical conclusions with CO lifetime studies [1, 2]. The molecular couplings are inputs, not an ab initio calculation of the fourteen-electron molecule.

Three results.

1. The electric dipole cannot connect pure singlet and triplet states. Spin–orbit coupling changes the states and generates a nonzero effective dipole.
2. The same mixing predicts an energy shift and a decay rate. The rate requires a photon calculation, which gives its absolute normalization and its ω^3 dependence.
3. Several virtual pathways interfere in a radiative amplitude. We construct two models with identical matter energies but different emission strengths, proving a limit on inference from line positions alone.

Prerequisites. Basic linear algebra and the Schrödinger equation. Kets, spin operators, oscillator quanta, perturbation amplitudes, and all rate conventions used below are defined as needed. No previous quantum-field-theory course is assumed. The calculation uses quantum mechanics of molecular states and harmonic-oscillator quantization of radiation.

Conventions. SI units, explicit $\hbar = h/(2\pi)$, and ε_0 the vacuum permittivity. Photon angular frequency is $\omega = (E_i - E_f)/\hbar$; ordinary frequency is $\nu = \omega/(2\pi)$; a spectroscopic wavenumber is $\widetilde{\nu} = (E_i - E_f)/(hc)$. A decay rate Γ has units s^{-1}; the corresponding energy width is $\hbar\Gamma$. We use free-space emission, electric-dipole order, and resolved molecular levels. Material environments, collisions, and external-field mixing are outside the model.

Notation

Symbol	Meaning
$H_0, H_{\mathrm{SO}}, \mathbf{d}$	Spin-independent molecular Hamiltonian; spin–orbit interaction; molecular electric-dipole operator.
S, Λ, Ω, J, M	Total electronic spin; electronic orbital projection magnitude; electronic total projection magnitude; molecular total angular momentum excluding nuclear spin, and its laboratory projection.
$\lvert t\rangle, \lvert s\rangle, \lvert g\rangle$	Unmixed triplet, bright singlet, and lower singlet levels. These are discrete rovibronic states in the model.
E_t, E_s, E_g, V	Their unperturbed energies and $V = \langle s\lvert H_{\mathrm{SO}}\rvert t\rangle$, an energy. In Section 2 choose its phase so $V \geq 0$; later it may be complex.
$\Delta, \mathcal{R}, \delta$	Positive bare detuning $E_s - E_t$; exact mixed splitting; positive downward shift $E_t - E_-$.
$\theta, w, E_\pm$	Mixing angle; singlet weight in the lower mixed state; its two exact eigenenergies.
$\mathcal{V}, \mathbf{k}, \lambda$	Photon normalization volume; wave vector; one of two transverse polarization labels. $\mathcal{V}$ is distinct from the coupling V.
$a, a^\dagger, \lvert 0_\gamma\rangle$	Photon annihilation and creation operators; photon vacuum. The electronic state label $a^3\Pi$ is unrelated to these operators.
$\mathbf{d}_{fi}, \Gamma_{i\to f}, \tau_i$	Dipole matrix element; partial radiative rate; lifetime. The total radiative rate sums over accessible final states.
R, v	Internuclear separation and vibrational quantum number when discussing actual CO wavefunctions.

1 Why the transition is forbidden in the first approximation

Aim. Identify the symmetry that suppresses the Cameron transition, then derive the interaction that relaxes it. The central distinction is between an operator selection rule and the composition of the states on which that operator acts.

1.1 States, amplitudes, and the nonrelativistic molecule

A ket $|\psi\rangle$ denotes a quantum state; its bra $\langle\psi|$ is its Hermitian conjugate. The inner product $\langle\phi|\psi\rangle$ is a complex overlap, and $\langle\psi|\psi\rangle = 1$ specifies normalization. A matrix element $\langle\phi|O|\psi\rangle$ is the amplitude associated with an operator O. For molecular wavefunctions the inner product integrates all spatial coordinates and sums all electron-spin components. Electronic states must also be antisymmetric under exchange of identical electrons.

Neutral CO has $Z_{\mathrm{C}} + Z_{\mathrm{O}} = 6 + 8 = 14$ electrons. Before relativistic and radiative corrections, its laboratory-coordinate Coulomb Hamiltonian is

$$H_0 = \sum_{A=\mathrm{C,O}} \frac{\mathbf{P}_A^2}{2M_A} + \sum_{i=1}^{14} \frac{\mathbf{p}_i^2}{2m_e} + V_{\mathrm{C}}, \tag{1.1}$$

$$V_{\mathrm{C}} = \frac{e^2}{4\pi\varepsilon_0} \left[\frac{Z_{\mathrm{C}} Z_{\mathrm{O}}}{|\mathbf{R}_{\mathrm{C}} - \mathbf{R}_{\mathrm{O}}|} - \sum_{A,i} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right]. \tag{1.2}$$

Here $e > 0$ is the elementary charge, m_e the electron mass, M_A a nuclear mass, and ε_0 the vacuum permittivity. Nuclear positions and momenta are $\mathbf{R}_A, \mathbf{P}_A = -\mathrm{i}\hbar\nabla_{R_A}$; electronic ones are $\mathbf{r}_i, \mathbf{p}_i = -\mathrm{i}\hbar\nabla_{r_i}$. The free center-of-mass motion can be separated; the labels below describe internal molecular states. Nuclear spins are zero for ^{12}C and ^{16}O.

An electron has spin $\mathbf{s}_i = (\hbar/2)\boldsymbol{\sigma}_i$, with $\boldsymbol{\sigma}_i$ its three Pauli matrices acting on that electron's two spin components. In the basis of spin up and spin down along z, the matrices for one electron are

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.3}$$

The subscript i on $\boldsymbol{\sigma}_i$ labels an electron, whereas x, y, z label matrix components. The total electronic spin and its eigenvalues are

$$\mathbf{S} = \sum_i \mathbf{s}_i, \qquad S^2|\alpha, S, M_S\rangle = \hbar^2 S(S+1)|\alpha, S, M_S\rangle, \qquad S_z|\alpha, S, M_S\rangle = \hbar M_S|\alpha, S, M_S\rangle. \tag{1.4}$$

The label α collects all other quantum numbers; z here is a laboratory axis. Since every term of H_0 acts on coordinates and is the identity in spin space,

$$[H_0, S^2] = [H_0, S_z] = 0. \tag{1.5}$$

Commuting Hermitian operators admit common eigenstates. We can therefore classify these nonrelativistic states by S, even when antisymmetry prevents writing their full wavefunction as one spatial factor times one spin factor. A singlet has $S = 0$ and multiplicity $2S + 1 = 1$; a triplet has $S = 1$ and multiplicity three.

Physical meaning. The Coulomb interaction can produce different singlet and triplet energies through the allowed antisymmetric spatial states. "Spin independent" does not mean that their energies must coincide. It means that H_0 cannot mix states with different total S.

1.2 Reading the molecular labels

For a fixed internuclear axis, let $\hbar\lambda$ and $\hbar\sigma$ be the projections of total electronic orbital angular momentum $\mathbf{L}$ and electronic spin $\mathbf{S}$ on that axis. In the useful approximate case in which both projections are well defined,

$$\Lambda = |\lambda|, \qquad \Omega = |\lambda + \sigma|. \tag{1.6}$$

The traditional orbital symbols Σ, Π, Δ stand for $\Lambda = 0, 1, 2$. The capital Σ in a term symbol is therefore an orbital label, not the spin projection σ used here. A term is written ${}^{2S+1}\Lambda_\Omega$ when Ω is specified. For Σ electronic states a superscript $+$ or $-$ gives the reflection symmetry in a plane containing the axis.

Thus $X\,{}^1\Sigma^+$ is a singlet with $\Lambda = 0$, whereas $a\,{}^3\Pi$ is a triplet with $\Lambda = 1$ and approximate components $\Omega = 0, 1, 2$. $A\,{}^1\Pi$ is a singlet with $\Omega = 1$. Letters X, a, A name electronic states; they are not operators. The Cameron system connects $a\,{}^3\Pi$ to $X\,{}^1\Sigma^+$. We focus on its $a\,{}^3\Pi_1$ component, for which mixing with ${}^1\Pi_1$ states supplies an instructive radiative pathway [1, 2].

A rovibronic level also carries a vibrational label v, total internal angular momentum J, laboratory projection M_J, and total parity $\pi = \pm 1$. Here J includes molecular rotation and electronic spin. The orbital projections in Eq. (1.6) are approximate molecular labels; J, M_J and parity are exact symmetry labels for the isolated field-free Hamiltonian considered here. Heteronuclear CO has no electronic gerade/ungerade label.

1.3 The electric dipole cannot change a pure total-spin sector

At wavelengths much longer than the molecular size, the leading electric interaction with light is the electric-dipole, or E1, interaction,

$$H_{\mathrm{E1}} = -\mathbf{d}\cdot\mathbf{E}, \qquad \mathbf{d} = -e\sum_i \mathbf{r}_i + e\sum_A Z_A \mathbf{R}_A. \tag{1.7}$$

The quantized electric field $\mathbf{E}$ will be constructed later. Since CO is neutral, translating all coordinates by the same constant leaves $\mathbf{d}$ unchanged. Each Cartesian component d_b acts only on positions, so

$$[S^2, d_b] = 0, \qquad b = x, y, z. \tag{1.8}$$

Take initial and final states with definite spins S_i, S_f. Insert them around the commutator, applying S^2 once to the bra and once to the ket:

$$\begin{aligned} 0 &= \langle f|[S^2, d_b]|i\rangle && (1.9)\\ &= \langle f|S^2 d_b|i\rangle - \langle f|d_b S^2|i\rangle && (1.10)\\ &= \hbar^2\big[S_f(S_f+1) - S_i(S_i+1)\big]\langle f|d_b|i\rangle. && (1.11) \end{aligned}$$

If $S_f \neq S_i$, the bracket does not vanish, hence

$$\boxed{\langle g, S=0|\mathbf{d}|t, S=1\rangle = \mathbf{0}.} \tag{1.12}$$

The result holds for fully antisymmetric many-electron states. It does not require a product ansatz for their spin and spatial parts.

Checkpoint. "Spin forbidden" means zero E1 amplitude between states of definite, different total electronic spin in this approximation. It does not mean that the real molecule has an exactly stable excited state. We must inspect the interactions omitted from H_0.

1.4 Where spin–orbit coupling comes from

The needed interaction is a relativistic correction to the electronic Hamiltonian. Its one-electron nuclear-field contribution is

$$H_{\mathrm{SO}}^{(1)} = \sum_i \frac{\hbar}{4m_e^2c^2}\,\boldsymbol{\sigma}_i \cdot \left[\nabla_i U_{\mathrm{nuc}}(\mathbf{r}_i) \times \mathbf{p}_i\right], \quad U_{\mathrm{nuc}}(\mathbf{r}) = -\frac{e^2}{4\pi\varepsilon_0}\sum_A \frac{Z_A}{|\mathbf{r}-\mathbf{R}_A|}. \tag{1.13}$$

Here c is the speed of light and U_{nuc} is potential *energy*, including the electron's negative charge. Full molecular spin–orbit calculations also require electron–electron relativistic terms and electronic correlation; Eq. (1.13) identifies the mechanism, not a quantitatively complete CO operator.

Derivation of the spin-dependent term from the Dirac equation. The following algebra explains both the coefficient and its sign. For one electron in a static electrostatic potential energy $U(\mathbf{r})$, the stationary Dirac equation in two-component blocks is

$$\begin{pmatrix} m_ec^2 + U & c\boldsymbol{\sigma}\cdot\mathbf{p} \\ c\boldsymbol{\sigma}\cdot\mathbf{p} & -m_ec^2 + U \end{pmatrix}\begin{pmatrix}\phi\\ \chi\end{pmatrix} = (m_ec^2+\epsilon)\begin{pmatrix}\phi\\ \chi\end{pmatrix}. \tag{1.14}$$

The two-spinors ϕ, χ are the large and small components in the nonrelativistic limit; ϵ is energy with the rest energy subtracted. The two row equations are

$$\epsilon\phi = U\phi + c\boldsymbol{\sigma}\cdot\mathbf{p}\,\chi, \qquad (2m_ec^2 + \epsilon - U)\chi = c\boldsymbol{\sigma}\cdot\mathbf{p}\,\phi. \tag{1.15}$$

Eliminate χ. The inverse factor is a position-dependent multiplication operator and must stay to the left of the momentum acting on ϕ:

$$\chi = (2m_ec^2 + \epsilon - U)^{-1} c\boldsymbol{\sigma}\cdot\mathbf{p}\,\phi. \tag{1.16}$$

In the nonrelativistic expansion,

$$(2m_ec^2 + \epsilon - U)^{-1} = \frac{1}{2m_ec^2}\left[1 - \frac{\epsilon - U}{2m_ec^2} + \cdots\right], \tag{1.17}$$

$$\epsilon\phi = \left[U + \frac{\mathbf{p}^2}{2m_e} - \frac{1}{4m_e^2c^2}(\boldsymbol{\sigma}\cdot\mathbf{p})(\epsilon - U)(\boldsymbol{\sigma}\cdot\mathbf{p}) + \cdots\right]\phi. \tag{1.18}$$

To isolate its spin term, use

$$\sigma_i\sigma_j = \delta_{ij} + \mathrm{i}\epsilon_{ijk}\sigma_k, \qquad p_iF = Fp_i - \mathrm{i}\hbar\partial_iF. \tag{1.19}$$

Here $F(\mathbf{r})$ is any scalar multiplication operator; ϵ_{ijk} is the antisymmetric Levi–Civita symbol with $\epsilon_{123} = 1$, and repeated Cartesian indices are summed. Since p_ip_j is symmetric in i, j,

$$(\boldsymbol{\sigma}\cdot\mathbf{p})F(\boldsymbol{\sigma}\cdot\mathbf{p}) = \sigma_i\sigma_j(Fp_ip_j - \mathrm{i}\hbar(\partial_iF)p_j) \tag{1.20}$$

$$= F\mathbf{p}^2 - \mathrm{i}\hbar\nabla F\cdot\mathbf{p} + \hbar\boldsymbol{\sigma}\cdot(\nabla F\times\mathbf{p}). \tag{1.21}$$

Set $F = \epsilon - U$, so $\nabla F = -\nabla U$. The last term in Eq. (1.21), multiplied by the negative prefactor in Eq. (1.18), gives

$$H_{\mathrm{SO}} = +\frac{\hbar}{4m_e^2c^2}\boldsymbol{\sigma}\cdot(\nabla U\times\mathbf{p}). \tag{1.22}$$

Only the spin-dependent term has been extracted here. Obtaining the complete normalized, energy-independent Pauli Hamiltonian additionally requires eliminating the residual energy dependence and normalizing the large component; it also produces spin-independent kinetic and Darwin corrections. We have not claimed to derive those terms.

For a central potential $U(r)$, $\nabla U = U'(r)\mathbf{r}/r$. With $\boldsymbol{\ell} = \mathbf{r}\times\mathbf{p}$ and $\mathbf{s} = \hbar\boldsymbol{\sigma}/2$,

$$\nabla U\times\mathbf{p} = \frac{U'(r)}{r}\boldsymbol{\ell}, \qquad H_{\mathrm{SO}} = \frac{1}{2m_e^2c^2r}\frac{\mathrm{d}U}{\mathrm{d}r}\boldsymbol{\ell}\cdot\mathbf{s}. \tag{1.23}$$

This familiar one-electron result is a useful check, including its positive sign for an attractive Coulomb potential $U = -C/r$ with $C > 0$.

1.5 Why individual spin–orbit terms can mix singlet and triplet

One must not replace the many-electron operator by a single constant times $\mathbf{L}\cdot\mathbf{S}$ when asking about different spin multiplicities. A global $\mathbf{L}\cdot\mathbf{S}$ commutes with S^2 and therefore cannot produce singlet–triplet mixing. Individual electronic spins with different orbital coefficients need not conserve total S.

Two spins already demonstrate the distinction. Define $|\uparrow\rangle, |\downarrow\rangle$ by $s_z|\uparrow\rangle = (\hbar/2)|\uparrow\rangle$ and $s_z|\downarrow\rangle = -(\hbar/2)|\downarrow\rangle$. Their normalized states with zero total spin projection are

$$|\mathcal{S}\rangle = \frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}}, \qquad |\mathcal{T}_0\rangle = \frac{|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle}{\sqrt{2}}. \tag{1.24}$$

Acting separately on the two spins gives

$$s_{1z}|\mathcal{T}_0\rangle = \frac{\hbar}{2\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle) = \frac{\hbar}{2}|\mathcal{S}\rangle, \tag{1.25}$$

$$s_{2z}|\mathcal{T}_0\rangle = -\frac{\hbar}{2}|\mathcal{S}\rangle, \tag{1.26}$$

$$\langle\mathcal{S}|(u_1 s_{1z} + u_2 s_{2z})|\mathcal{T}_0\rangle = \frac{\hbar}{2}(u_1 - u_2). \tag{1.27}$$

Here u_1, u_2 are scalar coefficients with dimensions of angular frequency if the weighted sum is an energy. The contributions cancel for equal coefficients, as they must for a multiple of total S_z. Distinct orbital matrix elements in spin–orbit interactions can instead supply unequal weights. This spin-space example illustrates an allowed mechanism; molecular matrix elements must still use antisymmetric electronic wavefunctions and the correct orbital symmetries.

Physical meaning. A singlet component in an otherwise triplet state can radiate through the spin-conserving dipole operator. The leading explanation changes the state, not the spin selection rule of $\mathbf{d}$.

1.6 Which states may be mixed, and which may radiate?

Spin–orbit coupling is a scalar under simultaneous spatial and spin rotations, and is even under parity. It can change S while preserving J, M_J and total parity. Accordingly our reduced Hamiltonian will mix levels *within one fixed* J, M_J, π *block*. Matching approximate $\Omega = 1$ labels is useful for the $a^3\Pi_1$–$^1\Pi_1$ example; it does not replace the exact symmetry conditions.

The E1 operator is instead a parity-odd angular-momentum vector. Let $\mathcal{P}$ invert all spatial coordinates, so $\mathcal{P}^\dagger \mathbf{d} \mathcal{P} = -\mathbf{d}$. A state of total parity π_i obeys $\mathcal{P}|i\rangle = \pi_i |i\rangle$, with $\pi_i = \pm 1$. Inserting parity operators into a matrix element gives

$$\pi_f \pi_i \langle f|\mathbf{d}|i\rangle = \langle f|\mathcal{P}^\dagger \mathbf{d} \mathcal{P}|i\rangle = -\langle f|\mathbf{d}|i\rangle, \qquad (1 + \pi_f \pi_i)\langle f|\mathbf{d}|i\rangle = \mathbf{0}. \tag{1.28}$$

Thus a nonzero E1 amplitude requires opposite total parities.

Under rotations the three dipole components transform like angular momentum one; this is what a *rank-one vector operator* means. With $\mathbf{J}$ the total internal angular-momentum operator, its defining Cartesian commutator is $[J_a, d_b] = \mathrm{i}\hbar \epsilon_{abc} d_c$. Applying the dipole operator to a state with angular momentum J_i therefore produces states in the angular-momentum product $1 \otimes J_i$. The angular-momentum addition rule yields

$$|J_i - 1| \le J_f \le J_i + 1, \tag{1.29}$$

in integer steps. For $J_i \ge 1$, this allows $\Delta J = 0, \pm 1$. For $J_i = 0$ the product is $1 \otimes 0 = 1$, so $J_f = 0$ is absent: a vector cannot connect two angular-momentum-zero states. These are necessary symmetry conditions, not guarantees of a nonzero radial or electronic matrix element.

The vector matrix elements used below retain the specified angular states; rotational factors are not silently set to one. Electronic Σ^+ reflection symmetry and total rovibronic parity must not be conflated.

Checkpoint. We have located the small effect that opens the transition: a relativistic interaction can mix compatible singlet and triplet levels. Next we calculate the resulting energy shift and dipole amplitude in the same two-state Hamiltonian. Their common origin does not make them the same observable.

2 One mixing interaction, two observable consequences

Aim. Solve an explicitly specified model exactly, then identify the approximation used in ordinary deperturbation theory.

2.1 Define the model before doing the algebra

Take three orthonormal *rovibronic* states of the spin-independent molecule:

$$|t\rangle:\ S=1, \qquad |s\rangle:\ S=0, \qquad |g\rangle:\ S=0, \qquad H_0|a\rangle = E_a|a\rangle \quad (a=t,s,g). \tag{2.1}$$

The excited states t, s have the same J, M_J, π and can mix. The lower level g has angular quantum numbers and opposite parity allowing $s \to g$ E1 emission. For now choose one specified magnetic sublevel for each ket. Set

$$E_s > E_t > E_g, \qquad \Delta = E_s - E_t > 0, \qquad V = \langle s|H_{\mathrm{SO}}|t\rangle, \qquad \mathbf{d}_{gs} = \langle g|\mathbf{d}|s\rangle \neq 0, \quad \mathbf{d}_{gt} = \mathbf{0}. \tag{2.2}$$

We retain only this off-diagonal spin–orbit coupling, absorb any diagonal corrections into E_t, E_s, and leave g unmixed. For emission we also require the physical lower mixed energy $E_- > E_g$; this must be checked if V becomes large. These are model assumptions. The general case with a dressed lower state will be derived in Section 4.3.

Hermiticity gives $\langle t|H_{\mathrm{SO}}|s\rangle = V^*$. The excited-state matrix in the ordered basis $(|t\rangle, |s\rangle)$ is

$$H = \begin{pmatrix} E_t & V^* \\ V & E_s \end{pmatrix}. \tag{2.3}$$

If $V = |V|\mathrm{e}^{\mathrm{i}\varphi}$, replace $|s\rangle$ by $\mathrm{e}^{\mathrm{i}\varphi}|s\rangle$. In that basis V becomes $|V|$ and $\mathbf{d}_{gs}$ becomes $\mathrm{e}^{\mathrm{i}\varphi}\mathbf{d}_{gs}$. No observable changes. Throughout this section use the real convention $V \geq 0$.

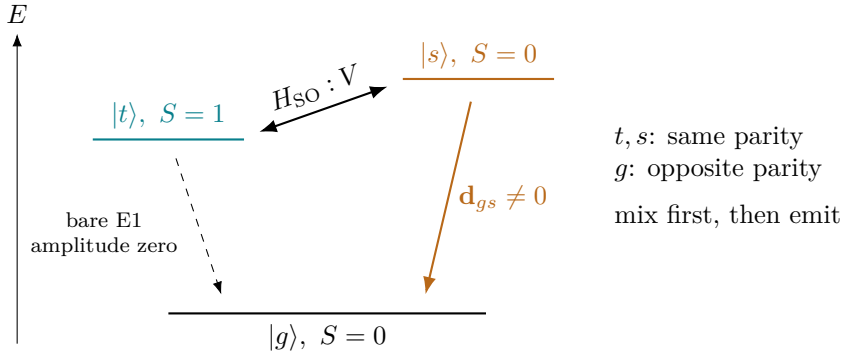


Figure. The declared three-level model, not a scale drawing of CO energies. Mixing preserves total parity; electric-dipole emission changes it.

2.2 Exact eigenvalues: level repulsion

An eigenvector with components c_t, c_s satisfies $(H - EI)c = 0$, where I is the identity matrix. A nonzero solution requires

$$0 = \det(H - EI) = (E_t - E)(E_s - E) - V^2, \tag{2.4}$$

$$E^2 - (E_t + E_s)E + E_tE_s - V^2 = 0. \tag{2.5}$$

Complete the square:

$$\left(E - \frac{E_t + E_s}{2}\right)^2 = \frac{(E_s - E_t)^2}{4} + V^2. \tag{2.6}$$

Define the positive exact splitting $\mathcal{R} = \sqrt{\Delta^2 + 4V^2}$. The two eigenvalues are

$$\boxed{E_\pm = \frac{E_t + E_s \pm \mathcal{R}}{2}.} \tag{2.7}$$

For positive Δ, E_- tends to E_t as $V \to 0$. Its downward shift is

$$\delta \equiv E_t - E_- = \frac{\mathcal{R} - \Delta}{2} = \frac{2V^2}{\mathcal{R} + \Delta} \geq 0, \qquad E_+ = E_s + \delta. \tag{2.8}$$

The trace $E_- + E_+ = E_t + E_s$ is conserved. This gives an immediate check on both signs.

2.3 Exact eigenvectors: the borrowed singlet weight

Use the second row of the eigenvalue equation:

$$V c_t + (E_s - E_-)c_s = 0 \quad \Longrightarrow \quad \frac{c_s}{c_t} = -\frac{V}{E_s - E_-} = -\frac{2V}{\Delta + \mathcal{R}}. \tag{2.9}$$

Choose a normalized real eigenvector and define its mixing angle by

$$|-\rangle = \cos\theta\,|t\rangle - \sin\theta\,|s\rangle, \qquad \tan\theta = \frac{2V}{\Delta + \mathcal{R}}, \quad 0 \leq \theta < \frac{\pi}{4}. \tag{2.10}$$

The orthogonal upper eigenvector is

$$|+\rangle = \sin\theta\,|t\rangle + \cos\theta\,|s\rangle. \tag{2.11}$$

Indeed their overlap is $\cos\theta\sin\theta - \sin\theta\cos\theta = 0$ and each norm is one. The same ratio gives

$$\cos 2\theta = \frac{\Delta}{\mathcal{R}}, \qquad \sin 2\theta = \frac{2V}{\mathcal{R}}, \qquad \boxed{w \equiv |\langle s|-\rangle|^2 = \sin^2\theta = \frac{1}{2}\left(1 - \frac{\Delta}{\mathcal{R}}\right).} \tag{2.12}$$

Here w is a probability, not an amplitude. Since $\mathcal{R} = \Delta + 2\delta$,

$$\boxed{w = \frac{\delta}{\Delta + 2\delta}.} \tag{2.13}$$

This exact identity is the first connection between a spectral shift and the state composition. It is conditional on the two-level model and its bare detuning.

Physical meaning. The interaction both moves the level and changes what the level is. Transition amplitudes must use the changed eigenvector, not just the changed eigenenergy.

2.4 The weak-coupling expansion and its boundary

Let $\eta = V/\Delta$ be the dimensionless expansion parameter. From $\sqrt{1+x} = 1 + x/2 - x^2/8 + O(x^3)$,

$$\mathcal{R} = \Delta\sqrt{1 + 4\eta^2} = \Delta[1 + 2\eta^2 - 2\eta^4 + O(\eta^6)], \tag{2.14}$$

$$\delta = \frac{V^2}{\Delta} - \frac{V^4}{\Delta^3} + O(V^6/\Delta^5), \tag{2.15}$$

$$w = \eta^2 - 3\eta^4 + O(\eta^6), \tag{2.16}$$

$$|-\rangle = \left(1 - \frac{\eta^2}{2} + O(\eta^4)\right)|t\rangle - \left(\eta + O(\eta^3)\right)|s\rangle. \tag{2.17}$$

The leading energy correction is second order, while the induced singlet *amplitude* is first order. The leading emission probability will therefore also be second order in V.

For $\Delta \to 0$ at fixed $V > 0$, perturbation theory fails but the exact answer is finite:

$$E_\pm \to E_t \pm V, \qquad |\mp\rangle \to \frac{|t\rangle \mp |s\rangle}{\sqrt{2}}, \qquad w \to \frac{1}{2}. \tag{2.18}$$

There is no divergent physical probability. The failure belongs to the expansion in V/Δ, which must be replaced by diagonalization. If both Δ and V vanish, the degenerate basis is not uniquely selected by this Hamiltonian.

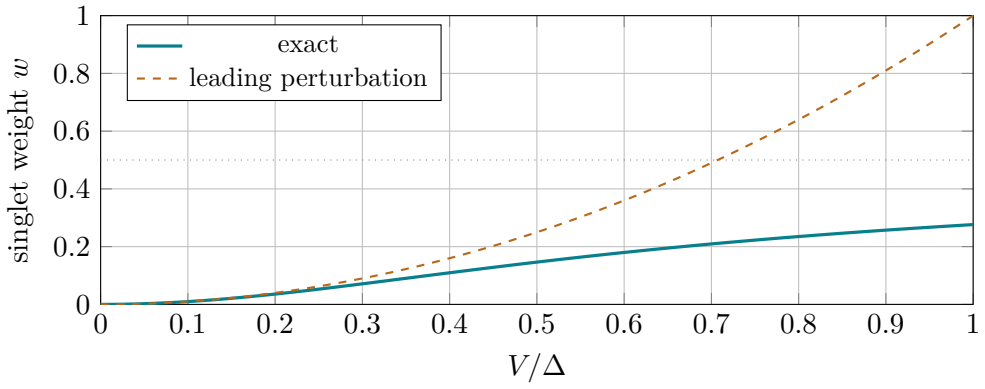


Figure. An analytic plot of Eq. (2.12). The perturbative weight $(V/\Delta)^2$ agrees near the origin but is not a probability formula for strong mixing. The exact lower-state weight remains below 1/2 for $\Delta > 0$.

2.5 Transform the observable as well as the Hamiltonian

Because $\mathbf{d}_{gt} = 0$, linearity gives

$$\mathbf{d}_{g-} = \langle g|\mathbf{d}|-\rangle = \cos\theta\, \mathbf{d}_{gt} - \sin\theta\, \mathbf{d}_{gs} = -\sin\theta\, \mathbf{d}_{gs}, \tag{2.19}$$

$$\mathbf{d}_{g+} = \sin\theta\, \mathbf{d}_{gt} + \cos\theta\, \mathbf{d}_{gs} = \cos\theta\, \mathbf{d}_{gs}. \tag{2.20}$$

Define the line strength for these specified states as $\mathcal{S}_{fi} = \mathbf{d}^*_{fi} \cdot \mathbf{d}_{fi} = \sum_{b=x,y,z} |(d_{fi})_b|^2$. Then

$$\mathcal{S}_{g-} = w\mathcal{S}_{gs}, \qquad \mathcal{S}_{g+} = (1-w)\mathcal{S}_{gs}, \qquad \boxed{\mathcal{S}_{g-} + \mathcal{S}_{g+} = \mathcal{S}_{gs}.} \tag{2.21}$$

This sum rule is a consequence of an orthonormal basis rotation. It conserves the unweighted dipole strength in this model. It does *not* conserve the sum of the two radiative rates, because their photon frequencies differ. Nor does it predict measured brightness without the populations of the two upper levels.

Checkpoint. Exact mixing has now supplied both δ and $\mathbf{d}_{g-}$. The missing ingredient is the conversion of a dipole amplitude into an absolute probability per unit time. That requires the available photon modes.

2.6 Eliminating the virtual level: an effective description

This calculation connects the matrix model with the energy-dependent effective Hamiltonians used in quantum theory. Restore complex V in Eq. (2.3). The coupled equations are

$$(E - E_t)c_t = V^* c_s, \qquad (E - E_s)c_s = V c_t. \tag{2.22}$$

For $E \neq E_s$, solve the second equation and substitute in the first:

$$c_s = \frac{V}{E - E_s} c_t, \qquad \left[E_t + \underbrace{\frac{|V|^2}{E - E_s}}_{\Sigma_t(E)} \right] c_t = E c_t. \tag{2.23}$$

$\Sigma_t(E)$ is the energy-dependent correction, often called a self-energy. It encodes an excursion from t into s and back into t. It has units of energy. Evaluating it at E_t gives the familiar second-order shift $-|V|^2/\Delta$; for $V \neq 0$, solving $E = E_t + \Sigma_t(E)$ exactly gives both roots in Eq. (2.7). When $V = 0$, this retained-state equation gives only E_t; the decoupled level E_s remains a separate state.

The eliminated component must also be retained in the observable:

$$\mathbf{d}_{gE} = \mathbf{d}_{gs} \frac{V}{E - E_s} c_t, \qquad |c_t|^2 = \frac{1}{1 + |V|^2/(E - E_s)^2}. \tag{2.24}$$

The normalization follows directly from $1 = |c_t|^2 + |c_s|^2$. At weak coupling $E \simeq E_t, c_t \simeq 1$, so

$$\mathbf{d}_{gt}^{\text{eff}} \simeq -\frac{V}{\Delta} \mathbf{d}_{gs}. \tag{2.25}$$

Thus eliminating an unobserved level changes both the effective Hamiltonian and the effective transition operator. Keeping one correction while discarding the other is inconsistent.

The pole and its weight. The tt entry of the inverse matrix $(EI - H)^{-1}$ is

$$G_{tt}(E) = \frac{1}{E - E_t - \Sigma_t(E)}. \tag{2.26}$$

Near a nondegenerate root E_* of its denominator,

$$G_{tt}(E) \simeq \frac{Z_t}{E - E_*}, \quad Z_t = \frac{1}{1 - \Sigma_t'(E_*)} = \frac{1}{1 + |V|^2/(E_* - E_s)^2} = |c_t|^2. \tag{2.27}$$

A pole is simply an eigenenergy; its residue Z_t is the probability that this eigenstate occupies the retained level t. This is a finite-dimensional example of the same resolvent language used for propagators and self-energies in field theory, without requiring a field-theory course.

3 A mixed molecule emits a photon: deriving the radiative rate

Physical meaning. Mixing has supplied a nonzero molecular dipole matrix element. We now calculate how that matrix element becomes a measurable decay rate. The only additional quantum system required is the electromagnetic field: each of its normal modes is a harmonic oscillator.

3.1 A radiation mode is a harmonic oscillator

Let ε_0 and μ_0 denote the vacuum permittivity and permeability, with $c^{-2} = \varepsilon_0\mu_0$. The energy of the free radiation field is

$$H_\gamma = \frac{1}{2}\int \mathrm{d}^3r \left[\varepsilon_0 \mathbf{E}^2 + \frac{\mathbf{B}^2}{\mu_0}\right]. \tag{3.1}$$

The molecular Hamiltonian already contains the electrostatic interactions among the charged particles. Here we quantize the additional, transverse radiation field. In Coulomb gauge its vector potential satisfies $\boldsymbol{\nabla}\cdot\mathbf{A} = 0$, and its free fields are $\mathbf{E} = -\partial_t\mathbf{A}$ and $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}$.

Derivation of the mode-coordinate normalization. The oscillator Hamiltonian and the normalized electric field follow from Eq. (3.1). Temporarily place the field in a cubic box of volume $\mathcal{V} = L^3$, with periodic boundaries. The box is a normalization device; the final result will not depend on $\mathcal{V}$. A convenient derivation starts with real transverse mode functions $\mathbf{f}_b(\mathbf{r})$, where b labels both the spatial mode and its polarization:

$$\int_{\mathcal{V}} \mathrm{d}^3r\, \mathbf{f}_b\cdot\mathbf{f}_{b'} = \delta_{bb'}, \qquad \boldsymbol{\nabla}\cdot\mathbf{f}_b = 0, \tag{3.2}$$

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\mathbf{f}_b = \frac{\omega_b^2}{c^2}\mathbf{f}_b. \tag{3.3}$$

The Kronecker symbol $\delta_{bb'}$ equals one for identical indices and zero otherwise. Expand

$$\mathbf{A}(\mathbf{r},t) = \frac{1}{\sqrt{\varepsilon_0}}\sum_b Q_b(t)\mathbf{f}_b(\mathbf{r}). \tag{3.4}$$

Orthogonality makes the electric-field energy $\frac{1}{2}\sum_b \dot{Q}_b^2$. Integration by parts, with the periodic boundary term vanishing, gives

$$\int \mathrm{d}^3r\, (\boldsymbol{\nabla}\times\mathbf{f}_b)\cdot(\boldsymbol{\nabla}\times\mathbf{f}_{b'}) = \int \mathrm{d}^3r\, \mathbf{f}_b\cdot(\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\mathbf{f}_{b'}) \tag{3.5}$$

$$= \frac{\omega_{b'}^2}{c^2}\delta_{bb'}. \tag{3.6}$$

Thus the magnetic energy is $\frac{1}{2}\sum_b \omega_b^2 Q_b^2$, and

$$H_\gamma = \sum_b \frac{1}{2}\left(P_b^2 + \omega_b^2 Q_b^2\right), \qquad P_b = \dot{Q}_b. \tag{3.7}$$

Quantizing each mode. The chosen normalization has absorbed the oscillator's mass into its coordinate Q_b. Quantization means promoting Q_b, P_b to operators obeying

$$[Q_b, P_{b'}] = \mathrm{i}\hbar\delta_{bb'}, \qquad [Q_b, Q_{b'}] = [P_b, P_{b'}] = 0. \tag{3.8}$$

Define the dimensionless annihilation and creation operators

$$a_b = \sqrt{\frac{\omega_b}{2\hbar}}Q_b + \frac{\mathrm{i}}{\sqrt{2\hbar\omega_b}}P_b, \qquad a_b^\dagger = \sqrt{\frac{\omega_b}{2\hbar}}Q_b - \frac{\mathrm{i}}{\sqrt{2\hbar\omega_b}}P_b, \tag{3.9}$$

$$[a_b, a_{b'}^\dagger] = \delta_{bb'}, \qquad H_\gamma = \sum_b \hbar\omega_b\left(a_b^\dagger a_b + \frac{1}{2}\right). \tag{3.10}$$

These identities follow by substitution into Eq. (3.7). An eigenstate of $a_b^\dagger a_b$ with integer eigenvalue $n_b \geq 0$ contains n_b photons in mode b; its excitation energy above that mode's ground state is $n_b\hbar\omega_b$. In particular,

$$a_b|n_b\rangle = \sqrt{n_b}|n_b-1\rangle, \qquad a_b^\dagger|n_b\rangle = \sqrt{n_b+1}|n_b+1\rangle. \tag{3.11}$$

The vacuum $|0\rangle$ has no photons in any mode. It obeys $a_b|0\rangle = 0$, whereas $a_b^\dagger|0\rangle = |1_b\rangle$ is nonzero. The additive zero-point energy does not enter the energy difference in the calculation below.

3.2 Plane waves and their normalization

The allowed wavevectors are $\mathbf{k} = (2\pi/L)(n_x, n_y, n_z)$ with integer n_x, n_y, n_z, and the nonzero modes have $\omega_\mathbf{k} = c|\mathbf{k}|$. For each $\mathbf{k}$, the two physical polarizations are labeled $\lambda = 1, 2$. Their unit vectors obey

$$\mathbf{k}\cdot\mathbf{e}_{\mathbf{k}\lambda} = 0, \qquad \mathbf{e}_{\mathbf{k}\lambda}^*\cdot\mathbf{e}_{\mathbf{k}\lambda'} = \delta_{\lambda\lambda'}. \tag{3.12}$$

Plane waves normalized in the box are $\mathbf{e}_{\mathbf{k}\lambda}e^{i\mathbf{k}\cdot\mathbf{r}}/\sqrt{\mathcal{V}}$. Passing from real standing waves to these complex traveling waves is a unitary change of oscillator basis, not a change of field normalization. For example, the normalized real pair $\sqrt{2/\mathcal{V}}\cos(\mathbf{k}\cdot\mathbf{r})$ and $\sqrt{2/\mathcal{V}}\sin(\mathbf{k}\cdot\mathbf{r})$ corresponds to

$$a_\mathbf{k} = \frac{a_c - i a_s}{\sqrt{2}}, \qquad a_{-\mathbf{k}} = \frac{a_c + i a_s}{\sqrt{2}}. \tag{3.13}$$

The subscripts c, s in this equation label the cosine and sine modes only. Substituting $P_b = -i\sqrt{\hbar\omega_b/2}(a_b - a_b^\dagger)$ into $\mathbf{E} = -\sum_b P_b\mathbf{f}_b/\sqrt{\varepsilon_0}$ then gives, at time zero,

$$\boxed{\mathbf{E}(\mathbf{r}) = i\sum_{\mathbf{k}\lambda}\sqrt{\frac{\hbar\omega_\mathbf{k}}{2\varepsilon_0\mathcal{V}}}\left[\mathbf{e}_{\mathbf{k}\lambda}a_{\mathbf{k}\lambda}e^{i\mathbf{k}\cdot\mathbf{r}} - \mathbf{e}_{\mathbf{k}\lambda}^* a_{\mathbf{k}\lambda}^\dagger e^{-i\mathbf{k}\cdot\mathbf{r}}\right].} \tag{3.14}$$

The two terms are Hermitian conjugates after including the factor i, so $\mathbf{E}$ is Hermitian. The commutator is $[a_{\mathbf{k}\lambda}, a_{\mathbf{k}'\lambda'}^\dagger] = \delta_{\mathbf{k}\mathbf{k}'}\delta_{\lambda\lambda'}$. The sum includes all wavevector directions and the two transverse polarizations once each.

Checkpoint. The coefficient in Eq. (3.14) is fixed by the electromagnetic energy and oscillator quantization. It is not an adjustable radiation strength. Its square is proportional to $\hbar\omega/\mathcal{V}$.

3.3 The one-photon transition matrix element

Let $|i\rangle$ and $|f\rangle$ be normalized molecular eigenstates of energies $E_i > E_f$, and define their positive transition angular frequency and vector dipole matrix element by

$$\omega_{if} = \frac{E_i - E_f}{\hbar}, \qquad \mathbf{d}_{fi} = \langle f|\mathbf{d}|i\rangle. \tag{3.15}$$

The electric-dipole operator is $\mathbf{d} = \sum_a q_a\boldsymbol{\rho}_a$, where q_a and $\boldsymbol{\rho}_a$ are the charge and internal position of constituent a. For neutral CO the total charge vanishes, making this operator independent of the internal origin. In SI units $\mathbf{d}_{fi}$ has units C m.

When $|\mathbf{k}|r_{\mathrm{mol}} \ll 1$, where r_{mol} is a characteristic molecular size, the field varies little across the molecule. We can derive the leading electric interaction by first treating the external radiation as classical. Set $\mathbf{A}(\mathbf{R}_0 + \boldsymbol{\rho}, t) \simeq \mathbf{A}(\mathbf{R}_0, t) \equiv \mathbf{A}(t)$ and choose its scalar potential to vanish. The nonrelativistic minimal-coupling Hamiltonian of the charges is

$$H_A(t) = \sum_a \frac{[\mathbf{p}_a - q_a\mathbf{A}(t)]^2}{2m_a} + V_{\mathrm{Coul}}, \qquad \mathbf{E}(t) = -\dot{\mathbf{A}}(t). \tag{3.16}$$

Here m_a and $\mathbf{p}_a$ are the constituent masses and momenta, and V_{Coul} contains their internal electrostatic interactions. Within this spatially uniform approximation take the gauge function $\chi(\boldsymbol{\rho}, t) = -\boldsymbol{\rho} \cdot \mathbf{A}(t)$. The transformed potentials are

$$\mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\chi = 0, \qquad \phi' = \phi - \partial_t \chi = \boldsymbol{\rho} \cdot \dot{\mathbf{A}} = -\boldsymbol{\rho} \cdot \mathbf{E}. \tag{3.17}$$

Equivalently the many-charge state transforms by $|\psi'\rangle = U(t)|\psi\rangle$, with

$$U(t) = \exp\left[-\frac{\mathrm{i}}{\hbar}\mathbf{d} \cdot \mathbf{A}(t)\right], \qquad U\mathbf{p}_a U^\dagger = \mathbf{p}_a + q_a \mathbf{A}(t), \tag{3.18}$$

$$H' = U H_A U^\dagger + \mathrm{i}\hbar \dot{U} U^\dagger = H_0 + \mathbf{d} \cdot \dot{\mathbf{A}}(t) = H_0 - \mathbf{d} \cdot \mathbf{E}(t). \tag{3.19}$$

Here H_0 is the nonrelativistic charge Hamiltonian with the external radiation removed. For the emission calculation below, H_{mol} instead denotes the molecular Hamiltonian including the spin–orbit mixing already derived. The term $\mathrm{i}\hbar\dot{U}U^\dagger$ follows by differentiating $|\psi'\rangle = U|\psi\rangle$ in the Schrödinger equation. Thus the sign and operator structure of the leading electric-dipole interaction follow directly from minimal coupling. This local dipole approximation does not remove a radiation field globally: spatial gradients and magnetic multipoles were neglected. In the quantized theory we retain the corresponding term linear in the radiation field,

$$H_{\mathrm{int}} = -\mathbf{d} \cdot \mathbf{E}(\mathbf{R}_0), \tag{3.20}$$

where $\mathbf{R}_0$ is the molecular center position. Higher electric multipoles and magnetic couplings are omitted here. We also neglect molecular recoil in the energy-conservation condition; it can be restored by including center-of-mass kinetic energies. The classical gauge derivation fixes this leading one-photon interaction; it is not a derivation of the complete multipolar quantum Hamiltonian, whose additional self-energy terms are outside the present transition calculation.

Initially there are no photons: $|I\rangle = |i; 0\rangle$. A final channel is $|F\rangle = |f; 1_{\mathbf{k}\lambda}\rangle$, containing one photon. Only the creation term in Eq. (3.14) contributes:

$$M_{\mathbf{k}\lambda} \equiv \langle f; 1_{\mathbf{k}\lambda}|H_{\mathrm{int}}|i; 0\rangle \tag{3.21}$$

$$= \mathrm{i}\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\varepsilon_0 \mathcal{V}}}(\mathbf{d}_{fi} \cdot \mathbf{e}^*_{\mathbf{k}\lambda})\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_0}, \tag{3.22}$$

$$|M_{\mathbf{k}\lambda}|^2 = \frac{\hbar\omega_{\mathbf{k}}}{2\varepsilon_0 \mathcal{V}} \left|\mathbf{d}_{fi} \cdot \mathbf{e}^*_{\mathbf{k}\lambda}\right|^2. \tag{3.23}$$

The photon matrix element is $\langle 1_{\mathbf{k}\lambda}|a^\dagger_{\mathbf{k}\lambda}|0\rangle = 1$. The phase involving $\mathbf{R}_0$ drops out of the single-molecule rate.

Physical meaning. Spontaneous emission is possible from the photon vacuum because the creation operator has a nonzero vacuum-to-one-photon matrix element. Replacing the field by its vacuum mean, $\langle 0|\mathbf{E}|0\rangle = 0$, would discard this transition altogether.

3.4 From finite-time amplitudes to the golden rule

We derive the rate rather than assuming it. Let $|N\rangle$ label a complete set of uncoupled molecule-plus-field states with total energies $\mathcal{E}_N$, so $(H_{\mathrm{mol}} + H_\gamma)|N\rangle = \mathcal{E}_N|N\rangle$. Expand the full state as

$$|\Psi(t)\rangle = \sum_N c_N(t)\mathrm{e}^{-\mathrm{i}\mathcal{E}_N t/\hbar}|N\rangle. \tag{3.24}$$

Insert this into $\mathrm{i}\hbar\partial_t|\Psi\rangle = (H_{\mathrm{mol}} + H_\gamma + H_{\mathrm{int}})|\Psi\rangle$. The derivatives of the exponential factors cancel the uncoupled Hamiltonian terms. Projecting with $\langle F|$ and multiplying by $\mathrm{e}^{\mathrm{i}\mathcal{E}_F t/\hbar}$ leaves

$$\mathrm{i}\hbar\dot{c}_F(t) = \sum_N \langle F|H_{\mathrm{int}}|N\rangle\, \mathrm{e}^{\mathrm{i}(\mathcal{E}_F - \mathcal{E}_N)t/\hbar} c_N(t). \tag{3.25}$$

Initially $c_I(0) = 1$ and all other coefficients vanish. To first order in H_{int}, substitute the zeroth-order values $c_I = 1, c_{N\neq I} = 0$ on the right and integrate. This is first-order time-dependent perturbation theory, also called the interaction-picture calculation. Since $\mathcal{E}_F - \mathcal{E}_I = E_f + \hbar\omega_{\mathbf{k}} - E_i$, the amplitude is

$$c_F^{(1)}(t) = -\frac{\mathrm{i}}{\hbar} M_{\mathbf{k}\lambda} \int_0^t \mathrm{d}t'\, \mathrm{e}^{\mathrm{i}\Delta E t'/\hbar}, \qquad \Delta E = E_f + \hbar\omega_{\mathbf{k}} - E_i. \tag{3.26}$$

The exponent uses the total final minus initial energy, including the emitted photon. Direct integration yields

$$\int_0^t \mathrm{d}t'\, \mathrm{e}^{\mathrm{i}\Delta E t'/\hbar} = t\, \mathrm{e}^{\mathrm{i}\Delta E t/(2\hbar)} \operatorname{sinc}\!\left(\frac{\Delta E t}{2\hbar}\right), \qquad \operatorname{sinc} z = \frac{\sin z}{z}, \tag{3.27}$$

$$P_F^{(1)}(t) = |c_F^{(1)}(t)|^2 = \frac{|M_{\mathbf{k}\lambda}|^2 t^2}{\hbar^2} \operatorname{sinc}^2\!\left(\frac{\Delta E t}{2\hbar}\right) = \frac{4|M_{\mathbf{k}\lambda}|^2}{\Delta E^2} \sin^2\!\left(\frac{\Delta E t}{2\hbar}\right). \tag{3.28}$$

The expression at $\Delta E = 0$ is its continuous limit. For a continuum of final states, the peak around $\Delta E = 0$ narrows as $\hbar/t$. Its integrated area divided by t remains finite:

$$\int_{-\infty}^{\infty} \mathrm{d}(\Delta E)\, \frac{4\sin^2(\Delta E t/2\hbar)}{t\,\Delta E^2} = \frac{2\pi}{\hbar}, \quad \frac{4\sin^2(\Delta E t/2\hbar)}{t\,\Delta E^2} \longrightarrow \frac{2\pi}{\hbar}\delta(\Delta E). \tag{3.29}$$

The last limit is a distribution: it holds after integration against a function smooth on the peak's energy scale. The Dirac delta $\delta(\Delta E)$ has units J^{-1} and enforces energy conservation. Summing final photon channels gives Fermi's golden rule,

$$\Gamma_{i\to f} = \frac{2\pi}{\hbar} \sum_{\mathbf{k}\lambda} |M_{\mathbf{k}\lambda}|^2 \delta(E_i - E_f - \hbar\omega_{\mathbf{k}}). \tag{3.30}$$

It is a rate in s^{-1}, not a transition amplitude or an energy width. The continuum limit is taken before the long-time limit. Physically the observation interval must resolve the smooth photon spectrum while the transition probability remains small; the expression $P \simeq \Gamma t$ cannot itself be extrapolated to arbitrarily large probabilities.

3.5 Counting photons: density of states and polarization sum

Each allowed wavevector occupies a cell of volume $(2\pi/L)^3$ in wavevector space. Consequently,

$$\sum_{\mathbf{k}} \longrightarrow \frac{\mathcal{V}}{(2\pi)^3} \int \mathrm{d}^3k = \frac{\mathcal{V}}{(2\pi)^3} \int_0^\infty k^2\, \mathrm{d}k \int \mathrm{d}\Omega_{\mathbf{n}}, \qquad \mathbf{n} = \frac{\mathbf{k}}{k}. \tag{3.31}$$

Here $\mathrm{d}\Omega_{\mathbf{n}}$ is solid angle and must not be confused with an angular frequency. The factor $\mathcal{V}$ cancels the inverse volume in Eq. (3.23).

For each emission direction, the two polarizations span the plane perpendicular to $\mathbf{n}$. In Cartesian components $r, s \in \{x, y, z\}$ their completeness relation is

$$\sum_{\lambda=1}^{2} e^*_{\mathbf{k}\lambda,r} e_{\mathbf{k}\lambda,s} = \delta_{rs} - n_r n_s. \tag{3.32}$$

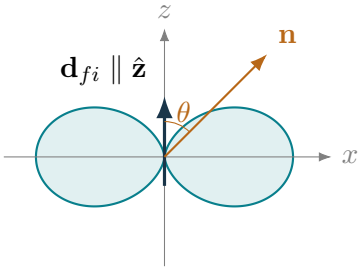
Therefore, allowing the dipole matrix element to be complex,

$$\sum_{\lambda} |\mathbf{d}_{fi} \cdot \mathbf{e}^*_{\mathbf{k}\lambda}|^2 = \sum_{r,s} d_{fi,r} d^*_{fi,s} (\delta_{rs} - n_r n_s) \tag{3.33}$$

$$= |\mathbf{d}_{fi}|^2 - |\mathbf{n} \cdot \mathbf{d}_{fi}|^2. \tag{3.34}$$

For a real dipole directed along z, this is $|\mathbf{d}_{fi}|^2 \sin^2\theta$. Radiation is transverse: there is no electric-dipole radiation along that dipole's axis. To integrate all directions, rotational symmetry requires $\int \mathrm{d}\Omega_{\mathbf{n}}\, n_r n_s = C\delta_{rs}$. Taking the trace gives $3C = \int \mathrm{d}\Omega_{\mathbf{n}}\, |\mathbf{n}|^2 = 4\pi$, so

$$\int \mathrm{d}\Omega_{\mathbf{n}} \sum_{\lambda} |\mathbf{d}_{fi}\cdot\mathbf{e}^*_{\mathbf{k}\lambda}|^2 = \left(4\pi - \frac{4\pi}{3}\right)|\mathbf{d}_{fi}|^2 = \frac{8\pi}{3}|\mathbf{d}_{fi}|^2. \tag{3.35}$$



Polar plot of angular rate $\mathrm{d}\Gamma/\mathrm{d}\Omega_{\mathbf{n}} \propto \sin^2\theta$
Zero along z; maximal in the plane perpendicular to z.

Figure. Electric-dipole radiation for a real transition matrix element along z. The drawn radius is proportional to the angular emission rate; revolving the two lobes around z gives the three-dimensional pattern. $\mathbf{d}_{fi}$ connects two different quantum states and is not an assumed permanent dipole. The general complex-vector result is Eq. (3.34).

Combining Eqs. (3.23), (3.30) and (3.31), without suppressing the normalization factors, gives

$$\Gamma_{i\to f} = \frac{2\pi}{\hbar}\frac{\hbar}{2\varepsilon_0\mathcal{V}}\frac{\mathcal{V}}{(2\pi)^3}\int_0^\infty \mathrm{d}k\, ck^3\delta(\hbar\omega_{if} - \hbar ck)\,\frac{8\pi}{3}|\mathbf{d}_{fi}|^2 \tag{3.36}$$

$$= \frac{1}{3\pi\varepsilon_0}|\mathbf{d}_{fi}|^2\int_0^\infty \mathrm{d}k\, ck^3\frac{1}{\hbar c}\delta\left(k - \frac{\omega_{if}}{c}\right) \tag{3.37}$$

$$\Gamma_{i\to f} = \boxed{\frac{\omega_{if}^3}{3\pi\varepsilon_0\hbar c^3}|\mathbf{d}_{fi}|^2.} \tag{3.38}$$

The factor $1/(\hbar c)$ is the Jacobian of the delta function; missing it would give both an incorrect coefficient and incorrect units. Indeed, $[\varepsilon_0] = \mathrm{C}^2/(\mathrm{J\,m})$ implies $[|\mathbf{d}|^2/(\varepsilon_0\hbar c^3)] = \mathrm{s}^2$, so Eq. (3.38) has units s^{-1}.

Checkpoint. The ω^3 law has two distinct origins: one power of ω comes from the squared one-photon field amplitude; two come from the number of available photon modes. A molecular transition dipole alone is insufficient to compare rates at different transition frequencies.

3.6 Magnetic sublevels, branches, and the total lifetime

Equation (3.38) uses specified normalized molecular states. When angular momentum is resolved, these include magnetic quantum numbers. For an unpolarized upper level with angular momentum J_i and a lower level with angular momentum J_f, define

$$\mathcal{D}^2_{if} \equiv \frac{1}{2J_i+1}\sum_{M_i=-J_i}^{J_i}\sum_{M_f=-J_f}^{J_f} |\langle f, J_f M_f|\mathbf{d}|i, J_i M_i\rangle|^2\,. \tag{3.39}$$

The factor $1/(2J_i+1)$ describes equal initial populations. The sum over M_f collects experimentally unresolved final sublevels; the vector norm already sums three Cartesian components. When the magnetic sublevels have the same transition frequency,

$$\Gamma_{i\to f} = \frac{\omega_{if}^3}{3\pi\varepsilon_0\hbar c^3}\mathcal{D}^2_{if}. \tag{3.40}$$

There is no additional orientation factor 1/3: the angular integration producing the $1/(3\pi)$ coefficient has already been done. If an external field splits the frequencies appreciably, one must use the frequency of each individual transition before summing.

Different lower levels are orthogonal final states. Their probabilities, and hence partial rates, add:

$$\Gamma_i^{\mathrm{rad}} = \sum_{f:E_f<E_i} \Gamma_{i\to f}, \qquad b_{if} = \frac{\Gamma_{i\to f}}{\Gamma_i^{\mathrm{rad}}}, \qquad \sum_f b_{if} = 1. \tag{3.41}$$

The dimensionless b_{if} is a branching fraction. The sum may include different vibrational and rotational levels. For each fixed final level, however, different virtual mixing paths contribute to one amplitude and must be added *before* its squared modulus is taken. This distinction will be essential for CO.

3.7 Why a rate produces an exponential, and what sets the linewidth

In weak coupling to a broad radiation continuum, a coarse time interval Δt can be chosen long enough to obtain the golden-rule rate but short compared with the lifetime. The conditional loss probability is then $\Gamma_i^{\mathrm{rad}}\Delta t$ to leading order. If $P_i(t)$ is the surviving excited-state probability,

$$P_i(t+\Delta t) = P_i(t)\left(1-\Gamma_i^{\mathrm{rad}}\Delta t\right), \tag{3.42}$$

$$\frac{\mathrm{d}P_i}{\mathrm{d}t} = -\Gamma_i^{\mathrm{rad}}P_i, \qquad P_i(t) = \mathrm{e}^{-\Gamma_i^{\mathrm{rad}}t}, \qquad \tau_i^{\mathrm{rad}} = \frac{1}{\Gamma_i^{\mathrm{rad}}}. \tag{3.43}$$

The last quantity is the radiative lifetime, also equal to $\int_0^\infty P_i(t)\,\mathrm{d}t$ in this approximation. Exponential decay is a weak-coupling, intermediate-time description, not an exact statement for all times. The finite-time amplitude in Eq. (3.28) already exhibits quadratic short-time behavior for each channel. Extremely late-time departures from an exponential are also possible. Neither issue is needed to interpret the ordinary radiative lifetime here.

With a stable lower state, the decaying transition amplitude has the form $\mathrm{e}^{-\mathrm{i}\omega_{if}t}\mathrm{e}^{-\Gamma_i^{\mathrm{rad}}t/2}$. Near the resonance, where the slowly varying photon prefactors can be evaluated at ω_{if}, its Fourier transform at observation angular frequency ω is

$$\mathcal{A}(\omega) \propto \int_0^\infty \mathrm{d}t\,\mathrm{e}^{\mathrm{i}(\omega-\omega_{if})t}\mathrm{e}^{-\Gamma_i^{\mathrm{rad}}t/2} = \frac{1}{\Gamma_i^{\mathrm{rad}}/2 - \mathrm{i}(\omega-\omega_{if})}, \tag{3.44}$$

$$|\mathcal{A}(\omega)|^2 \propto \frac{1}{(\omega-\omega_{if})^2 + (\Gamma_i^{\mathrm{rad}}/2)^2}. \tag{3.45}$$

This is a Lorentzian line shape. At half its maximum intensity the angular-frequency detuning is $\pm\Gamma_i^{\mathrm{rad}}/2$. Therefore its natural full width at half maximum (FWHM) is

$$\Delta\omega_{\mathrm{FWHM}} = \Gamma_i^{\mathrm{rad}}, \qquad \Delta\nu_{\mathrm{FWHM}} = \frac{\Gamma_i^{\mathrm{rad}}}{2\pi}, \qquad \Delta E_{\mathrm{FWHM}} = \hbar\Gamma_i^{\mathrm{rad}}, \qquad \Delta\widetilde{\nu}_{\mathrm{FWHM}} = \frac{\Gamma_i^{\mathrm{rad}}}{2\pi c}. \tag{3.46}$$

Here $\nu = \omega/(2\pi)$ is ordinary frequency in hertz; $\widetilde{\nu} = \nu/c$ is wavenumber, in cm^{-1} if c is in $\mathrm{cm\,s}^{-1}$. These widths assume a stable lower state and neglect other broadening. A measured line can be much wider because of Doppler, instrumental or collisional effects. Likewise, a measured population lifetime equals τ_i^{rad} only when additional loss mechanisms are negligible; independent extra loss rates add to Γ_i^{rad}.

Checkpoint. We have completed the field-theory step using harmonic oscillators and first-order perturbation theory. A molecular matrix element predicts a partial radiative rate; summing distinct final states gives the inverse radiative lifetime. Mixing paths to the *same* final state interfere, while distinct final states contribute separate rates.

4 From spectral shifts to radiative lifetimes

Aim. Join the two completed calculations, then test which inference survives beyond the three-level model.

4.1 The exact rate–shift relation in the declared model

Apply Eq. (3.38) to the lower mixed level:

$$\omega_{-g} = \frac{E_- - E_g}{\hbar} > 0, \qquad \Gamma_{-\to g} = \frac{\omega_{-g}^3}{3\pi\varepsilon_0\hbar c^3} w|\mathbf{d}_{gs}|^2. \tag{4.1}$$

Define a reference rate for the *unmixed* bright level to the *same* lower state:

$$\omega_{sg} = \frac{E_s - E_g}{\hbar}, \qquad \Gamma^{(0)}_{s\to g} = \frac{\omega_{sg}^3}{3\pi\varepsilon_0\hbar c^3}|\mathbf{d}_{gs}|^2. \tag{4.2}$$

Dividing cancels the dipole magnitude and the radiation constants:

$$\boxed{\frac{\Gamma_{-\to g}}{\Gamma^{(0)}_{s\to g}} = \left(\frac{\omega_{-g}}{\omega_{sg}}\right)^3 \frac{\delta}{\Delta + 2\delta}.} \tag{4.3}$$

This is our main solved result. It is exact in V within the declared matter Hamiltonian, and leading order in the coupling to radiation. We assume distinct, narrow mixed levels, with radiative energy widths small compared with their separation. The radiative interaction does not need to be included in the Hermitian level diagonalization at this accuracy.

At weak mixing $\delta \ll \Delta$,

$$\frac{\Gamma_{-\to g}}{\Gamma^{(0)}_{s\to g}} \simeq \left(\frac{\omega_{-g}}{\omega_{sg}}\right)^3 \frac{\delta}{\Delta} \simeq \left(\frac{\omega_{tg}}{\omega_{sg}}\right)^3 \frac{|V|^2}{\Delta^2}, \qquad \omega_{tg} = \frac{E_t - E_g}{\hbar}. \tag{4.4}$$

Replacing ω_{-g} by ω_{tg} changes only higher-order terms when the shift is also small compared with the photon energy. Even when w is known, the *frequency ratio* cannot be dropped unless the two transition frequencies are close.

For unresolved magnetic sublevels, replace $|\mathbf{d}_{gs}|^2$ by the correctly summed and averaged $\mathcal{D}^2_{sg}$ of Eq. (3.39). A scalar spin–orbit interaction has the same mixing coefficient for all M_J within a fixed J block, so the corresponding averaged ratio retains this form.

Checkpoint. A weak singlet admixture suppresses the rate by its probability w, multiplied by a photon phase-space factor. A spectral shift supplies w only after a mixing model and its bare detuning have been established. The absolute lifetime still requires an allowed dipole strength and all final branches.

4.2 Why one line shift does not by itself measure a lifetime

A measured line gives a difference between *physical* energies. In contrast, $\delta = E_t - E_-$ refers to an unperturbed energy E_t supplied by a deperturbation model. It is not a second independently observed level. Several interaction models can fit a limited set of positions.

Even in the three-level model, Eq. (4.3) requires:

1. the bare detuning Δ and the inferred shift δ;
2. a reference dipole strength or reference *partial* rate;

3. the two transition frequencies and consistent angular-state definitions.

If several lower levels are accessible, the total rate is

$$\Gamma_-^{\mathrm{rad}} = \frac{w}{3\pi\varepsilon_0\hbar c^3} \sum_{g:E_g<E_-} \omega_{-g}^3 |\mathbf{d}_{gs}|^2. \tag{4.5}$$

There is generally no single frequency factor converting the total bright-state rate into this total rate. A lower level accessible from s may even lie above E_- and be inaccessible from the mixed state. For an ensemble with upper population N_i, the emitted photon count rate in branch $i \to f$ is $N_i\Gamma_{i\to f}$ before detector factors. An observed line intensity therefore also depends on population and experimental conditions.

4.3 Both initial and final states can acquire admixtures

Derivation of the general first-order transition amplitude. Let W be a weak spin–orbit perturbation, and let $|i\rangle, |f\rangle, |n\rangle, |m\rangle$ be orthonormal eigenstates of H_0. All energies in this subsection are unperturbed. For nondegenerate levels with small admixtures, the first-order state correction follows by projecting $(H_0 + W)(|i\rangle + |\delta i\rangle) = (E_i + \delta E_i)(|i\rangle + |\delta i\rangle)$ onto $\langle n|$ with $n \neq i$:

$$(E_n - E_i)\langle n|\delta i\rangle = -W_{ni}, \qquad W_{ni} = \langle n|W|i\rangle. \tag{4.6}$$

Therefore

$$|\widetilde{i}\rangle = |i\rangle + \sum_{n\neq i} |n\rangle \frac{W_{ni}}{E_i - E_n} + O(W^2), \tag{4.7}$$

$$\langle\widetilde{f}| = \langle f| + \sum_{m\neq f} \frac{W_{fm}}{E_f - E_m} \langle m| + O(W^2). \tag{4.8}$$

The second line is the Hermitian conjugate of the corrected final ket; its matrix element is W_{fm}, not W_{mf}. If the bare dipole $\mathbf{d}_{fi}$ vanishes, the induced amplitude is

$$\boxed{\mathbf{d}_{fi}^{(1)} = \sum_{n\neq i} \frac{\mathbf{d}_{fn} W_{ni}}{E_i - E_n} + \sum_{m\neq f} \frac{W_{fm}\mathbf{d}_{mi}}{E_f - E_m}.} \tag{4.9}$$

In the first sum a triplet initial state acquires singlet components before the spin-conserving E1 transition. In the second sum the singlet final state acquires triplet components. Symmetry eliminates many terms, but neglect of the second sum is an additional approximation, not a universal selection rule.

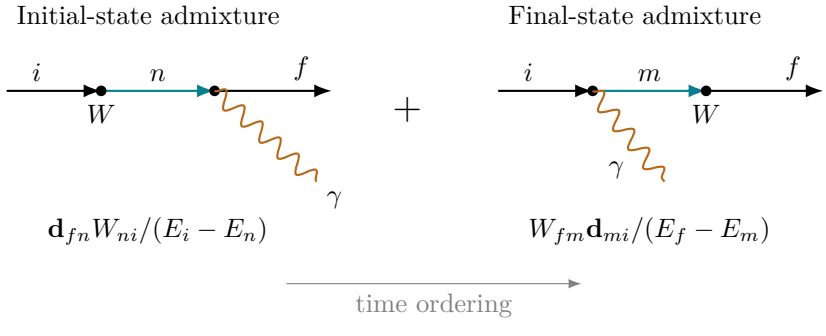


Figure. Two time-ordered contributions to one emission amplitude. Straight segments label molecular intermediate states; wavy segments denote the emitted photon. They are perturbative pathway diagrams, not a claim that the virtual molecular level is populated as an on-shell state.

The same denominators follow from second-order time-ordered perturbation theory for matter plus photons. If W acts first, the intermediate state is $|n;0\rangle$, giving $E_i - E_n$. If the photon is emitted first, the intermediate state is $|m;1_\gamma\rangle$, giving

$$E_i - (E_m + \hbar\omega) = E_f - E_m, \qquad \hbar\omega = E_i - E_f. \tag{4.10}$$

Thus state dressing and the two ordered emission pathways are the same calculation in two forms. Small denominators signal the need to diagonalize the nearly degenerate subspace before treating emission.

4.4 Several singlet pathways: shifts add, amplitudes interfere

Return to the reduced assumption that the lower state is unmixed. Suppose t couples to singlet levels s_n, with $V_n = \langle s_n|W|t\rangle$ and $\Delta_n = E_{s_n} - E_t > 0$. To leading nonzero order,

$$E_{\widetilde{t}} - E_t = -\sum_n \frac{|V_n|^2}{\Delta_n}, \tag{4.11}$$

$$\mathbf{d}_{g\widetilde{t}} = -\sum_n \frac{\mathbf{d}_{gs_n} V_n}{\Delta_n}. \tag{4.12}$$

The first expression follows by projecting the second-order eigenvalue equation onto $\langle t|$; insert $\langle s_n|\delta t\rangle = -V_n/\Delta_n$ to obtain $\delta E_t = \sum_n V_n^* \langle s_n|\delta t\rangle$. The second expression follows by inserting the corrected ket in $\langle g|\mathbf{d}|\widetilde{t}\rangle$.

Write $\mathbf{A}_n = -\mathbf{d}_{gs_n} V_n/\Delta_n$. The line strength is

$$\left|\sum_n \mathbf{A}_n\right|^2 = \sum_n |\mathbf{A}_n|^2 + 2\ \mathrm{Re} \sum_{n<m} \mathbf{A}_n^* \cdot \mathbf{A}_m. \tag{4.13}$$

The cross terms can enhance or suppress a rate. Energy shifts contain $|V_n|^2$ and no dipole factors; they do not determine these cross terms. A small contribution to a level shift can still matter for a radiative amplitude if its allowed dipole is large. Levels below t can also contribute in the general expression; their energy denominators have the opposite sign.

This phase dependence is physical, not an arbitrary sign convention. Under $|s_n\rangle \mapsto \mathrm{e}^{\mathrm{i}\chi_n}|s_n\rangle$,

$$\mathbf{d}_{gs_n} \mapsto \mathrm{e}^{\mathrm{i}\chi_n}\mathbf{d}_{gs_n}, \qquad V_n \mapsto \mathrm{e}^{-\mathrm{i}\chi_n} V_n, \qquad \mathbf{d}_{gs_n} V_n \mapsto \mathbf{d}_{gs_n} V_n. \tag{4.14}$$

Only the invariant products enter the amplitude.

4.5 An exact counterexample: identical energies, different emission

To prove the information limit without a weak-coupling expansion, take two degenerate singlets and one triplet:

$$H_{\mathrm{ex}} = \begin{pmatrix} E_t & V & V \\ V & E_s & 0 \\ V & 0 & E_s \end{pmatrix}_{(t,s_1,s_2)}, \qquad V \text{ real}, \quad E_s > E_t. \tag{4.15}$$

Define the orthonormal combinations

$$|B\rangle = \frac{|s_1\rangle + |s_2\rangle}{\sqrt{2}}, \qquad |D\rangle = \frac{|s_1\rangle - |s_2\rangle}{\sqrt{2}}. \tag{4.16}$$

Here B, D label the combination coupled to t and its uncoupled partner; their brightness will depend on the dipole operator. Direct multiplication gives

$$\langle B|H_{\mathrm{ex}}|t\rangle = \sqrt{2}V, \qquad \langle D|H_{\mathrm{ex}}|t\rangle = 0, \qquad H_{\mathrm{ex}} = \begin{pmatrix} E_t & \sqrt{2}V & 0 \\ \sqrt{2}V & E_s & 0 \\ 0 & 0 & E_s \end{pmatrix}_{(t,B,D)}. \tag{4.17}$$

Thus its spectrum is E_s and

$$E_\pm = \frac{E_t + E_s \pm \sqrt{(E_s - E_t)^2 + 8V^2}}{2}. \tag{4.18}$$

The lower mixed eigenstate contains only t, B, with singlet weight $w_B = \frac{1}{2}[1 - \Delta/\sqrt{\Delta^2 + 8V^2}]$.

Now compare two hypothetical transition operators with the *same* matter Hamiltonian and lower level:

	Constructive case	Destructive case
Bare dipoles	$\mathbf{d}_{gs_1} = \mathbf{d},\ \mathbf{d}_{gs_2} = \mathbf{d}$	$\mathbf{d}_{gs_1} = \mathbf{d},\ \mathbf{d}_{gs_2} = -\mathbf{d}$
Dipole from B	$\mathbf{d}_{gB} = \sqrt{2}\mathbf{d}$	$\mathbf{d}_{gB} = \mathbf{0}$
Lower mixed strength	$2w_B\lvert\mathbf{d}\rvert^2$	0

All three excited-state energies are identical in the two cases, as are the magnitudes of the individual bare dipoles. The lower mixed state nevertheless has a nonzero E1 rate to g in one case and zero in the other.

These are different Hamiltonian–observable models, not a rephasing of one model: changing a basis phase would also change the corresponding V, leaving every dipole–coupling product invariant. The zero concerns the selected E1 channel in this restricted model; it does not prove that any real molecular state has an infinite lifetime.

Checkpoint. Line positions alone do not determine transition strengths. A deperturbation analysis constrains the Hamiltonian, while radiative observables also probe transition operators and interference. A consistent model should be tested against both kinds of information.

4.6 What the model represents in actual CO

The pathway $a^3\Pi_1 \leftrightarrow A^1\Pi_1 \rightarrow X^1\Sigma^+$ is a useful first approximation. Higher $^1\Pi$ states also contribute to the induced transition moment, and rotational and vibrational structure select the individual channels [1, 2]. The $a^3\Pi_\Omega$ components therefore need not have the same lifetime. An $\Omega = 1$ example is not a formula for every $a^3\Pi$ level.

To connect the model parameters to molecular wavefunctions, let $\Phi_\alpha(\mathbf{r}; R)$ be a normalized electronic state at fixed internuclear separation R, and let $\chi_{\alpha v}(R)$ be its radial vibrational wavefunction, normalized with measure $\mathrm{d}R$. The radial coordinate convention absorbs the usual radial measure into χ. Define electronic matrix elements at fixed R by

$$h^{\mathrm{SO}}_{\alpha\beta}(R) = \langle\Phi_\alpha(R)|H_{\mathrm{SO}}|\Phi_\beta(R)\rangle, \qquad \mathbf{d}^{\mathrm{el}}_{g\alpha}(R) = \langle\Phi_g(R)|\mathbf{d}|\Phi_\alpha(R)\rangle. \tag{4.19}$$

Within this channel expansion, the radial parts of the rovibronic inputs are

$$V_{\alpha v,\beta v'} = \int_0^\infty \mathrm{d}R\, \chi^*_{\alpha v}(R) h^{\mathrm{SO}}_{\alpha\beta}(R) \chi_{\beta v'}(R), \tag{4.20}$$

$$\mathbf{d}^{\mathrm{radial}}_{gv'',\alpha v} = \int_0^\infty \mathrm{d}R\, \chi^*_{gv''}(R) \mathbf{d}^{\mathrm{el}}_{g\alpha}(R) \chi_{\alpha v}(R). \tag{4.21}$$

The second equation is a molecule-fixed electronic vector integral. Converting it into the laboratory vector matrix elements used in the rates additionally requires the rotational wavefunctions and their angular factors. No universal factor equal to one has been assumed. The first equation likewise presumes the compatible electronic and rotational angular channel has been selected.

If $\mathbf{d}^{\mathrm{el}}(R)$ can be approximated by a constant vector $\mathbf{d}_0$ over the vibrational support, the radial strength factorizes:

$$|\mathbf{d}^{\mathrm{radial}}_{gv'',\alpha v}|^2 \simeq |\mathbf{d}_0|^2 \left| \int_0^\infty \mathrm{d}R\, \chi^*_{gv''}(R) \chi_{\alpha v}(R) \right|^2. \tag{4.22}$$

The squared overlap is the Franck–Condon factor. Strong variation or cancellation of an effective intercombination dipole can invalidate the constant-dipole approximation.

The intermediate index n in Eq. (4.9) can denote complete electronic, vibrational and rotational channels. Replacing that sum by a single electronic energy gap is a further approximation. Our exact two-state solution is exact for its chosen finite-dimensional model, not an exact fourteen-electron prediction.

4.7 A literature benchmark and the remaining scientific question

The $a^3\Pi_1(v = 0, J = 1)$ radiative lifetime provides a concrete test:

Result for this specified level	Lifetime
2007 trapped-molecule measurement [1]	2.63 ± 0.03 ms
2007 calculation retaining only the $A^1\Pi$ borrowing channel [1]	3.16 ms
2021 relativistic Fock-space coupled-cluster calculation [2]	2.59 ms

The 2007 study examined higher singlet contributions; the later relativistic calculation agrees closely with the measured lifetime of this level. Thus the earlier discrepancy should not be presented as an unresolved anomaly. It illustrates the quantitative importance of the molecular transition moment beyond a single borrowing state.

The finite-model analytical problem has a complete solution. The broader modeling question is how to choose and constrain a reduced set of molecular states that predicts both positions and radiative observables. The derivation identifies the needed information: spin–orbit couplings, electronic dipoles, vibrational overlaps, angular factors, and their relative phases. Quantifying these for CO requires molecular calculations or data beyond the finite model.

5 Conclusions and report questions

5.1 What has been established

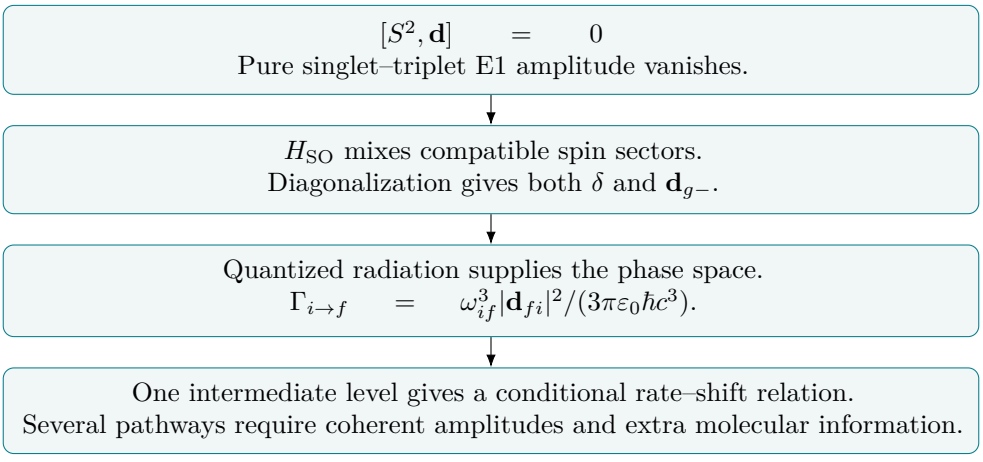


Figure. The completed chain of reasoning. Every step answers the same question: how does the interaction seen in deperturbed energies appear in the emitted light?

1. A selection rule is a statement about an operator and the symmetries of its states. Relativistic mixing permits E1 radiation without changing the spin independence of the leading dipole operator.
2. A consistent effective description transforms transition operators together with energies. Eliminated virtual states remain visible in both.
3. The radiative calculation fixes the absolute rate and the ω^3 factor. Distinct final states contribute separate probabilities; distinct virtual paths to one final state interfere.
4. A single-state borrowing model predicts Eq. (4.3). A many-state model explains why fitting positions alone need not predict lifetimes or intensities.

Applying these results to deperturbation work means asking which state admixtures a fit determines, which transition moments remain external inputs, and which additional measurements could distinguish otherwise similar models.

5.2 Questions for the report

The following six questions define the report. Develop the physical arguments and re-derive the central calculations, distinguishing results of the finite model from quantitative claims about CO.

1. **Why can a spin-forbidden transition radiate?** Re-derive the singlet–triplet E1 selection rule and explain how microscopic spin–orbit interactions make it possible for physical eigenstates to radiate. Clarify why replacing these interactions by a single scalar multiple of $\mathbf{L} \cdot \mathbf{S}$ cannot describe singlet–triplet mixing.

2. **How are level shifts and state composition related?** Solve the two-state mixing problem and re-derive the exact relation between the energy shift and the borrowed singlet weight. Analyze the weak-mixing and near-degeneracy limits, identifying where nondegenerate perturbation theory fails.

3. **How does a transition dipole determine an absolute emission rate?** Re-derive the free-space E1 rate from the quantized radiation field and time-dependent perturbation theory. Account for the normalization, polarization sum and photon phase space, and explain the physical origin of the ω^3 dependence.

4. **Under what assumptions does a spectral shift predict a lifetime?** Derive the rate–shift relation in the one-intermediate-state model, treating the Hamiltonian and transition operator consistently. Explain which additional information is needed to pass from a partial rate to a total radiative lifetime and to an observed line intensity.

5. **Why do several intermediate states change the inference?** Re-derive the induced dipole amplitude when both initial and final states acquire admixtures. Use the exact interference example to explain how identical matter spectra can coexist with different emission strengths, keeping arbitrary basis phases distinct from physical relative amplitudes.

6. **What does the CO benchmark establish?** Assess the single-$A^1\Pi$ borrowing picture against the specified-level lifetime measurements and later relativistic calculation. Explain which molecular inputs and additional observables are needed to extend the analysis to other $a^3\Pi(v, J, \Omega)$ levels, and what deperturbed line positions alone leave undetermined.

Scope of the calculation

The two-state mixing is exact within the finite model. The induced many-state dipole is perturbative in spin–orbit mixing away from degeneracies. Photon emission is calculated to leading electric-dipole order in free space. Radiative lifetimes assume the absence of competing losses; the Lorentzian width assumes a stable lower level. This finite-model analysis does not compute CO electronic wavefunctions, relativistic couplings or molecular transition moments from scratch.

References

[1] J. J. Gilijamse, S. Hoekstra, S. A. Meek, M. Metsälä, S. Y. T. van de Meerakker, G. Meijer and G. C. Groenenboom, *The radiative lifetime of metastable CO ($a^3\Pi, v = 0$)*, J. Chem. Phys. **127**, 221102 (2007). doi:10.1063/1.2813888; arXiv:0710.2240. Read the measured level-specific lifetimes and the comparison of one- and multiple-singlet borrowing calculations.

[2] N. S. Mosyagin, A. V. Oleynichenko, A. Zaitsevskii, A. V. Kudrin, E. A. Pazyuk and A. V. Stolyarov, *Ab initio relativistic treatment of the $a^3\Pi - X^1\Sigma^+$, $a'^3\Sigma^+ - X^1\Sigma^+$ and $A^1\Pi - X^1\Sigma^+$ systems of the CO molecule*, J. Quant. Spectrosc. Radiat. Transfer **263**, 107532 (2021). doi:10.1016/j.jqsrt.2021.107532. The open preprint, arXiv:2010.08849, has the narrower title *Ab initio relativistic treatment of the intercombination $a^3\Pi - X^1\Sigma^+$ Cameron system of the CO molecule*. See its transition moments and Table III for level-resolved lifetimes.

[3] P. A. M. Dirac, *The Quantum Theory of the Emission and Absorption of Radiation*, Proc. R. Soc. Lond. A **114**, 243–265 (1927). doi:10.1098/rspa.1927.0039. Historical primary source for treating emission with a quantized radiation field; the modern SI derivation used here is supplied in Section 3.

Fields, symmetries and particles

Lecture 1

Fields, symmetries and particles

In this lecture

- Why a relativistic quantum theory needs fields; relativistic notation and scalar fields.
- The action of a free real scalar field and the Klein–Gordon equation from the variational principle, with every boundary term displayed; plane waves and the dispersion relation.
- Noether's theorem for space-time translations: the energy–momentum tensor and the conserved four-momentum, whose time component becomes the Hamiltonian.
- Canonical quantisation: the harmonic oscillator as a template, the equal-time commutators from a lattice, and the algebra of the mode operators $a_{\vec{p}}$, $a^\dagger_{\vec{p}}$, derived from them.
- The particle interpretation: H_0 and $\vec{P}$ count quanta of energy $E_{\vec{p}}$ and momentum $\vec{p}$; vacuum energy and normal ordering; Fock space and Bose symmetry.
- Relativistic normalisation: the invariant measure $d\Pi_{\vec{p}}$, the overlap $\langle\vec{q}|\vec{p}\rangle$ and the matrix element $\langle 0|\phi(x)|\vec{p}\rangle = e^{-ip\cdot x}$ used in the scattering calculation of Lectures 2 and 3.

Builds on: special relativity and quantum mechanics (harmonic oscillator, Heisenberg picture); no previous knowledge of field theory is assumed.
In the book: Lund lecture notes, Lectures 2–6; book (R. Pasechnik, *Introduction to QFT*), Chapters 3–4 (relativistic notation: Chapter 2).

Our first task is to explain what a particle is in quantum field theory. We start from an action, quantise its normal modes, and show, by calculation rather than by assumption, that one quantum carries the energy $\sqrt{|\vec{p}|^2 + m^2}$ and the momentum $\vec{p}$. Here the field is free: we construct the states that enter and leave a collision. Lecture 2 lets these states interact, and Lecture 3 turns the resulting amplitude into a cross section (Figure 1.1).

1.1 Why fields? Relativistic notation

A route through the three lectures. The course follows one thread:

$$\underbrace{\mathcal{A}_0[\phi] \longrightarrow H_0 \longrightarrow \phi(x),\ |\vec{p}\rangle}_{\text{Lecture 1}} \longrightarrow \underbrace{D_F,\ \mathsf{S},\ \mathcal{M}}_{\text{Lecture 2}} \longrightarrow \underbrace{d\sigma}_{\text{Lecture 3}} . \tag{1.1}$$

Here $\mathcal{A}_0$ is the free-field action, H_0 the Hamiltonian, $\phi(x)$ the field, which will become an operator, and $|\vec{p}\rangle$ a one-particle state of momentum $\vec{p}$. Later, D_F denotes a time-ordered correlation function, S the scattering operator, $\mathcal{M}$ its invariant amplitude and σ a cross section. These results do not depend on these later objects.

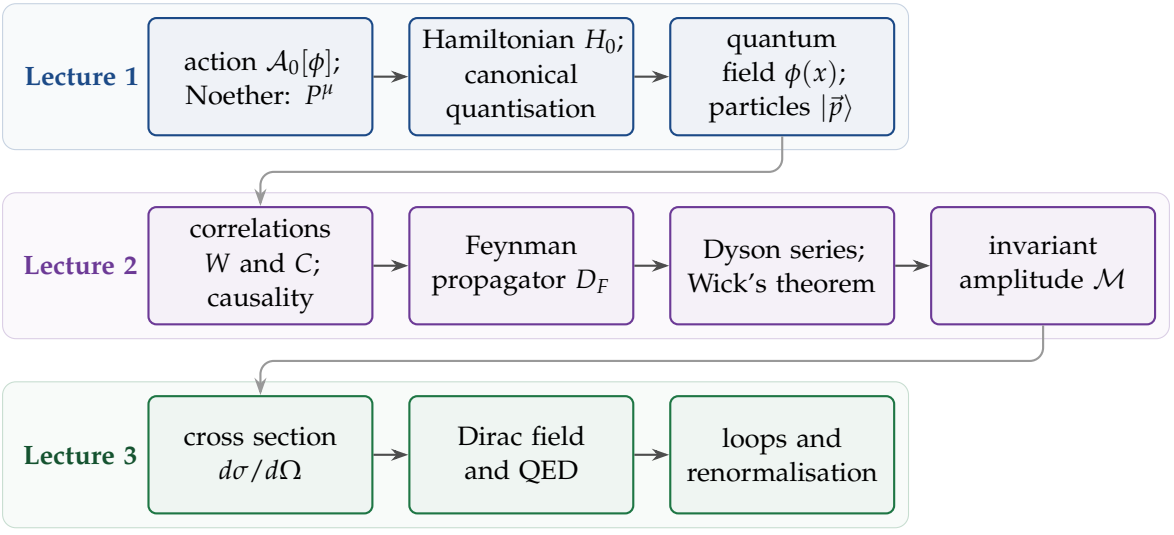


Figure 1.1: Road map of the course. Lecture 1, *Fields, symmetries and particles*, builds the free quantum field and its particle states. Lecture 2, *Causality and perturbative interactions*, adds correlation functions and causality, the Feynman propagator and the perturbative expansion, and ends with the invariant amplitude $\mathcal{M}$. Lecture 3, *From amplitudes to observables*, turns $\mathcal{M}$ into a cross section and opens the Dirac field, QED and loops.

Why a field? A wave function for a fixed number of particles is useful as long as the production of new particles can be neglected. Relativistic interactions, however, connect sectors with different particle numbers, and the space of states must accommodate all of them:

$$\mathcal{F} = \mathcal{F}_0 \oplus \mathcal{F}_1 \oplus \mathcal{F}_2 \oplus \cdots , \tag{1.2}$$

where $\mathcal{F}_n$ denotes the n-particle sector. The free theory will construct these sectors; interactions can connect them when the interaction and the conservation laws permit. The localisation scale gives a useful estimate of when this happens:

$$\Delta x \sim L, \qquad \Delta p \gtrsim \frac{1}{2L}, \qquad E_{\text{probe}} \sim \Delta p \quad (\Delta p \gg m). \tag{1.3}$$

At L of the order of the Compton wavelength m^{-1}, the energy needed to resolve the particle becomes comparable to its rest energy, and particle creation may become relevant; the actual threshold also depends on the process and on four-momentum conservation. Relativity adds a second requirement, causality. In a one-particle relativistic theory the propagation amplitude leaks outside the light cone, and there this is a genuine problem (Lund lecture notes, Lecture 1); in quantum field theory a non-zero amplitude outside the light cone is not by itself a demonstration of superluminal signalling, because the relevant test is the commutator of local observables, which we derive in Lecture 2.

Convention (units, metric and indices). We use $\hbar = c = 1$, so that time and length have the units of inverse energy, and the metric $(+,-,-,-)$:

$$x^\mu = (t, x^1, x^2, x^3) = (t, \vec{x}), \qquad p^\mu = (E, \vec{p}), \qquad g_{\mu\nu} = \operatorname{diag}(1,-1,-1,-1), \tag{1.4}$$
$$x_\mu = g_{\mu\nu} x^\nu = (t, -\vec{x}), \qquad p \cdot x = p_\mu x^\mu = Et - \vec{p}\cdot\vec{x},$$

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = (\partial_t, \nabla), \qquad \partial^\mu = (\partial_t, -\nabla), \qquad \Box = \partial_\mu \partial^\mu = \partial_t^2 - \nabla^2. \tag{1.5}$$

Repeated Greek indices are summed over $0, 1, 2, 3$; Latin indices i, j run over $1, 2, 3$. A dot between three-vectors, $\vec{p} \cdot \vec{x}$, is the Euclidean scalar product; a dot between four-vectors uses the metric. The d'Alembertian $\Box$ is a differential operator. The minus signs in $\Box$ and in $p \cdot x$ come from the metric and are fixed once and for all.

Lorentz transformations and scalar fields. A Lorentz transformation $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ preserves the interval:

$$\begin{aligned} x'^\mu g_{\mu\nu} x'^\nu &= x^\rho \left(\Lambda^\mu{}_\rho \, g_{\mu\nu} \, \Lambda^\nu{}_\sigma\right) x^\sigma = x^\rho g_{\rho\sigma} x^\sigma, \\ \Lambda^{\mathsf{T}} g \Lambda &= g, \qquad |\det \Lambda| = 1, \qquad d^4x' = d^4x. \end{aligned} \tag{1.6}$$

A scalar field assigns to each event a number that acquires no index under a Lorentz transformation:

$$\phi'(x') = \phi(x), \qquad \phi'(x) = \phi(\Lambda^{-1}x), \qquad \partial'_\mu \phi'(x') = (\Lambda^{-1})^\nu{}_\mu \, \partial_\nu \phi(x). \tag{1.7}$$

Consequently $\partial_\mu\phi\,\partial^\mu\phi$ and ϕ^2 are scalars. The sign of the interval $x^2 = t^2 - |\vec{x}|^2$ is invariant and classifies the separation of an event from the origin as time-like ($x^2 > 0$), light-like ($x^2 = 0$) or space-like ($x^2 < 0$) (Figure 1.2a). For space-like separations even the time order of two events depends on the frame (Figure 1.2b); this is why causality will be formulated, in Lecture 2, as a condition on observables at space-like separation. When we transform positive-energy states we restrict ourselves to proper orthochronous Lorentz transformations ($\det\Lambda = 1$, $\Lambda^0{}_0 \geq 1$), which preserve the future mass shell.

1.2 The action and the field equation

We obtain the field equation by varying a local relativistic action, with every boundary term displayed. The action specifies the model. Lorentz invariance and locality constrain it but do not determine it; we choose the simplest stable, free, real scalar field and ask what this choice predicts. For a real classical field $\phi(x)$ with mass parameter $m > 0$,

$$\begin{aligned} \mathcal{A}_0[\phi] &= \int d^4x \, \mathcal{L}_0, & \mathcal{L}_0 &= \tfrac{1}{2}\,\partial_\mu\phi\,\partial^\mu\phi - \tfrac{1}{2}m^2\phi^2 \\ &= \int dt \int d^3x \, \mathcal{L}_0, & \mathcal{L}_0 &= \tfrac{1}{2}\dot{\phi}^2 - \tfrac{1}{2}(\nabla\phi)^2 - \tfrac{1}{2}m^2\phi^2. \end{aligned} \tag{1.8}$$

Here $\dot{\phi} = \partial_t\phi$, and $\mathcal{L}_0$ is a Lagrangian density: the ordinary Lagrangian is $L_0(t) = \int d^3x\,\mathcal{L}_0$. The subscript 0 stands for "free". "Local" means that the fields and their derivatives in each term are evaluated at the same space-time point.

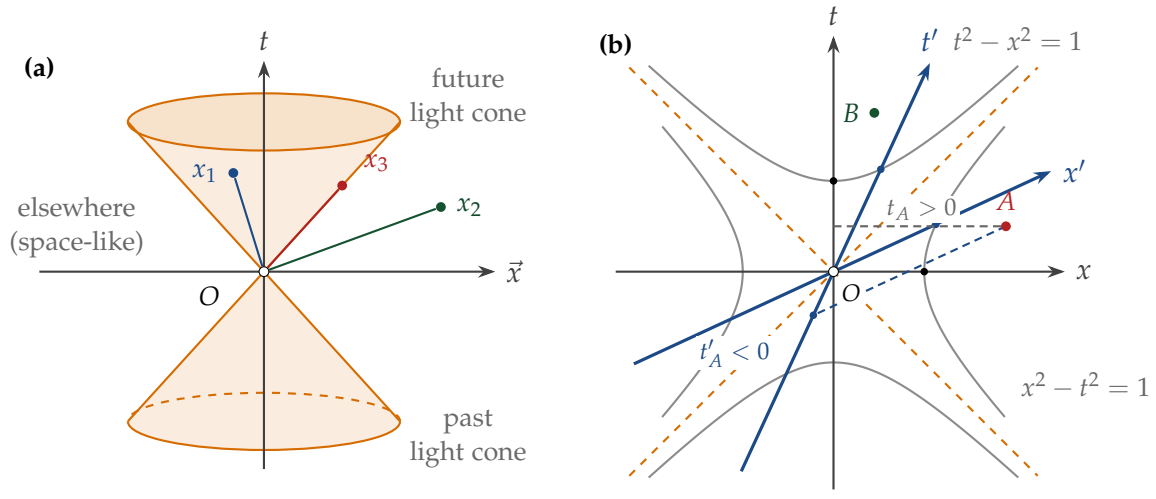


Figure 1.2: (a) The light cone of the origin O (two space dimensions suppressed into the horizontal axis, one shown in perspective by the ellipses). The sign of $x^2 = t^2 - |\vec{x}|^2$ is Lorentz invariant: x_1 is time-like ($x_1^2 > 0$), x_2 space-like ($x_2^2 < 0$) and x_3 light-like ($x_3^2 = 0$) separated from O. (b) Minkowski diagram of an observer moving with velocity $\beta = \tanh 0.5 \approx 0.46$ along x (blue axes, tilted towards the dashed light cone); the hyperbolae $t^2 - x^2 = \pm 1$ mark unit time and length on both sets of axes. The time-like separated event B happens after O for both observers. The space-like separated event A happens after O in the unprimed frame ($t_A > 0$, grey line of simultaneity) but before O in the primed frame ($t'_A < 0$, blue line of simultaneity, parallel to the x' axis).

The variational principle. We vary the field while keeping its values on the boundary fixed:

$$\begin{aligned} \phi \longmapsto \phi + \delta\phi, \qquad & \delta(\partial_\mu \phi) = \partial_\mu \delta\phi, \\ \delta\mathcal{A}_0 &= \int d^4x \left[\partial^\mu \phi\, \partial_\mu \delta\phi - m^2 \phi\, \delta\phi\right] \\ &= \int d^4x \left\{\partial_\mu\left(\partial^\mu \phi\, \delta\phi\right) - \left(\Box\phi + m^2\phi\right) \delta\phi\right\} \\ &= \underbrace{\int_{\partial R} d\Sigma_\mu\, \partial^\mu \phi\, \delta\phi}_{0} - \int_R d^4x\, (\Box + m^2)\phi\, \delta\phi. \end{aligned} \tag{1.9}$$

R is the integration region and $d\Sigma_\mu$ the oriented surface element of its boundary ∂R (for a slab between two times see Figure 1.4b). Taking $\delta\phi = 0$ on the boundary eliminates the first term, and arbitrary variations in the interior then give

Klein–Gordon equation.

$$(\Box + m^2)\, \phi = 0. \tag{1.10}$$

More generally, for a Lagrangian density $\mathcal{L}(\phi, \partial_\mu \phi)$,

$$\delta\mathcal{L} = \left[\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\right] \delta\phi + \partial_\mu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\, \delta\phi\right], \tag{1.11}$$

so stationarity of the action yields the Euler–Lagrange equation, the vanishing of the first square bracket. A classical field that satisfies it is said to be *on shell*; for a free plane wave this means $p^2 = m^2$.

Plane waves and the dispersion relation. For $\phi(x) = A\, e^{-ip\cdot x}$,

$$\begin{aligned} \partial_\mu \phi &= -ip_\mu \phi, & \Box\phi &= -p_\mu p^\mu \phi, \\ (\Box + m^2)\phi &= (-p^2 + m^2)\phi = 0, & p^2 &= (p^0)^2 - |\vec{p}|^2 = m^2, \\ p^0 &= \pm E_{\vec{p}}, & E_{\vec{p}} &= \sqrt{|\vec{p}|^2 + m^2} > 0. \end{aligned} \tag{1.12}$$

The four-momenta lie on the two sheets of the mass shell (Figure 1.3a). A real classical solution contains both frequency components, $e^{-iE_{\vec{p}}t}$ and its complex conjugate. After quantisation the negative-frequency component will multiply a creation operator: it creates a quantum of positive energy, not a particle of negative energy. The dispersion relation has the familiar limits (Figure 1.3b)

$$\begin{aligned} |\vec{p}| \ll m: \quad & E_{\vec{p}} = m\sqrt{1 + |\vec{p}|^2/m^2} = m + \frac{|\vec{p}|^2}{2m} + \mathcal{O}\big(|\vec{p}|^4/m^3\big), \\ |\vec{p}| \gg m: \quad & E_{\vec{p}} = |\vec{p}| + \frac{m^2}{2|\vec{p}|} + \mathcal{O}\big(m^4/|\vec{p}|^3\big). \end{aligned} \tag{1.13}$$

The normal-mode frequency thus has the correct non-relativistic and ultra-relativistic limits. To show that it is a *particle energy* we need the Hamiltonian and the quantum states, derived below (Section 1.7).

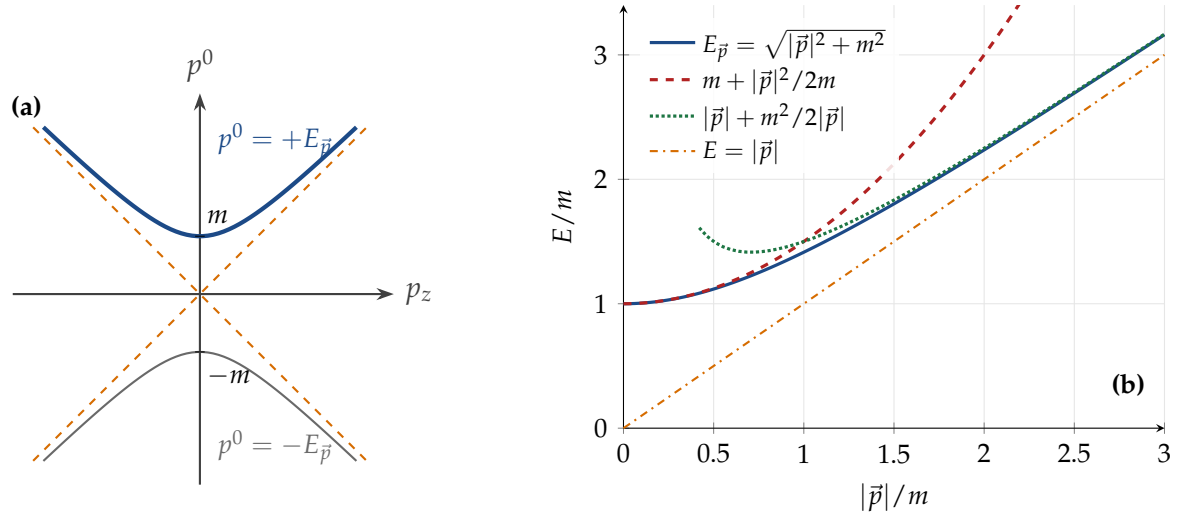


Figure 1.3: (a) The mass shell $p^2 = m^2$ in the (p_z, p^0) plane ($\vec{p}_\perp = 0$): two sheets $p^0 = \pm E_{\vec{p}}$ inside the light cone (dashed). The positive-energy sheet (blue) is mapped onto itself by proper orthochronous Lorentz transformations, and $\theta(p^0)\,\delta(p^2 - m^2)$ selects it in the invariant measure (1.76). (b) The dispersion relation $E_{\vec{p}} = \sqrt{|\vec{p}|^2 + m^2}$ with its non-relativistic and ultra-relativistic approximations (1.13); for $|\vec{p}| \gg m$ it approaches the light cone $E = |\vec{p}|$.

Units. Write $[X]$ for the mass dimension of X, the power of energy that it carries. A quantum phase is e^{iA}, so $[\mathcal{A}] = 0$ and

$$\begin{aligned} [x^\mu] &= -1, & [\partial_\mu] &= 1, & [d^4x] &= -4, \\ [\mathcal{L}_0] &= 4, & 2 + 2[\phi] &= 4, & [\phi] &= 1, \quad [m] = 1. \end{aligned} \tag{1.14}$$

Both terms of the Lagrangian therefore have dimension four. A quadratic Lagrangian gives a linear field equation, whose Fourier modes do not scatter off one another; the interaction $-\lambda\phi^4/4!$ of Lecture 2 has $[\lambda] = 0$ in four space-time dimensions. Lorentz invariance alone does not define a model: we have also chosen a real scalar, its mass, locality, a positive kinetic term and a quadratic, free action. We now quantise this model.

1.3 Translation symmetry and the energy–momentum tensor

The action does not depend explicitly on the origin of time or position. Shifting a complete field configuration by a constant four-vector therefore leaves its action unchanged, up to boundary terms, and Noether's theorem turns this statement into a local energy–momentum conservation law. Energy conservation is thus a consequence of the time-translation symmetry of the action, not a second postulate; the conserved charge will become the quantum Hamiltonian.

The Noether current. Let ϵ^ν be a constant infinitesimal displacement. At a fixed coordinate x, the actively shifted configuration is

$$\begin{aligned} \phi'(x) &= \phi(x-\epsilon) = \phi(x) - \epsilon^\nu \partial_\nu \phi(x) + \mathcal{O}(\epsilon^2), \\ \delta\phi &= -\epsilon^\nu \partial_\nu \phi, \qquad \delta\mathcal{L}_0 = -\epsilon^\nu \partial_\nu \mathcal{L}_0 = \partial_\mu B^\mu, \\ B^\mu &= -\epsilon^\nu \delta^\mu{}_\nu \mathcal{L}_0. \end{aligned} \tag{1.15}$$

Here $\delta^\mu{}_\nu$ is the unit tensor and $\delta\phi$ the variation of the field. On shell, (1.11) gives a second expression for the same change:

$$\delta\mathcal{L}_0 = \partial_\mu \big(K^\mu \delta\phi\big), \qquad K^\mu \equiv \frac{\partial \mathcal{L}_0}{\partial(\partial_\mu \phi)} = \partial^\mu \phi. \tag{1.16}$$

Subtracting the two expressions, we obtain the current and its conservation law:

$$j^\mu = K^\mu \delta\phi - B^\mu = -\epsilon^\nu \big(K^\mu \partial_\nu \phi - \delta^\mu{}_\nu \mathcal{L}_0\big), \qquad \partial_\mu j^\mu = 0. \tag{1.17}$$

Since the four components of ϵ^ν are arbitrary, $j^\mu = -\epsilon^\nu T^\mu{}_\nu$ contains four conserved currents, one for each ν:

Energy–momentum tensor.

$$T^\mu{}_\nu = \partial^\mu\phi\,\partial_\nu\phi - \delta^\mu{}_\nu \mathcal{L}_0, \qquad T^{\mu\nu} = \partial^\mu\phi\,\partial^\nu\phi - g^{\mu\nu}\mathcal{L}_0, \qquad \partial_\mu T^{\mu\nu} = 0. \tag{1.18}$$

The minus sign in $j^\mu = -\epsilon^\nu T^\mu{}_\nu$ follows from our convention $\delta\phi = -\epsilon\cdot\partial\phi$; with it, T^{00} is positive (below).

A direct check. Conservation can be verified without invoking the theorem a second time:

$$\begin{aligned} \partial_\mu T^\mu{}_\nu &= (\Box\phi)\,\partial_\nu\phi + \partial^\mu\phi\,\partial_\mu\partial_\nu\phi - \partial_\nu\mathcal{L}_0, \\ \partial_\nu \mathcal{L}_0 &= \partial^\rho\phi\,\partial_\nu\partial_\rho\phi - m^2\phi\,\partial_\nu\phi, \\ \partial_\mu T^\mu{}_\nu &= \underbrace{(\Box\phi + m^2\phi)}_{0}\,\partial_\nu\phi = 0. \end{aligned} \tag{1.19}$$

The mixed second-derivative terms cancel because partial derivatives commute. The result holds on solutions of the field equation, not for an arbitrary field configuration.

From local conservation to conserved charges. For each fixed ν the conservation law is a continuity equation, with density $T^{0\nu}$ and current density $T^{i\nu}$. Integrated over a spatial volume V, it reads

$$\begin{aligned} &\partial_0 T^{0\nu} + \partial_i T^{i\nu} = 0, \\ &\frac{d}{dt}\int_V d^3x\, T^{0\nu} = -\int_V d^3x\, \partial_i T^{i\nu} = -\oint_{\partial V} dS\, n_i T^{i\nu}, \end{aligned} \tag{1.20}$$

where dS is the area element of the boundary ∂V and $\vec{n}$ its outward unit normal (Figure 1.4a). The first index of $T^{\mu\nu}$ distinguishes density ($\mu = 0$) from flux ($\mu = i$); the second labels the conserved quantity that is transported. The local law does not keep the content of a finite region fixed: it says that every change is accounted for by the flux through the boundary. If the flux vanishes at spatial infinity, or if opposite faces cancel under periodic boundary conditions, then (Figure 1.4b)

$$P^\nu = \int d^3x\, T^{0\nu}, \qquad \frac{dP^\nu}{dt} = 0. \tag{1.21}$$

Report Problem 1 reconstructs this argument together with its boundary assumptions. The relevant components are

$$\begin{aligned} T^{00} &= \dot\phi^2 - \mathcal{L}_0 = \tfrac{1}{2}\dot\phi^2 + \tfrac{1}{2}(\nabla\phi)^2 + \tfrac{1}{2}m^2\phi^2, \\ T^{0i} &= \partial^0\phi\,\partial^i\phi = -\dot\phi\,\partial_i\phi, \\ H_0 = P^0 &= \int d^3x\, T^{00}, \qquad \vec{P} = -\int d^3x\, \dot\phi\,\nabla\phi. \end{aligned} \tag{1.22}$$

For a real classical field and $m^2 > 0$ every term in H_0 is non-negative: the gradient term is the energy cost of spatial variation, and the mass term is a local restoring term. This positivity is a property of the chosen model.

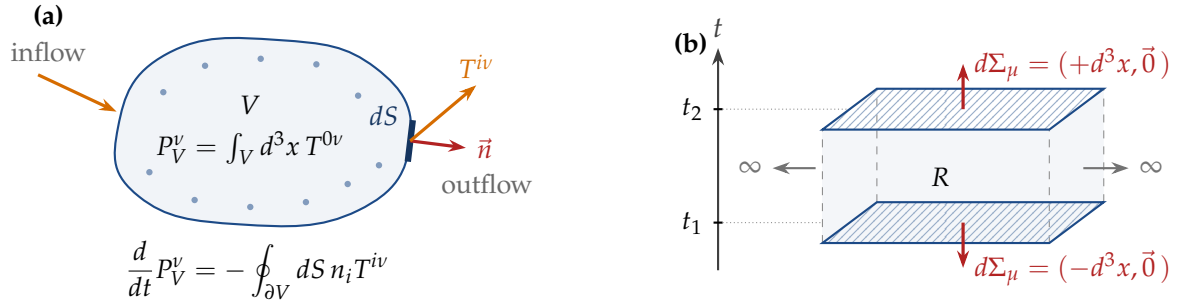


Figure 1.4: Local conservation of energy and momentum. (a) For each ν, $T^{0\nu}$ is the density of P^ν (dots) and $T^{i\nu}$ its current density. Only the component $n_i T^{i\nu}$ normal to the boundary, with outward normal $\vec{n}$ and area element dS, changes the content P^ν_V of the volume V, (1.20): outflow decreases it, inflow increases it. (b) Integrated over the slab R between t_1 and t_2, $\int_R d^4x\, \partial_\mu T^{\mu\nu} = \oint_{\partial R} d\Sigma_\mu\, T^{\mu\nu} = P^\nu(t_2) - P^\nu(t_1)$, because the fields vanish on the lateral surfaces at spatial infinity; hence P^ν is conserved. The oriented surface elements $d\Sigma_\mu$ are those of the boundary term in (1.9).

1.4 Hamiltonian form and canonical quantisation

We now bring the classical theory into the Hamiltonian form, which ordinary quantum mechanics knows how to quantise.

Canonical momentum and Hamiltonian. The momentum conjugate to the field value at the point $\vec{x}$ is a density:

$$\begin{aligned} \pi(t,\vec{x}) &\equiv \frac{\partial \mathcal{L}_0}{\partial \dot{\phi}(t,\vec{x})} = \dot{\phi}(t,\vec{x}), \\ \mathcal{H}_0 &= \pi\dot{\phi} - \mathcal{L}_0 = \tfrac{1}{2}\pi^2 + \tfrac{1}{2}(\nabla\phi)^2 + \tfrac{1}{2}m^2\phi^2, \\ H_0[\phi,\pi] &= \int d^3x\, \mathcal{H}_0. \end{aligned} \tag{1.23}$$

The field momentum π is conjugate to the field value at one point, whereas the total momentum P^i of (1.22) generates translations of the entire field configuration. The dimensions are $[\pi] = 2$, $[\mathcal{H}_0] = 4$ and $[H_0] = 1$.

Functional derivatives and Hamilton's equations. The functional derivative is defined by the first variation of a functional $F[\phi,\pi]$:

$$\delta F = \int d^3x \left[\frac{\delta F}{\delta\phi(\vec{x})}\,\delta\phi(\vec{x}) + \frac{\delta F}{\delta\pi(\vec{x})}\,\delta\pi(\vec{x}) \right]. \tag{1.24}$$

For the Hamiltonian we integrate the gradient term by parts in space:

$$\begin{aligned} \delta H_0 &= \int d^3x \left[\pi\,\delta\pi + \nabla\phi\cdot\nabla\delta\phi + m^2\phi\,\delta\phi \right] \\ &= \int d^3x \left[\pi\,\delta\pi + (-\nabla^2\phi + m^2\phi)\,\delta\phi \right], \\ \frac{\delta H_0}{\delta\pi} &= \pi, \qquad \frac{\delta H_0}{\delta\phi} = -\nabla^2\phi + m^2\phi. \end{aligned} \tag{1.25}$$

The discarded surface term vanishes under the same conditions as those used for the conserved charges. Hamilton's equations then give

$$\begin{aligned} \dot{\phi} &= \frac{\delta H_0}{\delta\pi} = \pi, \\ \dot{\pi} &= -\frac{\delta H_0}{\delta\phi} = \nabla^2\phi - m^2\phi, \\ \ddot{\phi} - \nabla^2\phi + m^2\phi &= 0 : \end{aligned} \tag{1.26}$$

the Hamiltonian and the Lagrangian descriptions give the same dynamics. The choice of a time slice hides the manifest Lorentz covariance but does not change the theory; we shall check the relativistic properties, including the normalisation of the states, after quantisation.

One quantum harmonic oscillator. The algebra that will recur at every momentum is that of the harmonic oscillator. We use a mechanical coordinate q and its momentum p_q, with the oscillator mass set to one (p_q is unrelated to the mode label $\vec{p}$):

$$[q, p_q] = i, \qquad H_{\text{osc}} = \tfrac{1}{2}p_q^2 + \tfrac{1}{2}\omega^2 q^2, \qquad \omega > 0. \tag{1.27}$$

The ladder operators and their inverse are

$$\begin{aligned} b &= \sqrt{\frac{\omega}{2}}\,q + \frac{i}{\sqrt{2\omega}}\,p_q, & b^\dagger &= \sqrt{\frac{\omega}{2}}\,q - \frac{i}{\sqrt{2\omega}}\,p_q, \\ q &= \frac{b+b^\dagger}{\sqrt{2\omega}}, & p_q &= -i\sqrt{\frac{\omega}{2}}\,(b - b^\dagger). \end{aligned} \tag{1.28}$$

With $A = \sqrt{\omega/2}$ and $B = 1/\sqrt{2\omega}$,

$$\begin{aligned} \left[b, b^\dagger\right] &= -iAB\ \left[q, p_q\right] + iAB\ \left[p_q, q\right] = 2AB = 1, \\ H_{\text{osc}} &= \frac{\omega}{4}\Big[-\left(b - b^\dagger\right)^2 + \left(b + b^\dagger\right)^2\Big] = \frac{\omega}{2}\left(bb^\dagger + b^\dagger b\right) = \omega\Big(b^\dagger b + \tfrac{1}{2}\Big); \end{aligned} \tag{1.29}$$

the terms b^2 and $(b^\dagger)^2$ cancel. With the number operator $N = b^\dagger b$,

$$\begin{aligned} &\left[N, b^\dagger\right] = b^\dagger, && [N, b] = -b, && \left[H_{\text{osc}}, b^\dagger\right] = \omega\, b^\dagger, \\ &b|0\rangle_{\text{osc}} = 0, && |n\rangle_{\text{osc}} = \frac{(b^\dagger)^n}{\sqrt{n!}}|0\rangle_{\text{osc}}, && H_{\text{osc}}|n\rangle_{\text{osc}} = \omega\big(n + \tfrac{1}{2}\big)|n\rangle_{\text{osc}}. \end{aligned} \tag{1.30}$$

The normalisation follows from $b\,(b^\dagger)^n = (b^\dagger)^n b + n\,(b^\dagger)^{n-1}$ by induction; the spectrum is shown in Figure 1.5. In the Heisenberg picture, $O_H(t) = e^{iHt} O_S\, e^{-iHt}$,

$$\begin{aligned} &\dot{O}_H = i\,[H, O_H]\,, && \dot{b}_H = -i\omega\, b_H, \\ &b_H(t) = b\, e^{-i\omega t}, && q_H(t) = \frac{b\, e^{-i\omega t} + b^\dagger e^{i\omega t}}{\sqrt{2\omega}}. \end{aligned} \tag{1.31}$$

Both signs of the frequency belong to an oscillator of positive energy: the term $e^{+i\omega t}$ accompanies the operator that raises the energy. We shall repeat this calculation with $\omega = E_{\vec{p}}$; the field is not one oscillator, however, but a continuum of them, labelled by the momentum.

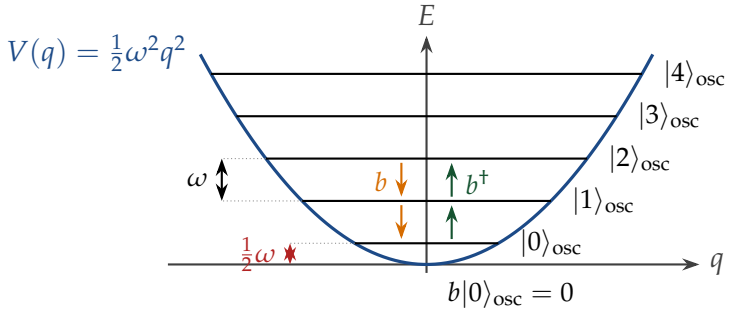


Figure 1.5: Spectrum of the oscillator (1.27): equally spaced levels $E_n = (n + \frac{1}{2})\omega$, raised by $b^\dagger$ and lowered by b. The ladder ends at the ground state, annihilated by b, whose energy $\frac{1}{2}\omega$ is the zero-point energy. The free field repeats this ladder for every momentum $\vec{p}$, with $\omega = E_{\vec{p}}$ (Section 1.5).

From a lattice to field commutators. Where does the Dirac δ-function in the field commutators come from? We temporarily replace space by cells of volume ℓ^3, with the field value $\phi_n(t)$ at site n (Figure 1.6). The discretised Lagrangian is

$$\begin{aligned} L_\ell &= \sum_n \ell^3 \left[\tfrac{1}{2}\dot{\phi}_n^2 - \tfrac{1}{2}\sum_{i=1}^{3}\left(\frac{\phi_{n+\hat{\imath}} - \phi_n}{\ell}\right)^2 - \tfrac{1}{2}m^2\phi_n^2\right], \\ p_n &= \frac{\partial L_\ell}{\partial \dot{\phi}_n} = \ell^3\dot{\phi}_n = \ell^3\pi_n, \qquad [\phi_n, p_k] = i\delta_{nk}, \\ [\phi_n, \pi_k] &= \frac{i}{\ell^3}\,\delta_{nk}. \end{aligned} \tag{1.32}$$

The cell volume is essential: p_n is an ordinary canonical momentum, whereas π_n approximates a momentum *density*. The gradient term couples neighbouring sites like springs. The continuum limit follows from the defining property of the δ-function acting on a test function f:

$$\sum_k \ell^3 \,\frac{\delta_{nk}}{\ell^3}\, f(\vec{x}_k) = f(\vec{x}_n),$$
$$\int d^3y\, \delta^{(3)}(\vec{x}-\vec{y})\, f(\vec{y}) = f(\vec{x}),$$
$$\frac{\delta_{nk}}{\ell^3} \longrightarrow \delta^{(3)}(\vec{x}-\vec{y}). \tag{1.33}$$

Hence the quantum fields obey the equal-time commutation relations:

Equal-time commutation relations.

$$[\phi(t,\vec{x}),\pi(t,\vec{y})] = i\,\delta^{(3)}(\vec{x}-\vec{y}), \qquad [\phi(t,\vec{x}),\phi(t,\vec{y})] = [\pi(t,\vec{x}),\pi(t,\vec{y})] = 0. \tag{1.34}$$

From now on ϕ and π are operators; we do not write hats. They are operator-valued distributions: integrated against smooth functions they give well-defined operators, while products at one point need a regulator or an ordering prescription, which we specify where they occur.

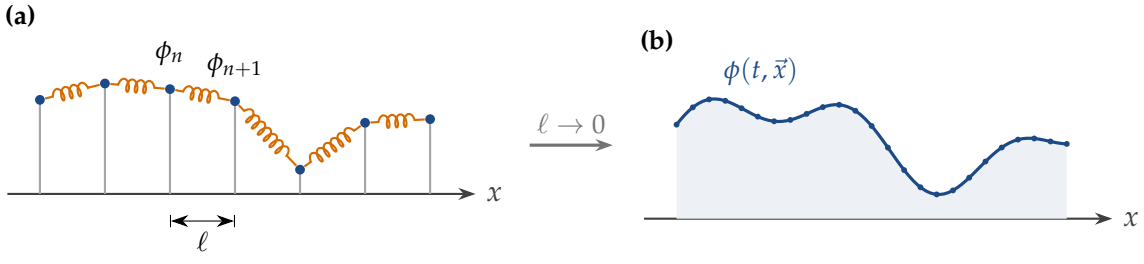


Figure 1.6: (a) The field on a spatial lattice of spacing ℓ (one direction shown): one degree of freedom ϕ_n per cell of volume ℓ^3, coupled to its neighbours by the gradient term of (1.32), like masses on springs. The canonical momentum of a site is $p_n = \ell^3\pi_n$, so that $[\phi_n,\pi_k] = i\delta_{nk}/\ell^3$. (b) As $\ell \to 0$ the site values become the field $\phi(t,\vec{x})$, sums $\sum_n \ell^3$ become integrals $\int d^3x$, and δ_{nk}/ℓ^3 becomes $\delta^{(3)}(\vec{x}-\vec{y})$, (1.33).

1.5 Mode expansion and the ladder algebra

We now diagonalise the free dynamics and identify the oscillators. With the spatial Fourier pair

$$\phi(t,\vec{x}) = \int \frac{d^3p}{(2\pi)^3}\, e^{i\vec{p}\cdot\vec{x}}\,\tilde{\phi}(t,\vec{p}), \qquad \tilde{\phi}(t,\vec{p}) = \int d^3x\, e^{-i\vec{p}\cdot\vec{x}}\,\phi(t,\vec{x}),$$
$$\int d^3x\, e^{i(\vec{p}-\vec{q})\cdot\vec{x}} = (2\pi)^3\delta^{(3)}(\vec{p}-\vec{q}), \tag{1.35}$$

the Klein–Gordon equation becomes

$$0 = \int \frac{d^3p}{(2\pi)^3}\, e^{i\vec{p}\cdot\vec{x}}\,\left[\partial_t^2 + |\vec{p}|^2 + m^2\right]\tilde{\phi}(t,\vec{p}),$$
$$0 = \ddot{\tilde{\phi}}(t,\vec{p}) + E_{\vec{p}}^2\,\tilde{\phi}(t,\vec{p}). \tag{1.36}$$

The Fourier waves are independent, so the coefficient vanishes for each $\vec{p}$: every Fourier mode is a harmonic oscillator of frequency $E_{\vec{p}}$. Reality of the classical field, or Hermiticity of the quantum field, requires $\tilde{\phi}(t,-\vec{p}) = \tilde{\phi}(t,\vec{p})^\dagger$. Guided by the oscillator (1.31), we expand

$$\begin{aligned}\phi(x) &= \int \frac{d^3p}{(2\pi)^3\sqrt{2E_{\vec{p}}}}\Big[\tilde{a}_{\vec{p}}\,e^{-ip\cdot x} + \tilde{a}^\dagger_{\vec{p}}\,e^{ip\cdot x}\Big],\\ \pi(x) = \dot{\phi}(x) &= -i\int \frac{d^3p}{(2\pi)^3}\sqrt{\frac{E_{\vec{p}}}{2}}\Big[\tilde{a}_{\vec{p}}\,e^{-ip\cdot x} - \tilde{a}^\dagger_{\vec{p}}\,e^{ip\cdot x}\Big], \qquad p^0 = E_{\vec{p}}.\end{aligned} \tag{1.37}$$

The tilde distinguishes this oscillator normalisation, used in the student assignment, from the covariant operators $a_{\vec{p}}$ introduced below. Each mode solves the field equation; the positive-frequency term is the one proportional to $e^{-iE_{\vec{p}}t}$, and the two terms together make $\phi^\dagger = \phi$. The form of the expansion is guessed from the oscillator, but the commutators of $\tilde{a}_{\vec{p}}$ are not guessed: the canonical algebra (1.34) determines them, including every factor of $2E_{\vec{p}}$ and 2π.

Inverting the expansion at fixed time. To shorten the calculation we write $A_{\vec{p}} = \sqrt{E_{\vec{p}}/2}$ and $B_{\vec{p}} = 1/\sqrt{2E_{\vec{p}}}$. At $t=0$, (1.35) gives

$$\begin{aligned}\int d^3x\, e^{-i\vec{p}\cdot\vec{x}}\,\phi(0,\vec{x}) &= \frac{\tilde{a}_{\vec{p}} + \tilde{a}^\dagger_{-\vec{p}}}{\sqrt{2E_{\vec{p}}}},\\ \int d^3x\, e^{-i\vec{p}\cdot\vec{x}}\,\pi(0,\vec{x}) &= -i\sqrt{\frac{E_{\vec{p}}}{2}}\,(\tilde{a}_{\vec{p}} - \tilde{a}^\dagger_{-\vec{p}}).\end{aligned} \tag{1.38}$$

Multiplying the first relation by $A_{\vec{p}}$ and the second by $iB_{\vec{p}}$, and adding them, removes the creation operator:

$$\begin{aligned}\tilde{a}_{\vec{p}} &= \int d^3x\, e^{-i\vec{p}\cdot\vec{x}}\,\big[A_{\vec{p}}\,\phi(0,\vec{x}) + iB_{\vec{p}}\,\pi(0,\vec{x})\big],\\ \tilde{a}^\dagger_{\vec{q}} &= \int d^3y\, e^{i\vec{q}\cdot\vec{y}}\,\big[A_{\vec{q}}\,\phi(0,\vec{y}) - iB_{\vec{q}}\,\pi(0,\vec{y})\big].\end{aligned} \tag{1.39}$$

Their commutator now follows from (1.34); all fields in the calculation are at $t=0$:

$$\begin{aligned}\Big[\tilde{a}_{\vec{p}}, \tilde{a}^\dagger_{\vec{q}}\Big] &= \int d^3x\, d^3y\; e^{-i\vec{p}\cdot\vec{x}+i\vec{q}\cdot\vec{y}}\\ &\qquad\times\Big[-iA_{\vec{p}}B_{\vec{q}}\,[\phi(\vec{x}),\pi(\vec{y})] + iB_{\vec{p}}A_{\vec{q}}\,[\pi(\vec{x}),\phi(\vec{y})]\Big]\\ &= \big(A_{\vec{p}}B_{\vec{q}} + B_{\vec{p}}A_{\vec{q}}\big)\int d^3x\, e^{i(\vec{q}-\vec{p})\cdot\vec{x}}\\ &= \frac{1}{2}\left(\sqrt{\frac{E_{\vec{p}}}{E_{\vec{q}}}} + \sqrt{\frac{E_{\vec{q}}}{E_{\vec{p}}}}\right)(2\pi)^3\delta^{(3)}(\vec{p}-\vec{q})\\ &= (2\pi)^3\delta^{(3)}(\vec{p}-\vec{q}).\end{aligned} \tag{1.40}$$

In the last step the δ-function sets $E_{\vec{q}} = E_{\vec{p}}$. Likewise,

$$\big[\tilde{a}_{\vec{p}}, \tilde{a}_{\vec{q}}\big] = \big(-A_{\vec{p}}B_{\vec{q}} + B_{\vec{p}}A_{\vec{q}}\big)(2\pi)^3\delta^{(3)}(\vec{p}+\vec{q}) = 0, \qquad \Big[\tilde{a}^\dagger_{\vec{p}}, \tilde{a}^\dagger_{\vec{q}}\Big] = 0, \tag{1.41}$$

where now the δ-function sets $\vec{q} = -\vec{p}$, which again makes the energies equal.

The covariant normalisation. Throughout these lectures we use the operators and the measure

$$a_{\vec{p}} \equiv \sqrt{2E_{\vec{p}}}\,\tilde{a}_{\vec{p}}, \qquad d\Pi_{\vec{p}} \equiv \frac{d^3p}{(2\pi)^3\,2E_{\vec{p}}}. \tag{1.42}$$

The same field and the same algebra then read

Mode expansion and ladder algebra.

$$\phi(x) = \int d\Pi_{\vec{p}}\left[a_{\vec{p}}\,e^{-ip\cdot x} + a^\dagger_{\vec{p}}\,e^{ip\cdot x}\right], \qquad p^0 = E_{\vec{p}}, \tag{1.43}$$

$$\left[a_{\vec{p}}, a^\dagger_{\vec{q}}\right] = (2\pi)^3\,2E_{\vec{p}}\,\delta^{(3)}(\vec{p}-\vec{q}), \qquad \left[a_{\vec{p}}, a_{\vec{q}}\right] = \left[a^\dagger_{\vec{p}}, a^\dagger_{\vec{q}}\right] = 0. \tag{1.44}$$

The conjugate momentum is

$$\pi(x) = \int d\Pi_{\vec{p}}\,(-iE_{\vec{p}})\left[a_{\vec{p}}\,e^{-ip\cdot x} - a^\dagger_{\vec{p}}\,e^{ip\cdot x}\right], \tag{1.45}$$

and at a general time the inversion reads

$$\begin{aligned} a_{\vec{p}} &= \int d^3x\, e^{ip\cdot x}\left[E_{\vec{p}}\,\phi(x) + i\pi(x)\right], \\ a^\dagger_{\vec{p}} &= \int d^3x\, e^{-ip\cdot x}\left[E_{\vec{p}}\,\phi(x) - i\pi(x)\right]. \end{aligned} \tag{1.46}$$

The operators on the left are time independent: the phases remove the free oscillation of the fields on the right, and all integrals are over one time slice.

Convention (normalisation of the mode operators). These lectures use the covariant operators $a_{\vec{p}}$ of the Lund lecture notes. The student assignment, like Peskin and Schroeder, uses the oscillator-normalised operators, written here with a tilde (the assignment and Peskin–Schroeder write them without it):

$$\tilde{a}_{\vec{p}} = \frac{a_{\vec{p}}}{\sqrt{2E_{\vec{p}}}}, \qquad \left[\tilde{a}_{\vec{p}}, \tilde{a}^\dagger_{\vec{q}}\right] = (2\pi)^3\delta^{(3)}(\vec{p}-\vec{q}), \qquad d\Pi_{\vec{p}} = \frac{d^3p}{(2\pi)^3\,2E_{\vec{p}}} \quad \text{(Lund notes: } \widetilde{dp}\text{)}. \tag{1.47}$$

The measure, the operators, the commutator and the states change together, never one symbol in one formula. In the two normalisations the field and the one-particle states read

$$\phi(x) = \int d\Pi_{\vec{p}}\left[a_{\vec{p}}\,e^{-ip\cdot x} + \text{h.c.}\right] = \int \frac{d^3p}{(2\pi)^3\sqrt{2E_{\vec{p}}}}\left[\tilde{a}_{\vec{p}}\,e^{-ip\cdot x} + \text{h.c.}\right], \tag{1.48}$$

$$|\vec{p}\rangle = a^\dagger_{\vec{p}}|0\rangle = \sqrt{2E_{\vec{p}}}\,\tilde{a}^\dagger_{\vec{p}}|0\rangle, \qquad \langle\vec{q}|\vec{p}\rangle = (2\pi)^3\,2E_{\vec{p}}\,\delta^{(3)}(\vec{p}-\vec{q}) = 2E_{\vec{p}}\,\langle 0|\tilde{a}_{\vec{q}}\,\tilde{a}^\dagger_{\vec{p}}|0\rangle. \tag{1.49}$$

The relativistically normalised state $|\vec{p}\rangle$ is the same in all three sources: the assignment (Problem 2) and Peskin–Schroeder define it as $\sqrt{2E_{\vec{p}}}\,\tilde{a}^\dagger_{\vec{p}}|0\rangle$. Bilinears agree as well, for example $\int d\Pi_{\vec{p}}\,a^\dagger_{\vec{p}}a_{\vec{p}} = \int \frac{d^3p}{(2\pi)^3}\,\tilde{a}^\dagger_{\vec{p}}\tilde{a}_{\vec{p}}$.

As a direct check we insert (1.43), (1.45) and (1.44) into the equal-time commutator. Only the

mixed commutators contribute, and the $\vec{q}$-integration sets $\vec{q}=\vec{p}$; with $\vec{r}=\vec{x}-\vec{y}$,

$$\begin{aligned}[\phi(t,\vec{x}),\pi(t,\vec{y})] &= \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}}\,(-iE_{\vec{q}})\Big(-\Big[a_{\vec{p}},a^\dagger_{\vec{q}}\Big]\; e^{-ip\cdot x+iq\cdot y}+\Big[a^\dagger_{\vec{p}},a_{\vec{q}}\Big]\; e^{ip\cdot x-iq\cdot y}\Big)\\ &= \int d\Pi_{\vec{p}}\; iE_{\vec{p}}\left(e^{-ip\cdot(x-y)}+e^{ip\cdot(x-y)}\right)=\frac{i}{2}\int\frac{d^3p}{(2\pi)^3}\left[e^{i\vec{p}\cdot\vec{r}}+e^{-i\vec{p}\cdot\vec{r}}\right]\\ &= i\,\delta^{(3)}(\vec{r})\,: \end{aligned} \tag{1.50}$$

each of the two terms supplies one half of the result. The dimensions are consistent as well: $[\phi]=1$, $[d\Pi_{\vec{p}}]=2$ and $[a_{\vec{p}}]=-1$.

The quantum equation of motion. Finally we verify that the canonical algebra and the Hamiltonian reproduce the original dynamics. For a time-independent Hamiltonian,

$$\begin{gathered}\phi_H(t,\vec{x})=e^{iH_0t}\phi_S(\vec{x})\,e^{-iH_0t},\qquad \pi_H(t,\vec{x})=e^{iH_0t}\pi_S(\vec{x})\,e^{-iH_0t},\\ \dot{O}_H=i\,[H_0,O_H]\,.\end{gathered} \tag{1.51}$$

This unitary transformation preserves the equal-time algebra. We suppress the subscript H and use $[AB,C]=A\,[B,C]+[A,C]\,B$. In the first equation only π^2 contributes:

$$\begin{aligned}[H_0,\phi(\vec{x})] &= \tfrac{1}{2}\int d^3y\;\left[\pi(\vec{y})^2,\phi(\vec{x})\right]\\ &= \tfrac{1}{2}\int d^3y\,\Big[-i\pi(\vec{y})\,\delta^{(3)}(\vec{y}-\vec{x})-i\,\delta^{(3)}(\vec{y}-\vec{x})\,\pi(\vec{y})\Big]\\ &= -i\pi(\vec{x}),\qquad \dot{\phi}=i\,[H_0,\phi]=\pi. \end{aligned} \tag{1.52}$$

In the second, the gradient and mass terms contribute:

$$\begin{aligned}[H_0,\pi(\vec{x})] &= i\int d^3y\,\Big[\partial_{y^i}\phi(\vec{y})\,\partial_{y^i}\delta^{(3)}(\vec{y}-\vec{x})+m^2\phi(\vec{y})\,\delta^{(3)}(\vec{y}-\vec{x})\Big]\\ &= i\big[-\nabla^2\phi(\vec{x})+m^2\phi(\vec{x})\big],\\ \dot{\pi}=i\,[H_0,\pi] &= \nabla^2\phi-m^2\phi,\qquad (\Box+m^2)\,\phi_H=0. \end{aligned} \tag{1.53}$$

The minus sign in the first term comes from integrating the derivative of the δ-function by parts. A regulator is understood for the local quadratic operator products, and a constant subtraction from H_0 leaves these commutators unchanged.

Three independent descriptions now agree: the variation of the action, the classical Hamilton equations and the quantum Heisenberg equations; the mode expansion solves this same equation. The field equation alone does not yet give a particle interpretation. For that we must find what the Hamiltonian does to the states created by the mode operators.

1.6 The Hamiltonian counts quanta

We now calculate the energy, rather than infer it from the frequency. We insert (1.43) into $H_0=\frac{1}{2}\int d^3x\,[\pi^2+(\nabla\phi)^2+m^2\phi^2]$; since H_0 is conserved, we may evaluate it at $t=0$, where

$$\begin{aligned}\phi &= \int d\Pi_{\vec{p}}\left(a_{\vec{p}}\,e^{i\vec{p}\cdot\vec{x}}+a^\dagger_{\vec{p}}\,e^{-i\vec{p}\cdot\vec{x}}\right),\\ \pi &= \int d\Pi_{\vec{p}}\,(-iE_{\vec{p}})\left(a_{\vec{p}}\,e^{i\vec{p}\cdot\vec{x}}-a^\dagger_{\vec{p}}\,e^{-i\vec{p}\cdot\vec{x}}\right),\\ \nabla\phi &= \int d\Pi_{\vec{p}}\,(i\vec{p})\left(a_{\vec{p}}\,e^{i\vec{p}\cdot\vec{x}}-a^\dagger_{\vec{p}}\,e^{-i\vec{p}\cdot\vec{x}}\right). \end{aligned} \tag{1.54}$$

Every product contains two independent momentum integrations, over $\vec{p}$ and $\vec{q}$. We collect equal operator orderings before performing the spatial integral (Table 1.1).

Term in H_0	$a_{\vec{p}}a_{\vec{q}}$ and $a^\dagger_{\vec{p}}a^\dagger_{\vec{q}}$	$a_{\vec{p}}a^\dagger_{\vec{q}}$ and $a^\dagger_{\vec{p}}a_{\vec{q}}$
π^2	$-E_{\vec{p}}E_{\vec{q}}$	$+E_{\vec{p}}E_{\vec{q}}$
$(\nabla\phi)^2$	$-\vec{p}\cdot\vec{q}$	$+\vec{p}\cdot\vec{q}$
$m^2\phi^2$	$+m^2$	$+m^2$
spatial integral	$(2\pi)^3\delta^{(3)}(\vec{p}+\vec{q})$	$(2\pi)^3\delta^{(3)}(\vec{p}-\vec{q})$

Table 1.1: Coefficients of the operator products in H_0 after inserting (1.54), and the δ-functions produced by the spatial integral of the phases. The column sums are C_{aa} and $C_{aa^\dagger}$ of (1.55).

The phases give $\delta^{(3)}(\vec{p}+\vec{q})$ for the first column and $\delta^{(3)}(\vec{p}-\vec{q})$ for the second, so the column sums C_{aa} and $C_{aa^\dagger}$ are needed only at $\vec{q}=-\vec{p}$ and at $\vec{q}=\vec{p}$, respectively:

$$\begin{aligned} C_{aa}\big|_{\vec{q}=-\vec{p}} &= -E_{\vec{p}}^2+|\vec{p}|^2+m^2 = 0, \\ C_{aa^\dagger}\big|_{\vec{q}=\vec{p}} &= E_{\vec{p}}^2+|\vec{p}|^2+m^2 = 2E_{\vec{p}}^2. \end{aligned} \tag{1.55}$$

The terms that would annihilate or create a pair of quanta cancel by the on-shell relation $E_{\vec{p}}^2=|\vec{p}|^2+m^2$. The surviving terms give

$$\begin{aligned} H_0 &= \tfrac{1}{2}\int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}}\; (2\pi)^3\delta^{(3)}(\vec{p}-\vec{q})\; 2E_{\vec{p}}^2\left(a_{\vec{p}}a^\dagger_{\vec{q}}+a^\dagger_{\vec{p}}a_{\vec{q}}\right) \\ &= \tfrac{1}{2}\int d\Pi_{\vec{p}}\; E_{\vec{p}}\left(a_{\vec{p}}a^\dagger_{\vec{p}}+a^\dagger_{\vec{p}}a_{\vec{p}}\right), \end{aligned} \tag{1.56}$$

where the factor $1/2E_{\vec{q}}$ in the measure $d\Pi_{\vec{q}}$ is essential in the last step.

Vacuum energy and normal ordering. The second form of (1.56) contains a common energy offset, which we now separate from the excitation energies. The free vacuum is defined by

$$a_{\vec{p}}|0\rangle = 0 \quad \text{for every } \vec{p}, \qquad \langle 0|0\rangle = 1. \tag{1.57}$$

Reordering the first term in (1.56) with (1.44),

$$\begin{aligned} a_{\vec{p}}a^\dagger_{\vec{p}} &= a^\dagger_{\vec{p}}a_{\vec{p}} + (2\pi)^3\, 2E_{\vec{p}}\,\delta^{(3)}(0), \\ H_0 &= \int d\Pi_{\vec{p}}\, E_{\vec{p}}\, a^\dagger_{\vec{p}}a_{\vec{p}} + E_0, \qquad E_0 = \tfrac{1}{2}\,\delta^{(3)}(0)\int d^3p\, E_{\vec{p}}. \end{aligned} \tag{1.58}$$

The symbol $\delta^{(3)}(0)$ has to be interpreted, not manipulated as an ordinary number. In a periodic box of volume V,

$$(2\pi)^3\delta^{(3)}(0) \longleftrightarrow V, \qquad E_0 = \frac{V}{2}\int\frac{d^3p}{(2\pi)^3}\,E_{\vec{p}} \longleftrightarrow \tfrac{1}{2}\sum_{\vec{p}} E_{\vec{p}}, \tag{1.59}$$

the sum of the zero-point energies $\frac{1}{2}E_{\vec{p}}$ of all the oscillators. A momentum cut-off $|\vec{p}|<\Lambda$ makes the sum finite; its leading behaviour for large Λ is

$$\begin{aligned} \frac{E_0}{V} &= \frac{1}{4\pi^2}\int_0^\Lambda dp\, p^2\sqrt{p^2+m^2}, \\ \frac{E_0}{V} &\sim \frac{1}{4\pi^2}\int_0^\Lambda dp\, p^3 = \frac{\Lambda^4}{16\pi^2} \qquad (\Lambda\gg m). \end{aligned} \tag{1.60}$$

The volume divergence and the high-momentum divergence have different origins, and neither changes an excitation energy after a common subtraction. Normal ordering, $:\cdots:$, places all creation operators to the left of all annihilation operators. Relative to the free vacuum,

$$H_0^{\text{sub}} \equiv :H_0: = H_0 - E_0 = \int d\Pi_{\vec{p}}\, E_{\vec{p}}\, a^\dagger_{\vec{p}} a_{\vec{p}}, \qquad H_0^{\text{sub}}|0\rangle = 0. \tag{1.61}$$

The superscript identifies the vacuum-subtracted energy; from now on we write H_0 for it. A constant commutes with every operator, so the Heisenberg equations are unchanged. This free-theory choice of the zero of energy does not settle the questions about vacuum energy that arise in gravity, or in the renormalisation of composite operators of an interacting theory.

1.7 Particles: energy, momentum and Fock space

The particle interpretation is established by the action of the operators on states. With (1.44) and the subtracted H_0,

$$\begin{aligned} \left[H_0, a^\dagger_{\vec{q}}\right] &= \int d\Pi_{\vec{p}}\, E_{\vec{p}}\, \left[a^\dagger_{\vec{p}} a_{\vec{p}}, a^\dagger_{\vec{q}}\right] \\ &= \int d\Pi_{\vec{p}}\, E_{\vec{p}}\, a^\dagger_{\vec{p}}\, (2\pi)^3\, 2E_{\vec{p}}\, \delta^{(3)}(\vec{p}-\vec{q}) = E_{\vec{q}}\, a^\dagger_{\vec{q}}, \\ \left[H_0, a_{\vec{q}}\right] &= -E_{\vec{q}}\, a_{\vec{q}}. \end{aligned} \tag{1.62}$$

For the state $|\vec{q}\rangle = a^\dagger_{\vec{q}}|0\rangle$ this gives

$$H_0|\vec{q}\rangle = \Big(\left[H_0, a^\dagger_{\vec{q}}\right] + a^\dagger_{\vec{q}} H_0 \Big)|0\rangle = E_{\vec{q}}|\vec{q}\rangle. \tag{1.63}$$

The quantum has the *positive* energy $E_{\vec{q}}$. This establishes, rather than assumes, the connection between the mode frequency and the energy.

Momentum. The Hermitian operator corresponding to the classical charge $\vec{P}$ of (1.22) is the symmetrised product

$$P^i = -\tfrac{1}{2} \int d^3x \left[\pi\, \partial_i \phi + (\partial_i \phi)\, \pi\right]. \tag{1.64}$$

For an explicit check we first expand the unsymmetrised $-\int d^3x\, \pi\, \partial_i\phi$ at $t = 0$, using (1.54):

$$\begin{aligned} -\int d^3x\, \pi\, \partial_i \phi = \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}}\, (2\pi)^3\, E_{\vec{p}}\, q^i \Big\{ &-\delta^{(3)}(\vec{p}+\vec{q})\, \left(a_{\vec{p}} a_{\vec{q}} + a^\dagger_{\vec{p}} a^\dagger_{\vec{q}}\right) \\ &+ \delta^{(3)}(\vec{p}-\vec{q})\, \left(a_{\vec{p}} a^\dagger_{\vec{q}} + a^\dagger_{\vec{p}} a_{\vec{q}}\right) \Big\}. \end{aligned} \tag{1.65}$$

The first term is proportional to $p^i\, a_{\vec{p}} a_{-\vec{p}}$ or its adjoint; it vanishes under $\vec{p} \to -\vec{p}$, since operators of the same type commute. The surviving term is Hermitian, and it is also what the symmetrised charge gives. Hence

$$\begin{aligned} P^i &= \tfrac{1}{2} \int d\Pi_{\vec{p}}\, p^i \left(a_{\vec{p}} a^\dagger_{\vec{p}} + a^\dagger_{\vec{p}} a_{\vec{p}}\right) = \int d\Pi_{\vec{p}}\, p^i\, a^\dagger_{\vec{p}} a_{\vec{p}}, \\ \left[P^i, a^\dagger_{\vec{q}}\right] &= q^i\, a^\dagger_{\vec{q}}, \qquad P^i|\vec{q}\rangle = q^i|\vec{q}\rangle. \end{aligned} \tag{1.66}$$

The possible vacuum term is an odd integral in a symmetric regulator and vanishes. Here p^i is the Cartesian component of the three-vector $\vec{p}$; the lower component $p_i = -p^i$ has the opposite sign. Collecting the four charges:

Energy and momentum of the quanta.

$$P^\mu = \int d\Pi_{\vec{p}}\; p^\mu\, a^\dagger_{\vec{p}} a_{\vec{p}}, \qquad P^\mu|\vec{p}\rangle = p^\mu|\vec{p}\rangle, \qquad p^2 = m^2. \tag{1.67}$$

A scalar field carries no spin index: its quanta are spin-zero particles of mass m. A full classification by representations of the Lorentz group is not needed here.

Fock space and Bose symmetry. Repeated creation generates the multiparticle states

$$|\vec{p}_1, \ldots, \vec{p}_n\rangle = a^\dagger_{\vec{p}_1} \cdots a^\dagger_{\vec{p}_n}|0\rangle. \tag{1.68}$$

Because the creation operators commute,

$$|\vec{p}, \vec{q}\rangle = |\vec{q}, \vec{p}\rangle : \tag{1.69}$$

the order of the labels does not distinguish physical states, and the quanta are bosons. With $a^\dagger_r \equiv a^\dagger_{\vec{p}_r}$ (only here), the commutator product rule gives the total four-momentum:

$$\begin{aligned} \left[P^\mu, a_1^\dagger \cdots a_n^\dagger\right] &= \sum_{r=1}^{n} a_1^\dagger \cdots \left[P^\mu, a_r^\dagger\right] \cdots a_n^\dagger = \left(\sum_{r=1}^{n} p_r^\mu\right) a_1^\dagger \cdots a_n^\dagger, \\ P^\mu|\vec{p}_1, \ldots, \vec{p}_n\rangle &= \left(\sum_{r=1}^{n} p_r^\mu\right)|\vec{p}_1, \ldots, \vec{p}_n\rangle. \end{aligned} \tag{1.70}$$

The particle-number operator is

$$\begin{aligned} N &= \int d\Pi_{\vec{p}}\, a^\dagger_{\vec{p}} a_{\vec{p}}, \\ \left[N, a^\dagger_{\vec{q}}\right] &= a^\dagger_{\vec{q}}, \qquad N|0\rangle = 0, \qquad N|\vec{p}_1, \ldots, \vec{p}_n\rangle = n|\vec{p}_1, \ldots, \vec{p}_n\rangle, \\ [H_0, N] &= 0. \end{aligned} \tag{1.71}$$

The last relation is a property of the free Hamiltonian. In an interacting real scalar theory particle number is not conserved: in ϕ^4 theory, for example, processes at higher energy can change it by an even number while conserving energy and momentum. The complete space of states is the bosonic Fock space

$$\mathcal{F} = \bigoplus_{n=0}^{\infty} \mathcal{F}_n, \qquad \mathcal{F}_n = \mathrm{Sym}\, \mathcal{F}_1^{\otimes n}, \tag{1.72}$$

where Sym keeps the states that are symmetric under exchange of the particle labels, and the sector $\mathcal{F}_0$ is spanned by the vacuum. In a finite box, an occupied single mode is normalised with $(b^\dagger_{\vec{p}})^n/\sqrt{n!}$ and the unit commutator $\left[b_{\vec{p}}, b^\dagger_{\vec{q}}\right] = \delta_{\vec{p}\vec{q}}$, exactly as for the oscillator (1.30); continuum momentum eigenstates carry δ-function normalisation instead.

Particles are quanta of the field. Particles are not added to the theory as a second kind of dynamical object. They are the discrete occupation quanta of the normal modes of the field: one oscillator ladder (Figure 1.5) for every momentum $\vec{p}$, with $\omega = E_{\vec{p}}$. For a real scalar field the single species is its own antiparticle; charged fields need additional structure, introduced later (the Dirac field in Lecture 3).

1.8 Relativistic normalisation and the field–particle link

We now fix the normalisation that propagators and scattering amplitudes require. For $|\vec{p}\rangle = a^\dagger_{\vec{p}}|0\rangle$ and $a_{\vec{q}}|0\rangle = 0$,

$$\begin{aligned}\langle\vec{q}|\vec{p}\rangle &= \langle 0|a_{\vec{q}}a^\dagger_{\vec{p}}|0\rangle = \langle 0|\left[a_{\vec{q}}, a^\dagger_{\vec{p}}\right]|0\rangle + \underbrace{\langle 0|a^\dagger_{\vec{p}}a_{\vec{q}}|0\rangle}_{0} \\ &= (2\pi)^3\, 2E_{\vec{p}}\,\delta^{(3)}(\vec{p}-\vec{q}).\end{aligned} \tag{1.73}$$

A momentum eigenstate has a δ-function norm; it is an idealisation of a normalisable wave packet. Formally,

$$\mathbb{1}_1 = \int d\Pi_{\vec{p}}\,|\vec{p}\rangle\langle\vec{p}| \tag{1.74}$$

is the unit operator on the one-particle sector $\mathcal{F}_1$ (on that sector only), since

$$\mathbb{1}_1|\vec{q}\rangle = \int \frac{d^3p}{(2\pi)^3\,2E_{\vec{p}}}\,|\vec{p}\rangle\,(2\pi)^3\,2E_{\vec{p}}\,\delta^{(3)}(\vec{p}-\vec{q}) = |\vec{q}\rangle. \tag{1.75}$$

Why this measure is relativistic. We start from a manifestly invariant four-dimensional integral, restricted to the future mass shell, the blue sheet in Figure 1.3a:

$$\begin{aligned}\int \frac{d^4p}{(2\pi)^3}\,\theta(p^0)\,\delta(p^2-m^2)\,F(p) &= \int \frac{d^3p}{(2\pi)^3}\int dp^0\,\theta(p^0)\,\delta\big((p^0)^2-E^2_{\vec{p}}\big)\,F(p^0,\vec{p}), \\ \delta\big((p^0)^2-E^2_{\vec{p}}\big) &= \frac{\delta(p^0-E_{\vec{p}})+\delta(p^0+E_{\vec{p}})}{2E_{\vec{p}}}, \\ \int \frac{d^4p}{(2\pi)^3}\,\theta(p^0)\,\delta(p^2-m^2)\,F(p) &= \int d\Pi_{\vec{p}}\,F(E_{\vec{p}},\vec{p}).\end{aligned} \tag{1.76}$$

Here θ is the step function; the restriction to positive energy is preserved by proper orthochronous Lorentz transformations. The factor $2E_{\vec{p}}$ in the norm of the states is the inverse of the factor in the measure, so the completeness relation (1.74) is frame independent. For a genuine one-particle wave packet,

$$|f\rangle = \int d\Pi_{\vec{p}}\,f(\vec{p})\,|\vec{p}\rangle, \qquad \langle f|f\rangle = \int d\Pi_{\vec{p}}\,|f(\vec{p})|^2; \tag{1.77}$$

choosing the last integral equal to one gives an ordinary normalised state.

The matrix element used in the scattering calculation. A free field changes the occupation by one unit, since it contains one creation or one annihilation operator. In particular,

$$\begin{aligned}\phi(x)|0\rangle &= \int d\Pi_{\vec{k}}\,e^{ik\cdot x}\,a^\dagger_{\vec{k}}|0\rangle = \int d\Pi_{\vec{k}}\,e^{ik\cdot x}\,|\vec{k}\rangle, \\ \langle 0|\phi(x)|\vec{p}\rangle &= \int d\Pi_{\vec{k}}\,e^{-ik\cdot x}\,\langle 0|a_{\vec{k}}a^\dagger_{\vec{p}}|0\rangle \\ &= \int d\Pi_{\vec{k}}\,e^{-ik\cdot x}\,(2\pi)^3\,2E_{\vec{k}}\,\delta^{(3)}(\vec{k}-\vec{p}) \\ &= e^{-ip\cdot x}.\end{aligned} \tag{1.78}$$

Similarly,

$$\langle \vec{p}|\phi(x)|0\rangle = e^{ip\cdot x}, \qquad \left[\phi(x), a^\dagger_{\vec{p}}\right] = e^{-ip\cdot x}, \qquad \left[a_{\vec{p}}, \phi(x)\right] = e^{ip\cdot x}, \tag{1.79}$$

where the right-hand sides multiply the unit operator. These plane-wave factors will attach incoming and outgoing particles to an interaction point; their normalisation was fixed by the preceding derivations.

Relativistic normalisation and the external-particle factor.

$$d\Pi_{\vec{p}} = \frac{d^3p}{(2\pi)^3\, 2E_{\vec{p}}}, \qquad \langle \vec{q}|\vec{p}\rangle = (2\pi)^3\, 2E_{\vec{p}}\, \delta^{(3)}(\vec{p}-\vec{q}), \qquad \langle 0|\phi(x)|\vec{p}\rangle = e^{-ip\cdot x}. \tag{1.80}$$

A point field acting on the vacuum, $\phi(x)|0\rangle$, is a distributional superposition of all momenta, with infinite norm. For a smooth spatial function $h(\vec{x})$ of compact support, the smeared operator

$$\phi_t(h) = \int d^3x\, h(\vec{x})\, \phi(t,\vec{x}) \tag{1.81}$$

creates a superposition of momenta weighted by the Fourier transform of h. This is the appropriate way to connect local field expressions to physical wave packets.

For the counting of two-particle states in Lecture 3 the same algebra gives, with $\langle \vec{k}_1, \vec{k}_2| = \langle 0|a_{\vec{k}_2} a_{\vec{k}_1}$ and $a_{\vec{k}_1}|0\rangle = 0$,

$$\begin{aligned}\langle \vec{k}_1, \vec{k}_2|\vec{p}_1, \vec{p}_2\rangle &= \langle 0|a_{\vec{k}_2}\Big(\left[a_{\vec{k}_1}, a^\dagger_{\vec{p}_1}\right]\, a^\dagger_{\vec{p}_2} + a^\dagger_{\vec{p}_1}\left[a_{\vec{k}_1}, a^\dagger_{\vec{p}_2}\right]\Big)|0\rangle \\ &= \langle \vec{k}_1|\vec{p}_1\rangle\langle \vec{k}_2|\vec{p}_2\rangle + \langle \vec{k}_1|\vec{p}_2\rangle\langle \vec{k}_2|\vec{p}_1\rangle,\end{aligned} \tag{1.82}$$

with the one-particle overlaps (1.73). The two terms are the two ways of pairing the annihilation operators with the creation operators. This is the origin of the exchange counting for identical particles, which will require care in Lecture 3.

1.9 Synthesis: what we have achieved

We have not yet described a collision. We have constructed and checked the quantum theory of the free particles that enter and leave one: they have positive relativistic energies, Bose symmetry and a fixed, invariant normalisation. These are the ingredients that a scattering calculation must use. The resulting identities are:

$$\begin{gathered}\mathcal{L}_0 = \tfrac{1}{2}\,\partial_\mu\phi\,\partial^\mu\phi - \tfrac{1}{2}m^2\phi^2, \qquad (\Box + m^2)\,\phi = 0, \\ d\Pi_{\vec{p}} = \frac{d^3p}{(2\pi)^3\,2E_{\vec{p}}}, \qquad E_{\vec{p}} = \sqrt{|\vec{p}|^2 + m^2}, \\ \phi(x) = \int d\Pi_{\vec{p}}\left(a_{\vec{p}}\, e^{-ip\cdot x} + a^\dagger_{\vec{p}}\, e^{ip\cdot x}\right), \\ \left[a_{\vec{p}}, a^\dagger_{\vec{q}}\right] = (2\pi)^3\, 2E_{\vec{p}}\, \delta^{(3)}(\vec{p}-\vec{q}), \qquad \left[a_{\vec{p}}, a_{\vec{q}}\right] = 0, \\ P^\mu = \int d\Pi_{\vec{p}}\; p^\mu\, a^\dagger_{\vec{p}} a_{\vec{p}}, \qquad a_{\vec{p}}|0\rangle = 0, \qquad P^\mu|0\rangle = 0, \\ |\vec{p}\rangle = a^\dagger_{\vec{p}}|0\rangle, \qquad P^\mu|\vec{p}\rangle = p^\mu|\vec{p}\rangle, \\ \langle \vec{q}|\vec{p}\rangle = (2\pi)^3\, 2E_{\vec{p}}\, \delta^{(3)}(\vec{p}-\vec{q}), \qquad \langle 0|\phi(x)|\vec{p}\rangle = e^{-ip\cdot x}.\end{gathered} \tag{1.83}$$

Here the energy is vacuum subtracted, and the field is a real, free, spin-zero field in four-dimensional flat space-time; these qualifications are part of the result.

Remark (three questions for the discussion).

(i) *Where did the particle energy come from?* The dispersion relation fixed a mode frequency; $\left[H_0, a^\dagger_{\vec{p}}\right] = E_{\vec{p}}\, a^\dagger_{\vec{p}}$ then proved that creating a quantum raises the energy by that amount.

(ii) *Why keep both frequency signs?* A Hermitian field needs both. The negative-frequency term multiplies a creation operator and creates a quantum of positive energy.

(iii) *What is still missing?* The commutator at unequal times and the propagator, and an interaction that couples the free states. Equal-time canonical quantisation alone gives neither the full causality argument nor a scattering amplitude; both are the subject of Lecture 2.

Summary

- Particle creation requires a Fock space of n-particle sectors; causality in quantum field theory concerns commutators of local observables (Lecture 2).
- The free real scalar field has $\mathcal{L}_0 = \frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - \frac{1}{2}m^2\phi^2$; the variational principle gives the Klein–Gordon equation, whose plane waves obey $p^2 = m^2$, $p^0 = \pm E_{\vec{p}}$.
- Noether's theorem for translations gives the conserved $T^{\mu\nu} = \partial^\mu\phi\,\partial^\nu\phi - g^{\mu\nu}\mathcal{L}_0$ and four-momentum $P^\nu = \int d^3x\, T^{0\nu}$, with $H_0 = P^0 \geq 0$.
- Canonical quantisation: $\pi = \dot{\phi}$ and $[\phi(t,\vec{x}), \pi(t,\vec{y})] = i\delta^{(3)}(\vec{x}-\vec{y})$ (the lattice explains the δ-function); every Fourier mode is an oscillator of frequency $E_{\vec{p}}$.
- Inverting the mode expansion gives $\left[a_{\vec{p}}, a^\dagger_{\vec{q}}\right] = (2\pi)^3\, 2E_{\vec{p}}\,\delta^{(3)}(\vec{p}-\vec{q})$ for $a_{\vec{p}} = \sqrt{2E_{\vec{p}}}\,\tilde{a}_{\vec{p}}$; measure, operators, commutator and states change together.
- H_0 contains no aa or $a^\dagger a^\dagger$ terms (on-shell cancellation). After the subtraction of the vacuum energy, $E_0/V \sim \Lambda^4/16\pi^2$, $P^\mu = \int d\Pi_{\vec{p}}\, p^\mu a^\dagger_{\vec{p}} a_{\vec{p}}$ with $P^0 = H_0$.
- $a^\dagger_{\vec{p}}$ creates a spin-zero boson of energy $E_{\vec{p}} > 0$ and momentum $\vec{p}$; multiparticle states are symmetric; particle number is conserved only in the free theory.
- $\langle\vec{q}|\vec{p}\rangle = (2\pi)^3\, 2E_{\vec{p}}\,\delta^{(3)}(\vec{p}-\vec{q})$ goes with the invariant measure $d\Pi_{\vec{p}}$; $\langle 0|\phi(x)|\vec{p}\rangle = e^{-ip\cdot x}$ is the external-particle factor of Lectures 2 and 3.

Report. Problem 1 (action, translation current and Hamiltonian; §§1.2–1.4) and Problem 2 (quantisation, mode algebra and one-particle interpretation; §§1.4–1.8) build on this lecture. State the boundary conditions and the normalisation you use; the convention box in §1.5 converts between the covariant operators of the lectures and the oscillator-normalised operators of the assignment.

Reading. Peskin–Schroeder, Sections 2.1–2.3; Lund lecture notes, Lectures 2–6; D. Tong, *Lectures on Quantum Field Theory*, Sections 1–2.

Next lecture: Causality and perturbative interactions: the two-point function and microcausality, the Feynman propagator, the Dyson series and Wick's theorem, and the tree amplitude of ϕ^4 theory.

Causality and perturbative interactions

Lecture 2

Causality and perturbative interactions

In this lecture

- The vacuum two-point function W from the ladder algebra, its Lorentz invariance, and an explicit proof that the field commutator $C(z) = W(z) - W(-z)$ vanishes at space-like separation.
- Causality needs antiparticles: for a complex field the commutator is the amplitude of a particle going from y to x minus that of an antiparticle going from x to y.
- The Feynman propagator D_F by contour integration, its Green equation, and the retarded response function G_R.
- The interaction picture, the Dyson series and Wick's theorem: time-ordered products reduced to propagators.
- The first complete interaction calculation: the tree amplitude $\mathcal{M} = -\lambda$ of $\phi\phi \to \phi\phi$ in ϕ^4 theory.

Builds on: Lecture 1 (free real scalar field, covariant ladder operators and Fock states, Noether currents).
In the book: Lund lecture notes, Lectures 7–8 and 15–16; book (R. Pasechnik, *Introduction to QFT*), Chapters 5 and 10–11.

How does a local quantum field produce a causal theory and a calculable scattering amplitude? Starting from the free field quantised in Lecture 1, we ask which correlations its vacuum contains, which of them govern a response, and how they enter an interaction calculation. The endpoint is the leading invariant amplitude for two identical scalar particles interacting through a weak quartic coupling,

$$\phi(p_1) + \phi(p_2) \longrightarrow \phi(k_1) + \phi(k_2), \qquad \mathcal{L}_{\text{int}} = -\frac{\lambda}{4!}\,\phi^4, \tag{2.1}$$

which Lecture 3 converts into a cross section. Each construction of the lecture supplies one step of the chain

$$\text{free field} \;\to\; \text{two-point functions} \;\to\; \text{time-ordered evolution} \;\to\; \text{contractions} \;\to\; \mathcal{M}.$$

All derivations are written out to support independent study.

Starting data. We use $\hbar = c = 1$, the metric $g_{\mu\nu} = \text{diag}(1, -1, -1, -1)$, $p \cdot x = p^0 t - \vec{p} \cdot \vec{x}$, and $m > 0$. With

$$E_{\vec{p}} = \sqrt{\vec{p}^{\,2} + m^2}, \qquad d\Pi_{\vec{p}} = \frac{d^3p}{(2\pi)^3\, 2E_{\vec{p}}}, \tag{2.2}$$

the free field, its ladder algebra and the vacuum are

$$\phi(x) = \int d\Pi_{\vec{p}}\,\left(a_{\vec{p}}\,e^{-ip\cdot x} + a^{\dagger}_{\vec{p}}\,e^{ip\cdot x}\right), \qquad p^0 = E_{\vec{p}}, \tag{2.3}$$

$$\left[a_{\vec{p}}, a^{\dagger}_{\vec{q}}\right] = (2\pi)^3\, 2E_{\vec{p}}\,\delta^{(3)}(\vec{p}-\vec{q}), \qquad \left[a_{\vec{p}}, a_{\vec{q}}\right] = \left[a^{\dagger}_{\vec{p}}, a^{\dagger}_{\vec{q}}\right] = 0, \tag{2.4}$$

$$a_{\vec{p}}|0\rangle = 0, \qquad \langle 0|0\rangle = 1. \tag{2.5}$$

Here $[A,B] = AB - BA$, and operators carry no hats. Products of fields at the same point require a regulator or an explicitly stated normal ordering; identities at separated points are understood in the sense of distributions.

Convention (normalisation bridge to the report). The written report uses oscillator-normalised ladder operators and calls them a. In the present notation they are $\tilde{a}_{\vec{p}} = a_{\vec{p}}/\sqrt{2E_{\vec{p}}}$, with $\left[\tilde{a}_{\vec{p}}, \tilde{a}^{\dagger}_{\vec{q}}\right] = (2\pi)^3\,\delta^{(3)}(\vec{p}-\vec{q})$, and

$$|\vec{p}\rangle = a^{\dagger}_{\vec{p}}|0\rangle = \sqrt{2E_{\vec{p}}}\;\tilde{a}^{\dagger}_{\vec{p}}|0\rangle.$$

The states and all observables are the same in both conventions; only the bookkeeping of the factors $2E_{\vec{p}}$ differs.

2.1 The vacuum two-point function

The elementary building block of the lecture follows directly from the ladder algebra. With $z = x - y$ we define the *Wightman function*

$$W(z) = \langle 0|\phi(x)\phi(y)|0\rangle, \tag{2.6}$$

a complex-valued distribution. We insert both mode expansions, keeping separate integration variables; the angle brackets denote vacuum expectation values:

$$W(x-y) = \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}}\,\Big[\langle a_{\vec{p}}a_{\vec{q}}\rangle\, e^{-ip\cdot x - iq\cdot y} + \langle a_{\vec{p}}a^{\dagger}_{\vec{q}}\rangle\, e^{-ip\cdot x + iq\cdot y} + \langle a^{\dagger}_{\vec{p}}a_{\vec{q}}\rangle\, e^{ip\cdot x - iq\cdot y} + \langle a^{\dagger}_{\vec{p}}a^{\dagger}_{\vec{q}}\rangle\, e^{ip\cdot x + iq\cdot y}\Big]. \tag{2.7}$$

The first and third terms vanish because the annihilation operator on the right annihilates $|0\rangle$, the fourth because $\langle 0|a^{\dagger}_{\vec{p}} = (a_{\vec{p}}|0\rangle)^{\dagger} = 0$. In the remaining term

$$\langle a_{\vec{p}}a^{\dagger}_{\vec{q}}\rangle = \langle\left[a_{\vec{p}}, a^{\dagger}_{\vec{q}}\right]\rangle + \langle a^{\dagger}_{\vec{q}}a_{\vec{p}}\rangle = (2\pi)^3\, 2E_{\vec{p}}\,\delta^{(3)}(\vec{p}-\vec{q}). \tag{2.8}$$

The $\vec{q}$-integration removes the δ-function, and the energy factor cancels against the one in the second invariant measure:

The Wightman function.

$$W(z) = \int d\Pi_{\vec{p}}\,\frac{2E_{\vec{p}}}{2E_{\vec{p}}}\,e^{-ip\cdot z} = \int d\Pi_{\vec{p}}\,e^{-ip\cdot z}, \tag{2.9}$$

$$W(-z) = W(z)^{*}, \qquad (\Box_z + m^2)\,W(z) = 0. \tag{2.10}$$

The field equation follows from $(\Box + m^2)\,e^{-ip\cdot z} = (-p^2 + m^2)\,e^{-ip\cdot z} = 0$ on the mass shell. The vacuum correlation is non-zero although $\langle 0|\phi|0\rangle = 0$: a vanishing mean does not mean vanishing fluctuations.

Positive- and negative-frequency parts. We split the field as

$$\phi^{(+)}(x) = \int d\Pi_{\vec{p}}\, a_{\vec{p}}\, e^{-ip\cdot x}, \qquad \phi^{(-)}(x) = \int d\Pi_{\vec{p}}\, a^\dagger_{\vec{p}}\, e^{ip\cdot x},$$
$$\phi = \phi^{(+)} + \phi^{(-)}, \qquad \big(\phi^{(+)}\big)^\dagger = \phi^{(-)}. \tag{2.11}$$

The term "positive frequency" refers to the time dependence e^{-iEt} with $E > 0$ and labels the annihilation part. The same integral as above gives

$$\Big[\phi^{(+)}(x), \phi^{(-)}(y)\Big] = W(x-y)\,\mathbb{1}, \qquad \Big[\phi^{(+)}(x), \phi^{(+)}(y)\Big] = \Big[\phi^{(-)}(x), \phi^{(-)}(y)\Big] = 0; \tag{2.12}$$

this identity will generate the Wick contractions of Section 2.6.

Lorentz invariance. The causality argument below evaluates W in a convenient inertial frame, which requires its Lorentz invariance. The invariant measure can be written in four-dimensional form:

$$\int \frac{d^4p}{(2\pi)^3}\,\theta(p^0)\,\delta(p^2-m^2)\,f(p) = \int \frac{d^3p}{(2\pi)^3}\int dp^0\,\theta(p^0)\,\frac{\delta(p^0-E_{\vec{p}})+\delta(p^0+E_{\vec{p}})}{2E_{\vec{p}}}\,f(p)$$
$$= \int d\Pi_{\vec{p}}\, f(E_{\vec{p}},\vec{p}). \tag{2.13}$$

The factor $1/2E_{\vec{p}}$ is the inverse derivative of $(p^0)^2 - E^2_{\vec{p}}$ at either root, and $\theta(p^0)$ keeps the positive root. Therefore

$$W(z) = \int \frac{d^4p}{(2\pi)^3}\,\theta(p^0)\,\delta(p^2-m^2)\,e^{-ip\cdot z}. \tag{2.14}$$

For a proper orthochronous Lorentz transformation Λ we change variables, $p = \Lambda q$. The factors behave as

$$d^4p = d^4q, \qquad p^2 = q^2, \qquad p\cdot\Lambda z = q\cdot z, \qquad \theta(p^0)\,\delta(p^2-m^2) = \theta(q^0)\,\delta(q^2-m^2), \tag{2.15}$$

where the last equality uses that Λ maps the future mass shell onto itself. Thus

$$W(\Lambda z) = W(z). \tag{2.16}$$

Equivalently, a scalar field transforms as $\phi'(x') = \phi(x)$ for $x' = \Lambda x$, and the free vacuum is Lorentz invariant.

2.2 The commutator and microcausality

We construct the commutator of two fields before interpreting it, using the ladder algebra directly and without taking an expectation value. Of the four commutators of ladder operators only the two mixed ones contribute; with $\big[a^\dagger_{\vec{p}}, a_{\vec{q}}\big] = -\big[a_{\vec{q}}, a^\dagger_{\vec{p}}\big]$ the $\vec{q}$-integration sets $\vec{q} = \vec{p}$:

$$C(z) \equiv [\phi(x),\phi(y)] = \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}}\,\Big(\Big[a_{\vec{p}}, a^\dagger_{\vec{q}}\Big]\, e^{-ip\cdot x+iq\cdot y} + \Big[a^\dagger_{\vec{p}}, a_{\vec{q}}\Big]\, e^{ip\cdot x-iq\cdot y}\Big)$$
$$= \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}}\,(2\pi)^3\,2E_{\vec{p}}\,\delta^{(3)}(\vec{p}-\vec{q})\,\Big(e^{-ip\cdot x+iq\cdot y} - e^{ip\cdot x-iq\cdot y}\Big)\,\mathbb{1}$$
$$= \int d\Pi_{\vec{p}}\,\Big(e^{-ip\cdot z} - e^{ip\cdot z}\Big)\,\mathbb{1} = \big[W(z) - W(-z)\big]\,\mathbb{1}. \tag{2.17}$$

The commutator is proportional to the identity: its coefficient is a c-number, independent of the state. We keep the symbol C for the operator commutator; in scalar kernels below its coefficient of $\mathbb{1}$ is understood. It satisfies

$$C(-z) = -C(z), \qquad C(z)^* = -C(z), \qquad (\Box_z + m^2)\, C(z) = 0, \tag{2.18}$$

where complex conjugation refers to the coefficient. The two operator orders have different vacuum correlations, $W(z)$ and $W(-z)$. Causality will depend on their difference, not on either correlation separately.

2.2.1 An explicit proof of microcausality

Let $z = (z^0, \vec{z})$ be space-like, $(z^0)^2 < |\vec{z}|^2$. A boost along $\vec{z}$ with velocity β gives

$$z'^0 = \gamma\big(z^0 - \beta\,|\vec{z}|\big), \qquad \beta = \frac{z^0}{|\vec{z}|}, \qquad |\beta| < 1, \tag{2.19}$$

so that $z'^0 = 0$ (Figure 2.1a). We set $\vec{r} = \vec{z}\,'$. Lorentz invariance permits the evaluation in this frame:

$$C(z) = C(z') = \int \frac{d^3p}{(2\pi)^3\, 2E_{\vec{p}}} \left(e^{i\vec{p}\cdot\vec{r}} - e^{-i\vec{p}\cdot\vec{r}}\right) \mathbb{1}. \tag{2.20}$$

In the second integral we substitute $\vec{p} \to -\vec{p}$. The domain and the measure are unchanged, since $E_{-\vec{p}} = E_{\vec{p}}$; the two terms become equal and cancel:

Microcausality.

$$[\phi(x), \phi(y)] = 0 \qquad \text{for} \quad (x-y)^2 < 0. \tag{2.21}$$

The conclusion holds throughout the open space-like region, not only on one equal-time slice. Hence the derivatives of the commutator vanish there too, for example $\big[\partial_\mu \phi(x), \phi(y)\big] = \partial_{x^\mu} C(x-y) = 0$. For local polynomials at separated points we use repeatedly $[AB, D] = A\,[B, D] + [A, D]\,B$: properly defined local observables commute at space-like separation. Physical observables are smeared over finite regions; products at a single point require their own definition. Inside the light cone the commutator is non-zero, as it must be: microcausality forbids a controllable influence between space-like separated regions, not dynamics or correlations.

2.2.2 Antiparticles and causality: the complex field

In the proof the two terms of the commutator, $W(z)$ and $W(-z)$, cancelled after a reflection of the integration variable. Their physical meaning becomes clear for a complex scalar field, which describes charged particles. Its mode expansion contains two independent sets of ladder operators,

$$\phi(x) = \int d\Pi_{\vec{p}}\, \big(a_{\vec{p}}\, e^{-ip\cdot x} + b^\dagger_{\vec{p}}\, e^{ip\cdot x}\big), \qquad \phi^\dagger(x) = \int d\Pi_{\vec{p}}\, \big(b_{\vec{p}}\, e^{-ip\cdot x} + a^\dagger_{\vec{p}}\, e^{ip\cdot x}\big), \tag{2.22}$$

with $p^0 = E_{\vec{p}}$,

$$\Big[a_{\vec{p}}, a^\dagger_{\vec{q}}\Big] = \Big[b_{\vec{p}}, b^\dagger_{\vec{q}}\Big] = (2\pi)^3\, 2E_{\vec{p}}\, \delta^{(3)}(\vec{p} - \vec{q}), \qquad \text{all other commutators zero}, \tag{2.23}$$

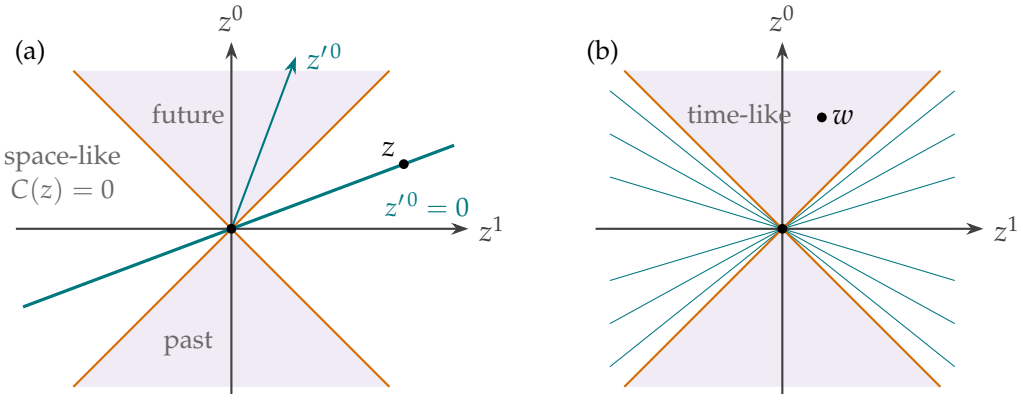


Figure 2.1: (a) The equal-time line $z'^0 = 0$ of the frame boosted with $\beta = z^0/|\vec{z}|$ along $\vec{z}$ (teal: the boosted axes) passes through the space-like event z. In that frame the commutator cancels by the reflection $\vec{p} \to -\vec{p}$, and by Lorentz invariance it vanishes in every frame. (b) The equal-time lines through the origin of all frames, $|\beta| < 1$, sweep out exactly the space-like region. A time-like event w cannot be brought onto an equal-time line with the origin by a physical boost, and there the commutator survives.

and a vacuum annihilated by both, $a_{\vec{p}}|0\rangle = b_{\vec{p}}|0\rangle = 0$. The operator $a^\dagger_{\vec{p}}$ creates a particle and $b^\dagger_{\vec{p}}$ an *antiparticle*: a quantum of the same mass and the opposite charge of the phase symmetry $\phi \to e^{i\alpha}\phi$. Hence ϕ lowers the charge by one unit (it annihilates a particle or creates an antiparticle), and $\phi^\dagger$ raises it.

The commutator. We repeat the steps of (2.17) for $\phi(x)$ and $\phi^\dagger(y)$. Of the four commutators, $\left[a_{\vec{p}}, b_{\vec{q}}\right]$ and $\left[b^\dagger_{\vec{p}}, a^\dagger_{\vec{q}}\right]$ vanish, because particle and antiparticle operators commute; in the remaining two we use $\left[b^\dagger_{\vec{p}}, b_{\vec{q}}\right] = -\left[b_{\vec{q}}, b^\dagger_{\vec{p}}\right]$ and integrate over $\vec{q}$:

$$\begin{aligned}
\left[\phi(x), \phi^\dagger(y)\right] &= \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}} \left[a_{\vec{p}}\, e^{-ip\cdot x} + b^\dagger_{\vec{p}}\, e^{ip\cdot x}\,,\; b_{\vec{q}}\, e^{-iq\cdot y} + a^\dagger_{\vec{q}}\, e^{iq\cdot y}\right] \\
&= \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}} \left(\left[a_{\vec{p}}, a^\dagger_{\vec{q}}\right]\, e^{-ip\cdot x + iq\cdot y} + \left[b^\dagger_{\vec{p}}, b_{\vec{q}}\right]\, e^{ip\cdot x - iq\cdot y}\right) \\
&= \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}}\, (2\pi)^3\, 2E_{\vec{p}}\, \delta^{(3)}(\vec{p} - \vec{q}) \left(e^{-ip\cdot x + iq\cdot y} - e^{ip\cdot x - iq\cdot y}\right) \mathbb{1} \\
&= \int d\Pi_{\vec{p}} \left(e^{-ip\cdot(x-y)} - e^{ip\cdot(x-y)}\right) \mathbb{1} = \left[W(x-y) - W(y-x)\right] \mathbb{1}. \qquad (2.24)
\end{aligned}$$

The complex field has the same commutator function $C(x-y)$ as the real field. The commutator of ϕ with itself contains only $\left[a_{\vec{p}}, a_{\vec{q}}\right]$, $\left[a_{\vec{p}}, b^\dagger_{\vec{q}}\right]$, $\left[b^\dagger_{\vec{p}}, a_{\vec{q}}\right]$ and $\left[b^\dagger_{\vec{p}}, b^\dagger_{\vec{q}}\right]$, all of which vanish:

$$[\phi(x), \phi(y)] = \left[\phi^\dagger(x), \phi^\dagger(y)\right] = 0 \qquad \text{for all } x, y. \qquad (2.25)$$

Two propagation amplitudes. The two orderings of the same pair of fields have a transparent meaning. Since $b_{\vec{q}}|0\rangle = 0$ and $\langle 0|b^\dagger_{\vec{p}} = 0$, the field $\phi^\dagger(y)$ acting on the vacuum can only create a particle, and $\langle 0|\phi(x)$ can only absorb one:

$$\phi^\dagger(y)|0\rangle = \int d\Pi_{\vec{q}}\, e^{iq\cdot y}\, a^\dagger_{\vec{q}}|0\rangle, \qquad \langle 0|\phi(x) = \int d\Pi_{\vec{p}}\, e^{-ip\cdot x}\, \langle 0|a_{\vec{p}}. \qquad (2.26)$$

Conversely, $\phi(x)|0\rangle = \int d\Pi_{\vec{p}}\, e^{ip\cdot x}\, b^\dagger_{\vec{p}}|0\rangle$ contains only an antiparticle, which $\langle 0|\phi^\dagger(y)$ absorbs. With (2.23) and the steps of (2.8),

$$\langle 0|\phi(x)\,\phi^\dagger(y)|0\rangle = \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}}\, \langle 0|a_{\vec{p}}a^\dagger_{\vec{q}}|0\rangle\, e^{-ip\cdot x+iq\cdot y} = \int d\Pi_{\vec{p}}\, e^{-ip\cdot(x-y)} = W(x-y),$$
$$\langle 0|\phi^\dagger(y)\,\phi(x)|0\rangle = \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}}\, \langle 0|b_{\vec{q}}b^\dagger_{\vec{p}}|0\rangle\, e^{-iq\cdot y+ip\cdot x} = \int d\Pi_{\vec{p}}\, e^{-ip\cdot(y-x)} = W(y-x). \quad (2.27)$$

The first is the amplitude for a particle created at y to be absorbed at x; the second, for an antiparticle created at x to be absorbed at y (Figure 2.2). Since the commutator is a c-number, it equals its vacuum expectation value:

$$\Big[\phi(x),\phi^\dagger(y)\Big] = \langle 0|\phi(x)\phi^\dagger(y)|0\rangle - \langle 0|\phi^\dagger(y)\phi(x)|0\rangle = \underbrace{W(x-y)}_{\text{particle } y\to x} - \underbrace{W(y-x)}_{\text{antiparticle } x\to y}. \quad (2.28)$$

Both processes transport the same charge: the particle carries one unit from y to x, the antiparticle minus one unit from x to y.

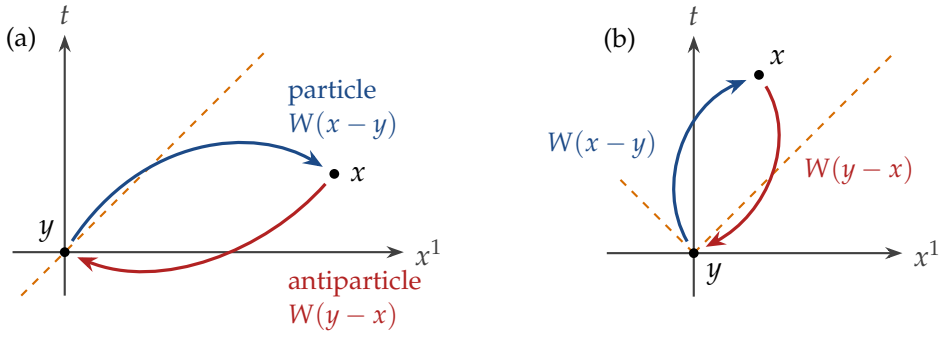


Figure 2.2: The commutator of the complex field as a difference of two propagation amplitudes, (2.28). Blue: $W(x-y)$, a particle created at y and absorbed at x. Red: $W(y-x)$, an antiparticle created at x and absorbed at y; it carries charge -1 from x to y, i.e. the same charge flow as the particle. (a) At space-like separation the two amplitudes are equal and cancel in C. (b) At time-like separation they differ, and the commutator survives.

Causality needs both. At space-like separation the two amplitudes are equal, $W(z) = W(-z)$, by the boost-and-reflection argument of Section 2.2.1. Neither of them vanishes (Section 2.2.3), but they cancel in the commutator. A field built from the particle part alone would violate microcausality:

$$\chi(x) = \int d\Pi_{\vec{p}}\, a_{\vec{p}}\, e^{-ip\cdot x}, \qquad \Big[\chi(x),\chi^\dagger(y)\Big] = \int d\Pi_{\vec{p}}\, d\Pi_{\vec{q}}\, \Big[a_{\vec{p}},a^\dagger_{\vec{q}}\Big]\, e^{-ip\cdot x+iq\cdot y} = W(x-y)\,\mathbb{1}, \quad (2.29)$$

which is non-zero outside the light cone. A local relativistic field needs both terms of (2.22): causality is restored by antiparticles.

The real field. For the real field $b_{\vec{p}} = a_{\vec{p}}$ and $\phi^\dagger = \phi$: the particle is its own antiparticle. The two terms of $C(z) = W(z) - W(-z)$ in (2.17) are then the same particle propagating

from y to x and from x to y, both described by one set of ladder operators. This is why the cancellation in the real case looked automatic. The Dirac field of Lecture 3 has the same structure: the antifermion operators $b_r^\dagger$ multiply its negative-frequency modes.

Causality is restored by antiparticles. The propagation amplitude $W(x-y)$ leaks out of the light cone, as does the amplitude of a single relativistic particle. What vanishes outside the light cone is the commutator: the amplitude for a particle to go from y to x minus the amplitude for its antiparticle – same mass, opposite charge – to go from x to y. A causal relativistic quantum theory needs antiparticles; for a real field every particle is its own antiparticle.

2.2.3 A non-zero correlation is not a signal

Microcausality is a statement about the commutator; the correlation W itself does not vanish outside the light cone. At equal times, with $p \equiv |\vec{p}|$ and $r \equiv |\vec{r}| > 0$ in the radial integral,

$$\begin{aligned} W(0,\vec{r}) &= \int \frac{d^3p}{(2\pi)^3\, 2\sqrt{\vec{p}^{\,2}+m^2}}\, e^{i\vec{p}\cdot\vec{r}} \\ &= \frac{1}{4\pi^2 r}\int_0^\infty dp\, \frac{p\,\sin(pr)}{\sqrt{p^2+m^2}}, \end{aligned} \tag{2.30}$$

where the angular integral is $2\pi\int_{-1}^{1} du\, e^{ipru} = 4\pi\sin(pr)/(pr)$. The massless limit is a simple explicit example. With a convergence factor $e^{-\eta p}$, $\eta > 0$,

$$\begin{aligned} \int_0^\infty dp\, e^{-\eta p}\sin(pr) &= \mathrm{Im}\int_0^\infty dp\, e^{-(\eta - ir)p} = \mathrm{Im}\,\frac{1}{\eta - ir} = \frac{r}{\eta^2+r^2}, \\ W_{m=0}(0,\vec{r}) &= \lim_{\eta\to 0^+}\frac{1}{4\pi^2(\eta^2+r^2)} = \frac{1}{4\pi^2 r^2} \neq 0. \end{aligned} \tag{2.31}$$

For $m > 0$ the integral is a modified Bessel function. Writing $p\sin(pr) = -\partial_r\cos(pr)$ and using $\int_0^\infty dp\,\cos(pr)/\sqrt{p^2+m^2} = K_0(mr)$ together with $K_0' = -K_1$,

$$W(0,\vec{r}) = -\frac{1}{4\pi^2 r}\,\partial_r K_0(mr) = \frac{m\,K_1(mr)}{4\pi^2 r}. \tag{2.32}$$

For $mr \ll 1$, $K_1(mr) \simeq 1/(mr)$ and the massless result is recovered; for $mr \gg 1$ the correlation decays as $W(0,\vec{r}) \simeq \frac{m^2}{4\pi^2}\sqrt{\pi/2}\, e^{-mr}/(mr)^{3/2}$, but it never vanishes (Figure 2.3). The scattering theory retains $m > 0$; the massless limit only illustrates the point. The cancellation in the space-like commutator requires the equality $W(z) = W(-z)$, not that either term should vanish.

A controlled local intervention. Let B be a Hermitian observable supported in one small region and A one supported in a space-like separated region. Consider the local unitary operation $U_B = e^{-i\varepsilon B}$, with a small real parameter ε. For any state,

$$\begin{aligned} \langle A\rangle_{\text{after}} &= \langle U_B^\dagger A\, U_B\rangle \\ &= \langle (1+i\varepsilon B)\, A\, (1-i\varepsilon B)\rangle + \mathcal{O}(\varepsilon^2) \\ &= \langle A\rangle + i\varepsilon\,\langle [B,A]\rangle + \mathcal{O}(\varepsilon^2). \end{aligned} \tag{2.33}$$

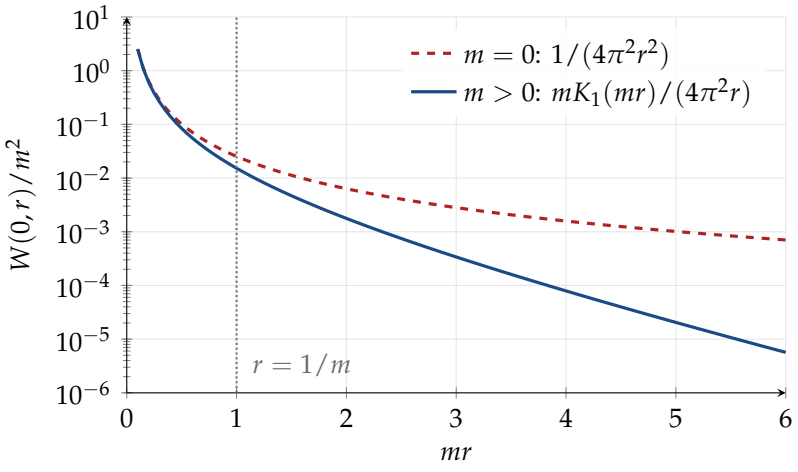


Figure 2.3: The equal-time Wightman function $W(0,r)$ of (2.32), on a logarithmic scale. Beyond the Compton wavelength $1/m$ the massive correlation (solid) falls off as $e^{-mr}/(mr)^{3/2}$, the massless one (dashed; here m only sets the unit of length) as $1/r^2$. Neither vanishes at space-like separation: microcausality rests on the equality $W(z) = W(-z)$, not on $W = 0$.

If $[B, A] = 0$, every nested commutator in the exact expansion of the exponential vanishes too, so the local operation changes no expectation value of A: a response involves a commutator. A pre-existing correlation $\langle AB \rangle$ need not vanish. Correlation measurements can reveal shared fluctuations, but they do not by themselves provide a controllable space-like signal. In the same way, $W(x-y)$ is the overlap of the states $\phi(x)|0\rangle$ and $\phi(y)|0\rangle$ created by the field (which themselves require smearing), not a detection probability; the operational test of causality is the change of a local observable.

2.3 Time ordering and the Feynman propagator

In perturbation theory operators appear ordered in time, and we need the corresponding two-point function. For a bosonic field the time-ordered product is

$$\mathrm{T}\{\phi(x)\phi(y)\} = \theta(z^0)\,\phi(x)\phi(y) + \theta(-z^0)\,\phi(y)\phi(x)\,: \tag{2.34}$$

the later time stands to the left; θ is the Heaviside step distribution. Its vacuum expectation value is the *Feynman propagator*

$$D_F(z) \equiv \langle 0|\mathrm{T}\{\phi(x)\phi(y)\}|0\rangle = \theta(z^0)\,W(z) + \theta(-z^0)\,W(-z), \tag{2.35}$$

defined without an additional factor of i (for other conventions see the box at the end of Section 2.3.2). At unequal times, with $z = (t, \vec{r})$,

$$D_F(t,\vec{r}) = \int \frac{d^3p}{(2\pi)^3}\,\frac{e^{i\vec{p}\cdot\vec{r}}}{2E_{\vec{p}}}\left[\theta(t)\,e^{-iE_{\vec{p}}t} + \theta(-t)\,e^{iE_{\vec{p}}t}\right], \tag{2.36}$$

where in the second term we changed $\vec{p} \to -\vec{p}$. For the complex field, (2.27) gives in the same way $\langle 0|\mathrm{T}\{\phi(x)\phi^\dagger(y)\}|0\rangle = D_F(x-y)$: for $x^0 > y^0$ a particle created at y is absorbed at x, for $y^0 > x^0$ an antiparticle created at x is absorbed at y. In both cases the quantum is created at the earlier point. We now verify that a simple four-dimensional Fourier integral reproduces this time-ordered definition.

2.3.1 The four-dimensional Fourier representation

Consider the energy integral

$$I_E(t) = \lim_{\epsilon\to 0^+} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{i\, e^{-i\omega t}}{\omega^2 - E^2 + i\epsilon}, \qquad E > 0, \tag{2.37}$$

where ϵ has mass dimension two. The two poles are at

$$\omega_+ = +E - \frac{i\epsilon}{2E} + \mathcal{O}(\epsilon^2), \qquad \omega_- = -E + \frac{i\epsilon}{2E} + \mathcal{O}(\epsilon^2)\,: \tag{2.38}$$

the positive-energy pole lies below the real axis, the negative-energy pole above it. Writing $\omega = u + iv$,

$$e^{-i\omega t} = e^{-iut}\, e^{vt}. \tag{2.39}$$

For $t > 0$ this decays for $v < 0$, for $t < 0$ for $v > 0$: the contour must be closed in the corresponding half-plane (Figure 2.4). The large semicircle contributes zero for $t \neq 0$, the limiting equality being understood as a distribution. The pole prescription fixes a boundary condition: $1/(p^2 - m^2)$ alone, without a rule for its poles, does not define a unique Green function.

Let $f(\omega) = i\, e^{-i\omega t}/(\omega^2 - E^2 + i\epsilon)$. For $t > 0$ the real line is traversed from left to right, and the lower semicircle closes the contour clockwise; hence

$$I_E(t > 0) = -i \operatorname*{Res}_{\omega=\omega_+} f(\omega) \;\longrightarrow\; -i\, \frac{i\, e^{-iEt}}{2E} = \frac{e^{-iEt}}{2E}. \tag{2.40}$$

The factor $-i$ is the $-2\pi i$ of the clockwise contour divided by the 2π of the measure. For $t < 0$ we close counterclockwise in the upper half-plane:

$$I_E(t < 0) = +i \operatorname*{Res}_{\omega=\omega_-} f(\omega) \;\longrightarrow\; +i\, \frac{i\, e^{iEt}}{-2E} = \frac{e^{iEt}}{2E}; \tag{2.41}$$

here the negative derivative of the denominator at the pole $-E$ supplies the second minus sign. Thus

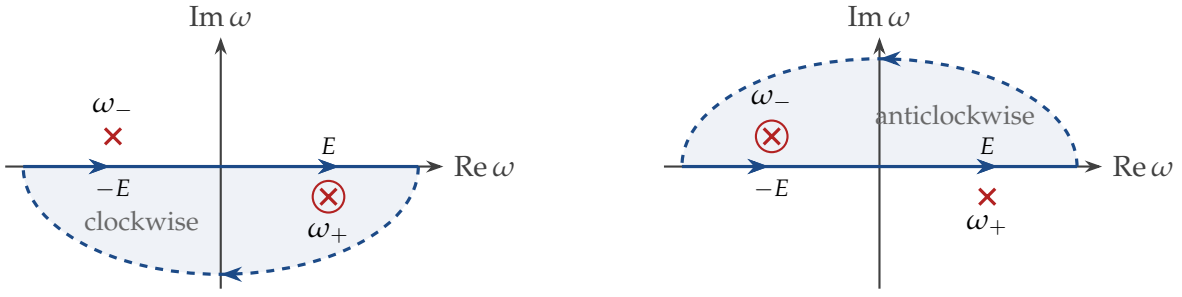
$$I_E(t) = \frac{\theta(t)\, e^{-iEt} + \theta(-t)\, e^{iEt}}{2E}. \tag{2.42}$$

With $p^2 - m^2 = (p^0)^2 - E_{\vec{p}}^2$ and $e^{-ip\cdot z} = e^{-ip^0 t}\, e^{i\vec{p}\cdot\vec{r}}$, the p^0-integral of a four-dimensional Fourier integral with this denominator is $I_{E_{\vec{p}}}(t)$. Comparison with (2.36) gives the four-momentum form of the propagator:

The Feynman propagator.

$$D_F(x-y) = \langle 0|\mathrm{T}\{\phi(x)\phi(y)\}|0\rangle = \lim_{\epsilon\to 0^+} \int \frac{d^4p}{(2\pi)^4} \frac{i\, e^{-ip\cdot(x-y)}}{p^2 - m^2 + i\epsilon}. \tag{2.43}$$

The integration variable p^0 now runs freely: inside this Fourier representation the momentum is not restricted to the mass shell, while the poles still encode the free-particle energies.



(a) $t > 0$: close below (b) $t < 0$: close above

Figure 2.4: The energy integral (2.37) in the complex ω plane. The poles $\omega_\pm = \pm(E - i\epsilon/2E)$ (crosses; their displacements from $\pm E$ exaggerated) lie below and above the real axis. (a) For $t > 0$ the contour is closed in the lower half-plane, clockwise, and encloses only ω_+: $I_E = e^{-iEt}/2E$. (b) For $t < 0$ it is closed in the upper half-plane, anticlockwise, and encloses only ω_-: $I_E = e^{iEt}/2E$. The large arcs (dashed) do not contribute.

2.3.2 Green functions: the contact term and the causal response

The sign of the Green equation is fixed by (2.43). In the distributional limit of the regulator,

$$(\Box_z + m^2)\, D_F(z) = \lim_{\epsilon\to 0^+} \int \frac{d^4p}{(2\pi)^4}\, \frac{-i\,(p^2 - m^2)}{p^2 - m^2 + i\epsilon}\, e^{-ip\cdot z} = -i\,\delta^{(4)}(z). \tag{2.44}$$

The remainder at finite ϵ is $i\epsilon D_F$, which vanishes in this limit. Since $[\phi] = 1$, we have $[D_F] = 2$, and both sides have mass dimension four.

The same source from the equal-time commutator. For fixed $\vec{r}$ let $f(t) = W(t,\vec{r})$ and $g(t) = W(-t,-\vec{r})$, so that $D_F = \theta(t)f + \theta(-t)g$ by (2.35). At $t = 0$,

$$\begin{aligned} f(0) - g(0) &= 0, \\ f'(0) - g'(0) &= -\frac{i}{2}\int \frac{d^3p}{(2\pi)^3}\left(e^{i\vec{p}\cdot\vec{r}} + e^{-i\vec{p}\cdot\vec{r}}\right) = -i\,\delta^{(3)}(\vec{r}), \end{aligned} \tag{2.45}$$

which is the statement $\langle 0|\left[\dot\phi(0,\vec{r}), \phi(0,\vec{0})\right]|0\rangle = -i\,\delta^{(3)}(\vec{r})$. A piecewise smooth function $F = \theta(t)f + \theta(-t)g$ satisfies

$$F'' = \theta(t)f'' + \theta(-t)g'' + \left[f'(0) - g'(0)\right]\delta(t) + \left[f(0) - g(0)\right]\delta'(t). \tag{2.46}$$

The δ' term vanishes. Away from $t = 0$ the homogeneous equations cancel the ordinary terms, leaving precisely $-i\,\delta(t)\,\delta^{(3)}(\vec{r})$, in agreement with (2.44). The result is sign-sensitive: it tests both the pole prescription and the canonical commutator.

Retarded means response only after the source. The real response kernel is built from the coefficient of the commutator,

$$G_R(z) = i\,\theta(z^0)\,C(z), \qquad (\Box_z + m^2)\,G_R(z) = \delta^{(4)}(z). \tag{2.47}$$

The equation follows from (2.46) with $C(0,\vec{r}) = 0$ and $\partial_t C(0,\vec{r}) = -i\,\delta^{(3)}(\vec{r})$. For a classical source coupled by $\mathcal{L}_J = J\phi$ the field equation is $(\Box + m^2)\phi = J$, so a retarded particular solution is

$$\phi_J(x) = \int d^4y\; G_R(x-y)\,J(y). \tag{2.48}$$

It vanishes before the source acts and outside its future light cone. D_F, instead, implements time ordering in amplitudes and is non-zero at space-like separation; the two definitions fix different boundary conditions.

Convention (notation dictionary). The report and these lectures use the symbols of the first column; the Lund lecture notes and Peskin–Schroeder use the others. The physics is identical.

Rzeszów lectures and report	Lund lecture notes	Peskin–Schroeder
$W(z) = \langle 0\|\phi(x)\phi(y)\|0\rangle$	$D(x-y)$	$D(x-y)$
$C(z) = [\phi(x), \phi(y)]$ (no i)	$\Delta(x-y)$ (no i)	$D(x-y) - D(y-x)$
$D_F(z) = \langle 0\|\mathrm{T}\{\phi(x)\phi(y)\}\|0\rangle$	$\Delta_F(x-y)$	$D_F(x-y)$
$G_R = i\theta(z^0)\, C$, $(\Box + m^2)G_R = \delta^{(4)}$	$\Delta_R = \theta\Delta = -iG_R$	$D_R = -iG_R$
$d\Pi_{\vec{p}}$	$\widetilde{dp}$	(no symbol)
covariant $a_{\vec{p}}$; report: $\tilde{a}_{\vec{p}}$	covariant $a_{\vec{p}}$	oscillator-normalised $a_{\vec{p}}$ $(= \tilde{a}_{\vec{p}})$

All three D_F obey $(\Box + m^2)D_F = -i\delta^{(4)}$, and so do Δ_R and D_R. The factor i in $G_R = i\Delta_R$ gives G_R a unit source, so that $\phi_J = \int G_R J$.

2.4 A local interaction and the interaction picture

To compute an amplitude we must specify a theory and an approximation. We take

$$\mathcal{L} = \frac{1}{2}\,\partial_\mu\phi\,\partial^\mu\phi - \frac{1}{2}\,m^2\phi^2 - \frac{\lambda}{4!}\,\phi^4, \qquad 0 < \lambda \ll 1. \tag{2.49}$$

The theory is invariant under $\phi \to -\phi$ and under translations. The interaction is local: all four fields meet at the same spacetime point. In four dimensions

$$[\mathcal{L}] = 4, \qquad [\phi] = 1, \qquad [\lambda] = 0, \tag{2.50}$$

and positive m^2 and λ make the classical potential $V(\phi) = m^2\phi^2/2 + \lambda\phi^4/4!$ bounded from below. There are no derivative interactions, so the canonical momentum remains $\pi = \dot{\phi}$, and the Legendre transform gives

$$\begin{aligned} \mathcal{H} &= \pi\dot{\phi} - \mathcal{L} = \frac{1}{2}\,\pi^2 + \frac{1}{2}\,(\nabla\phi)^2 + \frac{1}{2}\,m^2\phi^2 + \frac{\lambda}{4!}\,\phi^4, \\ H &= H_0 + H_{\mathrm{int}}, \qquad H_{\mathrm{int}} = \frac{\lambda}{4!}\int d^3x\;\phi^4(\vec{x}). \end{aligned} \tag{2.51}$$

Thus $H_{\mathrm{int}} = -\int d^3x\,\mathcal{L}_{\mathrm{int}}$; this holds because the interaction contains no derivatives.

Why the free-field solution is insufficient. The full classical field equation is nonlinear,

$$(\Box + m^2)\,\phi = -\frac{\lambda}{3!}\,\phi^3, \tag{2.52}$$

where the factor $1/3!$ comes from differentiating $\phi^4/4!$. A superposition of free plane waves does not solve this equation. The perturbative strategy keeps H_0 exactly and expands the evolution caused by H_{int} in powers of λ: instead of giving the interacting field a free mode expansion, we combine free operators with an interaction-dependent evolution of the states.

For the quantum tree calculation we define the interaction by normal ordering with respect to the free vacuum,

$$H_{\rm int} = \frac{\lambda}{4!}\int d^3x\, :\phi(\vec{x})^4:. \tag{2.53}$$

Normal ordering puts creation operators to the left without adding commutator terms, and so removes contractions of fields within this single vertex. At the order considered here it makes the four-particle calculation particularly transparent; it does not remove quantum corrections in general.

2.4.1 The interaction picture

We start from quantum mechanics with a time-independent Hamiltonian; all pictures agree at $t = 0$. In the Schrödinger picture

$$i\partial_t|\Psi(t)\rangle_S = H|\Psi(t)\rangle_S, \qquad |\Psi(t)\rangle_S = e^{-iHt}|\Psi(0)\rangle. \tag{2.54}$$

In the Heisenberg picture

$$|\Psi\rangle_H = e^{iHt}|\Psi(t)\rangle_S, \qquad O_H(t) = e^{iHt}\,O_S\,e^{-iHt}\ : \tag{2.55}$$

the state is fixed, and the operator carries the full dynamics. Expectation values agree, because the exponentials cancel between bras, operators and kets. For the interaction picture we transform only with the solvable free Hamiltonian:

$$\begin{aligned} |\Psi(t)\rangle_I &= e^{iH_0t}|\Psi(t)\rangle_S, \\ O_I(t) &= e^{iH_0t}\,O_S\,e^{-iH_0t}, \qquad H_I(t) = e^{iH_0t}\,H_{\rm int}\,e^{-iH_0t}. \end{aligned} \tag{2.56}$$

Differentiating the state explicitly,

$$\begin{aligned} i\partial_t|\Psi(t)\rangle_I &= -H_0\,e^{iH_0t}|\Psi(t)\rangle_S + e^{iH_0t}H|\Psi(t)\rangle_S \\ &= e^{iH_0t}\,(H - H_0)\,e^{-iH_0t}|\Psi(t)\rangle_I \\ &= H_I(t)|\Psi(t)\rangle_I. \end{aligned} \tag{2.57}$$

We used $\left[H_0, e^{iH_0t}\right] = 0$; H_0 and $H_{\rm int}$ need not commute.

The field is free in the interaction picture. From Lecture 1, $\left[H_0, a_{\vec{p}}\right] = -E_{\vec{p}}\,a_{\vec{p}}$. With $A(t) = e^{iH_0t}a_{\vec{p}}\,e^{-iH_0t}$,

$$\dot{A}(t) = i\,e^{iH_0t}\left[H_0, a_{\vec{p}}\right]e^{-iH_0t} = -iE_{\vec{p}}\,A(t), \qquad A(t) = a_{\vec{p}}\,e^{-iE_{\vec{p}}t}.$$

Likewise a creation operator acquires $e^{+iE_{\vec{p}}t}$. Therefore

$$\begin{aligned} \phi_I(x) &= \int d\Pi_{\vec{p}}\left(a_{\vec{p}}\,e^{-ip\cdot x} + a^\dagger_{\vec{p}}\,e^{ip\cdot x}\right), \\ H_I(t) &= \frac{\lambda}{4!}\int d^3x\, :\phi_I(x)^4:. \end{aligned} \tag{2.58}$$

All the free two-point functions derived above apply to ϕ_I. The interacting Heisenberg field is $\phi_H(t,\vec{x}) = U_I(t,0)^\dagger\,\phi_I(t,\vec{x})\,U_I(t,0)$, with the pictures matched at $t = 0$ and U_I defined in the next section. The subscript I labels the picture: operators evolve with H_0 and solve the free field equation, states evolve with $H_I(t)$.

2.5 The Dyson series

The evolution operator $U_I(t,t_0)$ of the interaction picture is defined by

$$|\Psi(t)\rangle_I = U_I(t,t_0)|\Psi(t_0)\rangle_I. \tag{2.59}$$

Since the initial state is arbitrary, (2.57) implies

$$i\partial_t\, U_I(t,t_0) = H_I(t)\, U_I(t,t_0), \qquad U_I(t_0,t_0) = \mathbb{1}. \tag{2.60}$$

Integrating from t_0 to t gives an exact integral equation,

$$U_I(t,t_0) = \mathbb{1} - i\int_{t_0}^{t} dt_1\, H_I(t_1)\, U_I(t_1,t_0). \tag{2.61}$$

We write $U_I = U_{[0]} + U_{[1]} + U_{[2]} + \cdots$, where $U_{[n]} = \mathcal{O}(\lambda^n)$ is the term of order n, and iterate:

$$\begin{aligned} U_{[0]}(t,t_0) &= \mathbb{1}, \\ U_{[1]}(t,t_0) &= -i\int_{t_0}^{t} dt_1\, H_I(t_1), \\ U_{[2]}(t,t_0) &= (-i)^2\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2\, H_I(t_1)\, H_I(t_2). \end{aligned} \tag{2.62}$$

The upper limit of the inner integral is t_1, because it came from $U_I(t_1,t_0)$. The complete second-order approximation is

$$U_I = \mathbb{1} - i\int_{t_0}^{t} dt_1\, H_I(t_1) - \int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2\, H_I(t_1)\, H_I(t_2) + \mathcal{O}(\lambda^3). \tag{2.63}$$

Later times stand to the left, and the minus sign of the second-order term is $(-i)^2 = -1$. A differential equation together with its initial condition has produced ordered integrals of non-commuting operators.

Unitarity. As an independent consistency check: since $H_I = H_I^\dagger$, the adjoint equation reads $\partial_t U_I^\dagger = +i\, U_I^\dagger H_I$. Hence

$$\begin{aligned} \partial_t\big(U_I^\dagger U_I\big) &= i\, U_I^\dagger H_I U_I - i\, U_I^\dagger H_I U_I = 0, \\ U_I^\dagger(t,t_0)\, U_I(t,t_0) &= \mathbb{1}. \end{aligned} \tag{2.64}$$

A truncated expansion preserves this equality only to its stated order: probabilities cannot be tested by keeping a term in one factor and dropping the terms of matching order in its conjugate.

2.5.1 Time ordering produces the factorial

Write $H_1 = H_I(t_1)$ and $H_2 = H_I(t_2)$. For two bosonic interaction Hamiltonians

$$\mathrm{T}\{H_1 H_2\} = \theta(t_1 - t_2)\, H_1 H_2 + \theta(t_2 - t_1)\, H_2 H_1. \tag{2.65}$$

Integrating over the square $t_0 < t_1, t_2 < t$, each step function selects one triangle (Figure 2.5):

$$\int_{t_0}^{t} dt_1\, dt_2\, \mathrm{T}\{H_1 H_2\} = \int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2\, H_1 H_2 + \int_{t_0}^{t} dt_2 \int_{t_0}^{t_2} dt_1\, H_2 H_1. \tag{2.66}$$

In the second term we rename the integration variables, $t_1 \leftrightarrow t_2$; its operator order then becomes exactly that of the first term. Thus

$$\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2 \; H_1 H_2 = \frac{1}{2!} \int_{t_0}^{t} dt_1 \, dt_2 \; \mathrm{T}\{H_1 H_2\}. \tag{2.67}$$

Relabelling dummy variables is legitimate even when $[H_1, H_2] \neq 0$; interchanging the operators at fixed time arguments would not be. For n times the cube $[t_0, t]^n$ splits into $n!$ ordered regions, one for each permutation, and time ordering maps each region onto $t_1 > t_2 > \cdots > t_n$. Therefore

$$\begin{aligned} U_{[n]}(t, t_0) &= (-i)^n \int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2 \cdots \int_{t_0}^{t_{n-1}} dt_n \; H_I(t_1) \cdots H_I(t_n) \\ &= \frac{(-i)^n}{n!} \int_{t_0}^{t} dt_1 \cdots dt_n \; \mathrm{T}\{H_I(t_1) \cdots H_I(t_n)\}. \end{aligned} \tag{2.68}$$

The time-ordered exponential is defined by this series:

Dyson series.

$$U_I(t, t_0) = \mathrm{T} \exp\left[-i \int_{t_0}^{t} dt' \, H_I(t') \right] \equiv \sum_{n=0}^{\infty} U_{[n]}(t, t_0). \tag{2.69}$$

If all the relevant $H_I(t')$ commute, the time ordering may be dropped, and the series becomes an ordinary exponential.

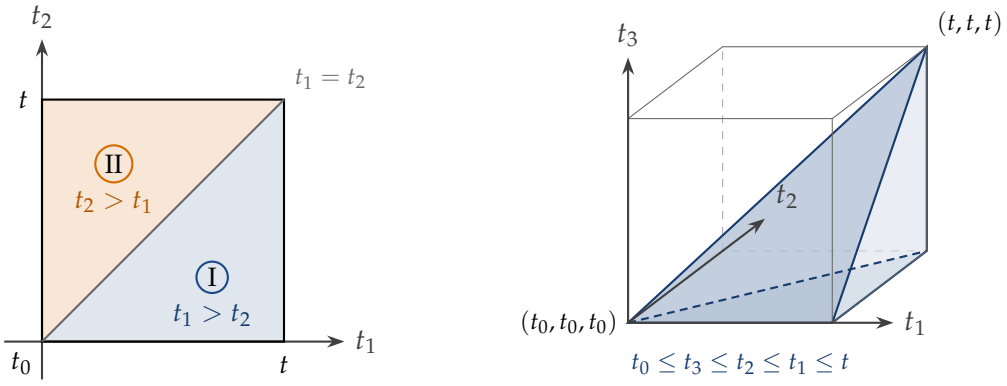


Figure 2.5: Left: the square $[t_0, t]^2$ of the second-order term. The nested integral of (2.62) covers triangle I ($t_1 > t_2$); triangle II is its mirror image under $t_1 \leftrightarrow t_2$. Because the T-product makes the integrand symmetric, the nested integral is half the integral over the square, (2.67). Right: for $n = 3$ the ordered region $t_0 \le t_3 \le t_2 \le t_1 \le t$ is one of $3! = 6$ congruent tetrahedra that fill the cube $[t_0, t]^3$, which gives the factor $1/3!$.

Space and time integrations together. Since $H_I(t) = \frac{\lambda}{4!} \int d^3x \, {:}\phi_I(x)^4{:}$,

$$U_I = \mathbb{1} + \frac{-i\lambda}{4!} \int d^4x \, {:}\phi_I(x)^4{:} + \frac{(-i\lambda)^2}{2!\,(4!)^2} \int d^4x \, d^4y \; \mathrm{T}\{{:}\phi_I(x)^4{:}\,{:}\phi_I(y)^4{:}\} + \cdots . \tag{2.70}$$

The time integrations run over the finite interval of U_I. No matrix element has been evaluated yet; the next task is to turn the time-ordered operator products into known free-field functions.

2.6 Wick's theorem

Wick's theorem converts time ordering into normal ordering plus the propagator already computed. In this section all fields are free interaction-picture fields, and we suppress the subscript I. Let $A = \phi(x) = A_+ + A_-$ and $B = \phi(y) = B_+ + B_-$, where $+$ denotes the annihilation parts $\phi^{(+)}$. If $x^0 > y^0$, then $\mathrm{T}\{AB\} = AB$, and with (2.12)

$$\begin{aligned} AB &= A_+B_+ + A_+B_- + A_-B_+ + A_-B_- \\ &= A_+B_+ + B_-A_+ + A_-B_+ + A_-B_- + [A_+, B_-] \\ &= :AB: + W(x-y)\,\mathbb{1}. \end{aligned} \tag{2.71}$$

The first four terms of the second line have all creation parts to the left; commuting two creation parts or two annihilation parts adds nothing. If $y^0 > x^0$, the same steps with BA give

$$\mathrm{T}\{AB\} = BA = :BA: + W(y-x)\,\mathbb{1} = :AB: + W(y-x)\,\mathbb{1}, \tag{2.72}$$

since normal products of bosonic free fields are symmetric, $:BA: = :AB:$. Combining the two time regions with (2.35):

Wick's theorem for two fields.

$$\mathrm{T}\{\phi(x)\phi(y)\} = :\phi(x)\phi(y): + D_F(x-y)\,\mathbb{1}. \tag{2.73}$$

Its vacuum expectation value removes the normal product and recovers the definition of D_F.

Contractions. A contraction is the c-number difference

$$\overbrace{\phi(x)\phi(y)} \equiv \mathrm{T}\{\phi(x)\phi(y)\} - :\phi(x)\phi(y): = D_F(x-y)\,\mathbb{1}. \tag{2.74}$$

It records the commutator terms generated when a time-ordered product is rewritten in normal order, and it acts as an algebraic replacement inside an operator calculation: the contracted pair disappears and is replaced by a propagator, while the remaining operators are normal ordered. For a normal product containing at least one field,

$$\langle 0|:\phi(x_1)\cdots\phi(x_n):|0\rangle = 0 \qquad (n \geq 1), \tag{2.75}$$

because each term has either an annihilation part on the right or a creation part on the left. A product with no remaining fields is the identity, with expectation value one.

2.6.1 Four fields: every term visible

Write $\phi_i = \phi(x_i)$ and $D_{ij} = D_F(x_i - x_j)$. At distinct points Wick's identity for four real scalar fields reads (Figure 2.6)

$$\begin{aligned} \mathrm{T}\{\phi_1\phi_2\phi_3\phi_4\} = {} & :\phi_1\phi_2\phi_3\phi_4: \\ & + D_{12}:\phi_3\phi_4: + D_{13}:\phi_2\phi_4: + D_{14}:\phi_2\phi_3: \\ & + D_{23}:\phi_1\phi_4: + D_{24}:\phi_1\phi_3: + D_{34}:\phi_1\phi_2: \\ & + \big(D_{12}D_{34} + D_{13}D_{24} + D_{14}D_{23}\big)\,\mathbb{1}. \end{aligned} \tag{2.76}$$

There is one term without contraction, $\binom{4}{2} = 6$ terms with one contraction, and three complete pairings; none carries an extra factorial.

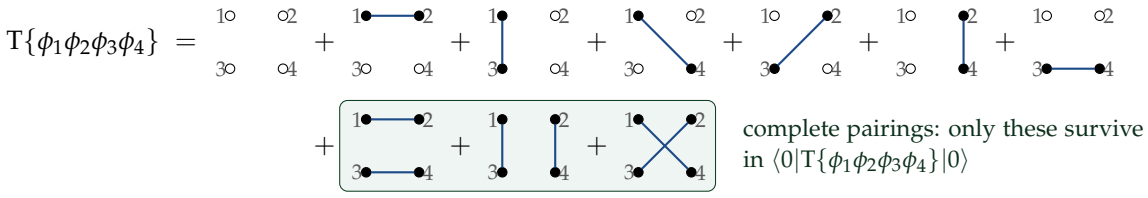


Figure 2.6: Wick's theorem for four fields in pictures, term by term as in (2.76). Each dot is a point x_i, each line a contraction $D_{ij} = D_F(x_i - x_j)$. Open circles are uncontracted fields, which remain in the normal product. Only the three complete pairings (shaded) contribute to the vacuum expectation value (2.78).

Where the enumeration comes from. Choose a definite time order, say $x_1^0 > x_2^0 > x_3^0 > x_4^0$, and split each field into creation and annihilation parts. Moving an annihilation part to the right uses

$$A_+ B_- = B_- A_+ + [A_+, B_-]\,.$$

At each crossing either both fields are retained, now normally ordered, or the crossed pair is replaced by its c-number commutator. A field removed in a contraction cannot be used again. Thus every selected pair contributes once, and the remaining fields undergo the same operation; the commutator is a c-number, so it introduces no further operator ordering. For example, the term with the pairs $(1,3)$ and $(2,4)$ is, in this time order,

$$\left[\phi_1^{(+)}, \phi_3^{(-)}\right] \left[\phi_2^{(+)}, \phi_4^{(-)}\right] = W(x_1 - x_3)\, W(x_2 - x_4) = D_{13} D_{24}. \tag{2.77}$$

Other time orders replace each pair by its time-ordered value D_{ij} and lead to the same formula (2.76). Taking the vacuum expectation value removes the uncontracted normal products:

$$\langle 0|\mathrm{T}\{\phi_1\phi_2\phi_3\phi_4\}|0\rangle = D_{12}D_{34} + D_{13}D_{24} + D_{14}D_{23}. \tag{2.78}$$

Odd free-vacuum correlators vanish, because a complete pairing always leaves one field over. For $2n$ fields the number of complete pairings is

$$(2n-1)(2n-3)\cdots 1 = \frac{(2n)!}{2^n\, n!}\,: \tag{2.79}$$

choose a partner for the first field, then for the first remaining field, and so on. This factorisation is a property of the free (Gaussian) vacuum; the interaction enters only through the Dyson series.

2.7 Asymptotic states and the scattering operator

We now define exactly which matrix element will be computed. A scattering experiment specifies the particles long before and long after their interaction; in perturbation theory these states are represented by free Fock states. Wave packets may be taken first if needed, sharp momenta being their distributional limit:

$$|i\rangle = |\vec{p}_1, \vec{p}_2\rangle = a^\dagger_{\vec{p}_1} a^\dagger_{\vec{p}_2} |0\rangle, \qquad |f\rangle = |\vec{k}_1, \vec{k}_2\rangle. \tag{2.80}$$

Each external four-momentum is on shell, $p_i^2 = k_i^2 = m^2$, with positive energy. The real scalar is neutral, and its quanta are identical bosons.

The free overlap. With the covariant delta kernel

$$\delta_\Pi(\vec{k},\vec{p}) = (2\pi)^3\, 2E_{\vec{p}}\,\delta^{(3)}(\vec{k}-\vec{p}), \qquad \int d\Pi_{\vec{k}}\;\delta_\Pi(\vec{k},\vec{p})\,f(\vec{k}) = f(\vec{p}), \tag{2.81}$$

commuting two annihilation operators through two creation operators gives

$$\begin{aligned}\langle f|i\rangle &= \langle 0|a_{\vec{k}_2}a_{\vec{k}_1}a^\dagger_{\vec{p}_1}a^\dagger_{\vec{p}_2}|0\rangle \\ &= \delta_\Pi(\vec{k}_1,\vec{p}_1)\,\delta_\Pi(\vec{k}_2,\vec{p}_2) + \delta_\Pi(\vec{k}_1,\vec{p}_2)\,\delta_\Pi(\vec{k}_2,\vec{p}_1).\end{aligned} \tag{2.82}$$

Both terms describe the incoming particles continuing freely, with the two possible labellings of the identical final particles; the overlap contains no interaction.

The scattering operator and the connected amplitude. In the usual perturbative construction

$$S = \lim_{t_+\to+\infty,\,t_-\to-\infty} U_I(t_+,t_-) = \mathrm{T}\exp\Big[-i\int_{-\infty}^{\infty} dt\; H_I(t)\Big]. \tag{2.83}$$

The limit is defined with adiabatic switching of the interaction at the time boundaries and the usual normalisation of the vacuum. At tree level the free and the physical one-particle normalisations can be matched directly; quantum corrections require further care. The connected invariant amplitude $\mathcal{M}$ is defined by

$$\langle f|(S-\mathbb{1})|i\rangle_{\mathrm{conn}} = i\,(2\pi)^4\,\delta^{(4)}(p_1+p_2-k_1-k_2)\;\mathcal{M}. \tag{2.84}$$

The subscript selects contributions joining all four external particles; vacuum factors and spectator propagation are excluded. The energy–momentum δ-function will emerge from the position integral; it is not inserted by hand.

2.8 The tree amplitude of $\phi\phi \to \phi\phi$

2.8.1 External contractions and their normalisation

Before counting at the vertex we check every external-state factor. For an incoming one-particle state

$$\begin{aligned}\phi^{(+)}(x)|\vec{p}\rangle &= \int d\Pi_{\vec{q}}\; e^{-iq\cdot x}\,a_{\vec{q}}a^\dagger_{\vec{p}}|0\rangle = \int d\Pi_{\vec{q}}\; e^{-iq\cdot x}\,\Big[a_{\vec{q}},a^\dagger_{\vec{p}}\Big]\,|0\rangle \\ &= \int \frac{d^3q}{(2\pi)^3\,2E_{\vec{q}}}\; e^{-iq\cdot x}\,(2\pi)^3\,2E_{\vec{p}}\,\delta^{(3)}(\vec{q}-\vec{p})|0\rangle = e^{-ip\cdot x}|0\rangle.\end{aligned} \tag{2.85}$$

Thus $\langle 0|\phi(x)|\vec{p}\rangle = e^{-ip\cdot x}$, and taking the adjoint,

$$\langle\vec{k}|\phi^{(-)}(x) = e^{ik\cdot x}\langle 0|. \tag{2.86}$$

No factor $1/\sqrt{2E}$ remains: the one present in an oscillator-normalised mode expansion is cancelled exactly by the covariant normalisation of the external states. For a two-particle ket the first annihilation can remove either particle:

$$\begin{aligned}a_{\vec{q}}\,a^\dagger_{\vec{p}_1}a^\dagger_{\vec{p}_2}|0\rangle &= \delta_\Pi(\vec{q},\vec{p}_1)|\vec{p}_2\rangle + \delta_\Pi(\vec{q},\vec{p}_2)|\vec{p}_1\rangle, \\ \phi^{(+)}(x)|\vec{p}_1,\vec{p}_2\rangle &= e^{-ip_1\cdot x}|\vec{p}_2\rangle + e^{-ip_2\cdot x}|\vec{p}_1\rangle.\end{aligned} \tag{2.87}$$

A second annihilation part at the same vertex gives

$$\big(\phi^{(+)}(x)\big)^2|i\rangle = \big(e^{-ip_1\cdot x}e^{-ip_2\cdot x} + e^{-ip_2\cdot x}e^{-ip_1\cdot x}\big)|0\rangle = 2!\, e^{-i(p_1+p_2)\cdot x}|0\rangle, \tag{2.88}$$

and by Hermitian conjugation

$$\langle f|\big(\phi^{(-)}(x)\big)^2 = 2!\, e^{i(k_1+k_2)\cdot x}\langle 0|. \tag{2.89}$$

Their product is

$$\langle f|\big(\phi^{(-)}(x)\big)^2\big(\phi^{(+)}(x)\big)^2|i\rangle = 2!\,2!\, e^{i(k_1+k_2-p_1-p_2)\cdot x}. \tag{2.90}$$

These factorials count operator assignments. The symmetry factor for identical particles in the final state is a different object; it belongs to the cross section (Lecture 3).

2.8.2 The quartic vertex

The first Dyson term gives

$$\langle f|S_{[1]}|i\rangle = \frac{-i\lambda}{4!}\int d^4x\, \langle f|{:}\phi_I(x)^4{:}|i\rangle. \tag{2.91}$$

We suppress I again. Normal ordering allows the binomial expansion

$${:}\phi^4{:} = (\phi^{(-)})^4 + 4\,(\phi^{(-)})^3\phi^{(+)} + 6\,(\phi^{(-)})^2(\phi^{(+)})^2 + 4\,\phi^{(-)}(\phi^{(+)})^3 + (\phi^{(+)})^4; \tag{2.92}$$

this is the definition of the normal product after grouping the equivalent strings (for products that are not normal ordered the binomial formula would fail). A term with r creation and $4-r$ annihilation parts, acting on a two-particle ket, leaves the particle number

$$2 - (4-r) + r = 2r - 2. \tag{2.93}$$

An overlap with a two-particle bra requires $2r-2=2$, so $r=2$; terms that try to annihilate more particles than are present vanish. The relevant term is therefore

$${:}\phi^4{:} \longrightarrow \binom{4}{2}(\phi^{(-)})^2(\phi^{(+)})^2, \tag{2.94}$$

and with (2.90)

$$\langle f|{:}\phi(x)^4{:}|i\rangle = \binom{4}{2}2!\,2!\, e^{i(k_1+k_2-p_1-p_2)\cdot x} = 4!\, e^{i(k_1+k_2-p_1-p_2)\cdot x}. \tag{2.95}$$

The same count in words: choose two of the four fields to absorb the incoming particles (six choices), assign their two labels (2!), then assign the outgoing labels (2!). Finally

$$\langle f|S_{[1]}|i\rangle_{\text{conn}} = -i\lambda\int d^4x\, e^{i(k_1+k_2-p_1-p_2)\cdot x} = -i\lambda\,(2\pi)^4\,\delta^{(4)}(p_1+p_2-k_1-k_2). \tag{2.96}$$

The 4! from the assignments cancels the $1/4!$ in the definition of the interaction. Comparison with (2.84) gives

Tree amplitude of $\phi\phi \to \phi\phi$.

$$i\mathcal{M}_{\text{tree}} = -i\lambda, \qquad \mathcal{M}_{\text{tree}} = -\lambda. \tag{2.97}$$

This is the first complete interaction calculation: the vertex rule is now a consequence of operator perturbation theory and state normalisation.

2.8.3 What the diagram records

The amplitude is constant at this order: no internal line remains after all four fields of the only vertex are attached to external states (Figure 2.7). Three objects must be kept apart:

Object	Value at the computed order
vertex factor	$-i\lambda$
connected matrix element of $S-\mathbb{1}$	$-i\lambda\,(2\pi)^4\,\delta^{(4)}(P_i-P_f)$
invariant amplitude	$\mathcal{M}_{\text{tree}}=-\lambda$

Here $P_i=p_1+p_2$ and $P_f=k_1+k_2$.

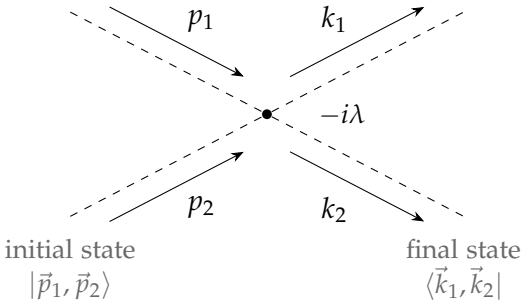


Figure 2.7: The tree contact diagram of $\phi\phi\to\phi\phi$; time runs from left to right. All four fields of the vertex $-i\lambda$ are attached to external particles, and the integral over the vertex position gives $(2\pi)^4\,\delta^{(4)}(p_1+p_2-k_1-k_2)$.

Why nothing disconnected was counted. The zeroth-order overlap (2.82) is the identity part of S and has been subtracted. A vacuum factor joins no external particle, and a spectator factor leaves an external particle disconnected from the interaction; neither belongs to the connected four-particle amplitude. In the normal-ordered first-order matrix element the only non-vanishing contribution uses all four vertex fields to absorb and emit the four external particles, so the direct calculation has selected precisely the desired connected term. If the interaction were not normal ordered, commutator terms within one vertex would also appear; they concern vacuum and lower-point contributions, not an extra tree contact amplitude. At higher orders normal ordering still permits contractions between different vertices; those corrections have been neither computed nor discarded here. In a more complicated graph a contraction between two interaction points supplies $D_F(x-y)$, or $i/(q^2-m^2+i0)$ in momentum space, with an internal momentum q that is integrated over and in general off shell; the external asymptotic particles, by contrast, are on shell.

2.9 Checks and the bridge to Lecture 3

Dimensions. Since $[\lambda]=0$, the four-scalar invariant amplitude is dimensionless. The vertex position integral has dimension -4, compensated by $[\phi^4]=4$ before the external matrix elements are taken. The overall $\delta^{(4)}$ has dimension -4, consistent with the covariant normalisation of the two-particle bra and ket.

Identical particles. The result is unchanged under $p_1 \leftrightarrow p_2$ or $k_1 \leftrightarrow k_2$. The field is bosonic, and the states were already symmetrised by the commuting creation operators; no extra factor belongs in $\mathcal{M}$ merely because the particles are identical.

Perturbative order. The amplitude is

$$\mathcal{M} = -\lambda + \mathcal{O}(\lambda^2), \qquad |\mathcal{M}|^2 = \lambda^2 + \mathcal{O}(\lambda^3) \tag{2.98}$$

when expanded at fixed kinematics and fixed coupling. The leading probability starts at second order in the coupling, although the amplitude required only one insertion of the interaction.

Conservation. The four-dimensional δ-function came from $\int d^4x\, e^{i(P_f - P_i)\cdot x}$. Translation invariance is the reason the amplitude conserves the total four-momentum, in accord with the translation current derived in Lecture 1.

What was achieved. The lecture has built the chain

$$\begin{aligned}
\text{ladder algebra} &\Rightarrow W(z) = \int d\Pi_{\vec{p}}\, e^{-ip\cdot z}, \\
\text{relativistic locality} &\Rightarrow C(z) = W(z) - W(-z) = 0 \quad (z^2 < 0), \\
\text{particles and antiparticles} &\Rightarrow \left[\phi(x), \phi^\dagger(y)\right] = W(x-y) - W(y-x), \\
\text{time ordering} &\Rightarrow D_F(p) = \frac{i}{p^2 - m^2 + i0}, \\
\text{Dyson expansion and operator contractions} &\Rightarrow \mathcal{M}_{\text{tree}} = -\lambda.
\end{aligned} \tag{2.99}$$

In the commutator lines the coefficient of $\mathbb{1}$ is understood. The causal response is instead $G_R = i\theta(t)\, C$, with a unit δ-function source. We can now calculate an amplitude from an explicitly defined quantum field theory, but not yet an event count or a cross section: the missing ingredients are the density of final states, the incident flux and the consistent counting of identical final configurations.

Summary

- $W(z) = \langle 0|\phi(x)\phi(y)|0\rangle = \int d\Pi_{\vec{p}}\, e^{-ip\cdot z}$ is Lorentz invariant and non-zero at space-like separation; at equal times $W(0,r) = mK_1(mr)/(4\pi^2 r)$, which tends to $1/(4\pi^2 r^2)$ for $m \to 0$.
- $C(z) = [\phi(x), \phi(y)] = [W(z) - W(-z)]\,\mathbb{1}$ is a c-number and vanishes for $z^2 < 0$: boost to equal times and reflect $\vec{p} \to -\vec{p}$. A response involves a commutator, so vacuum correlations supply no space-like signal.
- Complex field: $\left[\phi(x), \phi^\dagger(y)\right] = W(x-y) - W(y-x)$, a particle from y to x minus an antiparticle from x to y. At space-like separation the two amplitudes are equal and cancel: causality is restored by antiparticles. A real field is its own antiparticle.
- $D_F = \theta(z^0)W(z) + \theta(-z^0)W(-z) = \int d^4p\,(2\pi)^{-4}\, i\, e^{-ip\cdot z}/(p^2 - m^2 + i\epsilon)$: the ω contour is closed below for $t > 0$ (pole $E - i\epsilon/2E$) and above for $t < 0$ (pole $-E + i\epsilon/2E$).
- $(\Box + m^2)D_F = -i\delta^{(4)}$, both in momentum space and from the equal-time commutator; $G_R = i\theta(z^0)C$ with $(\Box + m^2)G_R = \delta^{(4)}$ is the causal response.
- Interaction picture: operators evolve with H_0 and are free fields; states evolve with $H_I(t)$, and $U_I = \mathrm{T}\exp[-i\int dt\, H_I]$, the $1/n!$ coming from the $n!$ time orderings.
- Wick: $\mathrm{T}\{\phi(x)\phi(y)\} = :\phi(x)\phi(y): + D_F(x-y)$; the vacuum value of four fields is $D_{12}D_{34} + D_{13}D_{24} + D_{14}D_{23}$, and $2n$ fields have $(2n)!/(2^n n!)$ complete pairings.
- $\langle f|(S - \mathbb{1})|i\rangle_{\text{conn}} = i(2\pi)^4\delta^{(4)}(P_i - P_f)\,\mathcal{M}$. External factors are $e^{\mp ip\cdot x}$ without $1/\sqrt{2E}$; the count $\binom{4}{2}\,2!\,2! = 4!$ cancels the $1/4!$, and $\mathcal{M}_{\text{tree}} = -\lambda$.

Report. Problem 3 builds on Sections 2.2 and 2.3.2: reconstruct the commutator cancellation and both checks of the Feynman Green equation, and state why vacuum correlations supply no space-like signal. Problem 4 builds on Sections 2.5 and 2.8: reconstruct the ordered second Dyson term and the vertex count $6 \times 2! \times 2! = 4!$, keeping the vertex, $S - \mathbb{1}$ and $\mathcal{M}$ apart. The report writes the oscillator-normalised $\tilde{a}$ as a; use the normalisation bridge at the start of this lecture when transferring a derivation.

Reading. Peskin–Schroeder, Sections 2.4, 4.2–4.4 and 4.6 (complex field: Problem 2.2); Lund lecture notes, Lectures 7–8 and 15–16; book, Chapters 5 and 10–11.

Next lecture: From amplitudes to observables: cross sections and decay rates, the Dirac field and the gauge principle, the QED Feynman rules and $e^+e^- \to \mu^+\mu^-$, and what the tree calculation leaves open.

From amplitudes to observables

Lecture 3

From amplitudes to observables

In this lecture

- The invariant master formula $d\sigma = |\mathcal{M}|^2\, d\Phi_2/\mathcal{F}$ for a measured rate, from box-normalised states, the squared δ-function and the incident flux.
- Two-body phase space, the counting of identical particles and the complete tree-level prediction $\sigma(\phi\phi \to \phi\phi) = \lambda^2/32\pi s$; decay rates and the width of $\Phi \to \phi\phi$.
- The Dirac equation from the relativistic dispersion relation: γ matrices, the spinors u and v, antiparticles and Fermi statistics.
- The gauge principle: from the conserved Dirac current to the QED Lagrangian and its vertex.
- The Feynman rules of QED and the first QED cross section: $\sigma = 4\pi\alpha^2/3s$ for $e^+e^- \to \mu^+\mu^-$.
- The first loop, power counting and renormalisation: what the tree approximation leaves open.

Builds on: Lectures 1 and 2 (covariant normalisation of states, Dyson series and Wick contractions, the tree amplitude $\mathcal{M} = -\lambda$).
In the book: Lund lecture notes, Lectures 9–15 and 17–20; book (R. Pasechnik, *Introduction to QFT*), Chapters 6–7 and 10–15.

How does an invariant amplitude become a number that an experiment measures? Lecture 2 ended with the leading amplitude of

$$\phi(p_1) + \phi(p_2) \longrightarrow \phi(k_1) + \phi(k_2), \qquad \mathcal{L} = \frac{1}{2}\,\partial_\mu\phi\,\partial^\mu\phi - \frac{1}{2}\,m^2\phi^2 - \frac{\lambda}{4!}\,\phi^4, \tag{3.1}$$

for a real scalar field with $m > 0$ and a weak coupling $\lambda > 0$; all external particles have mass m, and the interaction Hamiltonian is normal ordered as in Lecture 2. The dynamics is known: the contact amplitude is $-\lambda$. The remaining factors come from the normalisation of the states, the number of available final states and the flux of incident particles. We derive each of them in a finite box and watch the box drop out of the final formula; the same steps give decay rates. In the second half we change the field content. A scalar field does not describe the spin, charge and statistics of an electron, so we construct the Dirac field and its gauge coupling, collect the Feynman rules of QED and use the master formula of the first half for the first QED cross section. The lecture ends with the first loop and with what the tree approximation leaves open. All derivations are written out to support independent study.

Starting data. We use $\hbar = c = 1$ and $g_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$. From Lectures 1 and 2,

$$d\Pi_{\vec{p}} = \frac{d^3p}{(2\pi)^3\, 2E_{\vec{p}}}, \qquad E_{\vec{p}} = \sqrt{\vec{p}^{\,2} + m^2},$$

$$\left[a_{\vec{p}}, a^\dagger_{\vec{q}}\right] = (2\pi)^3\, 2E_{\vec{p}}\, \delta^{(3)}(\vec{p}-\vec{q}), \qquad |\vec{p}\rangle = a^\dagger_{\vec{p}}|0\rangle, \tag{3.2}$$

$$\phi(x) = \int d\Pi_{\vec{p}}\, \left(a_{\vec{p}}\, e^{-ip\cdot x} + a^\dagger_{\vec{p}}\, e^{ip\cdot x}\right). \tag{3.3}$$

With $P_i = p_1 + p_2$ and $P_f = k_1 + k_2$, the connected matrix element defines the invariant amplitude, computed at tree level in Lecture 2:

$$\langle \vec{k}_1, \vec{k}_2 | (S - \mathbb{1}) | \vec{p}_1, \vec{p}_2 \rangle_{\text{conn}} = i\,(2\pi)^4\, \delta^{(4)}(P_i - P_f)\, \mathcal{M}, \qquad \mathcal{M}_{\text{tree}} = -\lambda. \tag{3.4}$$

The factor 4! from attaching the four external fields cancelled the $1/4!$ of the interaction. We now turn this amplitude into a probability per unit time and per incident flux.

Convention (normalisation bridge to the report). The written report uses oscillator-normalised ladder operators and calls them a; in the present notation they are $\tilde{a}_{\vec{p}} = a_{\vec{p}}/\sqrt{2E_{\vec{p}}}$. This changes intermediate factors, not physical predictions. In particular, the box-normalised states of Section 3.1 are the same in both conventions: $|\vec{p}\rangle/\sqrt{2E_{\vec{p}}V} = \tilde{a}^\dagger_{\vec{p}}|0\rangle/\sqrt{V}$.

3.1 What a cross section measures

Let n_1 and n_2 be the particle number densities of two dilute, collinear beams (Figure 3.1a). For an overlap volume V_{ov} and the closing speed $v_{\text{cl}} = |v_1 - v_2|$, the elementary event-rate relation is

$$R = n_1 n_2\, V_{\text{ov}}\, v_{\text{cl}}\, \sigma \equiv \mathcal{L}_{\text{lum}}\, \sigma, \qquad [R] = 1, \quad [\mathcal{L}_{\text{lum}}] = 3, \quad [\sigma] = -2. \tag{3.5}$$

Square brackets denote mass dimensions, R is the number of events per unit time and $\mathcal{L}_{\text{lum}}$ the luminosity. The relation assumes that the densities are effectively constant over the overlap region; for non-uniform beams one integrates the product of the densities over space. The cross section has the dimension of an area: it is the effective size of one particle as seen by the flux of the other beam.

To derive the microscopic factor we place one particle of each incoming momentum in a large periodic box of volume $V = L^3$ (Figure 3.1b), in which the momenta are $2\pi\vec{n}/L$ with $\vec{n} \in \mathbb{Z}^3$. In the box the continuum shorthand $(2\pi)^3\delta^{(3)}(\vec{0})$ stands for V; this follows from the finite-box normalisation and is not a numerical value of a δ-distribution. For distinct momentum modes, with $E_a \equiv E_{\vec{p}_a}$,

$$\langle \vec{p} | \vec{p} \rangle = 2E_{\vec{p}}\, V, \qquad |\vec{p}\rangle_V = \frac{|\vec{p}\rangle}{\sqrt{2E_{\vec{p}}V}}, \qquad |\vec{p}_1, \vec{p}_2\rangle_V = \frac{|\vec{p}_1, \vec{p}_2\rangle}{\sqrt{(2E_1 V)(2E_2 V)}}. \tag{3.6}$$

There is one target particle and one incoming particle per volume V, so the incident number flux per target particle is

$$j_{\text{inc}} = \frac{v_{\text{cl}}}{V}, \qquad dR = j_{\text{inc}}\, d\sigma. \tag{3.7}$$

The dimensions check: $[j_{\text{inc}}] = 3$ and $[d\sigma] = -2$ give a rate of dimension one.

State counting. The allowed momenta form a lattice with cell volume $(2\pi)^3/V$. Hence

$$\sum_{\vec{k}} \longrightarrow V\int \frac{d^3k}{(2\pi)^3}, \qquad dN_f = \frac{V\,d^3k_1}{(2\pi)^3}\,\frac{V\,d^3k_2}{(2\pi)^3}. \tag{3.8}$$

For now these are *labelled* outgoing momenta; the double description of a state of identical particles is removed explicitly in Section 3.3.2. The amplitude refers to specified external states, whereas a detector accepts a range of them: the density of states (3.8) accounts for this range.

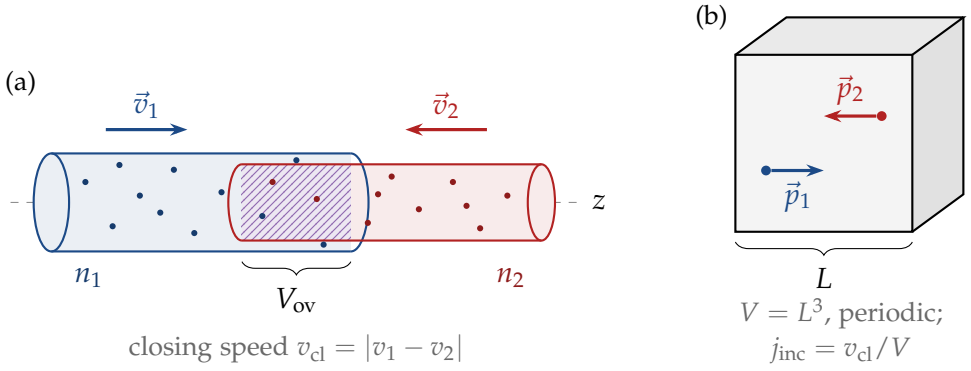


Figure 3.1: (a) Two dilute beams with densities n_1 and n_2 moving along z with velocities $\vec{v}_1$ and $\vec{v}_2$; events occur in the overlap volume $V_{\rm ov}$ (hatched) at the rate $R = n_1 n_2 V_{\rm ov} v_{\rm cl}\sigma$. (b) The idealisation used in the derivation: one particle of each momentum in a periodic box of volume $V = L^3$, so that the incident flux per target particle is $j_{\rm inc} = v_{\rm cl}/V$.

3.1.1 The squared δ-function: keep the regulator until it cancels

The square of the matrix element (3.4) contains the square of a δ-function. We keep a finite observation time T and the finite volume V until they cancel. With $\Delta E = E_f - E_i$ let

$$K_T(\Delta E) = \int_{-T/2}^{T/2} dt\; e^{i\Delta E\, t} = \frac{2\sin(\Delta E\, T/2)}{\Delta E}; \tag{3.9}$$

its value at $\Delta E = 0$ is the limit T. A rate requires $|K_T|^2/T$, not $\delta(\Delta E)^2$. For a smooth test function f we change variables to $u = \Delta E\, T/2$:

$$\int \frac{d\Delta E}{2\pi}\,\frac{|K_T(\Delta E)|^2}{T}\,f(\Delta E) = \int_{-\infty}^{\infty}\frac{du}{\pi}\left(\frac{\sin u}{u}\right)^2 f(2u/T) \xrightarrow[T\to\infty]{} f(0). \tag{3.10}$$

The normalisation follows without a table of integrals:

$$\int \frac{d\omega}{2\pi}\,|K_T(\omega)|^2 = \int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt'\;\delta(t-t') = T. \tag{3.11}$$

Together with the concentration near $\Delta E = 0$, this gives, in the distributional sense,

$$\frac{|K_T(\Delta E)|^2}{T} \longrightarrow 2\pi\,\delta(\Delta E), \tag{3.12}$$

the energy-conserving factor of Fermi's golden rule: the time factor becomes a transition rate.

For space, let $\vec{q} = \vec{P}_f - \vec{P}_i$ and

$$K_V(\vec{q}) = \int_V d^3x\, e^{-i\vec{q}\cdot\vec{x}}. \tag{3.13}$$

For periodic-box momenta $K_V = V\,\delta_{\vec{q},\vec{0}}$, and $|K_V|^2 = V\,K_V$ exactly. In the continuum

$$|K_V(\vec{q})|^2 \;\longrightarrow\; V\,(2\pi)^3\delta^{(3)}(\vec{q}), \qquad \frac{|K_T K_V|^2}{T} \;\longrightarrow\; V\,(2\pi)^4\delta^{(4)}(P_f - P_i). \tag{3.14}$$

Each arrow is the limit of a large box and a long time, taken after the integration over the relevant final-state cell; no product of singular distributions is ever formed. The one remaining factor V will cancel the normalisation of the incident flux: an infinite normalisation volume cannot survive in a measured cross section.

> **Remark (wave packets).** This is the perturbative long-time construction of a rate. Localised wave packets lead to the same master formula without idealised plane waves (Lund lecture notes, Lecture 19); the box is used here because it exposes every normalisation factor explicitly.

3.2 From the box amplitude to the master formula

Denote the outgoing energies by $\omega_a = E_{\vec{k}_a}$, $a = 1, 2$. Between the box-normalised states (3.6), and with the δ-function of (3.4) in its regulated form $K_T K_V$, the transition amplitude is

$$A_{fi}^{(V,T)} = \frac{i\mathcal{M}\,K_T(\Delta E)\,K_V(\vec{q})}{\sqrt{(2E_1V)(2E_2V)(2\omega_1V)(2\omega_2V)}}. \tag{3.15}$$

For the contact amplitude $\mathcal{M}$ does not vary across the narrow energy regulator. We square, divide by T, use (3.14) and multiply by the number of final states (3.8):

$$\begin{aligned} dR_{\text{ord}} &= \lim \frac{|A_{fi}^{(V,T)}|^2}{T}\,\frac{V\,d^3k_1}{(2\pi)^3}\,\frac{V\,d^3k_2}{(2\pi)^3} \\ &= \frac{|\mathcal{M}|^2}{(2E_1)(2E_2)\,V}\,(2\pi)^4\delta^{(4)}(P_i - P_f)\,\frac{d^3k_1}{(2\pi)^3\,2\omega_1}\,\frac{d^3k_2}{(2\pi)^3\,2\omega_2} \\ &\equiv \frac{|\mathcal{M}|^2}{4E_1E_2V}\,d\Phi_2. \end{aligned} \tag{3.16}$$

Of the four factors V in the denominator, one is cancelled by the squared δ-function and two by the state counting. The subscript "ord" indicates that all ordered final-momentum labels are integrated over. Dividing by $j_{\text{inc}} = v_{\text{cl}}/V$,

$$d\sigma_{\text{ord}} = \frac{|\mathcal{M}|^2}{4E_1E_2\,v_{\text{cl}}}\,d\Phi_2, \tag{3.17}$$

and all factors of V and T have cancelled.

The flux in invariant form. For collinear beams $v_a = p_a^z/E_a$ and $m_a^2 = E_a^2 - (p_a^z)^2$; we allow different masses m_1, m_2, as needed later for e^+e^- collisions. Expanding the invariant combination,

$$\begin{aligned} (p_1\cdot p_2)^2 - m_1^2m_2^2 &= (E_1E_2 - p_1^zp_2^z)^2 - \big(E_1^2 - (p_1^z)^2\big)\big(E_2^2 - (p_2^z)^2\big) \\ &= (E_2p_1^z - E_1p_2^z)^2 = E_1^2E_2^2\,(v_1 - v_2)^2. \end{aligned} \tag{3.18}$$

The flux factor $4E_1E_2v_{\text{cl}}$ is therefore a Lorentz invariant:

Master formula for $2 \to 2$ cross sections.

$$d\sigma_{\text{ord}} = \frac{|\mathcal{M}|^2}{\mathcal{F}}\, d\Phi_2, \qquad \mathcal{F} = 4\sqrt{(p_1\cdot p_2)^2 - m_1^2 m_2^2} = 4E_1E_2\, v_{\text{cl}}, \tag{3.19}$$

$$d\Phi_2 = (2\pi)^4\, \delta^{(4)}(P_i - k_1 - k_2)\, \frac{d^3k_1}{(2\pi)^3\, 2\omega_1}\, \frac{d^3k_2}{(2\pi)^3\, 2\omega_2}. \tag{3.20}$$

In the ϕ^4 example $m_1 = m_2 = m$. For non-collinear beams the velocity factor becomes $v_{\text{M}} = \sqrt{(\vec{v}_1 - \vec{v}_2)^2 - (\vec{v}_1 \times \vec{v}_2)^2}$, with $\mathcal{F} = 4E_1E_2v_{\text{M}}$. We work in the centre-of-mass frame, where the beams are collinear; the invariant expression for $\mathcal{F}$ is the one to retain.

3.3 Centre-of-mass kinematics and two-body phase space

In the centre-of-mass (CM) frame the total four-momentum and the incoming momenta are

$$P_i^\mu = (\sqrt{s}, \vec{0}), \qquad s = (p_1 + p_2)^2, \qquad p_1 = (E, \vec{p}), \quad p_2 = (E, -\vec{p}). \tag{3.21}$$

For elastic equal-mass scattering $k_1 = (E, \vec{k})$ and $k_2 = (E, -\vec{k})$ (Figure 3.2). Energy conservation fixes the common magnitude:

$$2E = \sqrt{s}, \qquad E^2 = p^2 + m^2 \quad\Longrightarrow\quad p \equiv |\vec{p}\,| = |\vec{k}| = \frac{1}{2}\sqrt{s - 4m^2},$$
$$\beta_{\text{cm}} \equiv \frac{p}{E} = \sqrt{1 - \frac{4m^2}{s}}, \qquad s > 4m^2. \tag{3.22}$$

β_{cm} is the speed of each particle in the CM frame, in units of c. We choose $\vec{p}_1$ along $+z$; the outgoing direction $\hat{k} = \vec{k}/|\vec{k}|$ has the polar angle θ and the azimuth φ, and $d\Omega = \sin\theta\, d\theta\, d\varphi$. The flux factor follows from $p_1 \cdot p_2 = E^2 + p^2 = (s - 2m^2)/2$:

$$(p_1\cdot p_2)^2 - m^4 = \frac{(s-2m^2)^2 - 4m^4}{4} = \frac{s\,(s-4m^2)}{4} \quad\Longrightarrow\quad \mathcal{F} = 2s\,\beta_{\text{cm}} = 4p\sqrt{s}. \tag{3.23}$$

The closing speed $v_{\text{cl}} = 2p/E = 2\beta_{\text{cm}}$ can approach 2: it is the rate at which the separation changes in this frame, not the speed of one particle seen from the other.

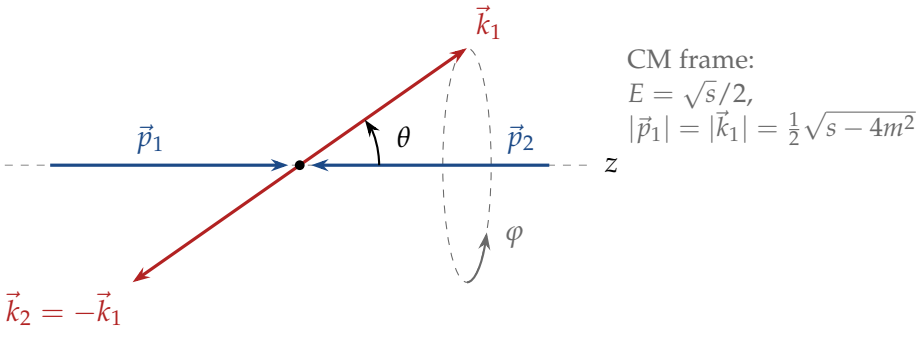


Figure 3.2: Kinematics of elastic $2 \to 2$ scattering in the CM frame. The incoming momenta (blue) are back to back along z, and so are the outgoing ones (red), with the same magnitude for equal masses. The final state is fixed by the direction of $\vec{k}_1$: the polar angle θ and the azimuth φ around the beam axis (dashed circle).

3.3.1 Two-body phase space: perform both δ-integrations

We begin with the labelled measure (3.20). It is Lorentz invariant, because each factor $d^3k/2\omega$ is the positive-energy on-shell measure of Lecture 1, and dimensionless: each measure has dimension two, the four-dimensional δ-function dimension minus four. The spatial δ-function sets $\vec{k}_2 = -\vec{k}_1$. With $\vec{k}_1 = \vec{k}$, $k = |\vec{k}|$ and $\omega(k) = \sqrt{k^2 + m^2}$,

$$\begin{aligned} d\Phi_2 &= \frac{(2\pi)^4}{(2\pi)^6} \frac{d^3k}{4\,\omega(k)^2}\, \delta\big(\sqrt{s} - 2\omega(k)\big) \\ &= \frac{1}{16\pi^2} \frac{k^2\, dk}{\omega(k)^2}\, \delta\big(\sqrt{s} - 2\omega(k)\big)\, d\Omega. \end{aligned} \tag{3.24}$$

There is one positive radial root,

$$k_* = \frac{1}{2}\sqrt{s - 4m^2}, \qquad \omega_* = \frac{\sqrt{s}}{2}, \tag{3.25}$$

and for a simple root

$$\delta\big(f(k)\big) = \frac{\delta(k - k_*)}{|f'(k_*)|}, \qquad f'(k) = -\frac{2k}{\omega(k)}. \tag{3.26}$$

Consequently

$$\int_0^\infty dk\, \frac{k^2}{\omega(k)^2}\, \delta\big(\sqrt{s} - 2\omega(k)\big) = \frac{k_*^2}{\omega_*^2}\, \frac{\omega_*}{2k_*} = \frac{k_*}{2\omega_*} = \frac{k_*}{\sqrt{s}}, \tag{3.27}$$

and

$$\frac{d\Phi_2}{d\Omega} = \frac{k_*}{16\pi^2\sqrt{s}} = \frac{\beta_{\mathrm{cm}}}{32\pi^2}, \qquad \int d\Phi_2 = \frac{\beta_{\mathrm{cm}}}{8\pi}. \tag{3.28}$$

The second result integrates over the full solid angle 4π; no identical-particle factor has been included yet. The factors have three origins: $(2\pi)^{-2}$ after the spatial δ-integration, $(4\omega^2)^{-1}$ from the normalisation of the states and $\omega_*/(2k_*)$ from the Jacobian of the radial δ-function. A factor missed in any one place changes the cross section.

3.3.2 Identical particles: count events once

For two scalar bosons the creation operators commute:

$$|\vec{k}_1, \vec{k}_2\rangle = a^\dagger_{\vec{k}_1} a^\dagger_{\vec{k}_2} |0\rangle = a^\dagger_{\vec{k}_2} a^\dagger_{\vec{k}_1} |0\rangle = |\vec{k}_2, \vec{k}_1\rangle. \tag{3.29}$$

Integrating over all labelled $(\vec{k}_1, \vec{k}_2)$ therefore counts every physical final state twice. Equivalently, the two-particle completeness relation reads

$$\mathbb{1}_2 = \frac{1}{2!} \int d\Pi_{\vec{k}_1}\, d\Pi_{\vec{k}_2}\, |\vec{k}_1, \vec{k}_2\rangle\langle \vec{k}_1, \vec{k}_2|. \tag{3.30}$$

To check the factor, apply the right-hand side to $|\vec{p}_1, \vec{p}_2\rangle$: the overlap $\langle \vec{k}_1, \vec{k}_2 | \vec{p}_1, \vec{p}_2\rangle$ has two terms, the direct and the exchanged pairing (Lecture 2), and the integral gives $2|\vec{p}_1, \vec{p}_2\rangle$ before the multiplication by $1/2!$.

Convention (identical particles in the final state). We keep the labelled final momenta over the full sphere and include the factor once:

$$d\sigma = \frac{1}{2!}\,\frac{|\mathcal{M}|^2}{\mathcal{F}}\,d\Phi_2, \qquad \sigma = \int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta\, d\theta\, \frac{d\sigma}{d\Omega}. \tag{3.31}$$

Then $d\sigma/d\Omega$ is a symmetrised event distribution: each physical pair contributes one half at each of its two opposite momentum directions. Counting one representative direction per pair over a hemisphere without the $1/2!$ gives the same σ, since $\frac{1}{2}\int_{4\pi} d\Omega\, f = \int_{\text{hemisphere}} d\Omega\, f$ for $f(\hat{k}) = f(-\hat{k})$. A single-particle inclusive distribution counts particles rather than events: it is twice ours and integrates to 2σ.

There is no factor $1/2!$ for the two incident particles: the specified incoming beam modes already define one collision state.

3.4 The scalar cross section and its limits

We now combine the three independently derived ingredients, (3.4), (3.23) and (3.28), with the counting (3.31):

Tree-level cross section of $\phi\phi \to \phi\phi$.

$$|\mathcal{M}_{\text{tree}}|^2 = \lambda^2, \qquad \mathcal{F} = 2s\,\beta_{\text{cm}}, \qquad \frac{d\Phi_2}{d\Omega} = \frac{\beta_{\text{cm}}}{32\pi^2}, \tag{3.32}$$

$$\frac{d\sigma}{d\Omega} = \frac{1}{2}\,\frac{\lambda^2}{2s\,\beta_{\text{cm}}}\,\frac{\beta_{\text{cm}}}{32\pi^2} = \frac{\lambda^2}{128\pi^2 s}, \tag{3.33}$$

$$\sigma = \frac{\lambda^2}{128\pi^2 s}\int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta\, d\theta = \frac{\lambda^2}{32\pi s} \qquad (s > 4m^2). \tag{3.34}$$

The suppression of the outgoing phase space cancels that of the incident flux: both are proportional to β_{cm} because the scattering is elastic and the masses are equal. For an inelastic channel the two velocity factors differ.

Physical checks. The contact amplitude has no preferred outgoing direction, so the distribution is isotropic. Since $[\lambda] = 0$ and $[s] = 2$, $[\sigma] = -2$, the dimension of an area. At fixed weak coupling the tree prediction falls as

$$\sigma \propto \frac{1}{s} \qquad (s \gg m^2). \tag{3.35}$$

For the elastic equal-mass process the $1/s$ form is in fact exact at this order; it does not show that loops may be neglected at arbitrarily high energies (Section 3.10).

Approach to the threshold from above.

$$s \downarrow 4m^2: \qquad \beta_{\text{cm}} \to 0, \qquad \mathcal{F} \to 0, \qquad \int d\Phi_2 \to 0,$$

$$\lim_{s\downarrow 4m^2} \sigma = \frac{\lambda^2}{128\pi m^2}, \qquad v_{\text{cl}}\,\sigma = 2\beta_{\text{cm}}\,\sigma \to 0. \tag{3.36}$$

The finite threshold value is the limit of a ratio. At exactly zero incident flux the ratio is not operationally defined by colliding stationary plane waves, and for fixed beam densities the event rate vanishes (Figure 3.3). We now have an observable prediction from a specified quantum field theory. A number requires m, λ and the collision energy, and the model is a scalar one, not a model of real electrons.

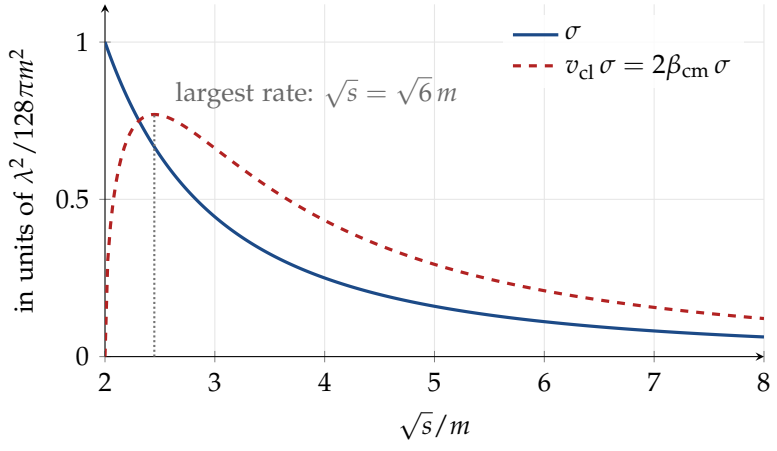


Figure 3.3: The tree-level cross section (3.34) of $\phi\phi \to \phi\phi$ (solid) and the event rate per unit beam densities and overlap volume, $v_{\rm cl}\sigma = R/(n_1 n_2 V_{\rm ov})$ in the CM frame (dashed), in units of the threshold value $\lambda^2/128\pi m^2$. The cross section tends to this finite value at threshold, $\sqrt{s} = 2m$, while the rate vanishes there with $\beta_{\rm cm}$.

3.4.1 Invariant kinematics

The Mandelstam invariants

$$s = (p_1 + p_2)^2, \qquad t = (p_1 - k_1)^2, \qquad u = (p_1 - k_2)^2 \tag{3.37}$$

describe the process in any frame. With $p_1 + p_2 = k_1 + k_2$ and $p_a^2 = k_a^2 = m^2$, expanding the three squares gives

$$s + t + u = 6m^2 + 2p_1 \cdot (p_2 - k_1 - k_2) = 6m^2 + 2p_1 \cdot (-p_1) = 4m^2, \tag{3.38}$$

so only two of the invariants are independent. In the CM frame

$$\begin{aligned} p_1 \cdot k_1 &= E^2 - p^2\cos\theta, \qquad p_1 \cdot k_2 = E^2 + p^2\cos\theta, \\ t &= 2m^2 - 2E^2 + 2p^2\cos\theta = -2p^2(1 - \cos\theta), \\ u &= 2m^2 - 2E^2 - 2p^2\cos\theta = -2p^2(1 + \cos\theta), \qquad s = 4(p^2 + m^2). \end{aligned} \tag{3.39}$$

Thus $-(s - 4m^2) \le t, u \le 0$. Exchanging the two identical outgoing particles sends $\cos\theta \to -\cos\theta$ and $t \leftrightarrow u$; the constant contact amplitude respects this symmetry. For massless particles $p^2 = s/4$, and

$$t = -\frac{s}{2}(1 - \cos\theta), \qquad u = -\frac{s}{2}(1 + \cos\theta) \qquad (m = 0), \tag{3.40}$$

which we shall need for $e^+e^- \to \mu^+\mu^-$.

The argument in five lines. The calculation is summarised by the following chain. Explain the reason for each arrow:

$$\begin{aligned} \mathcal{L}_{\text{int}} = -\lambda\phi^4/4! \;&\longrightarrow\; \mathcal{M} = -\lambda, \\ |\vec{p}\rangle_V = |\vec{p}\rangle/\sqrt{2E_{\vec{p}}V} \;&\longrightarrow\; dR_{\text{ord}} = \frac{|\mathcal{M}|^2}{4E_1E_2V}\,d\Phi_2, \\ j_{\text{inc}} = v_{\text{cl}}/V \;&\longrightarrow\; d\sigma_{\text{ord}} = \frac{|\mathcal{M}|^2}{\mathcal{F}}\,d\Phi_2, \\ |\vec{k}_1,\vec{k}_2\rangle = |\vec{k}_2,\vec{k}_1\rangle \;&\longrightarrow\; d\sigma = \tfrac{1}{2}\,d\sigma_{\text{ord}}, \\ \mathcal{F} = 2s\beta_{\text{cm}}, \quad d\Phi_2 = \frac{\beta_{\text{cm}}}{32\pi^2}\,d\Omega \;&\longrightarrow\; \sigma = \frac{\lambda^2}{32\pi s}. \end{aligned} \tag{3.41}$$

3.5 Decay rates

The same box steps give the decay rate of an unstable particle A of mass M. Its partial decay rate $\Gamma(A \to f)$ is the number of decays into the final state f per unit time divided by the number of particles A, in the rest frame of A. Summed over all channels, the total rate Γ gives the exponential law $N(t) = N(0)\,e^{-\Gamma t}$ and the lifetime $\tau = 1/\Gamma$. The initial state is now one normalised particle at rest, $P = (M,\vec{0})$, and the final state contains n particles with energies ω_a:

$$|P\rangle_V = \frac{|P\rangle}{\sqrt{2MV}}, \qquad A_{fi}^{(V,T)} = \frac{i\mathcal{M}\,K_T(\Delta E)\,K_V(\vec{q})}{\sqrt{(2MV)\prod_{a=1}^{n}(2\omega_a V)}}, \tag{3.42}$$

where $\mathcal{M}$ is defined as in (3.4), with a single particle in the initial state. We square, divide by T with (3.14) and multiply by the number of final states:

$$\begin{aligned} d\Gamma &= \lim \frac{|A_{fi}^{(V,T)}|^2}{T} \prod_{a=1}^{n} \frac{V\,d^3k_a}{(2\pi)^3} \\ &= \frac{|\mathcal{M}|^2\,V}{2MV}\,(2\pi)^4\delta^{(4)}\!\left(P - \textstyle\sum_a k_a\right) \prod_{a=1}^{n} \frac{d^3k_a}{(2\pi)^3\,2\omega_a} \equiv \frac{|\mathcal{M}|^2}{2M}\,d\Phi_n. \end{aligned} \tag{3.43}$$

Each final-state normalisation $1/V$ cancels against its counting factor V, and the factor V of the squared δ-function cancels the normalisation of the single initial particle. No flux is needed: the rate per decaying particle is itself the observable. In a frame in which A moves with energy E_A, the steps give $1/2E_A$ instead of $1/2M$; since $|\mathcal{M}|^2$ and $d\Phi_n$ are invariant, $\Gamma(E_A) = (M/E_A)\,\Gamma$, which is relativistic time dilation.

Two-body decays. For $A \to 1+2$ with masses m_1 and m_2 we repeat the steps of Section 3.3.1 with $\sqrt{s} \to M$. The spatial δ-function sets $\vec{k}_2 = -\vec{k}_1 \equiv -\vec{k}$; with $k = |\vec{k}|$ and $\omega_a(k) = \sqrt{k^2 + m_a^2}$,

$$\begin{aligned} d\Phi_2 &= \frac{(2\pi)^4}{(2\pi)^6}\,\frac{d^3k}{4\,\omega_1\omega_2}\,\delta\big(M - \omega_1 - \omega_2\big) \\ &= \frac{1}{16\pi^2}\,\frac{k^2\,dk}{\omega_1\omega_2}\,\delta\big(M - \omega_1(k) - \omega_2(k)\big)\,d\Omega, \\ \left|\frac{d(\omega_1+\omega_2)}{dk}\right| &= \frac{k}{\omega_1} + \frac{k}{\omega_2} = \frac{k\,(\omega_1+\omega_2)}{\omega_1\omega_2} = \frac{k\,M}{\omega_1\omega_2} \qquad \text{at the root, } \omega_1+\omega_2 = M. \end{aligned} \tag{3.44}$$

At the root $k = |\vec{p}^{\,*}|$ the radial integral therefore gives $(|\vec{p}^{\,*}|^2/\omega_1\omega_2)\,(\omega_1\omega_2/|\vec{p}^{\,*}|M)$, and

$$\frac{d\Phi_2}{d\Omega} = \frac{|\vec{p}^{\,*}|}{16\pi^2 M}. \tag{3.45}$$

The root follows from squaring $k_2 = P - k_1$, which gives $m_2^2 = M^2 - 2M\omega_1 + m_1^2$:

$$\begin{aligned} \omega_1 &= \frac{M^2 + m_1^2 - m_2^2}{2M}, \\ |\vec{p}^{\,*}|^2 = \omega_1^2 - m_1^2 &= \frac{(M^2 + m_1^2 - m_2^2)^2 - 4M^2 m_1^2}{4M^2} = \frac{\lambda(M^2, m_1^2, m_2^2)}{4M^2}, \end{aligned} \tag{3.46}$$

with the Källén function $\lambda(x,y,z) = x^2 + y^2 + z^2 - 2xy - 2yz - 2zx$. All invariants built from P, k_1 and k_2 are fixed by the masses (for example $2k_1 \cdot k_2 = M^2 - m_1^2 - m_2^2$), so for a spinless A, or after the average over its spin, $|\mathcal{M}|^2$ is a constant and the angular integral gives 4π:

$$\Gamma(A \to 1+2) = \frac{1}{2M}\,|\mathcal{M}|^2\,\frac{4\pi\,|\vec{p}^{\,*}|}{16\pi^2 M} = \frac{|\vec{p}^{\,*}|}{8\pi M^2}\,|\mathcal{M}|^2, \tag{3.47}$$

times $\frac{1}{2}$ for identical final particles, which are counted as in Section 3.3.2.

Example: $\Phi \to \phi\phi$. Let a heavy scalar Φ of mass $M > 2m$ couple to ϕ through

$$\mathcal{L}_{\text{int}} = -\frac{\mu}{2}\,\Phi\,\phi^2, \qquad [\mu] = 1. \tag{3.48}$$

The first Dyson term contains $:\Phi\phi^2: \supset (\phi^{(-)})^2\,\Phi^{(+)}$. As in Lecture 2, $\Phi^{(+)}(x)$ absorbs the incoming Φ with the factor $e^{-iP\cdot x}$, and the two outgoing particles are emitted by the two fields $\phi^{(-)}(x)$ in 2! ways:

$$\begin{aligned} \langle \vec{k}_1, \vec{k}_2 | S_{[1]} | P \rangle &= \frac{-i\mu}{2} \int d^4x\, \langle \vec{k}_1, \vec{k}_2 | :\Phi\phi^2: | P \rangle = \frac{-i\mu}{2}\, 2! \int d^4x\, e^{i(k_1+k_2-P)\cdot x} \\ &= -i\mu\,(2\pi)^4\,\delta^{(4)}(P - k_1 - k_2). \end{aligned} \tag{3.49}$$

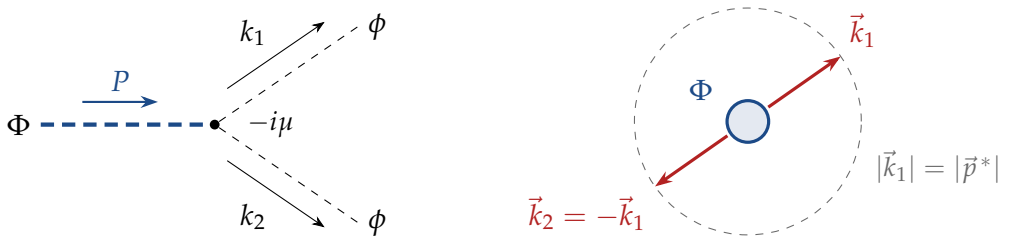
Hence $\mathcal{M} = -\mu$ (Figure 3.4). With $|\vec{p}^{\,*}| = \frac{1}{2}\sqrt{M^2 - 4m^2}$ and the factor $\frac{1}{2}$ for the identical final particles,

$$\Gamma(\Phi \to \phi\phi) = \frac{1}{2}\,\frac{\mu^2\,|\vec{p}^{\,*}|}{8\pi M^2} = \frac{\mu^2\,|\vec{p}^{\,*}|}{16\pi M^2} = \frac{\mu^2}{32\pi M}\sqrt{1 - \frac{4m^2}{M^2}}. \tag{3.50}$$

The width has the dimension of a rate, $[\Gamma] = 2 + 1 - 2 = 1$, and vanishes at the threshold $M = 2m$.

Decay rates.

$$\begin{aligned} d\Gamma &= \frac{|\mathcal{M}|^2}{2M}\,d\Phi_n, \qquad |\vec{p}^{\,*}| = \frac{\lambda^{1/2}(M^2, m_1^2, m_2^2)}{2M}, \\ \Gamma(A \to 1+2) &= \frac{|\vec{p}^{\,*}|}{8\pi M^2}\,|\mathcal{M}|^2 \qquad \left(\times \tfrac{1}{2} \text{ for identical final particles}\right). \end{aligned} \tag{3.51}$$



(a) $\Phi \to \phi\phi$: $\mathcal{M} = -\mu$ (b) rest frame, $P = (M, \vec{0})$

Figure 3.4: (a) The decay $\Phi \to \phi\phi$ at first order: the vertex $-i\mu$ follows from $\mathcal{L}_{\text{int}} = -\frac{\mu}{2}\Phi\phi^2$ (heavy dashed line: Φ; dashed: ϕ). (b) In the rest frame of the decaying particle the final momenta are back to back with the fixed magnitude $|\vec{p}^*| = \lambda^{1/2}(M^2, m_1^2, m_2^2)/2M$; only their direction (dashed sphere) is integrated over, which gives $d\Phi_2/d\Omega = |\vec{p}^*|/16\pi^2 M$.

3.6 The Dirac equation

We now change the field content. The logic from action to amplitude survives, but a scalar field does not describe the spin, charge and statistics of an electron. We derive the elementary Dirac and gauge structures needed to recognise QED; the general theory of spinor representations and the canonical quantisation of the Dirac field are quoted. The first task is an equation linear in the time and space derivatives whose solutions obey the relativistic dispersion relation (Figure 3.5). Let $m_f > 0$ be the fermion mass, distinct from the scalar mass m.

We start from a Schrödinger-like equation for a field with several components,

$$i\partial_t \psi = h_D \psi, \qquad h_D = \vec{\alpha}_D \cdot (-i\nabla) + \beta_D m_f. \tag{3.52}$$

The matrices α_D^i and β_D are constant and Hermitian; they act on the components of ψ, not on position. We require $h_D^2 = (-\nabla^2 + m_f^2)\,\mathbb{1}$, so that every component obeys the Klein–Gordon equation. In momentum space, with $h_D(\vec{p}) = \vec{\alpha}_D \cdot \vec{p} + \beta_D m_f$ and the Cartesian components p_i of $\vec{p}$ (summed over repeated indices),

$$\begin{aligned} h_D(\vec{p})^2 &= \alpha_D^i \alpha_D^j p_i p_j + m_f p_i (\alpha_D^i \beta_D + \beta_D \alpha_D^i) + m_f^2 \beta_D^2 \\ &= \frac{1}{2}\left\{\alpha_D^i, \alpha_D^j\right\} p_i p_j + m_f p_i \left\{\alpha_D^i, \beta_D\right\} + m_f^2 \beta_D^2. \end{aligned} \tag{3.53}$$

Equality with $\vec{p}^{\,2} + m_f^2$ for every momentum requires

$$\left\{\alpha_D^i, \alpha_D^j\right\} = 2\delta^{ij}\,\mathbb{1}, \qquad \left\{\alpha_D^i, \beta_D\right\} = 0, \qquad \beta_D^2 = \mathbb{1}, \tag{3.54}$$

where $\{A, B\} = AB + BA$. Ordinary numbers cannot satisfy these relations; matrices are necessary. The relations also restrict their size N. Since $(\alpha_D^1)^2 = \mathbb{1}$, cyclicity of the trace gives $\operatorname{Tr}\beta_D = \operatorname{Tr}(\alpha_D^1 \alpha_D^1 \beta_D) = -\operatorname{Tr}(\alpha_D^1 \beta_D \alpha_D^1) = -\operatorname{Tr}\beta_D$, and the same argument with $\beta_D^2 = \mathbb{1}$ gives $\operatorname{Tr}\alpha_D^i = 0$. Each of the four matrices squares to $\mathbb{1}$ and so has the eigenvalues ± 1; being traceless, it has as many of each, and N is even. For $N = 2$ at most three matrices anticommute pairwise, for example the Pauli matrices, which leaves no room for β_D. The smallest possibility for a massive fermion is $N = 4$, realised explicitly in Section 3.6.1.

We define

$$\gamma^0 = \beta_D, \qquad \gamma^i = \beta_D \alpha_D^i. \tag{3.55}$$

For example,

$$\left\{\gamma^0,\gamma^i\right\} = \beta_D^2\alpha_D^i + \beta_D\alpha_D^i\beta_D = \alpha_D^i - \beta_D^2\alpha_D^i = 0,$$
$$\left\{\gamma^i,\gamma^j\right\} = \beta_D\alpha_D^i\beta_D\alpha_D^j + \beta_D\alpha_D^j\beta_D\alpha_D^i = -\left\{\alpha_D^i,\alpha_D^j\right\} = -2\delta^{ij}\,\mathbb{1}. \tag{3.56}$$

Therefore $\{\gamma^\mu,\gamma^\nu\} = 2g^{\mu\nu}\mathbb{1}$. Multiplying (3.52) from the left by β_D gives the Dirac equation:

The Dirac equation.

$$(i\gamma^\mu\partial_\mu - m_f)\,\psi = 0, \qquad \{\gamma^\mu,\gamma^\nu\} = 2g^{\mu\nu}\,\mathbb{1}. \tag{3.57}$$

We check the square directly:

$$(i\gamma^\mu\partial_\mu + m_f)(i\gamma^\nu\partial_\nu - m_f)\,\psi = -\big(\gamma^\mu\gamma^\nu\partial_\mu\partial_\nu + m_f^2\big)\psi = -(\Box + m_f^2)\,\psi = 0. \tag{3.58}$$

Since partial derivatives commute, only the symmetric part $\frac{1}{2}\{\gamma^\mu,\gamma^\nu\} = g^{\mu\nu}$ of $\gamma^\mu\gamma^\nu$ contributes. Every component obeys the Klein–Gordon equation, but the first-order Dirac equation also relates the components.

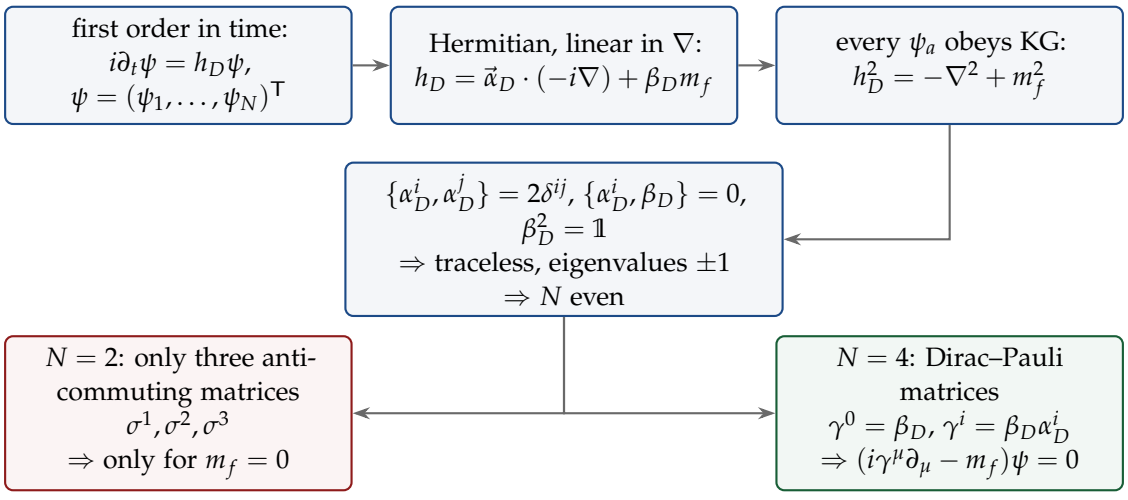


Figure 3.5: The logic of the Dirac construction. The requirement that a first-order equation reproduce the relativistic dispersion relation turns the coefficients of h_D into anticommuting matrices; their properties force an even number of components, and four are needed for a massive particle.

3.6.1 Explicit matrices and spin

The Pauli matrices are

$$\sigma^1 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \qquad \sigma^3 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}. \tag{3.59}$$

They obey $\sigma^i\sigma^j = \delta^{ij}\mathbb{1}_2 + i\epsilon^{ijk}\sigma^k$, where $\epsilon^{123} = 1$ is the completely antisymmetric symbol. Hence

$$(\vec{\sigma}\cdot\vec{p})^2 = p_ip_j\big(\delta^{ij}\mathbb{1}_2 + i\epsilon^{ijk}\sigma^k\big) = \vec{p}^{\,2}\,\mathbb{1}_2. \tag{3.60}$$

An explicit solution of the Dirac algebra is the Dirac–Pauli representation:

$$\beta_D = \begin{pmatrix} \mathbb{1}_2 & 0 \\ 0 & -\mathbb{1}_2 \end{pmatrix}, \qquad \alpha_D^i = \begin{pmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{pmatrix},$$
$$\gamma^0 = \begin{pmatrix} \mathbb{1}_2 & 0 \\ 0 & -\mathbb{1}_2 \end{pmatrix}, \qquad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}. \tag{3.61}$$

For instance $(\gamma^i)^2 = -\mathbb{1}_4$ follows from $(\sigma^i)^2 = \mathbb{1}_2$. Also

$$(\gamma^0)^\dagger = \gamma^0, \qquad (\gamma^i)^\dagger = -\gamma^i, \qquad (\gamma^\mu)^\dagger = \gamma^0\gamma^\mu\gamma^0; \tag{3.62}$$

these relations will fix the appropriate adjoint in the Lagrangian (Section 3.7.1).

What is a spinor? ψ has four components but is not a four-vector. Under a Lorentz transformation $x' = \Lambda x$ its components mix as $\psi'(x') = S_D(\Lambda)\,\psi(x)$, with a matrix satisfying

$$S_D^{-1}\gamma^\mu S_D = \Lambda^\mu{}_\nu\,\gamma^\nu, \tag{3.63}$$

which ensures the covariance of (3.57). The construction of S_D for general boosts and rotations belongs to the representation theory of the Lorentz group and is quoted here (Lund lecture notes, Lectures 9–10). For rotations the intrinsic generators in this basis are

$$S_i = \frac{1}{2}\begin{pmatrix} \sigma^i & 0 \\ 0 & \sigma^i \end{pmatrix}, \qquad \vec{S}^2 = \frac{3}{4}\,\mathbb{1}_4. \tag{3.64}$$

The Pauli identity gives $[S_i, S_j] = i\epsilon^{ijk}S_k$, and at rest S_z has the eigenvalues $\pm\frac{1}{2}$: the field carries spin $\frac{1}{2}$. The four components accommodate two spin states for particles and two for antiparticles.

3.6.2 Plane-wave solutions

For the fermion let $E_f = \sqrt{\vec{p}^{\,2} + m_f^2}$, $p = (E_f, \vec{p})$ and $\not{p} \equiv \gamma^\mu p_\mu$. For $\psi = u(p)\,e^{-ip\cdot x}$ the Dirac equation becomes

$$(\not{p} - m_f)\,u = 0, \qquad h_D(\vec{p})\,u = E_f\,u. \tag{3.65}$$

We write $u = (u_+, u_-)^{\mathsf{T}}$ in two-component blocks. In the Dirac–Pauli representation the matrix equation is

$$\begin{pmatrix} m_f & \vec{\sigma}\cdot\vec{p} \\ \vec{\sigma}\cdot\vec{p} & -m_f \end{pmatrix}\begin{pmatrix} u_+ \\ u_- \end{pmatrix} = E_f \begin{pmatrix} u_+ \\ u_- \end{pmatrix}. \tag{3.66}$$

The lower row gives $u_- = (\vec{\sigma}\cdot\vec{p})\,u_+/(E_f + m_f)$. Substituting this into the upper row and using (3.60) gives $\big(m_f - E_f + \vec{p}^{\,2}/(E_f + m_f)\big)u_+ = 0$, which holds because $E_f^2 - m_f^2 = \vec{p}^{\,2}$. We choose two-component spinors with $\xi_r^\dagger\xi_s = \delta_{rs}$, $r, s = 1, 2$. Then

$$u_r(p) = \sqrt{E_f + m_f}\begin{pmatrix} \xi_r \\ \dfrac{\vec{\sigma}\cdot\vec{p}}{E_f + m_f}\,\xi_r \end{pmatrix}, \qquad u_r^\dagger u_s = \Big(E_f + m_f + \frac{\vec{p}^{\,2}}{E_f + m_f}\Big)\delta_{rs} = 2E_f\,\delta_{rs}. \tag{3.67}$$

The two independent choices of ξ_r correspond to the two spin states.
The opposite frequency is $\psi = v(p)\,e^{+ip\cdot x}$, still with $p^0 = E_f > 0$. Since $i\partial_\mu e^{+ip\cdot x} = -p_\mu e^{+ip\cdot x}$,

$$(\not{p} + m_f)\,v = 0, \qquad h_D(-\vec{p})\,v = -E_f\,v. \tag{3.68}$$

The upper row of the block equation gives $v_+ = (\vec{\sigma}\cdot\vec{p})\,v_-/(E_f+m_f)$, and for $\eta_r^\dagger\eta_s=\delta_{rs}$

$$v_r(p) = \sqrt{E_f+m_f}\begin{pmatrix} \dfrac{\vec{\sigma}\cdot\vec{p}}{E_f+m_f}\,\eta_r \\ \eta_r \end{pmatrix}, \qquad v_r^\dagger v_s = 2E_f\,\delta_{rs}. \tag{3.69}$$

At rest $u_r=\sqrt{2m_f}\,(\xi_r,0)^{\mathsf T}$ and $v_r=\sqrt{2m_f}\,(0,\eta_r)^{\mathsf T}$; in this basis the upper and lower blocks should not be identified with the right- and left-handed chiralities. A negative-frequency mode is a solution needed to expand the field; the creation and annihilation operators of the next section show that it describes antiparticles of positive energy.

3.7 The Dirac field: antiparticles and Fermi statistics

The free Dirac field is a superposition of both kinds of plane waves,

$$\psi(x) = \sum_{r=1}^{2}\int \frac{d^3p}{(2\pi)^3\,2E_f}\Big[a_r(\vec{p})\,u_r(p)\,e^{-ip\cdot x} + b_r^\dagger(\vec{p})\,v_r(p)\,e^{+ip\cdot x}\Big], \qquad p^0=E_f. \tag{3.70}$$

Here a_r annihilates a fermion and $b_r^\dagger$ creates an antifermion; both kinds of physical excitations carry $p=(E_f,\vec{p})$ with $E_f>0$. As for the complex scalar field of Lecture 2, the antiparticle operators multiply the negative-frequency modes. We state the fermionic quantisation rule rather than deriving the canonical construction (Lund lecture notes, Lectures 11–12):

$$\Big\{a_r(\vec{p}),a_s^\dagger(\vec{q})\Big\} = \Big\{b_r(\vec{p}),b_s^\dagger(\vec{q})\Big\} = (2\pi)^3\,2E_f\,\delta_{rs}\,\delta^{(3)}(\vec{p}-\vec{q}), \tag{3.71}$$

with all other anticommutators zero. This rule is consistent with positive energy, positive norm and locality; the full spin–statistics theorem is deeper than the one-mode check below.

Convention (Dirac field and charge: notation dictionary). The fermion operators are called a_r and b_r as in Peskin–Schroeder and the Lund notes. Compared with the scalar dictionary of Lecture 2:

Rzeszów lectures and report	Lund lecture notes	Peskin–Schroeder
$a_r(\vec{p})$, $b_r(\vec{p})$, covariant (3.71)	$a^r_{\vec{p}}$, $b^r_{\vec{p}}$, covariant	$a^s_{\vec{p}}$, $b^s_{\vec{p}}$, non-covariant
$S_D(\Lambda)$	$\Lambda_{1/2}$	$\Lambda_{\frac12}$
charge q, $D_\mu=\partial_\mu+iqA_\mu$, vertex $-iq\gamma^\mu$; $q=-e_0$ for e^-	$e=-\lvert e\rvert$, vertex $-ie\gamma^\mu$	as Lund

One normalised mode exposes the sign. We suppress momenta and spin and take $\{a,a^\dagger\}=\{b,b^\dagger\}=1$. Inserting the positive- and negative-frequency modes into the free energy $\int d^3x\,\psi^\dagger h_D\psi$ gives the one-mode structure

$$\begin{aligned} H_{\text{mode}} &= E_f\,(a^\dagger a - b\,b^\dagger), \\ b\,b^\dagger = 1-b^\dagger b \quad&\Longrightarrow\quad H_{\text{mode}} = E_f\,(a^\dagger a + b^\dagger b) - E_f, \\ :H_{\text{mode}}: &= E_f\,(N_a+N_b), \qquad N_a=a^\dagger a,\quad N_b=b^\dagger b. \end{aligned} \tag{3.72}$$

Both excitation energies are positive. With bosonic commutators instead, $b\,b^\dagger=1+b^\dagger b$ would leave the wrong sign for the energy of an antiparticle. Pauli exclusion is immediate:

$$\Big\{a^\dagger,a^\dagger\Big\}=0 \implies (a^\dagger)^2=0, \qquad N_a^2 = a^\dagger(1-a^\dagger a)\,a = N_a, \tag{3.73}$$

so the eigenvalues of the one-mode number operator are 0 and 1. For a field whose particles have the charge q, the normal-ordered one-mode charge is

$$Q_{\text{mode}} = q\,(N_a - N_b). \tag{3.74}$$

Positive energy and opposite charge are separate statements: a single Dirac field contains particles and antiparticles, both of positive energy and with opposite charges.

3.7.1 Adjoint, action and conserved current

The Dirac adjoint $\bar\psi = \psi^\dagger\gamma^0$ is a row spinor. We take the Hermitian conjugate of (3.57), multiply on the right by γ^0 and use $(\gamma^\mu)^\dagger\gamma^0 = \gamma^0\gamma^\mu$:

$$\begin{aligned} -i(\partial_\mu\psi^\dagger)(\gamma^\mu)^\dagger - m_f\psi^\dagger &= 0,\\ -i(\partial_\mu\bar\psi)\gamma^\mu - m_f\bar\psi &= 0,\\ i(\partial_\mu\bar\psi)\gamma^\mu + m_f\bar\psi &= 0. \end{aligned} \tag{3.75}$$

The free Lagrangian is

$$\mathcal{L}_D = \bar\psi\,(i\gamma^\mu\partial_\mu - m_f)\,\psi. \tag{3.76}$$

Varying $\bar\psi$ gives the Dirac equation. Varying ψ and integrating its derivative term by parts gives the adjoint equation; boundary terms vanish under the same assumptions as in Lecture 1.

The vector current without its charge factor is $j^\mu_N = \bar\psi\gamma^\mu\psi$. With $\gamma^\mu\partial_\mu\psi = -im_f\psi$ and $(\partial_\mu\bar\psi)\gamma^\mu = im_f\bar\psi$,

$$\partial_\mu j^\mu_N = (\partial_\mu\bar\psi)\gamma^\mu\psi + \bar\psi\gamma^\mu\partial_\mu\psi = im_f\bar\psi\psi - im_f\bar\psi\psi = 0. \tag{3.77}$$

Its time component is $j^0_N = \psi^\dagger\psi$ before normal ordering. In the quantised theory the normal-ordered integrated charge counts particles minus antiparticles, as in (3.74).

The symmetry behind the current. For a real constant α the global phase change

$$\psi \to e^{-iq\alpha}\psi, \qquad \bar\psi \to \bar\psi\, e^{+iq\alpha} \tag{3.78}$$

leaves $\mathcal{L}_D$ unchanged. Noether's theorem identifies the current $J^\mu = q\,j^\mu_N$: with the infinitesimal change $\delta\psi = -iq\alpha\psi$,

$$\frac{\partial\mathcal{L}_D}{\partial(\partial_\mu\psi)}\,\delta\psi = i\bar\psi\gamma^\mu(-iq\alpha\psi) = \alpha\,J^\mu. \tag{3.79}$$

Conserved energy came from translations in Lecture 1; conserved charge comes from a phase symmetry: the same Noether argument for different field content. Finally, $2[\psi]+1=4$ gives $[\psi]=3/2$ and $[j^\mu_N]=3$; the charge q will be dimensionless.

3.8 The gauge principle and the QED Lagrangian

Let the phase depend smoothly on position, $\alpha = \alpha(x)$. The mass term remains invariant, but the derivative acts on the phase:

$$\begin{aligned} \partial_\mu\psi' &= \partial_\mu\big(e^{-iq\alpha}\psi\big) = e^{-iq\alpha}\big[\partial_\mu\psi - iq(\partial_\mu\alpha)\psi\big],\\ \bar\psi'\,i\gamma^\mu\partial_\mu\psi' &= \bar\psi\,i\gamma^\mu\partial_\mu\psi + q\,(\partial_\mu\alpha)\,\bar\psi\gamma^\mu\psi. \end{aligned} \tag{3.80}$$

The additional term shows that the free Dirac Lagrangian alone does not have this local symmetry. We introduce a four-potential A_μ and define

$$D_\mu = \partial_\mu + iqA_\mu, \qquad A'_\mu = A_\mu + \partial_\mu\alpha, \qquad \psi' = e^{-iq\alpha}\psi. \tag{3.81}$$

Now the cancellation, without suppressing the derivative of the phase:

$$\begin{aligned} D'_\mu\psi' &= \left[\partial_\mu + iqA_\mu + iq\,\partial_\mu\alpha\right]\left(e^{-iq\alpha}\psi\right) \\ &= e^{-iq\alpha}\left[\partial_\mu\psi - iq(\partial_\mu\alpha)\psi + iqA_\mu\psi + iq(\partial_\mu\alpha)\psi\right] = e^{-iq\alpha}\,D_\mu\psi. \end{aligned} \tag{3.82}$$

Thus $D_\mu\psi$ transforms with the same phase as ψ, and

$$\bar\psi'\,(i\gamma^\mu D'_\mu - m_f)\,\psi' = \bar\psi\,(i\gamma^\mu D_\mu - m_f)\,\psi. \tag{3.83}$$

Expanding the covariant derivative exhibits the interaction:

$$\begin{aligned} \bar\psi\,(i\gamma^\mu D_\mu - m_f)\,\psi &= \bar\psi\,(i\gamma^\mu\partial_\mu - m_f)\,\psi - q\,\bar\psi\gamma^\mu\psi\,A_\mu, \\ \mathcal{L}_{\text{int}} &= -J^\mu A_\mu, \qquad J^\mu = q\,\bar\psi\gamma^\mu\psi. \end{aligned} \tag{3.84}$$

The minus sign comes from $i^2 = -1$ and is fixed by our convention for D_μ (Figure 3.6).

Convention (sign of the charge). The same signed charge q appears in the phase, the covariant derivative and the current; changing one sign convention without the others spoils the cancellation (3.82). For the electron (and the muon) $q = -e_0$ with $e_0 > 0$, and the interaction reads $\mathcal{L}_{\text{int}} = +e_0\,\bar\psi\gamma^\mu\psi\,A_\mu$.

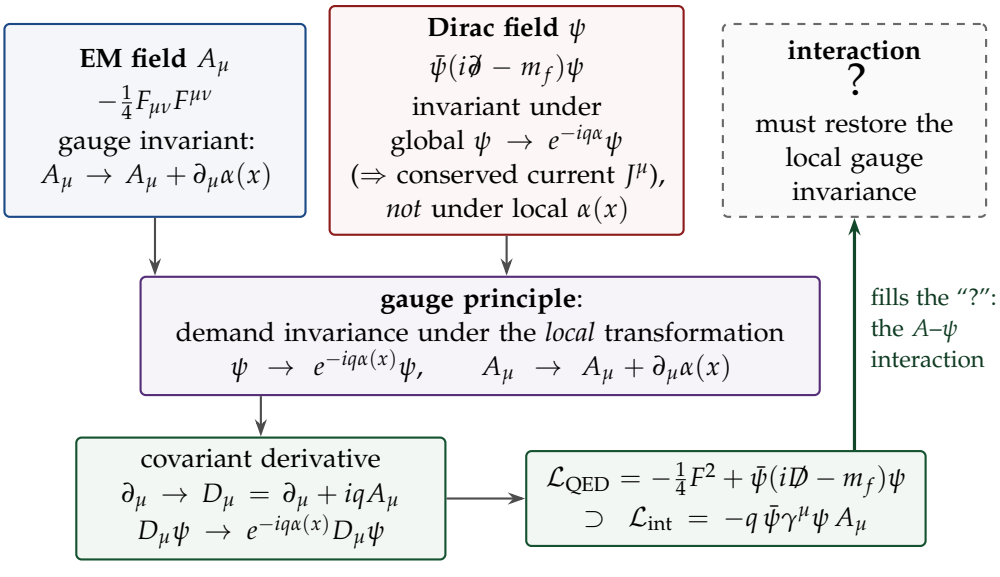


Figure 3.6: The gauge principle. The Maxwell term is gauge invariant, the Dirac term only under global phase changes. Requiring invariance under local phase transformations forces the covariant derivative, and this generates the interaction term of QED with the conserved current of (3.79).

3.8.1 The QED action and Maxwell's equations

The field strength is gauge invariant:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \qquad F'_{\mu\nu} = F_{\mu\nu} + \partial_\mu\partial_\nu\alpha - \partial_\nu\partial_\mu\alpha = F_{\mu\nu}. \tag{3.85}$$

The local minimal Lagrangian for one charged Dirac field is therefore

The QED Lagrangian.

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} + \bar\psi\,(i\gamma^\mu\partial_\mu - m_f)\,\psi - q\,\bar\psi\gamma^\mu\psi\,A_\mu. \tag{3.86}$$

Its three terms describe the electromagnetic field, the free matter field and their interaction. Gauge-equivalent potentials describe the same field strength; A_μ itself is not a gauge-invariant observable. We vary A_ν, keeping the matter fields fixed:

$$\begin{aligned}\delta\mathcal{L}_{\text{EM+int}} &= -\frac{1}{2}F^{\mu\nu}\big(\partial_\mu\delta A_\nu - \partial_\nu\delta A_\mu\big) - J^\nu\delta A_\nu = -F^{\mu\nu}\,\partial_\mu\delta A_\nu - J^\nu\delta A_\nu,\\ \delta\mathcal{A} &= \int d^4x\,\big(\partial_\mu F^{\mu\nu} - J^\nu\big)\,\delta A_\nu \quad\Longrightarrow\quad \partial_\mu F^{\mu\nu} = J^\nu.\end{aligned} \tag{3.87}$$

Antisymmetry gives $\partial_\nu\partial_\mu F^{\mu\nu} = 0$, so the source must obey charge conservation, $\partial_\nu J^\nu = 0$. With $A^\mu = (\Phi_{\text{el}}, \vec{A})$ and $F^{i0} = E^i$, the $\nu = 0$ equation is $\nabla\cdot\vec{E} = J^0$, which confirms that q has the stated physical sign. From $[F^2] = 4$ we obtain $[A] = 1$, and then $[\bar\psi\psi A] = 4$ gives $[q] = 0$. Local gauge symmetry also permits interactions of higher mass dimension, so it does not fix the theory uniquely; (3.86) is the minimal field content and coupling.

The interaction from a symmetry. The free Dirac Lagrangian is invariant under global phase changes, which gives the conserved current J^μ. Invariance under local phase changes cannot be achieved by the matter field alone: it requires a vector field that shifts by a gradient, and it fixes the coupling of this field to the conserved current, with the same q in the phase, the covariant derivative and the vertex. The minimal QED interaction is dictated by the symmetry.

3.9 QED Feynman rules and $e^+e^- \to \mu^+\mu^-$

The QED Lagrangian has been built; now we use it. We collect its Feynman rules and compute the simplest QED process. Only the photon propagator and the evaluation of the spin sums are quoted; everything else follows from the steps of Lecture 2 and of the first half of this lecture.

3.9.1 Feynman rules

The photon propagator. The quadratic part of $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ cannot be inverted, because it vanishes for the pure-gauge fields $A_\mu = \partial_\mu\alpha$; adding the gauge-fixing term $-\frac{1}{2}(\partial_\mu A^\mu)^2$ (Feynman gauge) makes it invertible without changing gauge-invariant predictions. The time-ordered two-point function of the photon is then

$$\langle 0|\mathrm{T}\{A_\mu(x)A_\nu(y)\}|0\rangle = \int \frac{d^4k}{(2\pi)^4}\,\frac{-i\,g_{\mu\nu}}{k^2 + i0}\,e^{-ik\cdot(x-y)}, \tag{3.88}$$

derived in the Lund lecture notes, Lecture 14. Up to the factor $-g_{\mu\nu}$ it is the massless scalar propagator of Lecture 2.

External particles. The external factors follow from the mode expansion (3.70) exactly as $e^{\mp ip\cdot x}$ followed for the scalar field in Lecture 2. For a fermion state $|p,r\rangle = a_r^\dagger(\vec{p})|0\rangle$ the anticommutator (3.71) gives

$$\langle 0|\psi(x)|p,r\rangle = \sum_s \int \frac{d^3q}{(2\pi)^3\,2E_f(\vec{q})}\, u_s(q)\, e^{-iq\cdot x}\, \langle 0|\left\{a_s(\vec{q}), a_r^\dagger(\vec{p})\right\}|0\rangle = u_r(p)\, e^{-ip\cdot x}, \quad (3.89)$$

where $E_f(\vec{q}) = \sqrt{\vec{q}^{\,2}+m_f^2}$; the antifermion term of ψ creates a second particle and does not contribute. By Hermitian conjugation $\langle p,r|\bar\psi(x)|0\rangle = \bar u_r(p)\, e^{ip\cdot x}$. For an antifermion state $|\bar p,r\rangle = b_r^\dagger(\vec{p})|0\rangle$ we need the adjoint field,

$$\bar\psi(x) = \psi^\dagger\gamma^0 = \sum_{r=1}^{2}\int \frac{d^3p}{(2\pi)^3\,2E_f}\left[a_r^\dagger(\vec{p})\,\bar u_r(p)\, e^{ip\cdot x} + b_r(\vec{p})\,\bar v_r(p)\, e^{-ip\cdot x}\right], \quad (3.90)$$

and find in the same way

$$\langle 0|\bar\psi(x)|\bar p,r\rangle = \bar v_r(p)\, e^{-ip\cdot x}, \qquad \langle \bar p,r|\psi(x)|0\rangle = v_r(p)\, e^{ip\cdot x}. \quad (3.91)$$

An external photon of momentum k contributes its polarisation vector, $\epsilon_\mu(k)$ when incoming and $\epsilon_\mu^*(k)$ when outgoing.

The vertex. The interaction enters the Dyson series as $i\int d^4x\,\mathcal{L}_{\rm int} = -iq\int d^4x\,\bar\psi\gamma^\mu\psi\,A_\mu$ (it contains no derivatives, so $H_{\rm int} = -\int d^3x\,\mathcal{L}_{\rm int}$ as in Lecture 2). Each vertex contributes $-iq\gamma^\mu$, and its position integral gives $(2\pi)^4\delta^{(4)}$ of the momenta entering it, as for the quartic vertex $-i\lambda$. For the electron and the muon the vertex is $-iq\gamma^\mu = +ie_0\gamma^\mu$.

The rules. The amplitude $i\mathcal{M}$ is the sum of all connected diagrams, each built from (Table 3.1): a factor $-iq\gamma^\mu$ and momentum conservation at each vertex; a propagator for each internal line, $-ig_{\mu\nu}/(k^2+i0)$ for a photon and $i(\not p + m_f)/(p^2-m_f^2+i0)$ for a fermion (Lund lecture notes, Lecture 18); the external factors (3.89)–(3.91). The Dirac indices are contracted along each fermion line, written against the direction of its arrow: a chain starts with $\bar u$ or $\bar v$ and ends with u or v. Loops add $\int d^4\ell/(2\pi)^4$, a factor -1 for each closed fermion loop, and diagrams that differ by the interchange of identical external fermions have a relative sign.

3.9.2 Electron–positron annihilation into muons

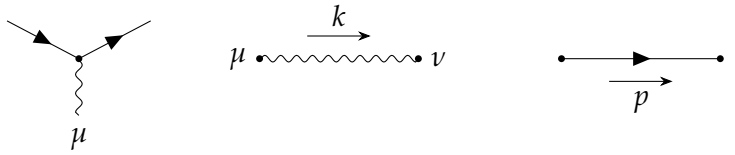
We consider

$$e^-(p_1,r_1) + e^+(p_2,r_2) \longrightarrow \mu^-(k_1,s_1) + \mu^+(k_2,s_2) \qquad \text{at } \sqrt{s} \gg m_\mu, \quad (3.92)$$

so that both masses may be neglected ($m_\mu \simeq 207\,m_e$). The electron and the muon are different fields with the same charge $q=-e_0$. At order e_0^2 there is a single diagram: the pair annihilates into a virtual photon of momentum $P = p_1+p_2$, $P^2 = s$, which produces the muon pair (Figure 3.7a). Reading each fermion line against its arrow,

$$\begin{aligned} i\mathcal{M} &= \left[\bar v(p_2)\,(-iq\gamma^\mu)\,u(p_1)\right]\frac{-i\,g_{\mu\nu}}{s}\left[\bar u(k_1)\,(-iq\gamma^\nu)\,v(k_2)\right] \\ &= (-iq)^2\left[\bar v(p_2)\gamma^\mu u(p_1)\right]\frac{-i\,g_{\mu\nu}}{s}\left[\bar u(k_1)\gamma^\nu v(k_2)\right], \end{aligned} \quad (3.93)$$

Table 3.1: Feynman rules of QED in Feynman gauge (time runs from left to right; the arrow on a fermion line shows the flow of fermion number). For the electron and the muon $q = -e_0$.



$-iq\gamma^\mu$	$\dfrac{-i\,g_{\mu\nu}}{k^2+i0}$	$\dfrac{i\,(\not{p}+m_f)}{p^2-m_f^2+i0}$
vertex	photon propagator	fermion propagator

external line	incoming	outgoing
fermion ($a_r^\dagger\|0\rangle$)	$u_r(p)$	$\bar{u}_r(p)$
antifermion ($b_r^\dagger\|0\rangle$)	$\bar{v}_r(p)$	$v_r(p)$
photon	$\epsilon_\mu(k)$	$\epsilon_\mu^*(k)$

with the spin labels suppressed. Since $(-iq)^2(-i) = iq^2 = ie_0^2$,

$$\mathcal{M} = \frac{e_0^2}{s}\,\big[\bar{v}(p_2)\gamma^\mu u(p_1)\big]\,\big[\bar{u}(k_1)\gamma_\mu v(k_2)\big]. \tag{3.94}$$

Unpolarised cross section. For unpolarised beams and a detector blind to spin, $|\mathcal{M}|^2$ in the master formula is averaged over the 2×2 initial spin states and summed over the final ones, $|\mathcal{M}|^2 \to \frac{1}{4}\sum_{\text{spins}}|\mathcal{M}|^2$. The spin sums $\sum_r u_r(p)\bar{u}_r(p) = \not{p} + m_f$ and $\sum_r v_r(p)\bar{v}_r(p) = \not{p} - m_f$ turn this into traces of products of γ matrices; their evaluation (Lund lecture notes, Lecture 20) is quoted:

$$\frac{1}{4}\sum_{\text{spins}}|\mathcal{M}|^2 = \frac{2e_0^4\,(t^2+u^2)}{s^2} = e_0^4\,(1+\cos^2\theta) \qquad (\sqrt{s}\gg m_\mu), \tag{3.95}$$

where θ is the angle between $\vec{p}_1$ and $\vec{k}_1$ in the CM frame. The second form follows from the massless invariants (3.40), $t^2+u^2 = \frac{s^2}{4}\big[(1-\cos\theta)^2+(1+\cos\theta)^2\big] = \frac{s^2}{2}\,(1+\cos^2\theta)$. With helicity eigenstates as spin basis only four of the sixteen spin combinations contribute, two with $|\mathcal{M}|^2 = e_0^4(1+\cos\theta)^2$ and two with $e_0^4(1-\cos\theta)^2$: massless fermions keep their helicity at the vector vertex.

The remaining factors come from the master formula (3.19) with all masses set to zero: $\mathcal{F} = 4\,p_1\cdot p_2 = 2s$, and $d\Phi_2/d\Omega = 1/32\pi^2$ from (3.28) with $\beta_{\text{cm}} = 1$. The final particles are different, so there is no factor $1/2!$. Hence

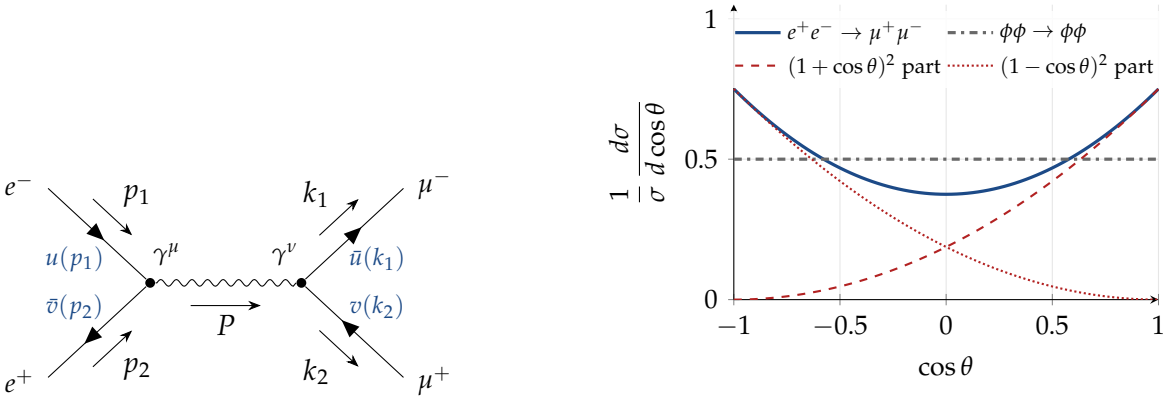
$$\frac{d\sigma}{d\Omega} = \frac{1}{2s}\,e_0^4\,(1+\cos^2\theta)\,\frac{1}{32\pi^2} = \frac{e_0^4\,(1+\cos^2\theta)}{64\pi^2 s}, \tag{3.96}$$

and with $\alpha = e_0^2/4\pi$ and $2\pi\int_{-1}^{1} d\cos\theta\,(1+\cos^2\theta) = 2\pi\cdot\frac{8}{3}$:

$e^+e^- \to \mu^+\mu^-$ at tree level, $\sqrt{s}\gg m_\mu$.

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4s}\,(1+\cos^2\theta), \qquad \sigma = \frac{4\pi\alpha^2}{3s}, \qquad \alpha = \frac{e_0^2}{4\pi} \simeq \frac{1}{137}. \tag{3.97}$$

With $(\hbar c)^2 = 0.3894\,\text{mb}\,\text{GeV}^2$ this is $\sigma = 86.8\,\text{nb}/(s/\text{GeV}^2)$, for example 0.87 nb at $\sqrt{s} = 10\,\text{GeV}$. The cross section falls as $1/s$, like the scalar result (3.34), because the coupling is dimensionless and the masses have been neglected. The angular distribution, however, differs (Figure 3.7b): the contact interaction of spinless particles gives an isotropic distribution, the spin-$\frac{1}{2}$ pairs produced by a vector current give $1 + \cos^2\theta$.



(a) $P = p_1 + p_2$, $P^2 = s$ (b) CM frame, $\sqrt{s} \gg m_\mu$

Figure 3.7: (a) The single tree diagram of $e^-(p_1)e^+(p_2) \to \mu^-(k_1)\mu^+(k_2)$ with the external spinors of Table 3.1; each fermion line is read against its arrow, $\bar{v}(p_2)\gamma^\mu u(p_1)$ and $\bar{u}(k_1)\gamma^\nu v(k_2)$, joined by the photon propagator $-ig_{\mu\nu}/s$. (b) Normalised angular distributions in the CM frame: $\frac{3}{8}(1+\cos^2\theta)$ for $e^+e^- \to \mu^+\mu^-$ at $\sqrt{s} \gg m_\mu$ (solid), the sum of the two helicity contributions $\frac{3}{16}(1 \pm \cos\theta)^2$ (red), compared with the isotropic $\phi\phi \to \phi\phi$ (dash-dotted).

3.10 What the tree approximation leaves open

3.10.1 The first scalar loop

We identify the first omitted quantum correction of $\phi\phi \to \phi\phi$, without attempting a loop evaluation. In the s channel, with $P = p_1 + p_2$, two quartic vertices are joined by two internal lines (Figure 3.8b). The internal momentum ℓ is not fixed by the external momenta and must be integrated over. With the momentum-space propagator $i/(\ell^2 - m^2 + i0)$ of Lecture 2,

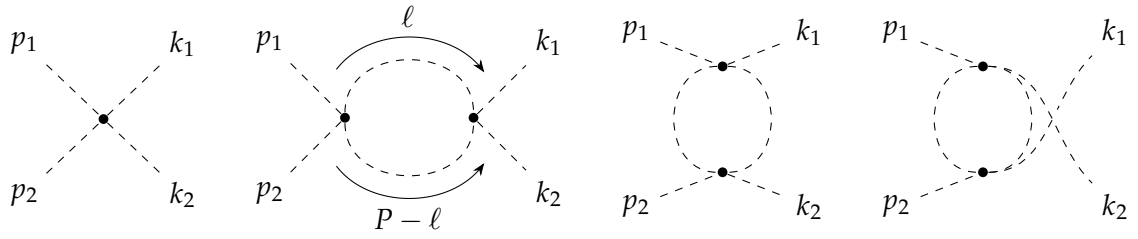
$$i\mathcal{M}_s^{(2)} = \frac{(-i\lambda)^2}{2} \int \frac{d^4\ell}{(2\pi)^4} \, \frac{i}{\ell^2 - m^2 + i0} \, \frac{i}{(P-\ell)^2 - m^2 + i0}. \tag{3.98}$$

The $\frac{1}{2}$ here is a factor of the diagram, unrelated to the counting of identical final particles in the phase space. We verify it from the second-order Dyson coefficient $1/\big(2!\,(4!)^2\big)$ of Lecture 2, for one specified pairing of the external particles into an s-channel diagram:

$$\begin{aligned} &\text{choose the two ordered external attachments at each vertex:} \quad (4\cdot 3)^2, \\ &\text{join the two remaining lines:} \quad 2!, \\ &\text{interchange which vertex receives the incoming pair:} \quad 2, \\ &\text{net weight:} \quad \frac{(4\cdot 3)^2 \cdot 2! \cdot 2}{2!\,(4!)^2} = \frac{1}{2}. \end{aligned} \tag{3.99}$$

The other two pairings give the t- and u-channel diagrams (Figure 3.8c,d); the connected four-point amplitude at order λ^2 contains all three. Normal ordering removes contractions

of two fields at the same vertex, not these contractions between different vertices. The same operator expansion that gave the tree prediction thus tells us exactly what was neglected: a loop is an integral over an unresolved internal momentum, not a new interaction added by hand.



(a) tree, $\mathcal{O}(\lambda)$ (b) s channel, $\mathcal{O}(\lambda^2)$ (c) t channel (d) u channel

Figure 3.8: (a) The tree contact diagram of $\phi\phi \to \phi\phi$ and (b)–(d) the three connected one-loop diagrams at order λ^2, each with the combinatorial weight $\frac{1}{2}$ (time runs from left to right; scalar lines dashed). In (b) the loop momentum is routed as in (3.98): ℓ and $P-\ell$ flow between the vertices, with $P = p_1 + p_2$. In (c) and (d) the bubble carries the momentum transfer $p_1 - k_1$ or $p_1 - k_2$.

3.10.2 Power counting, renormalisation and weak coupling

For a large Euclidean loop momentum ℓ_E the two denominators of (3.98) behave as ℓ_E^4; the Wick-rotated (Euclidean) integration serves here only to examine large momenta. In four dimensions

$$d^4\ell_E = \ell_E^3\, d\ell_E\, d\Omega_3, \qquad \int d\Omega_3 = 2\pi^2. \tag{3.100}$$

Between a fixed reference momentum $\mu_0 > 0$ and a large cut-off Λ_{UV}, therefore,

$$\int \frac{d^4\ell_E}{(2\pi)^4}\,\frac{1}{\ell_E^4} \;\sim\; \frac{1}{8\pi^2}\int_{\mu_0}^{\Lambda_{\mathrm{UV}}} \frac{d\ell_E}{\ell_E} = \frac{1}{8\pi^2}\,\log\frac{\Lambda_{\mathrm{UV}}}{\mu_0}. \tag{3.101}$$

This is a logarithmic ultraviolet divergence: a finite next-order prediction cannot be obtained by integrating to infinity and ignoring it.

What renormalisation does. A regulated calculation writes the coupling as

$$\lambda_0 = \lambda_R + \delta\lambda, \qquad \mathcal{L}_{\mathrm{int}} = -\frac{\lambda_R}{4!}\,\phi^4 - \frac{\delta\lambda}{4!}\,\phi^4. \tag{3.102}$$

Here λ_0 is a regulator-dependent bare parameter, λ_R is defined by a specified measurement or subtraction condition, and $\delta\lambda$ is a counterterm. To this order

$$\mathcal{M} = -\lambda_R + \mathcal{M}_{\mathrm{loops}}(s,t,u;\Lambda_{\mathrm{UV}}) - \delta\lambda + \cdots. \tag{3.103}$$

The leading divergence at large loop momentum does not depend on the external momenta, so it has the structure of the four-field vertex and can be absorbed into $\delta\lambda$. The finite momentum dependence remains a prediction once the coupling has been fixed by one measurement. Full renormalisation also treats the mass and the normalisation of the field; these steps are not performed here.

When is the leading term informative? Write $\mathcal{M} = -\lambda_R + \mathcal{M}^{(2)} + \cdots$, where $\mathcal{M}^{(2)}$ is of order λ_R^2 after renormalisation. Then

$$|\mathcal{M}|^2 = \lambda_R^2 - 2\lambda_R \,\mathrm{Re}\,\mathcal{M}^{(2)} + \mathcal{O}(\lambda_R^4). \tag{3.104}$$

The tree cross section is of order λ_R^2, and its first correction generally of order λ_R^3. Perturbation theory requires the correction to be small *in the observable and the kinematics of interest*: a small coupling alone may not control large logarithms of widely separated scales. Resummation reorganises such enhanced terms; genuinely strong coupling or bound states can require non-perturbative methods. We have derived a lowest-order prediction, identified its first omitted contribution and seen why measured parameters are needed beyond tree level; running couplings and strongly coupled theories are beyond this course.

3.11 Synthesis, research connections and the report

We consolidate the calculation by reconstructing the prediction and naming for each step its input, operation and output (Table 3.2): no single formula carried the argument. The action supplied the dynamics, symmetry the constraints, quantisation the states, and scattering theory the connection to measurement.

Table 3.2: The argument of the course in ten steps.

Input	Operation	Output
relativistic action	vary the field; integrate by parts	field equation
translation symmetry	apply Noether's theorem	conserved energy and momentum
oscillator modes	impose canonical commutators	Fock states and particle energies
vacuum and fields	time order and contract	Feynman propagator
interaction Hamiltonian	expand the evolution; count contractions	invariant amplitude
external-state normalisation	divide the transition rate by the incident flux	master cross-section formula
on-shell final states	integrate the energy–momentum δ-functions	phase space
identical quanta	count each physical final state once	event cross section
one unstable particle	the same box steps, without a flux	decay rate
local phase invariance	covariant derivative; read off the vertex	QED Feynman rules, $e^+e^- \to \mu^+\mu^-$

Self-study questions. Check your understanding:

1. Compare the 4! in $\mathcal{L}_{\text{int}}$, the loop factor $\frac{1}{2}$ and the final-state $1/2!$. Which operation produced each? They are three distinct combinatorial issues.
2. Replace the covariant ladder operator by $\tilde{a} = a/\sqrt{2E}$. Which state factors change? The invariant amplitude and the cross section must not change.
3. Consider D_F and the space-like commutator of Lecture 2. Why are non-zero correlations alone not a controllable signal?
4. Let s approach $4m^2$ from above. Why can a finite cross section coexist with a vanishing event rate?

What was derived and what was quoted. The chain from the scalar action to the cross section and to the decay rate was derived. The Dirac algebra, the plane-wave spinors, the local gauge cancellation, the vertex and the external factors of QED were derived, and so was the $e^+e^- \to \mu^+\mu^-$ cross section from the master formula. The general spinor transformations $S_D(\Lambda)$, the fermionic anticommutators and the spin–statistics theorem, the photon propagator and the evaluation of the spin traces were used as explicitly identified input. Loop evaluation and renormalisation remain for further study.

Research connections. The framework is transferable, but the scalar model is not a quantitative replacement for each student's physical system:

- **Particle scattering and proton structure.** The route from a process to an amplitude and an observable persists in QCD; quark and gluon fields, gauge structure, hadronic matrix elements and factorisation add essential physics.
- **Molecular radiation and spectroscopy.** Matter states, interaction matrix elements and photon phase space determine transition rates, as in the decay formula; molecular wave functions and selection rules must be supplied before a line intensity is predicted.
- **Effective descriptions of materials.** A Dirac-like matrix equation can describe low-energy quasiparticle bands; its velocity, parameters and range of validity are material-dependent, and the algebraic resemblance does not make a crystal Lorentz invariant.
- **Correlations and fluctuations.** The distinction between correlation and response is useful beyond particle physics; commutators and time ordering add information absent from a classical correlation.

Choose one connection, name the extra physical input and identify one claim that this calculation cannot establish.

The written report. It consists of five connected problems prepared by the three lectures:

Problem	What the report should demonstrate
1	scalar dynamics, the translation current and the energy, with boundary conditions
2	quantisation of the modes; how normalisation and the Hamiltonian give particle states
3	correlations and the propagator; microcausality versus vacuum correlation
4	the Dyson expansion and the contact amplitude with explicit contractions
5	phase space, flux, counting and limits; the QED gauge construction and one scientific application

We started from a local action and obtained rates per incident flux and per decaying particle. Explain what a particle state, an internal propagator and a scattering amplitude each contribute, and what extra work is needed when a simple perturbative model stops being sufficient.

Summary

- Box normalisation $|\vec{p}\rangle_V = |\vec{p}\rangle/\sqrt{2E_{\vec{p}}V}$; $|K_T|^2/T \to 2\pi\delta(\Delta E)$ and $|K_T K_V|^2/T \to V(2\pi)^4\delta^{(4)}$ give $dR_{\text{ord}} = |\mathcal{M}|^2 d\Phi_2/(4E_1E_2V)$, and the flux $j_{\text{inc}} = v_{\text{cl}}/V$ removes V.
- Master formula $d\sigma = |\mathcal{M}|^2 d\Phi_2/\mathcal{F}$, with the invariant flux factor $\mathcal{F} = 4E_1E_2v_{\text{cl}} = 4[(p_1\cdot p_2)^2 - m_1^2m_2^2]^{1/2}$; in the CM frame $\mathcal{F} = 2s\beta_{\text{cm}}$ and $d\Phi_2/d\Omega = \beta_{\text{cm}}/32\pi^2$.
- Scalar scattering, with a factor $1/2!$ for the identical final particles and none for the incident ones: $d\sigma/d\Omega = \lambda^2/128\pi^2 s$ and $\sigma = \lambda^2/32\pi s$, isotropic. The threshold limit $\lambda^2/128\pi m^2$ is finite, while the rate vanishes with β_{cm}; $s+t+u = 4m^2$.
- Decays: $d\Gamma = |\mathcal{M}|^2 d\Phi_n/2M$; for two bodies $\Gamma = |\vec{p}^{\,*}|\,|\mathcal{M}|^2/8\pi M^2$, times $\frac{1}{2}$ for identical particles, with $|\vec{p}^{\,*}| = \lambda^{1/2}(M^2, m_1^2, m_2^2)/2M$. For $\Phi \to \phi\phi$, $\mathcal{M} = -\mu$ and $\Gamma = \mu^2|\vec{p}^{\,*}|/16\pi M^2$.
- Dirac: $h_D^2 = \vec{p}^{\,2} + m_f^2$ forces anticommuting 4×4 matrices, and $(i\gamma^\mu\partial_\mu - m_f)\psi = 0$ with $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$. The spinors obey $u^\dagger u = v^\dagger v = 2E_f$; anticommuting a_r, b_r give positive energies, Pauli exclusion and the charge $q(N_a - N_b)$.
- Gauge principle: local phase invariance requires $D_\mu = \partial_\mu + iqA_\mu$ and gives the QED Lagrangian with $\mathcal{L}_{\text{int}} = -q\bar{\psi}\gamma^\mu\psi A_\mu$; Maxwell's equations $\partial_\mu F^{\mu\nu} = J^\nu$ follow.
- QED Feynman rules: the vertex $-iq\gamma^\mu$ with $q = -e_0$, the photon propagator $-ig_{\mu\nu}/(k^2 + i0)$ in Feynman gauge, and the external factors u, $\bar{u}$, $\bar{v}$, v, ϵ, ϵ^*.
- The first QED cross section, $e^+e^- \to \mu^+\mu^-$ at $\sqrt{s} \gg m_\mu$: from $\frac{1}{4}\sum|\mathcal{M}|^2 = e_0^4(1+\cos^2\theta)$ follow $d\sigma/d\Omega = \alpha^2(1+\cos^2\theta)/4s$ and $\sigma = 4\pi\alpha^2/3s$.
- First loop: the s-, t- and u-channel bubbles have weight $\frac{1}{2}$ each. The logarithmic divergence $(8\pi^2)^{-1}\log(\Lambda_{\text{UV}}/\mu_0)$ is absorbed into $\delta\lambda$; the tree result holds while the corrections are small in the observable of interest.

Report. Problem 5 builds on this lecture: the phase space and the flux (Sections 3.2 and 3.3), the counting of identical particles (Section 3.3.2), the limits of Section 3.4, the gauge construction of Section 3.8 and one application (Section 3.11). Problems 1–4 rest on Lectures 1 and 2. The report writes the oscillator-normalised $\tilde{a}$ as a; see the normalisation bridge at the start of this lecture.

Reading. Peskin–Schroeder, Sections 3.2–3.5, 4.5, 4.8, 5.1 and 10.1–10.2; Lund lecture notes, Lectures 9–15 and 17–20; book, Chapters 6–7 and 10–15.

Problems for independent study

Introduction to Quantum Field Theory

Student assignment

Roman Pasechnik

Doctoral School, University of Rzeszów

The five problems develop a single calculation: from the action of a real scalar field to its quantized particle states, causal commutators, scattering amplitude and measurable cross section. A short final question connects these methods to the field content and gauge symmetry of quantum electrodynamics (QED).

Instructions

Submit one individual report in English containing solutions to all five problems. Give intermediate reasoning, define any additional notation, state approximations, and explain the physical meaning of your results. Use the conventions below consistently. If a reference uses different conventions, explain the conversion. Earlier results may be cited within your report; you need not derive the same identity twice.

The lecture notes and other cited literature may be consulted. Reconstruct the arguments in your own work and acknowledge sources. Calculations may be handwritten and scanned legibly or typeset. Programming and numerical simulation are not required. There is no oral examination or presentation. The submission deadline and any revision arrangements will be communicated by the lecturer.

Each problem carries 20 marks: 12 for analytical reasoning and correctness, 5 for physical interpretation, and 3 for checks, conventions and clarity. The total is 100 marks. A pass requires at least 50 marks overall and at least 8 marks on each problem. Subject to these conditions, the grade bands are 90–100: 5.0; 80–89: 4.5; 70–79: 4.0; 60–69: 3.5; 50–59: 3.0. Failure to meet the pass conditions gives grade 2.0.

Common conventions

Work in four-dimensional Minkowski space with $\hbar = c = 1$ and

$$\begin{aligned} &g_{\mu\nu} = \operatorname{diag}(1,-1,-1,-1), \qquad x^\mu = (t, \boldsymbol{x}), \qquad p^\mu = (p^0, \boldsymbol{p}), \\ &p \cdot x = p^0 t - \boldsymbol{p}\cdot\boldsymbol{x}, \qquad \partial_\mu = \frac{\partial}{\partial x^\mu}, \qquad \Box = \partial_\mu \partial^\mu = \partial_t^2 - \nabla^2. \end{aligned} \tag{1}$$

Repeated Greek indices are summed over $0, 1, 2, 3$; spatial indices run over $1, 2, 3$. Raise and lower indices with $g_{\mu\nu}$. Bold symbols denote spatial vectors. The real scalar has mass $m > 0$. On shell, write $p^0 = E_{\boldsymbol{p}} = \sqrt{\boldsymbol{p}^2 + m^2}$, with positive energy.

Use the Fourier convention

$$f(x) = \int \frac{\mathrm{d}^4 p}{(2\pi)^4}\, \mathrm{e}^{-\mathrm{i}p\cdot x} \widetilde{f}(p), \qquad \int \mathrm{d}^4 x\, \mathrm{e}^{\mathrm{i}p\cdot x} = (2\pi)^4 \delta^{(4)}(p). \tag{2}$$

Here $\delta^{(n)}$ is the n-dimensional Dirac delta distribution and $\mathrm{i}^2 = -1$. The mass dimension $[X]$ of a quantity is its power of mass in natural units. Fields in Problem 1 are classical; from Problem 2 onward they are operators. All field identities are understood as distributions, or after smearing with smooth test functions.

Problem 1 Action, symmetry and conserved energy

20 marks

Purpose. Establish the classical dynamics and conserved quantities that will be quantized in Problem 2.

The free real scalar field $\phi(x)$ has action

$$\mathcal{A}_0[\phi] = \int \mathrm{d}^4x\, \mathcal{L}_0, \qquad \mathcal{L}_0 = \frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - \frac{1}{2}m^2\phi^2. \tag{3}$$

The symbol $\mathcal{A}_0$ denotes the action, not the scattering operator S used later. Assume variations vanish on the boundary. For integrated conserved charges, assume fields decrease sufficiently rapidly at spatial infinity, or use periodic boundary conditions in a finite box.

(a) Field equation and dimensions. Vary $\mathcal{A}_0$ explicitly, including the integration by parts, to obtain the Euler–Lagrange equation for ϕ. Derive the dispersion relation of a plane-wave solution and identify the positive-frequency branch. Determine the mass dimensions of ϕ, m and $\mathcal{L}_0$ from the requirement that the action is dimensionless.

(b) Translation symmetry. For an infinitesimal constant translation use

$$\delta\phi(x) = -\epsilon^\nu\partial_\nu\phi(x), \qquad \partial_\mu\epsilon^\nu = 0. \tag{4}$$

Derive the canonical energy–momentum tensor $T^\mu{}_\nu$ from Noether's theorem, choosing its overall convention so that the translation current is $j^\mu = -\epsilon^\nu T^\mu{}_\nu$. Compute $\partial_\mu T^\mu{}_\nu$ explicitly and use the field equation to establish local conservation. Explain which boundary assumption is needed to obtain time-independent charges

$$P^\nu = \int \mathrm{d}^3\boldsymbol{x}\, T^{0\nu}(t,\boldsymbol{x}). \tag{5}$$

(c) Hamiltonian and physical meaning. Define the canonical momentum and Hamiltonian density by

$$\pi(x) = \frac{\partial\mathcal{L}_0}{\partial\dot\phi(x)}, \qquad \mathcal{H}_0 = \pi\dot\phi - \mathcal{L}_0, \qquad \dot\phi = \partial_t\phi. \tag{6}$$

Calculate π, $\mathcal{H}_0$ and T^{00}, and compare $H_0 = \int \mathrm{d}^3\boldsymbol{x}\,\mathcal{H}_0$ with P^0. Determine whether the classical energy is bounded below for the stated $m^2 > 0$, and interpret its separate terms. Check the mass dimension of H_0.

Discussion to include. What is the difference between the field equation, a local conservation equation and a conserved integrated charge? Explain why the boundary conditions matter even when the local equation is satisfied. Integrate this discussion into your solution rather than writing a separate essay.

Problem 2 From field modes to particle states

20 marks

Purpose. Construct the particle interpretation of the free scalar field.

Promote ϕ and π to Hermitian operators and impose the equal-time canonical commutation relations

$$\begin{aligned} &[\phi(t,\boldsymbol{x}),\pi(t,\boldsymbol{y})] = \mathrm{i}\delta^{(3)}(\boldsymbol{x}-\boldsymbol{y}), \\ &[\phi(t,\boldsymbol{x}),\phi(t,\boldsymbol{y})] = 0, \qquad\qquad [\pi(t,\boldsymbol{x}),\pi(t,\boldsymbol{y})] = 0. \end{aligned} \tag{7}$$

The commutator is $[A,B] = AB - BA$. Use the mode expansion

$$\phi(x) = \int \frac{\mathrm{d}^3\boldsymbol{p}}{(2\pi)^3}\frac{1}{\sqrt{2E_{\boldsymbol{p}}}}\left[a_{\boldsymbol{p}}\mathrm{e}^{-\mathrm{i}p\cdot x} + a_{\boldsymbol{p}}^\dagger \mathrm{e}^{\mathrm{i}p\cdot x}\right], \qquad p^0 = E_{\boldsymbol{p}}. \tag{8}$$

The dagger denotes the Hermitian adjoint. The operators $a_{\boldsymbol{p}}$ and $a_{\boldsymbol{p}}^\dagger$ are time-independent mode coefficients; their algebra is to be determined. The vacuum is defined by $a_{\boldsymbol{p}}|0\rangle = 0$ for every $\boldsymbol{p}$, with $\langle 0|0\rangle = 1$. For bosons, normal ordering $:\mathcal{O}:$ places every creation operator to the left of every annihilation operator without adding commutator terms.

(a) Mode algebra. Obtain inverse relations for the mode coefficients from ϕ and π at one fixed time. Use them with (7) to derive $[a_{\boldsymbol{p}}, a_{\boldsymbol{q}}^\dagger]$, $[a_{\boldsymbol{p}}, a_{\boldsymbol{q}}]$ and $[a_{\boldsymbol{p}}^\dagger, a_{\boldsymbol{q}}^\dagger]$. Insert the resulting algebra back into (8) to check the field–momentum commutator, retaining all factors of 2π and $2E_{\boldsymbol{p}}$.

(b) Energy and particle number. Insert the modes into the Hamiltonian from Problem 1. Show how the terms containing two annihilation operators or two creation operators cancel. Separate the remaining vacuum contribution from the operator-dependent part, then obtain $:H_0:$ and interpret normal ordering in this free theory. Define

$$N = \int \frac{\mathrm{d}^3\boldsymbol{p}}{(2\pi)^3} a_{\boldsymbol{p}}^\dagger a_{\boldsymbol{p}}. \tag{9}$$

Compute $[:H_0:, a_{\boldsymbol{p}}^\dagger]$ and $[N, a_{\boldsymbol{p}}^\dagger]$ and use them to determine the energy and particle number of $a_{\boldsymbol{p}}^\dagger|0\rangle$. Use a finite box and a momentum cutoff if needed to keep the vacuum contribution well defined before normal ordering.

(c) Relativistic normalization. Define the covariantly normalized state

$$|\boldsymbol{p}\rangle = \sqrt{2E_{\boldsymbol{p}}}\, a_{\boldsymbol{p}}^\dagger|0\rangle. \tag{10}$$

Calculate $\langle \boldsymbol{q}|\boldsymbol{p}\rangle$ and $\langle 0|\phi(x)|\boldsymbol{p}\rangle$. Explain why a sharp-momentum state has distributional normalization rather than unit norm. Relate its normalization to the Lorentz-invariant measure by evaluating the p^0 integral in

$$\int \frac{\mathrm{d}^4p}{(2\pi)^3}\,\theta(p^0)\delta(p^2-m^2)\,f(p), \tag{11}$$

for a test function f. Here θ is the Heaviside step distribution.

Problem 3 Correlations, causality and the propagator

20 marks

Purpose. Distinguish vacuum correlations from causal influence and connect the field operator to a Green function.

Use the free field and vacuum from Problem 2 and set $z = x - y$. Define

$$\begin{aligned} W(z) &= \langle 0|\phi(x)\phi(y)|0\rangle, \\ C(z) &= [\phi(x), \phi(y)], \\ D_F(z) &= \langle 0|\mathcal{T}\{\phi(x)\phi(y)\}|0\rangle. \end{aligned} \tag{12}$$

These definitions contain *no additional factor of* i. For a bosonic field, time ordering is defined by

$$\mathcal{T}\{\phi(x)\phi(y)\} = \theta(x^0 - y^0)\phi(x)\phi(y) + \theta(y^0 - x^0)\phi(y)\phi(x). \tag{13}$$

The identity operator is denoted by $\mathbf{1}$. Useful distributional rules are $\mathrm{d}\theta(t)/\mathrm{d}t = \delta(t)$ and $\int \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t} = 2\pi\delta(\omega)$. An i0 prescription means the limit of a positive regulator tending to zero. Under a proper orthochronous Lorentz transformation $x' = \Lambda x$, the scalar field transforms as $\phi'(x') = \phi(x)$ and the vacuum is invariant. A retarded response describes the change caused by an applied source and vanishes at times earlier than the source.

(a) Compute the correlations. Starting from the mode expansion, derive momentum-integral representations for $W(z)$ and $C(z)$. Show explicitly that the free-field commutator is a c-number distribution times $\mathbf{1}$, and express it in terms of W. State which properties of the vacuum and mode algebra were used.

(b) Test microcausality. For spacelike separation $z^2 < 0$, choose a Lorentz frame in which $z^0 = 0$. Use the scalar-field transformation law and the measure from (11) to justify this choice in the calculation, and show that $C(z)$ vanishes. Does the same argument require $W(z)$ to vanish? Explain the distinction between a correlation and a signal that can be controlled by a local intervention.

(c) Construct the Feynman Green function. Obtain $D_F(z)$ from (13) and express it as a four-momentum Fourier integral using (2). Determine the numerator and the pole prescription; state which side of the real energy axis contains the positive- and negative-energy poles. Verify the time-ordering boundary condition by performing the energy integral, either by residues or by a convergent time-transform argument.

(d) Check the defining equation. Apply $\Box_x + m^2$ to $D_F(x - y)$ and determine the complete delta-function source, including its sign and factor of i. Independently recover the contact term by differentiating the step functions in (13) and using the equal-time commutators. Check the mass dimensions on both sides. Explain why a Feynman propagator and a retarded response function are different objects; a separate contour calculation for the retarded function is not required.

Problem 4 From an interaction to a scattering amplitude

20 marks

Purpose. Derive a Feynman vertex and an amplitude from operator perturbation theory rather than assuming a diagram rule.

Consider the interaction-picture scalar field ϕ_I, which obeys the free equation and has the mode expansion (8). Introduce a weak positive coupling λ through

$$\mathcal{L}_{\text{int}} = -\frac{\lambda}{4!}\phi^4, \qquad H_I(t) = \frac{\lambda}{4!}\int \mathrm{d}^3\boldsymbol{x}\ :\phi_I(t,\boldsymbol{x})^4: . \tag{14}$$

Normal ordering is with respect to the free vacuum. For the tree calculation here it removes first-order self-contractions. It does not exclude loops at higher orders. The evolution operator is specified by

$$\mathrm{i}\frac{\partial U_I(t,t_0)}{\partial t} = H_I(t)U_I(t,t_0), \qquad U_I(t_0,t_0) = \mathbf{1}. \tag{15}$$

Use the usual adiabatic scattering limit $\mathsf{S} = U_I(+\infty,-\infty)$; the interaction is smoothly switched off at the time boundaries before taking this limit.

For generic distinct momenta define the two-particle asymptotic states by

$$|\boldsymbol{p}_1,\boldsymbol{p}_2\rangle = \sqrt{2E_{\boldsymbol{p}_1}\,2E_{\boldsymbol{p}_2}}\,a^\dagger_{\boldsymbol{p}_1}a^\dagger_{\boldsymbol{p}_2}|0\rangle. \tag{16}$$

The analogous definition holds for outgoing momenta k_1, k_2. All four particles have mass m and positive on-shell energy. Define the connected invariant amplitude by

$$\langle \boldsymbol{k}_1,\boldsymbol{k}_2|(\mathsf{S}-\mathbf{1})|\boldsymbol{p}_1,\boldsymbol{p}_2\rangle_{\text{conn}} = \mathrm{i}(2\pi)^4\delta^{(4)}(p_1+p_2-k_1-k_2)\,\mathcal{M}. \tag{17}$$

"Connected" selects the interaction joining all four external particles, excluding free propagation and vacuum factors. Wick's theorem may be used: a time-ordered product of free fields is the sum of normal-ordered products with all possible contractions; a scalar contraction is D_F as defined in Problem 3. External contractions can be obtained from Problem 2.

(a) Convert (15) to an integral equation and iterate it through second order in λ. Derive the time-ordered form and its factorial from the ordered integration region. Keep the operator ordering explicit; do not assume that H_I at different times commutes.

(b) Evaluate the first-order connected matrix element in (17). Exhibit the external contractions, their multiplicity, the space-time integral and all normalization factors. Determine the corresponding $\mathcal{M}$ and draw its diagram with labeled external momenta.

(c) Determine the mass dimension of λ and check the dimension and exchange symmetry of the amplitude. Distinguish the diagram's vertex factor, the matrix element of $\mathsf{S}-\mathbf{1}$, and $\mathcal{M}$. Identify what additional ingredients would be needed for a quantum correction; evaluating a loop integral is not part of this problem.

Problem 5 From the amplitude to an observable

20 marks

Purpose. Complete the scalar prediction and connect the framework to the introductory QED material.

Use the amplitude obtained in Problem 4. Define the Mandelstam invariants

$$s = (p_1 + p_2)^2, \qquad t = (p_1 - k_1)^2, \qquad u = (p_1 - k_2)^2. \tag{18}$$

Assume $s > 4m^2$. In the centre-of-mass frame, take $\boldsymbol{p}_1$ along the positive z axis and let θ be the angle between $\boldsymbol{p}_1$ and $\boldsymbol{k}_1$. The azimuthal angle is φ, with $\mathrm{d}\Omega = \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi$.

For covariantly normalized states, the invariant incident flux and two-body phase-space measure are

$$\mathcal{F} = 4\sqrt{(p_1 \cdot p_2)^2 - m^4}, \tag{19}$$

$$\mathrm{d}\Phi_2 = (2\pi)^4 \delta^{(4)}(p_1 + p_2 - k_1 - k_2) \frac{\mathrm{d}^3\boldsymbol{k}_1}{(2\pi)^3\, 2E_{\boldsymbol{k}_1}} \frac{\mathrm{d}^3\boldsymbol{k}_2}{(2\pi)^3\, 2E_{\boldsymbol{k}_2}}. \tag{20}$$

You may take the scattering-probability relation

$$\mathrm{d}\sigma = \frac{1}{2!\,\mathcal{F}} |\mathcal{M}|^2 \mathrm{d}\Phi_2 \tag{21}$$

as the starting formula. The outgoing integration variables are labeled and cover the full sphere $0 \le \theta \le \pi$, $0 \le \varphi < 2\pi$. The explicit $1/2!$ compensates for the identical final particles. In this assignment $\mathrm{d}\sigma/\mathrm{d}\Omega$ means this symmetrized full-sphere distribution; its integral counts each scattering event once. Do not add another symmetry factor or restrict to a hemisphere in the same calculation.

(a) Kinematics and phase space. Derive the incoming and outgoing momentum magnitudes in the centre-of-mass frame. Reduce (20) explicitly by performing the momentum delta integration and the remaining radial energy delta integration. Obtain $\mathrm{d}\Phi_2/\mathrm{d}\Omega$ and reduce (19) in the same frame. Track the Jacobian and every factor of 2π. Verify the on-shell relation among s, t, u.

(b) Differential and total cross section. Insert your amplitude into (21), obtain $\mathrm{d}\sigma/\mathrm{d}\Omega$ and integrate it over the stated angular domain. Explain its angular dependence and check its mass dimension. Explain how the counting would change if each identical final configuration were instead represented only once by a hemisphere convention; do not mix the two conventions.

(c) Limits and validity. Determine the energy dependence for $s \gg m^2$ at fixed weak coupling. Discuss the limit $s \downarrow 4m^2$ from above, including the incident flux and the distinction between a cross section and an event rate. State the order retained in perturbation theory and which physical effects the calculation does not determine. Explain one circumstance in which higher-order or nonperturbative methods would be necessary.

Problem 5 continued

A short connection to QED

This part concerns the field content and local symmetry of QED; no spinor solutions, photon quantization or QED cross-section calculation is required. For one charged Dirac field use

$$\mathcal{L}_{\mathrm{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(\mathrm{i}\gamma^{\mu}D_{\mu} - m_f)\psi, \tag{22}$$
$$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}, \qquad D_{\mu} = \partial_{\mu} + \mathrm{i}qA_{\mu}, \qquad \bar{\psi} = \psi^{\dagger}\gamma^{0}.$$

Here ψ is a four-component matter spinor, A_{μ} is the electromagnetic four-potential, m_f is the fermion mass (distinct from the scalar mass m), and q is its signed electromagnetic coupling. The constant matrices γ^{μ} obey

$$\{\gamma^{\mu}, \gamma^{\nu}\} = 2g^{\mu\nu}\mathbf{1}_4, \qquad \{A, B\} = AB + BA, \tag{23}$$

where $\mathbf{1}_4$ is the four-by-four identity matrix. Consider the local transformation with a real smooth function $\alpha(x)$,

$$\psi'(x) = \mathrm{e}^{-\mathrm{i}q\alpha(x)}\psi(x), \qquad A'_{\mu}(x) = A_{\mu}(x) + \partial_{\mu}\alpha(x). \tag{24}$$

(d) Fields, interaction and gauge symmetry. Identify the free matter and electromagnetic terms in (22), then expand D_{μ} to isolate the interaction term, including its sign. Use (24) to check how $D_{\mu}\psi$ transforms and explain why the full Lagrangian is gauge invariant. Explain why replacing D_{μ} by ∂_{μ} would not have the same local symmetry. Interpret the role of the electromagnetic potential in coupling the fields.

(e) Scientific interpretation. In one or two short paragraphs, discuss one connection between the field-theory framework and particle scattering, molecular radiation or an effective description of matter. Identify which ideas from the course carry over and what additional physics would be required for a realistic prediction. Include the scientific or technological significance of the example and support factual statements with a reference.

Parts (a)–(e) carry 20 marks in total, including the QED and application discussion.

Reading

[1] Roman Pasechnik, *Introduction to Quantum Field Theory*, course lecture notes. Relevant topics are the action principle and Noether's theorem; canonical scalar quantization; causality and propagators; interaction-picture perturbation theory; scalar scattering and phase space; and the QED gauge principle.

[2] David Tong, *Lectures on Quantum Field Theory*, University of Cambridge. Selected sections on classical fields, free fields, interacting fields and QED. `https://www.damtp.cam.ac.uk/user/tong/qft.htm`

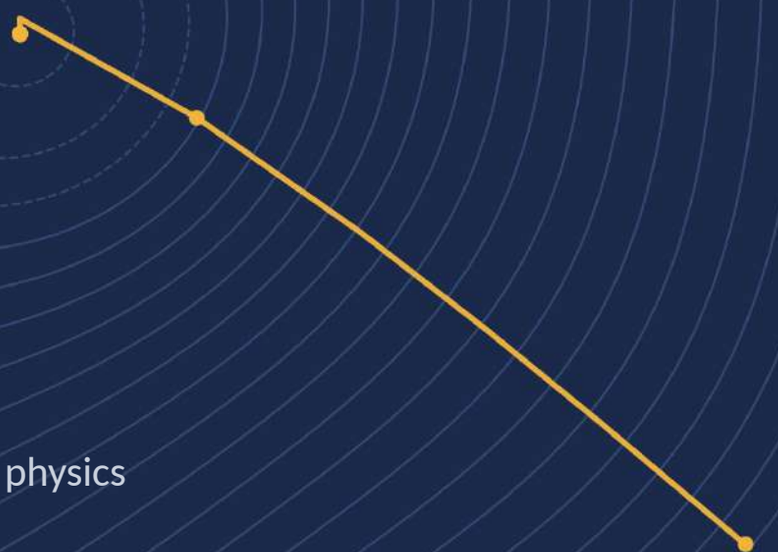
Modern machine learning and AI for physicists: How machines learn

OPEN LECTURES · UNIVERSITY OF RZESZÓW · OCTOBER 2026

Modern Machine Learning and AI for Physicists

Day 1 — How machines learn

Roman Pasechnik · Lund University



Day 1 How machines learn · Day 2 How modern AI works · Day 3 What AI does for physics

What does it mean for a machine to learn — and why should a physicist care?

2024 · Nobel Prize in Physics

J. Hopfield and G. Hinton, “for foundational discoveries and inventions that enable machine learning with artificial neural networks”: memory as an energy landscape, learning with Boltzmann statistics.

2024 · Nobel Prize in Chemistry

D. Baker for computational protein design; D. Hassabis and J. Jumper for protein structure prediction. AlphaFold has predicted more than 200 million protein structures.

2026 · Nobel Prize in Physics

F. Halzen for IceCube and the discovery of high-energy neutrinos of astrophysical origin. IceCube’s 2023 neutrino image of the Milky Way (4.5σ) relied on deep-learning event selection.

Three Nobel Prizes since 2024 involve learning machines — two for the methods themselves, one for a discovery that relied on them — and, as this course will show, physics is part of how they work.

The bottleneck has moved from collecting data to extracting physics from it

Data have outgrown human inspection

- **LHC:** 40 million bunch crossings per second; the trigger keeps about 1 in 40 000
- **Rubin Observatory:** about 20 TB of sky images every night, up to 10 million alerts
- **Spectroscopy, microscopy, synchrotrons:** automated instruments record far more spectra and images than anyone can look at
- **Biology:** more than 200 million predicted protein structures (AlphaFold)

What machine learning does for us

Classify
signal or background? which phase?

Regress
energy of a particle shower; a material property

Cluster
group spectra or images without labels

Generate
fast simulation of detector events or molecules

Detect anomalies
rare events, defects, new physics

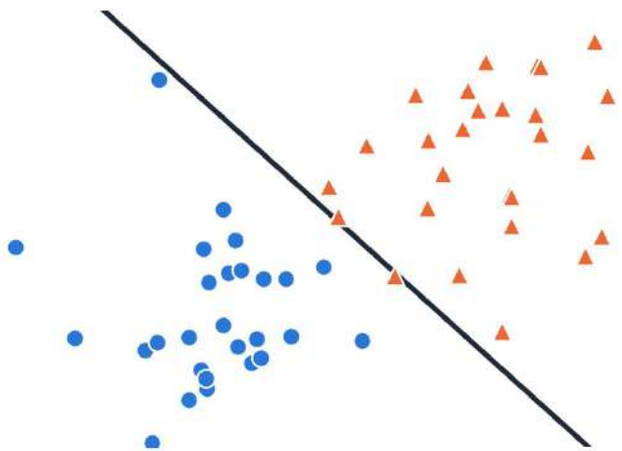
Infer
parameters with error bars from a simulator

Machine learning is a new instrument: like a microscope or a telescope, it lets us see structure in data that we could not see before.

Learning means improving at a task with experience — in three different settings

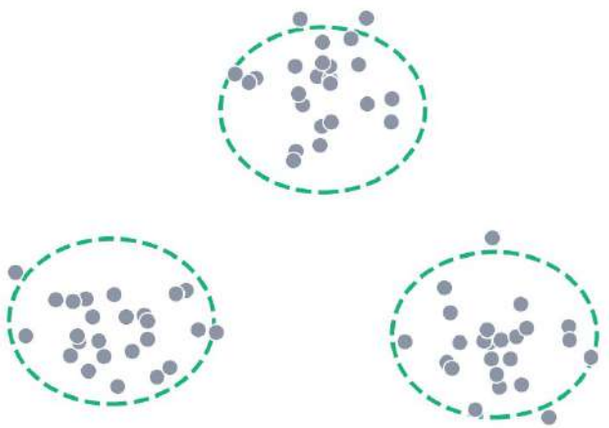
> *"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."* — T. Mitchell, Machine Learning (1997)

labelled examples → rule



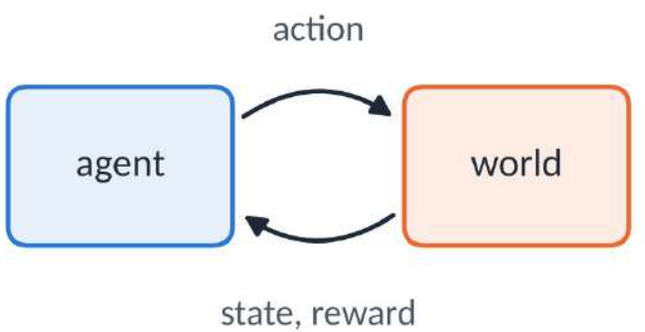
Supervised: Pairs (x, y) are given; learn $f(x) \approx y$. Example: spectrum → composition; jet → quark or gluon.

unlabelled data → structure



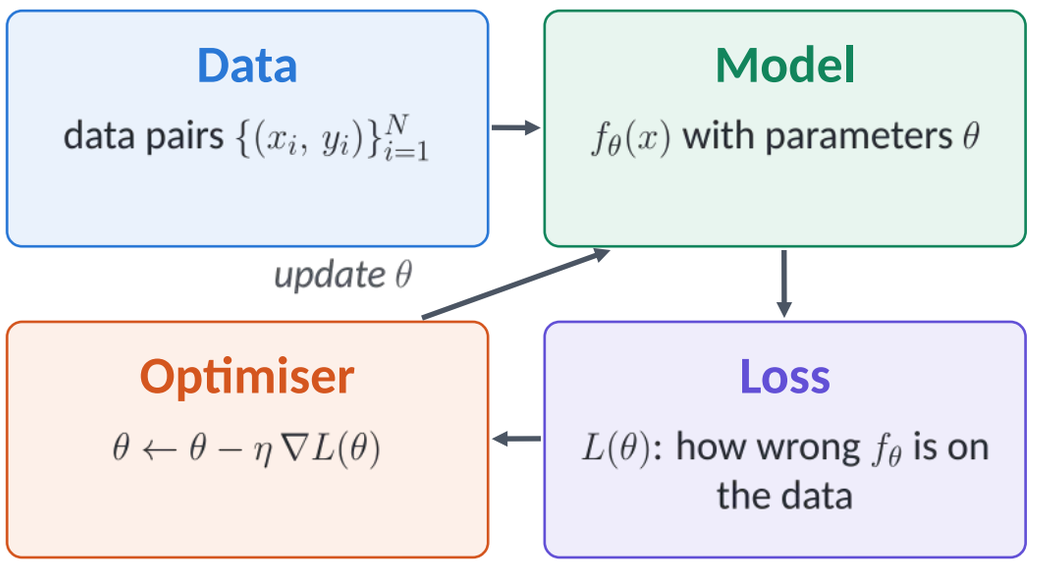
Unsupervised: Only x is given; find clusters, low-dimensional structure or the density $p(x)$. Example: sorting microscopy images; finding phases.

trial and error → policy

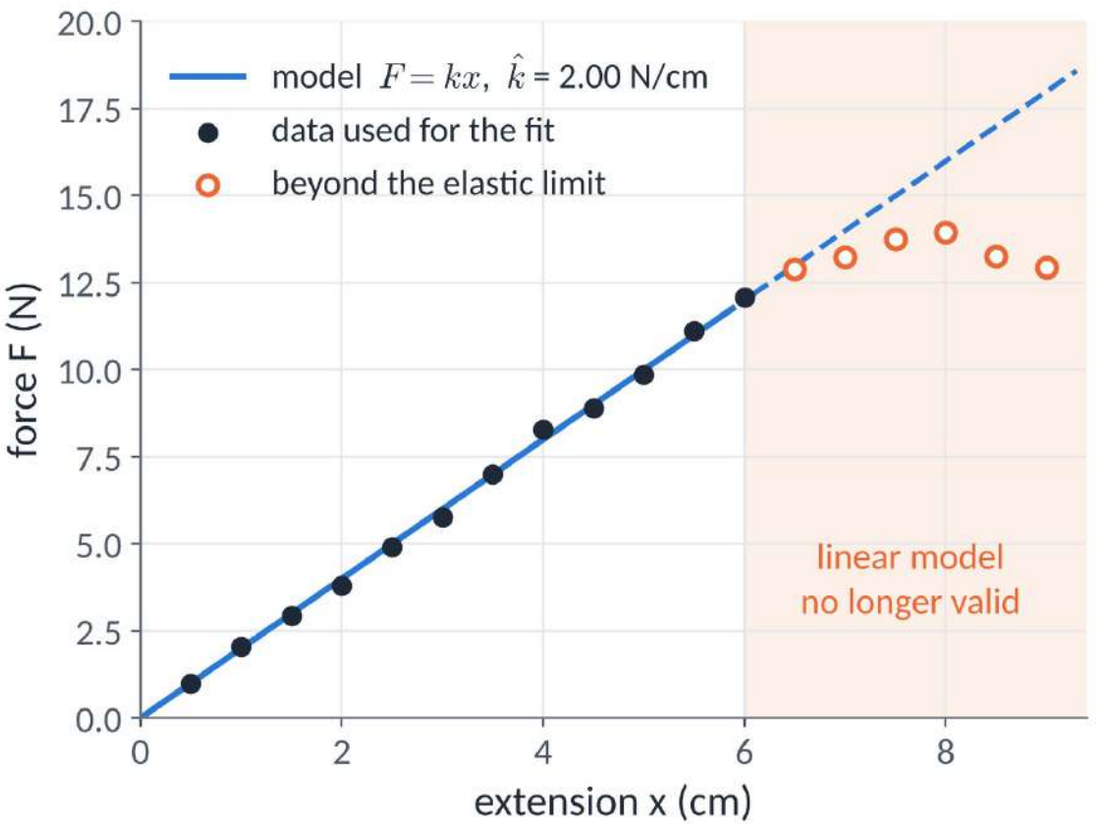


Reinforcement: Act, observe a reward, improve the policy. Example: shaping the plasma of a tokamak (Degrave et al., 2022).

Every learning method is four choices: data, model, loss and optimiser



repeat until the loss stops falling, then check on new data



- **Example — Hooke's law:** data (x, F), model $F = kx$, squared error; least squares gives $\hat{k} = 2.00 \pm 0.01$ N/cm.
- **Every choice is an assumption:** beyond the elastic limit the linear model is simply wrong — no optimiser can fix that.

Key equation 1: learning is the minimisation of the average loss on the data

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \ell\big(f_\theta(x_i),\, y_i\big)$$

$\hat{\theta}$

the learned parameters — the outcome of training

$\arg\min_{\theta}$

search the whole model family for the best θ

$\frac{1}{N}\sum_{i=1}^{N}$

average over the N training examples: the ‘empirical risk’

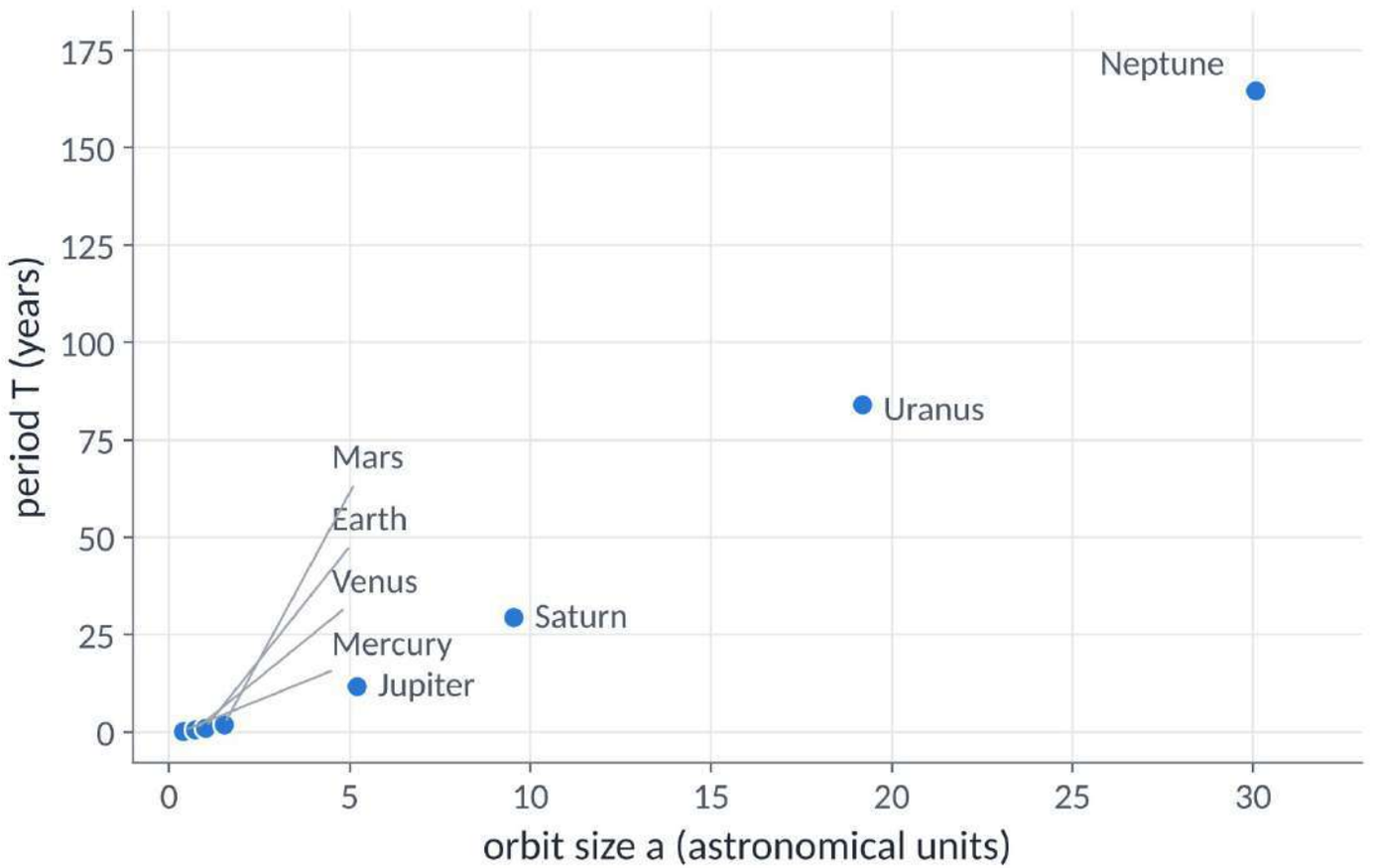
$\ell\big(f_\theta(x_i), y_i\big)$

the loss: how wrong the prediction is for example i

$$\underbrace{R(\theta) = \mathbb{E}_{(x,y)\sim p_{\text{data}}}\, \ell\big(f_\theta(x), y\big)}_{\text{what we want: error on new data}} \approx \underbrace{\frac{1}{N}\sum_{i=1}^{N} \ell\big(f_\theta(x_i), y_i\big)}_{\text{what we can compute}}$$

Hidden assumption: new data come from the same distribution as the training data (independent, identically distributed).

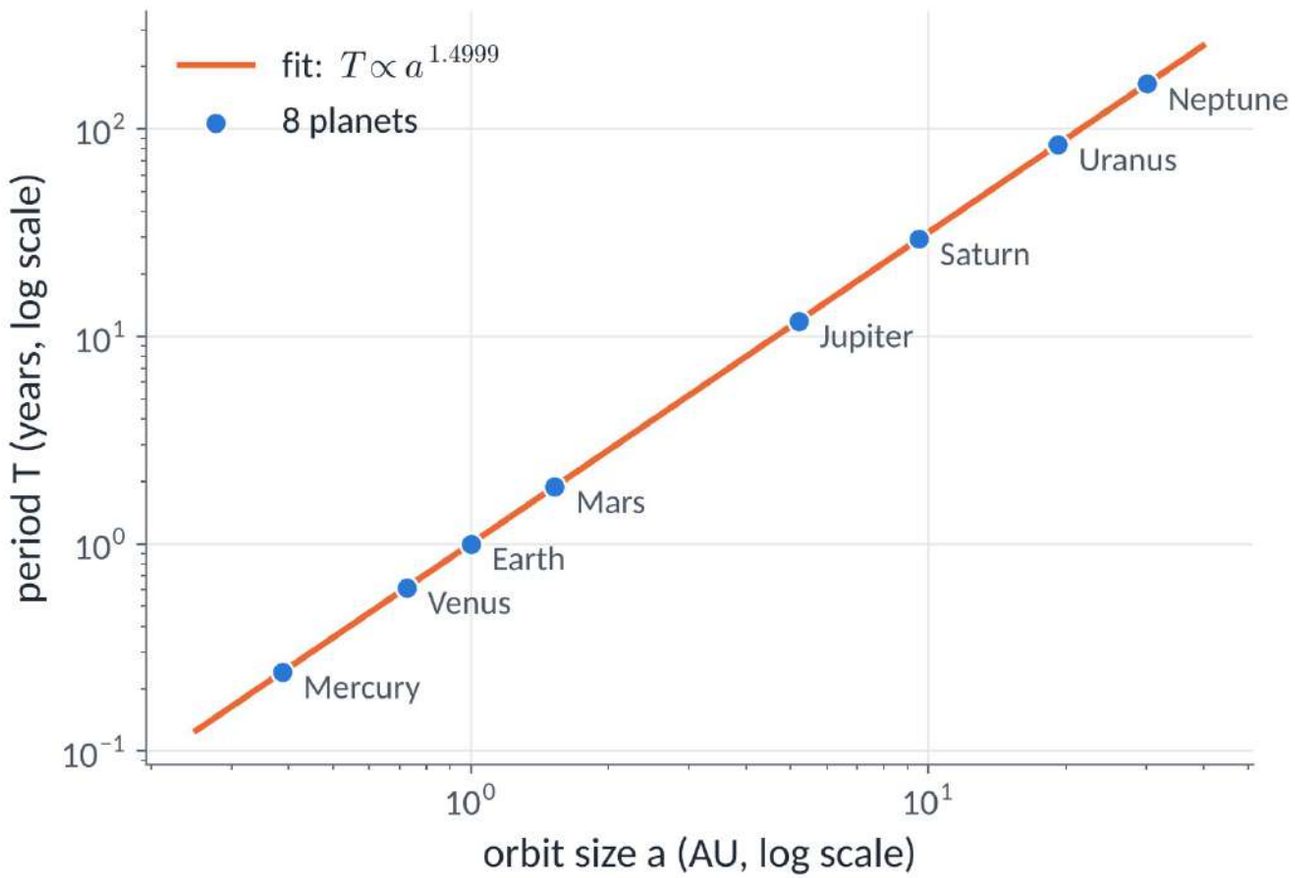
Predict: how does a planet's orbital period depend on the size of its orbit?



planet	a (AU)	T (years)
Mercury	0.38710	0.24085
Venus	0.72333	0.61520
Earth	1.00000	1.00002
Mars	1.52371	1.88085
Jupiter	5.20289	11.8626
Saturn	9.53667	29.4475
Uranus	19.1892	84.0168
Neptune	30.0699	164.791

Your guess: if $T \propto a^p$, what is p?
Hint: try a different way of plotting the same numbers.

In log–log axes the eight planets fall on a straight line: Kepler's third law



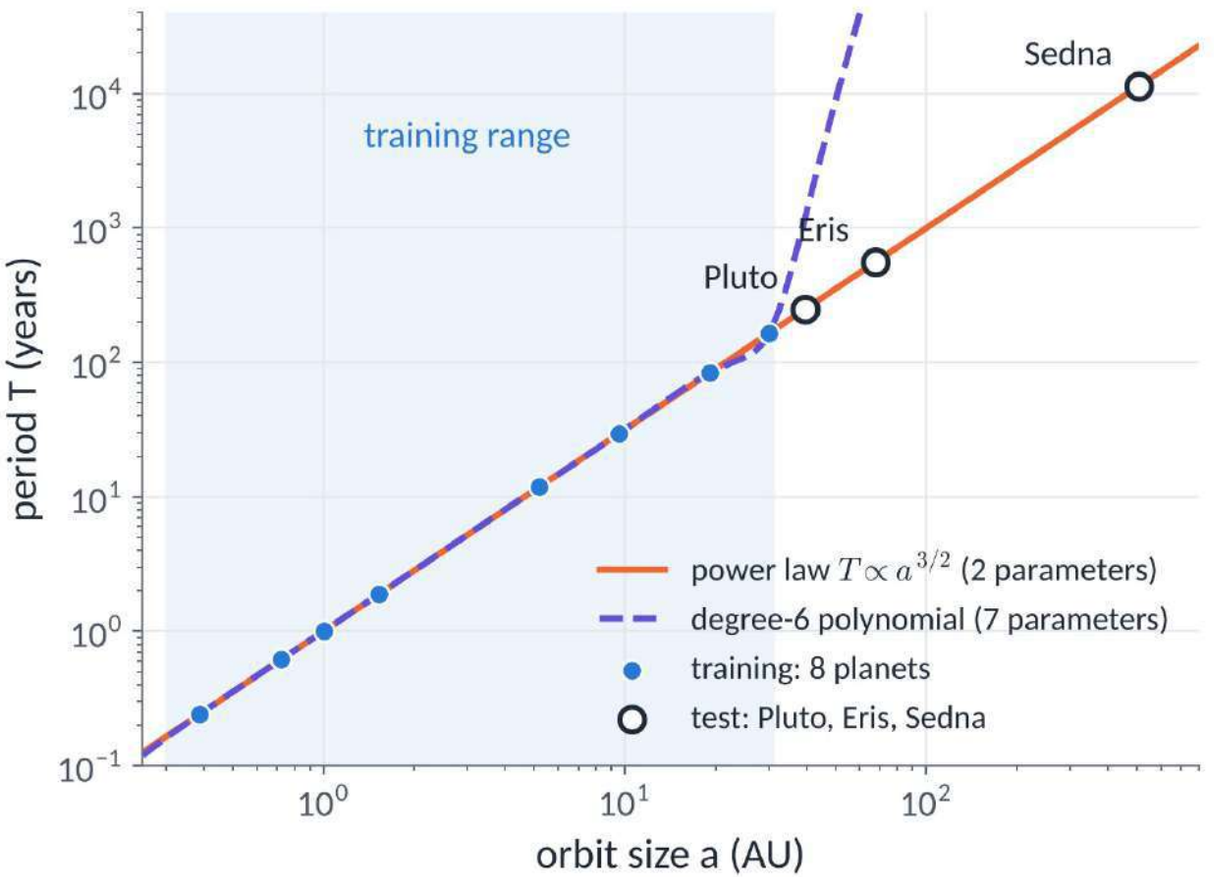
$$\log T = p \, \log a + c$$

- **Model:** a straight line in $(\log a, \log T)$, linear in the two parameters p and c.
- **Loss and optimiser:** squared error, solved exactly by least squares.
- **Result:** $p = 1.4999$, i.e. $T^2 \propto a^3$ to about one part in 10^4.
- **The key step was the representation:** logarithms turned a curve into a line.

Kepler found the law in 1619. Newton explained it in 1687: $p = 3/2$ follows from a $1/r^2$ force.

Two models agree on the training data and disagree completely beyond it

- **Both fit the eight planets:** a power law (2 parameters) and a degree-6 polynomial (7 parameters, within 0.2%).
- **Pluto:** power law 248 yr, polynomial $1\,163$ yr; observed 248 yr.
- **Eris, Sedna:** the power law stays right (559 and $\approx 11\,400$ yr); the polynomial says $98\,000$ yr and 5×10^{10} yr.
- **Why:** the power law encodes structure (Newton's $1/r^2$ force); a generic fit is reliable only where it has seen data.



A good fit is not an explanation. Physics-informed machine learning (Day 3) combines theory and data.

Three things to remember from Block 1

1. Learning = choosing a function from a model family by minimising the average loss on data: empirical risk minimisation.

2. Every method is four choices — data, model, loss, optimiser — and each choice is an assumption about the world.

3. The goal is to predict new data, not to fit old data. How well a model generalises is the real question — the subject of Block 2.

Three polynomials fit the same ten measurements. Which one will predict new data best?



The degree-9 polynomial passes through every point — zero training error. Is it the best model?

Only data that the model has never seen can judge it



fit the parameters θ

choose the model and its hyperparameters

final, unbiased estimate — used once

Leakage: test information sneaking into training

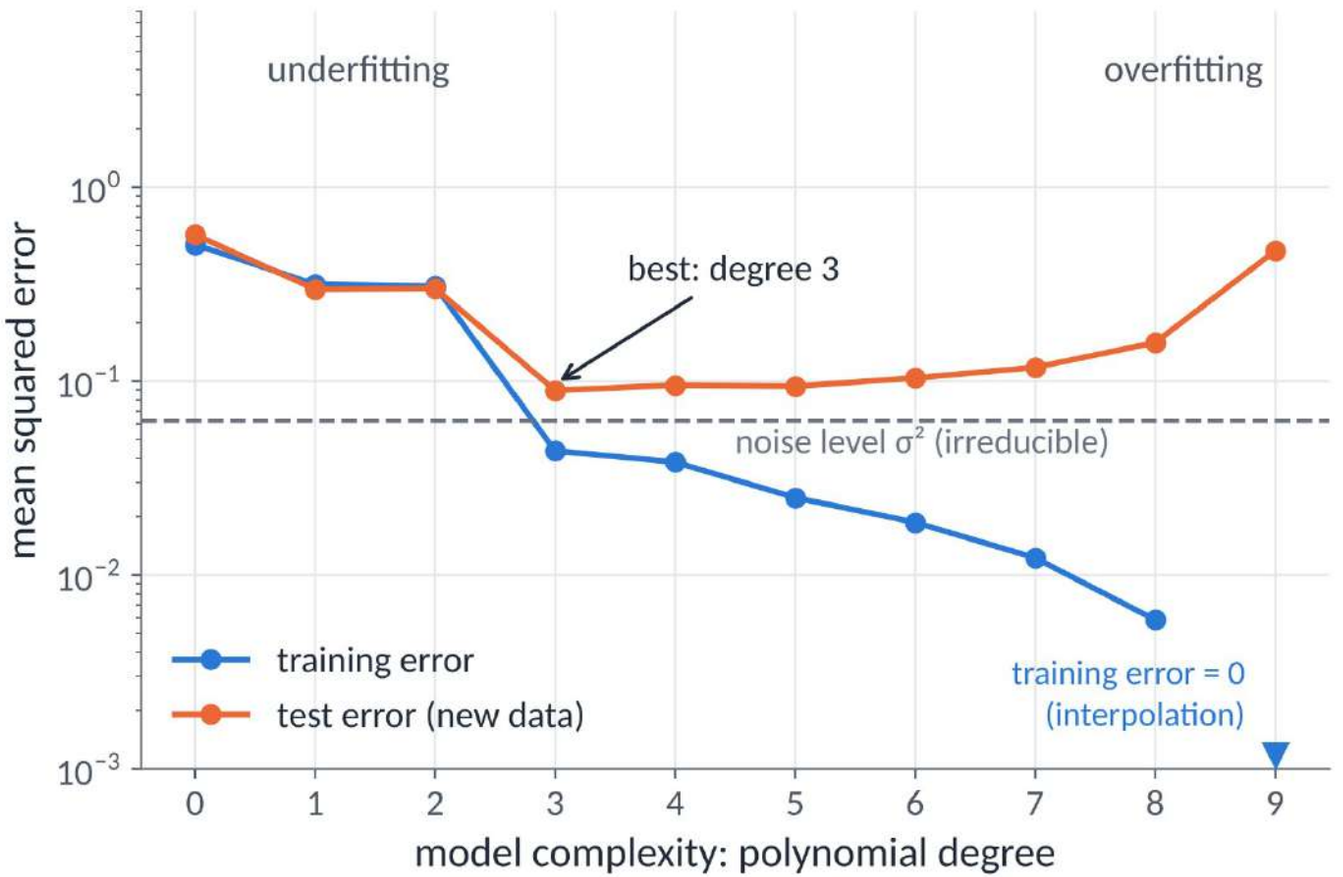
- normalising or selecting features with all data before splitting
- the same sample, run or specimen in both sets
- tuning the model until the test score looks good

Each one makes the test score optimistic.

The physicist's version: blind analysis

Particle physicists fix the whole analysis before looking at the signal region — ‘unblinding’ happens once, at the end. Machine learning needs the same discipline: the test set is opened once.

Training error always falls with complexity; the error on new data has a minimum



- **Setting:** 10 measurements of $\sin 2\pi x$ with noise $\sigma = 0.25$, averaged over many datasets.
- **Training error** falls monotonically and reaches zero at degree 9 (interpolation).
- **Test error** is lowest at degree 3 (0.090) and rises to 0.47 at degree 9.
- **Underfitting:** the model family cannot represent the truth.
- **Overfitting:** the model memorises the noise.
- On new data no model beats the noise level $\sigma^2 = 0.0625$; a training error below it means memorised noise.

Capacity must match the amount and quality of the data.

Key equation 2: the prediction error splits into bias squared, variance and noise

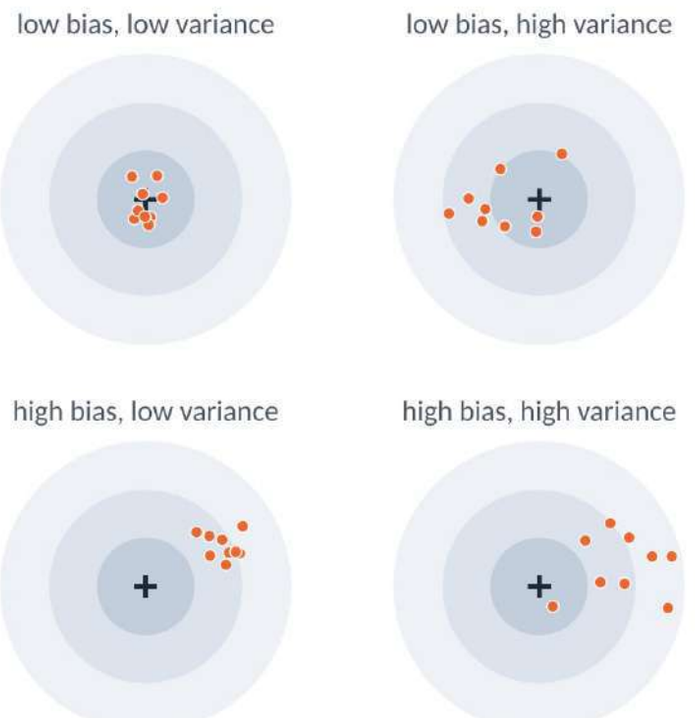
$$\mathbb{E}\big[(y - \hat{f}(x))^2\big] = \underbrace{\big(\mathbb{E}\hat{f}(x) - f(x)\big)^2}_{\text{bias}^2} + \underbrace{\mathbb{E}\big[(\hat{f}(x) - \mathbb{E}\hat{f}(x))^2\big]}_{\text{variance}} + \underbrace{\sigma^2}_{\text{noise}}$$

bias^2 Systematic error: the average model misses the truth. Typical of models that are too rigid.

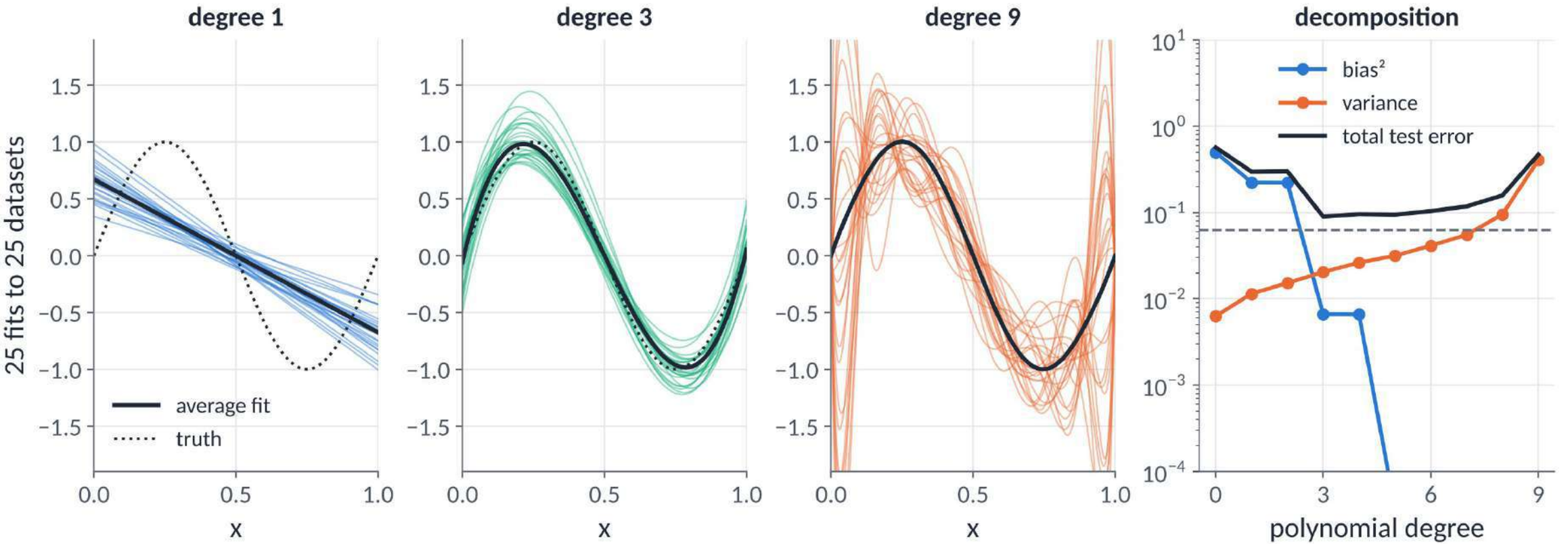
variance Statistical error: the fit changes from one training set to the next. Typical of models that are too flexible.

noise σ^2 Irreducible scatter of the measurements themselves. No model can remove it.

Physics lens: bias ↔ systematic error, variance ↔ statistical error.



Measured: refit the same model to many independent datasets



Degree 1: fits agree with each other but miss the truth: $\text{bias}^2 = 0.22$, variance $= 0.01$

Degree 3: right on average, little scatter: total error 0.090

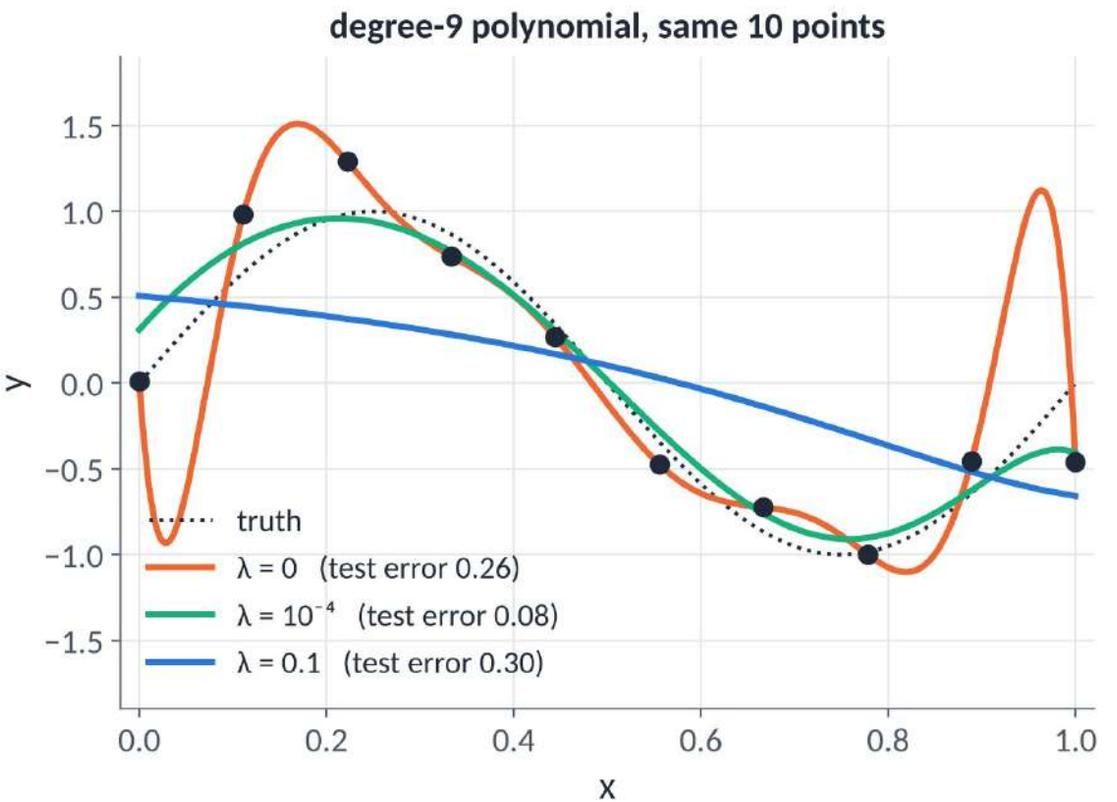
Degree 9: right on average but wildly scattered: variance $= 0.41$

Regularisation trades a little bias for a large reduction in variance

$$L_\lambda(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \theta^\top \phi(x_i)\right)^2 + \lambda \left\|\theta\right\|_2^2$$

$$\text{lasso:} \quad \ldots + \lambda \left\|\theta\right\|_1 = \ldots + \lambda \sum_k |\theta_k|$$

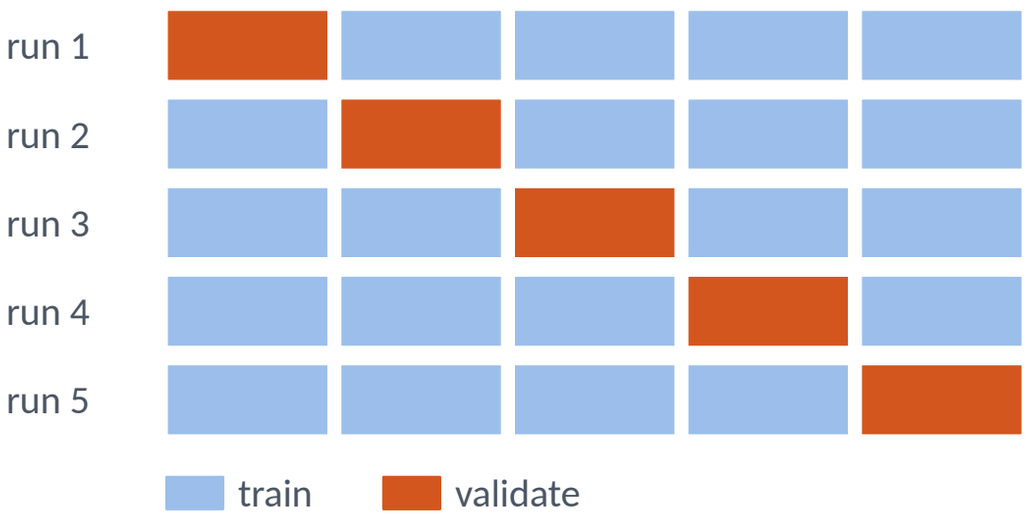
- **Ridge (L_2):** penalises large coefficients (not the intercept) — smoother fits.
- **Lasso (L_1):** sets many coefficients exactly to zero — sparse, interpretable models (equation discovery, Day 3).
- **Other regularisers:** early stopping, data augmentation (e.g. symmetries), dropout, smaller models.
- **λ is a hyperparameter,** chosen on validation data (next slide).



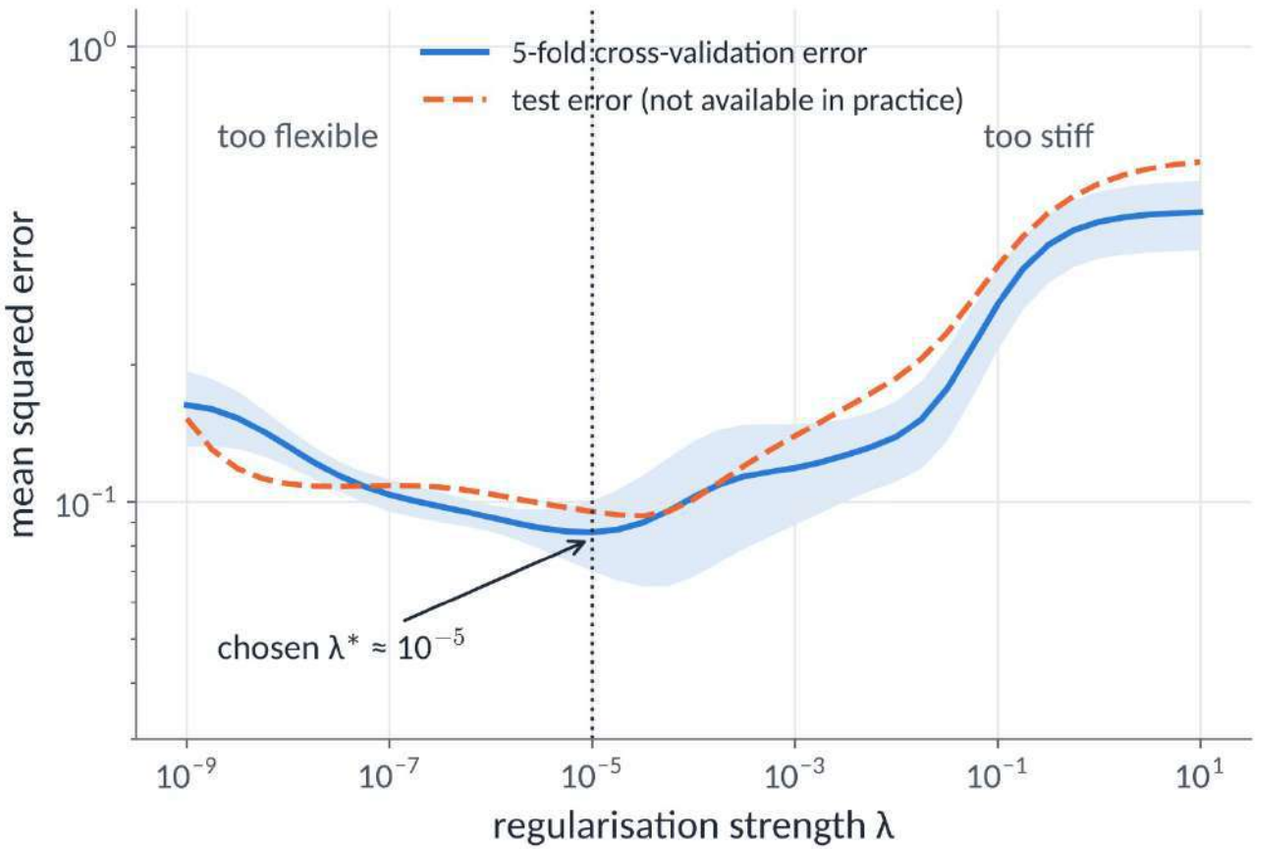
Physics lens: the penalty is a prior belief about θ (Block 4).

Cross-validation chooses the regularisation strength from the training data alone

5-fold cross-validation

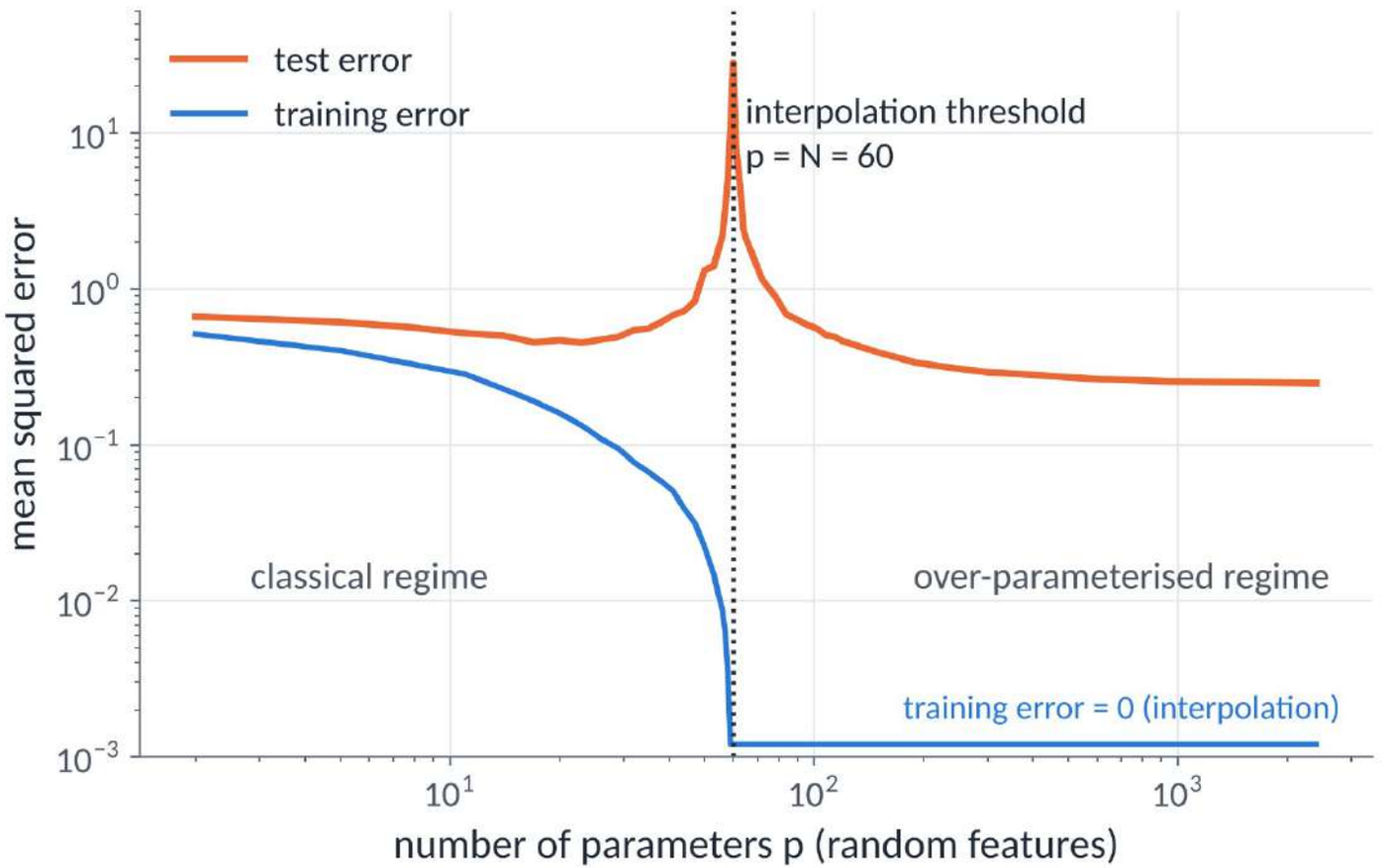


- For each λ: train 5 times, each time validating on the held-out fold; average the 5 errors.
- Pick λ^* with the lowest average; refit on all training data; report once on the test set.



15 points, degree-9 polynomial: CV picks $\lambda^* = 10^{-5}$; test error 0.095, close to the best possible 0.093 (0.152 without regularisation).

A modern surprise: beyond the interpolation threshold the error can fall again

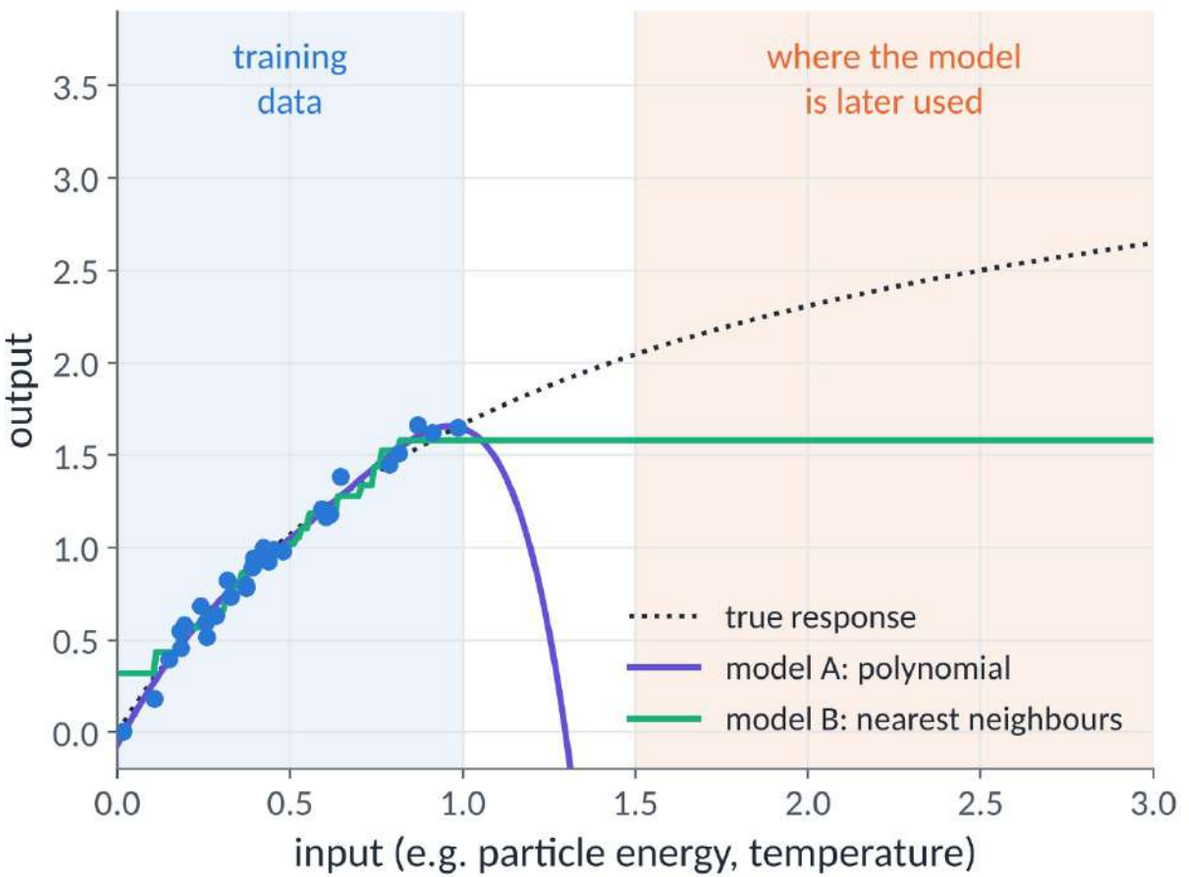


- **Model:** random ReLU features, $N = 60$ training points, p parameters, minimum-norm least squares.
- $\boldsymbol{p < N}$**:** the classical U-shape; best test error 0.45 at $p \approx 23$.
- $\boldsymbol{p = N}$**:** the model can just interpolate; the fit is forced to be wild and the error peaks.
- $\boldsymbol{p \gg N}$**:** of all the perfect fits, the smoothest (minimum-norm) one is chosen; the error falls to 0.25 — below the classical optimum.
- Gradient descent started at zero finds exactly this fit: an implicit regulariser. Large neural networks work in this regime.

Belkin, Hsu, Ma & Mandal, PNAS 116, 15849 (2019)

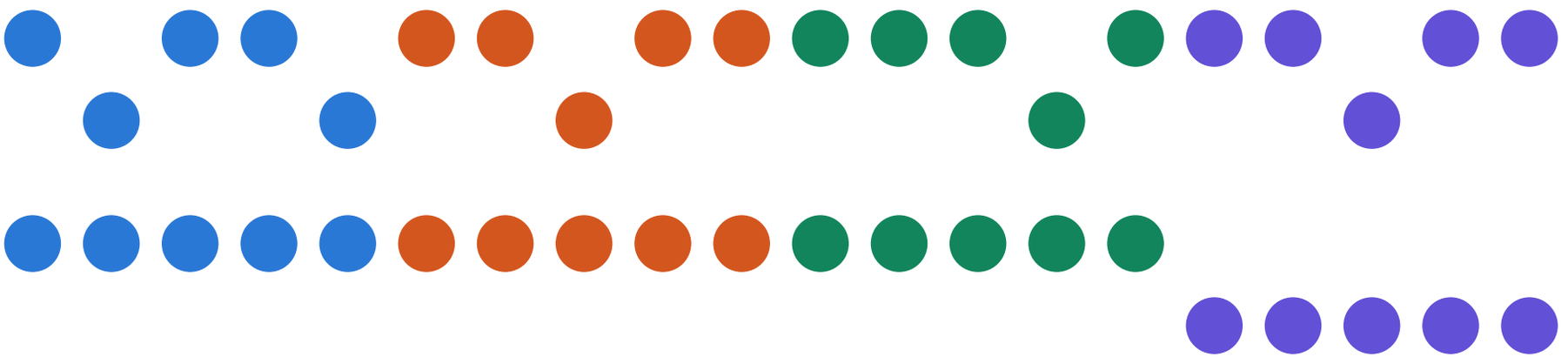
When the data distribution shifts, a model can fail however well it was validated

- Both models fit the training range (inputs 0–1) equally well and pass validation.
- Used at inputs 1.5–3, they disagree completely — and both are wrong.
- **In physics:** training on simulation and applying to real detector data; calibrating at one temperature, sample or composition and using another; instrument drift.
- **Remedies:** validation data that look like the target domain; include the expected variations in training; quantify the uncertainty (Day 3).



A validation score holds only for the distribution it was measured on.

Optional: when samples are correlated, a random split overestimates performance



colour = specimen (4 specimens × 5 measurements); lower row = test set

- Measurements from the same specimen, run, day or simulation share hidden properties.
- A random split puts near-duplicates into training and test, so the test score is optimistic.
- **Group k-fold:** keep each group inside one fold.
- **Time series:** train on the past, test on the future.
- **Nested cross-validation:** an inner loop chooses hyperparameters, an outer loop estimates performance.

Optional · expert corner: why can a model that interpolates noisy data still generalise?

$$\hat{\theta} = \arg\min \left\{ \|\theta\|_2 : \ \Phi\theta = y \right\} = \Phi^\top \left(\Phi\Phi^\top \right)^{-1} y$$

- With $p > N$ parameters, $\Phi\theta = y$ has infinitely many solutions; the minimum-norm one is unique.
- Gradient descent on the squared loss started at $\theta = 0$ converges exactly to it: its updates $\Phi^\top(\Phi\theta - y)$ never leave the row space of Φ.
- Near $p = N$ the smallest singular value of Φ goes to zero, so the norm of the solution and the variance blow up: the peak of double descent.
- For $p \gg N$ the noise can be absorbed by many weak directions that hardly affect predictions: ‘benign overfitting’ (Bartlett, Long, Lugosi & Tsigler, PNAS 2020).

Three things to remember from Block 2

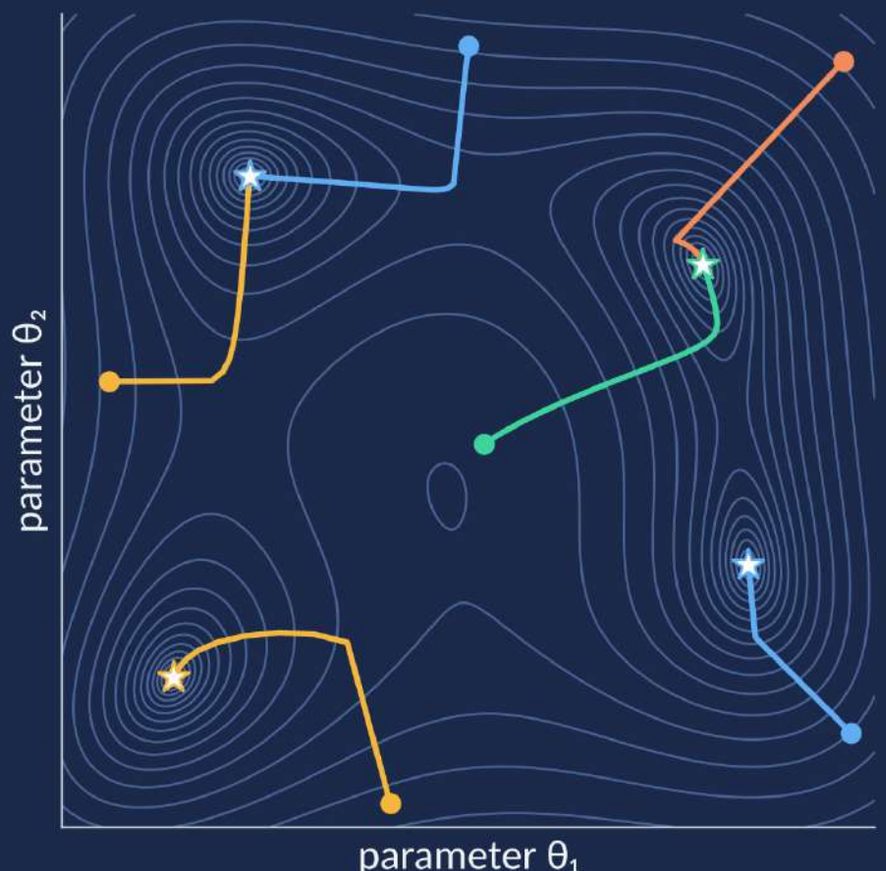
1. Training error flatters. Only held-out data judge a model — and the test set is opened once.

2. Error $=$ bias^2 $+$ variance $+$ noise $(\text{systematic} + \text{statistical} + \text{irreducible})$. Regularisation and cross-validation set the balance.

3. Heavily over-parameterised models can still generalise (double descent), but a validation score holds only for the distribution it was measured on.

How do we find the lowest point of a landscape with millions of dimensions — in the fog?

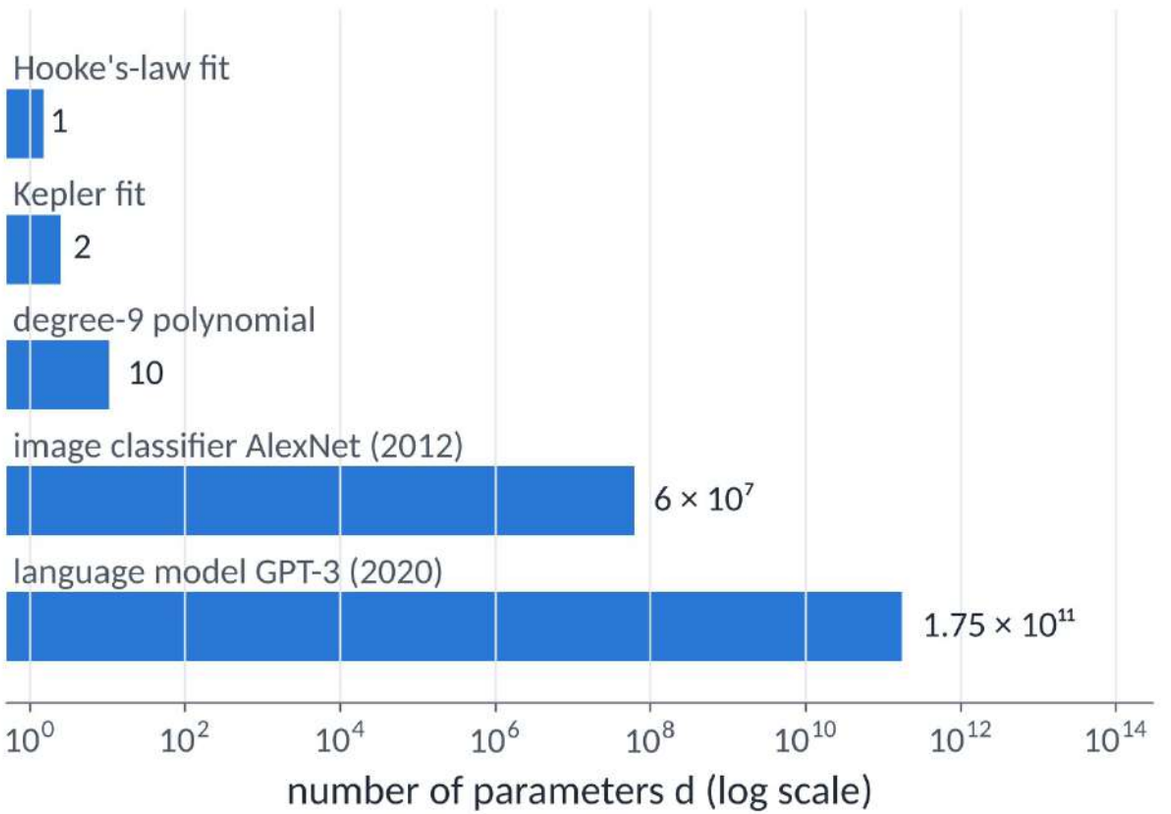
- The loss $L(\theta)$ is a landscape over parameter space; training means finding a deep valley.
- We never see the landscape — only the slope where we stand: the gradient.
- Six walkers from different starting points end up in four different valleys: where you start matters.
- In machine learning θ has from a few to 10^{12} dimensions.



For linear least squares an exact solution exists — but it does not scale

$$\nabla_\theta \tfrac{1}{N}\|y - X\theta\|^2 = -\tfrac{2}{N}X^\top(y - X\theta) = 0 \quad \Rightarrow \quad \hat{\theta} = (X^\top X)^{-1}X^\top y$$

- **Cost of the exact solution:** forming $X^\top X$ takes $\propto Nd^2$ operations, solving the system $\propto d^3$.
- $d = 10^8$ **parameters:** $d^3 = 10^{24}$ operations — about 30 years at 10^{15} operations per second.
- **Nonlinear models** (logistic regression, neural networks) have no closed-form solution at all.
- **Non-convex losses** have many minima and saddles.
- **Data arrive** in batches or streams.



We need an iterative method that uses only local information: the gradient.

Key equation 3: gradient descent — take a small step downhill, and repeat

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t)$$

θ_t

current parameters; t counts the steps

η

learning rate: the size of each step

$\nabla_\theta L(\theta_t)$

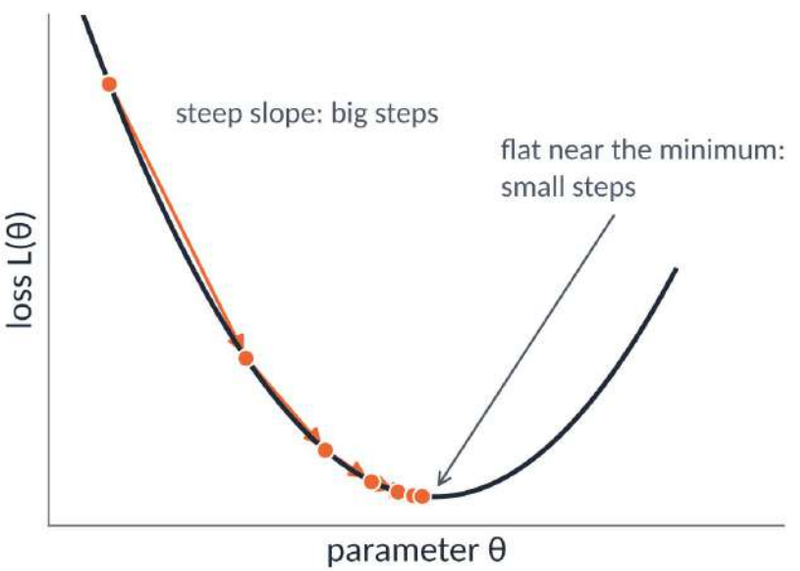
gradient: steepest ascent; we step against it

Physics lens: gradient flow is overdamped motion

$$m\,\ddot{\theta} + \gamma\,\dot{\theta} = -\nabla L \quad \xrightarrow{m \to 0} \quad \gamma\,\dot{\theta} = -\nabla L$$

$$\frac{d\theta}{dt} = -\nabla L(\theta) \qquad \Rightarrow \qquad \frac{dL}{dt} = -\|\nabla L\|^2 \leq 0$$

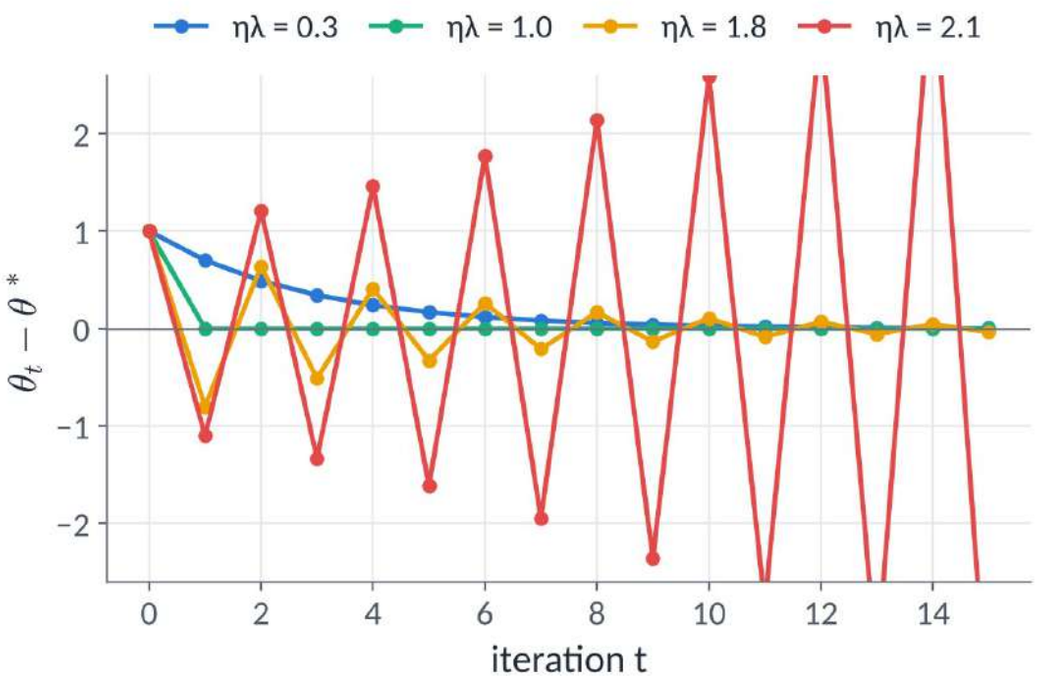
A particle sliding down the potential L through a viscous fluid; gradient descent is the Euler step of this flow with $dt = \eta$. Along the flow the loss can only decrease: L is a Lyapunov function.

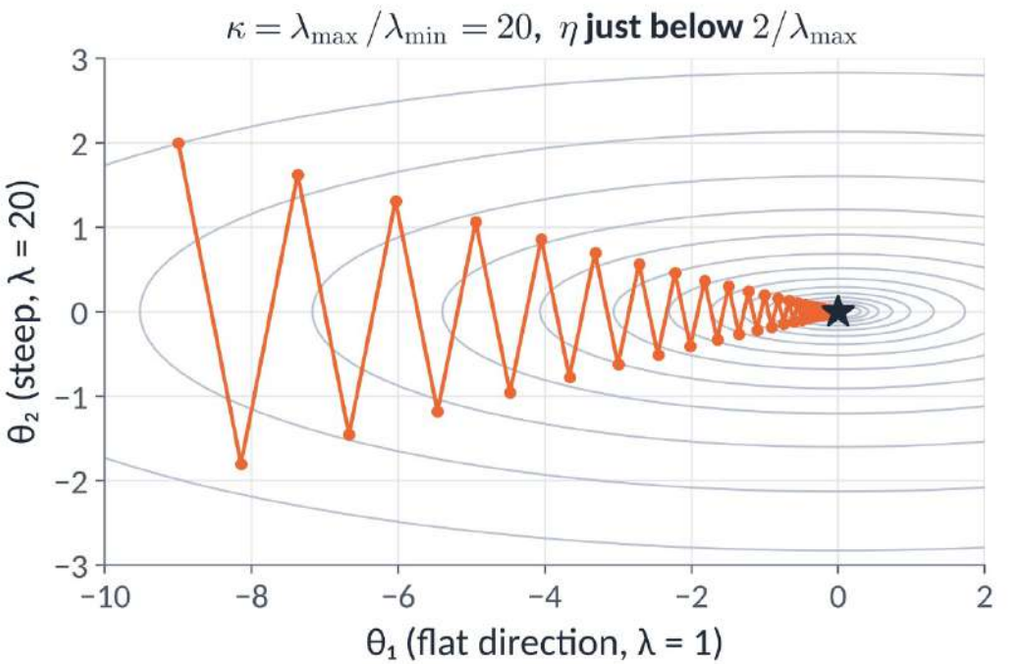


Stability: the learning rate must respect the curvature of the loss

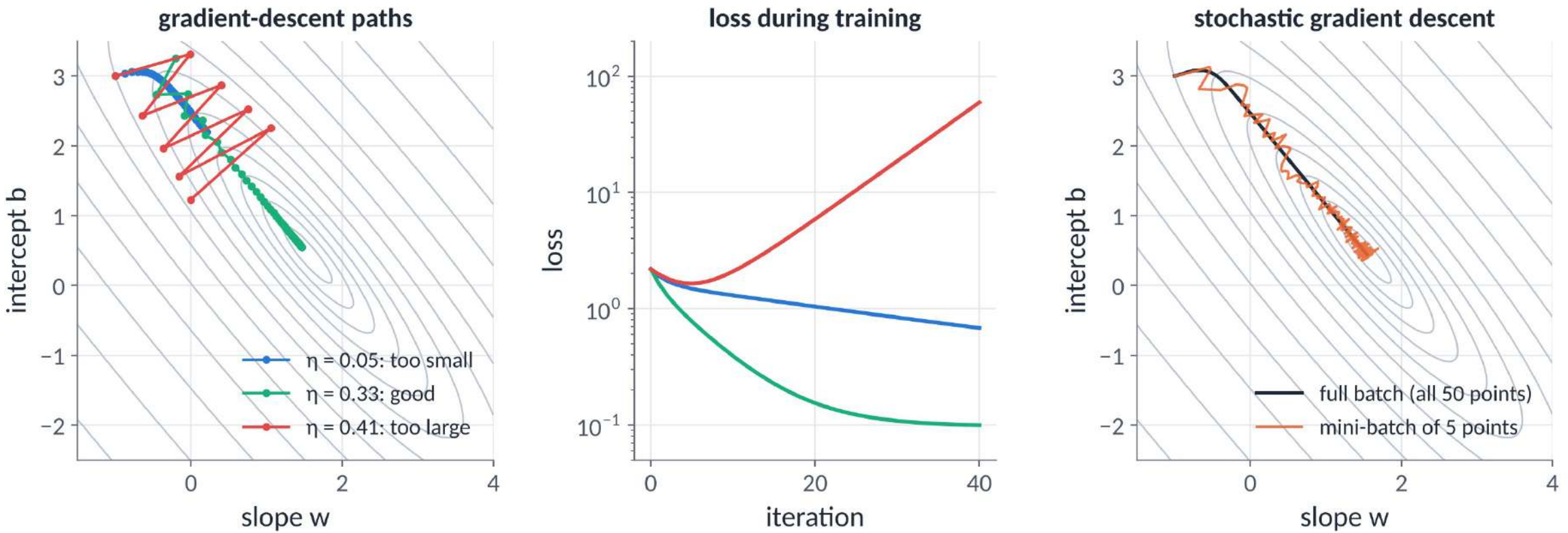
$$\theta_{t+1} - \theta^* = (1 - \eta\lambda)\,(\theta_t - \theta^*) \quad \Rightarrow \quad \text{converges iff } \; 0 < \eta < \frac{2}{\lambda}$$

$$\text{many directions:} \quad \eta < \frac{2}{\lambda_{\max}}, \qquad \text{steps} \sim \kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$$





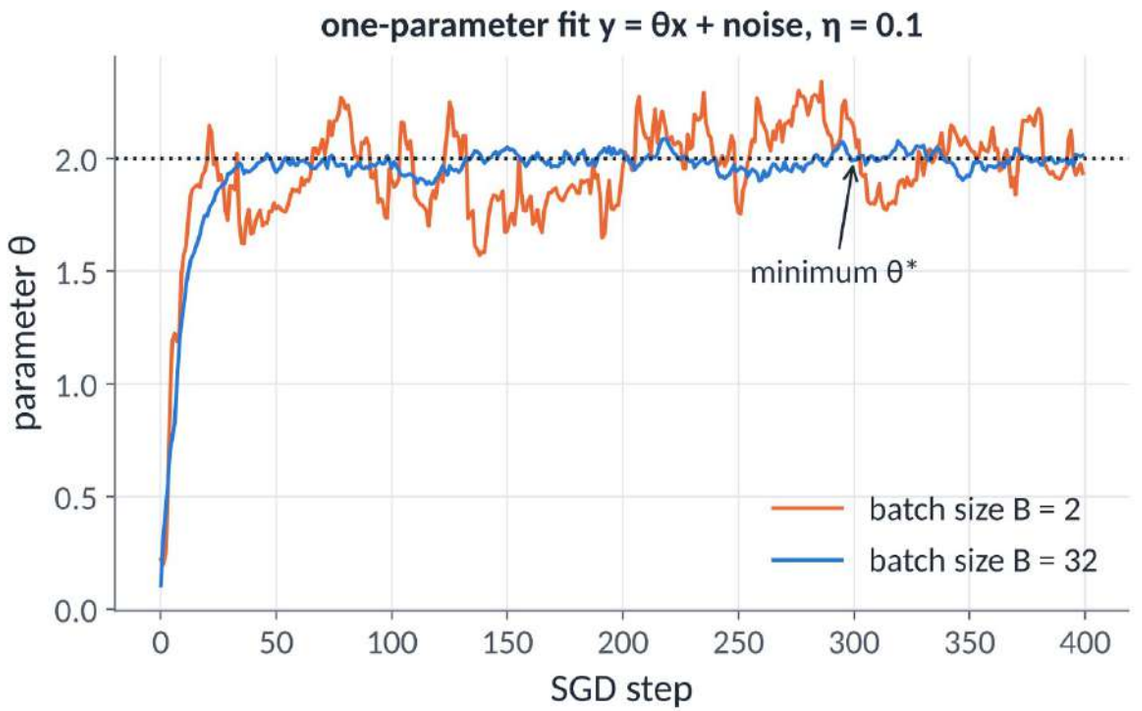
Gradient descent in practice: too small crawls, too large diverges, mini-batches add noise

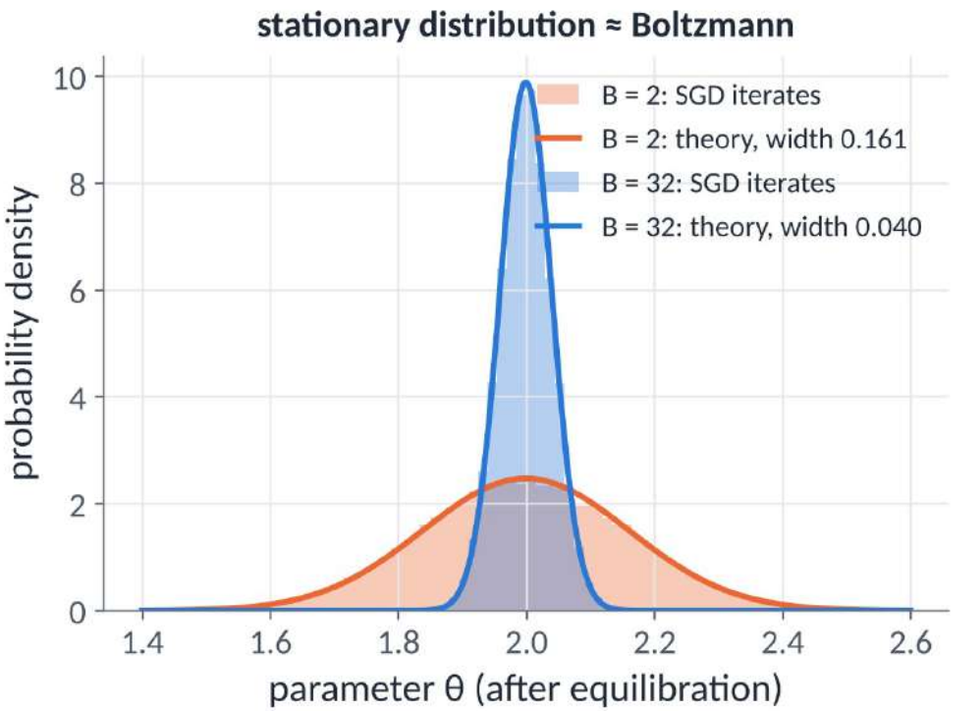


Straight-line fit to 50 points. Largest curvature $\lambda_{\max} = 5.02$, so the limit is $\eta < 2/\lambda_{\max} = 0.40$: $\eta = 0.41$ diverges, $\eta = 0.33$ reaches the noise floor in 40 steps, $\eta = 0.05$ is still far away. The zig-zag comes from $\kappa = \lambda_{\max}/\lambda_{\min} \approx 21$.

Mini-batch noise acts like a temperature: SGD samples a Boltzmann-like distribution

$$\theta_{t+1} = \theta_t - \eta\,\nabla L_B(\theta_t), \qquad \nabla L_B = \nabla L + \xi, \quad \mathrm{Var}\,\xi \simeq \frac{\sigma_g^2}{B} \qquad d\theta = -\nabla L\,dt + \sqrt{2T_{\mathrm{eff}}}\,dW, \qquad T_{\mathrm{eff}} = \frac{\eta\,\sigma_g^2}{2B}$$





One-parameter linear fit; σ_g^2 = variance of a single-example gradient. Widths: measured 0.165 ($B = 2$) and 0.041 ($B = 32$), theory 0.161 and 0.040. Small batches are 'hot', large batches 'cold'.

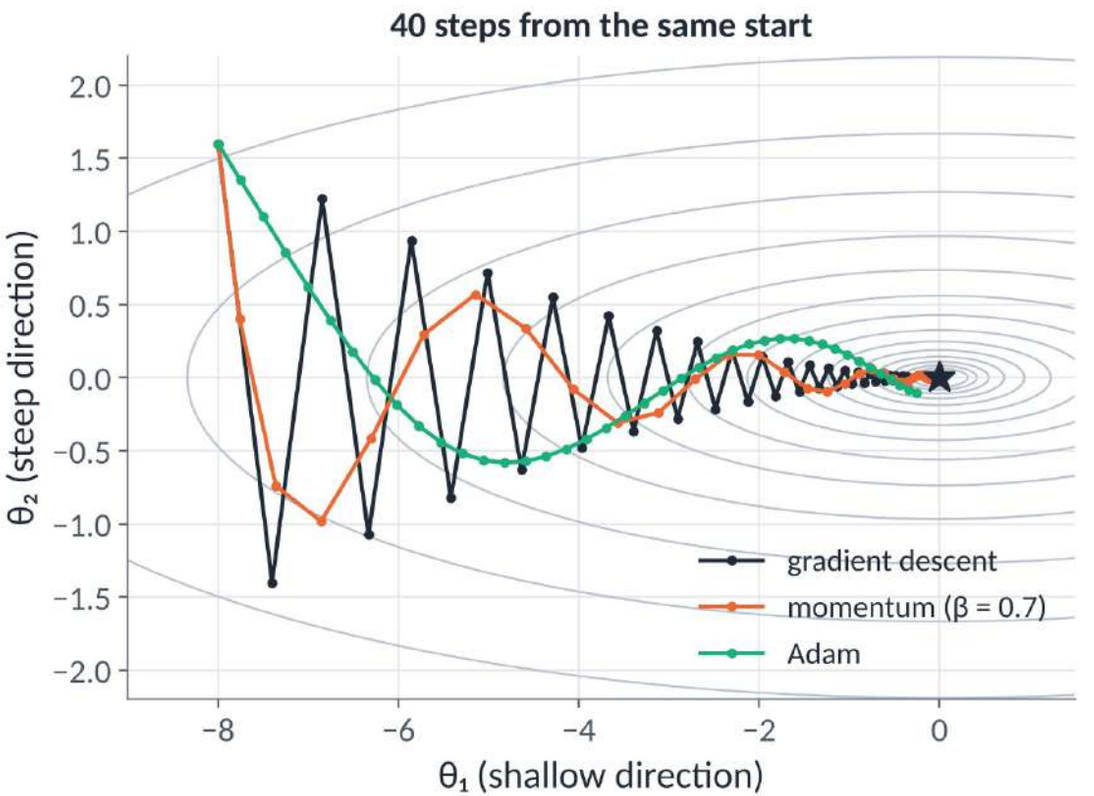
Momentum and Adam: a heavy ball with friction, and a step size for every parameter

$$v_{t+1} = \beta\, v_t - \eta\, \nabla L(\theta_t), \qquad \theta_{t+1} = \theta_t + v_{t+1}$$

$$\theta_{t+1} = \theta_t - \eta\, \frac{\hat{m}_t}{\sqrt{\hat{s}_t} + \epsilon}$$

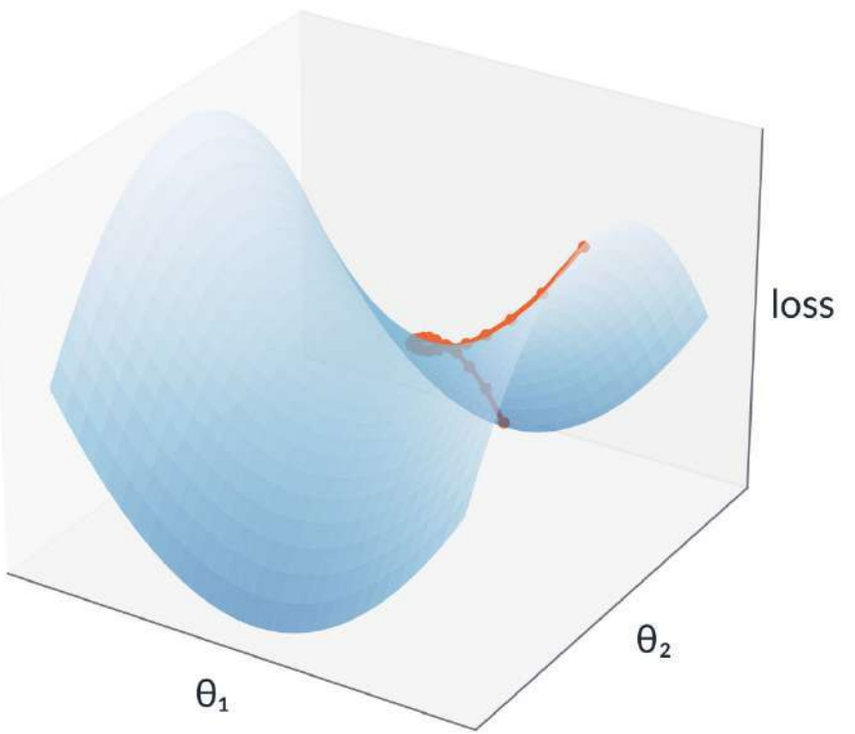
$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\, g_t, \;\; \hat{m}_t = m_t/(1-\beta_1^t),$$
$$s_t = \beta_2 s_{t-1} + (1-\beta_2)\, g_t^2, \quad \hat{s}_t = s_t/(1-\beta_2^t)$$

- **Momentum:** v accumulates past gradients — Newton's law with friction ($\beta \leftrightarrow 1 -$ friction $\times$ time step): faster along shallow valleys, less zig-zag across steep ones.
- **Adam:** m and s average the gradient g_t and its square (hats: start-up correction); dividing by $\sqrt{s}$ gives each parameter its own step size. Defaults $\eta = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ (Kingma & Ba, 2015).
- **Here, 40 steps, hand-tuned η** (GD 0.075; momentum 0.03, $\beta = 0.7$; Adam 0.25): distance to the minimum 0.35, 0.004, 0.28.



One algorithm at every scale — and in high dimensions the obstacles are saddles, not minima

- **The same gradient step** fits spectroscopic Hamiltonians (Levenberg–Marquardt blends it with Gauss–Newton), tunes variational quantum circuits and trains language models with 10^{11} parameters.
- **A critical point** has $\nabla L = 0$; the signs of the Hessian eigenvalues decide its type.
- **A minimum has all d eigenvalues non-negative.** In high dimensions, most critical points at high loss are saddles (Dauphin et al., 2014).
- **Near a saddle the gradient is tiny:** training stalls on a plateau, then escapes along a downhill direction. Noise and momentum help.



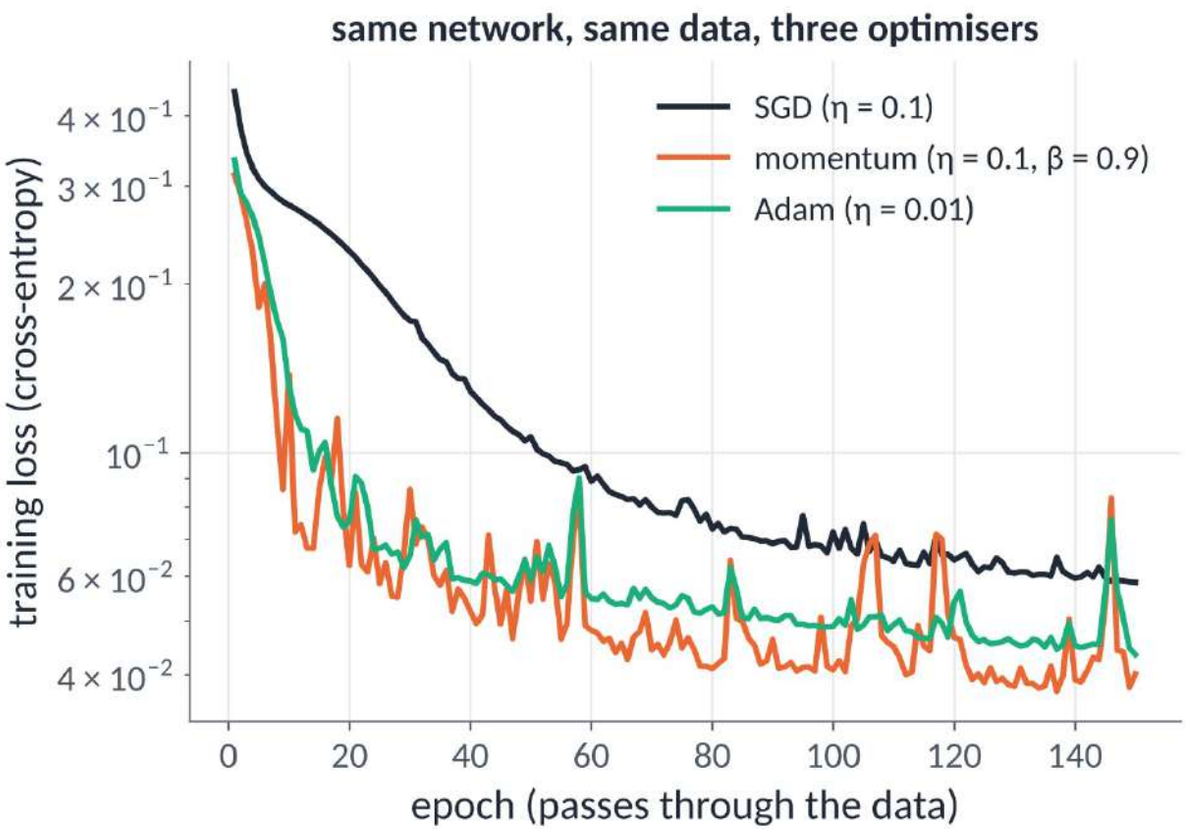
Optional · expert corner: the stationary state of SGD and the edge of stability

$$p(\theta) \;\propto\; e^{-L(\theta)/T_{\text{eff}}} \qquad\qquad \lambda_{\max}\big(\nabla^2 L\big) \;\approx\; \frac{2}{\eta}$$

- Near a minimum and for isotropic noise, SGD samples $p(\theta) \propto \exp(-L/T_{\text{eff}})$. Real gradient noise is anisotropic and depends on θ: the stationary state is then still Gaussian near the minimum, but not exactly Boltzmann (Mandt, Hoffman & Blei, 2017).
- A larger T_{eff} (large η, small B) prefers wide, flat minima — often linked to better generalisation, still an open research question.
- Edge of stability (Cohen et al., 2021): in full-batch training of neural networks the sharpness $\lambda_{\max}(\nabla^2 L)$ grows until it reaches $\approx 2/\eta$, then hovers there — training settles at the stability limit derived for quadratics.

Optional: the same small network trained with three optimisers

- A small neural network (2–16–16–1; networks are introduced in Block 5) on the two-moons data; mini-batches of 32 points.
- Average loss over the last 10 of 150 passes: SGD 0.060, momentum 0.047, Adam 0.051.
- Momentum and Adam get lower, faster; Adam used a 10× smaller learning rate ($\eta = 0.01$) than the other two.
- There is no universally best optimiser; Adam with a modest learning-rate search is a robust first choice.



Three things to remember from Block 3

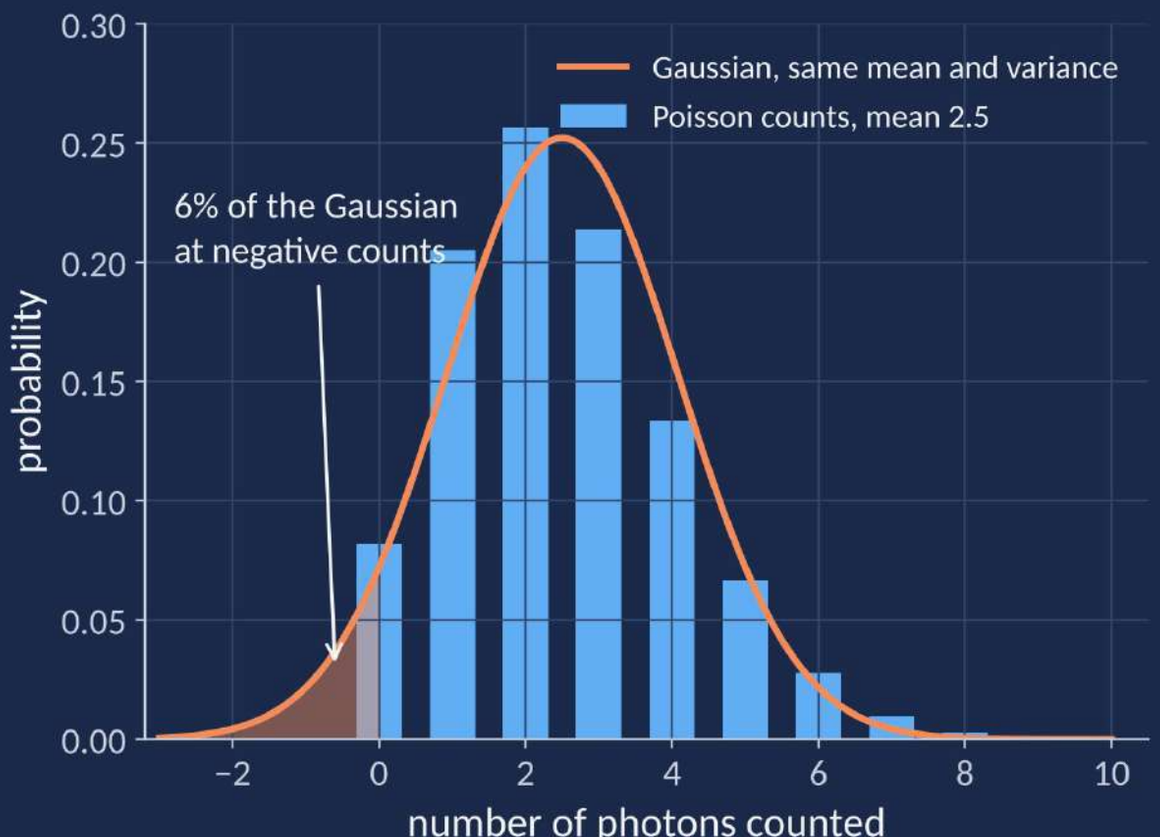
1. Training is gradient descent on a loss landscape — overdamped motion. The learning rate is limited by the curvature: $\eta < 2/\lambda_{\max}$.

2. Mini-batch noise acts as a temperature $T_{\text{eff}} \propto \eta/B$: SGD explores and samples a Boltzmann-like distribution around minima.

3. Momentum and Adam add inertia and per-parameter step sizes. In high dimensions, saddles and plateaus — not local minima — slow training down.

Why squared errors? And which loss should we use when we count photons?

- Least squares is so familiar that we rarely ask where it comes from.
- For counts, a Gaussian puts 6% of its probability on negative numbers of photons.
- The loss function is not a convention: it is a statement about the noise in the data.



Maximum likelihood turns a noise model into a loss function

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta} \prod_i p(y_i \mid x_i, \theta) = \arg\min_{\theta} \left[-\sum_i \log p(y_i \mid x_i, \theta) \right]$$

noise model for y	$-\log p(y \mid f)$, up to constants	loss function
Gaussian, width σ	$(y-f)^2/2\sigma^2$	squared error: least squares, χ^2
Laplace, width b	$\lvert y-f\rvert/b$	absolute error: robust to outliers
Poisson, mean f	$f - y\log f$	Poisson deviance: counting experiments
Bernoulli, probability f	$-y\log f - (1-y)\log(1-f)$	binary cross-entropy: classification

Physics lens: a χ^2 fit is maximum likelihood for Gaussian errors — physicists have always been doing machine learning without calling it that.

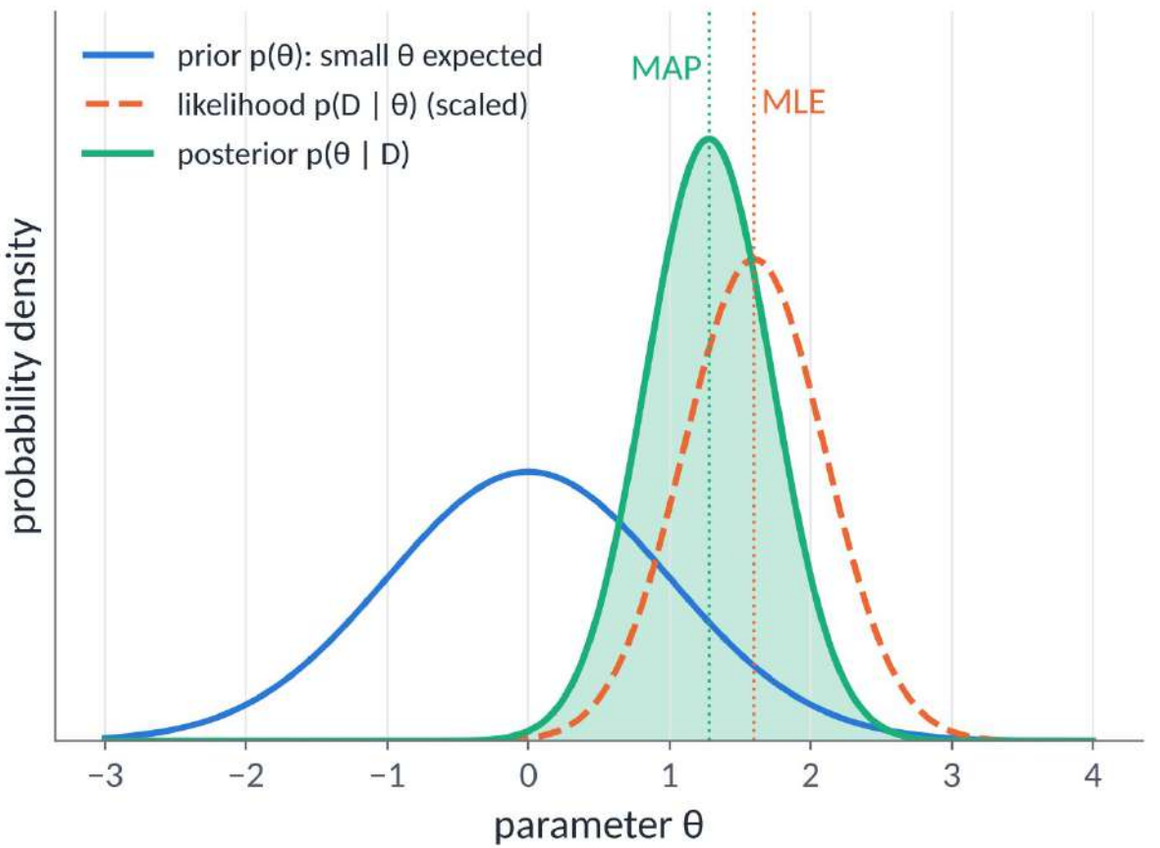
A regulariser is a prior: maximum a posteriori = loss + penalty

$$p(\theta \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \theta)\, p(\theta)}{p(\mathcal{D})}$$

$$-\log p(\theta \mid \mathcal{D}) = \underbrace{-\log p(\mathcal{D} \mid \theta)}_{\text{loss}} \underbrace{-\log p(\theta)}_{\text{regulariser}} + \text{const}$$

$$\theta \sim \mathcal{N}(0, \tau^2): \quad \hat{\theta}_{\text{MAP}} = \arg\min_{\theta} \left[\frac{1}{N} \sum_i \left(y_i - f_\theta(x_i) \right)^2 + \frac{\sigma^2}{N\tau^2} \|\theta\|^2 \right]$$

- **Gaussian prior ⇒ ridge (L2);** Laplace prior ⇒ lasso (L1).
- $\lambda = \sigma^2/(N\tau^2)$**:** the more data, the less the prior matters.
- **In the figure:** $\text{MAP} = 1.28$ lies between the prior mean 0 and the maximum-likelihood value 1.60.

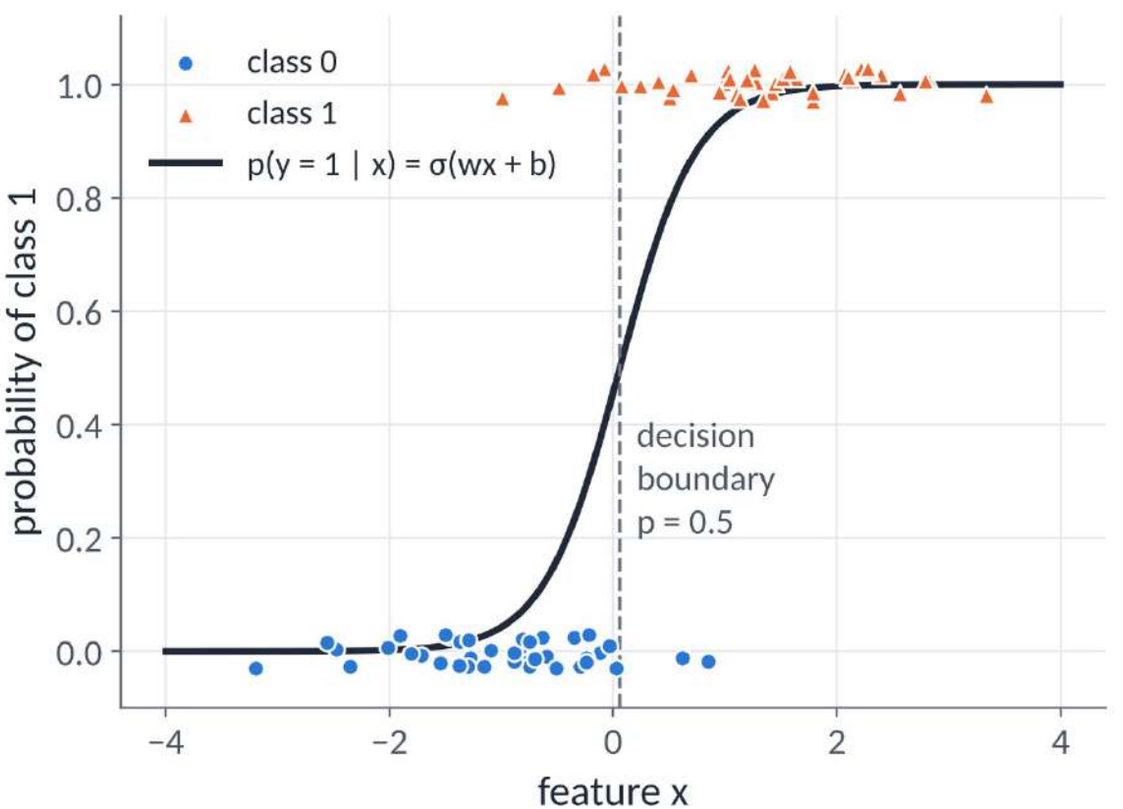


Logistic regression: a linear score, a sigmoid, and the cross-entropy loss

$$p(y{=}1 \mid x) = \sigma(w^\top x + b), \qquad \sigma(z) = \frac{1}{1+e^{-z}}$$

$$L = -\frac{1}{N}\sum_i \left[y_i \log p_i + (1-y_i)\log(1-p_i)\right], \qquad \nabla_w L = \frac{1}{N}\sum_i (p_i - y_i)\, x_i$$

- The score $w^\top x + b$ is squashed into a probability between 0 and 1.
- The decision boundary $p = 0.5$ is the hyperplane $w^\top x + b = 0$.
- The gradient is '(prediction − label) × input': simple and intuitive.
- **Why not squared error?** With a sigmoid it is non-convex and gives tiny gradients for confidently wrong predictions; cross-entropy is convex in w and b.



Key equation 4: the softmax classifier is the Boltzmann distribution

$$p(k \mid x) = \frac{e^{z_k(x)/T}}{\sum_j e^{z_j(x)/T}}$$

$z_k(x)$

score (logit) of class k: plays the role of minus an energy

T

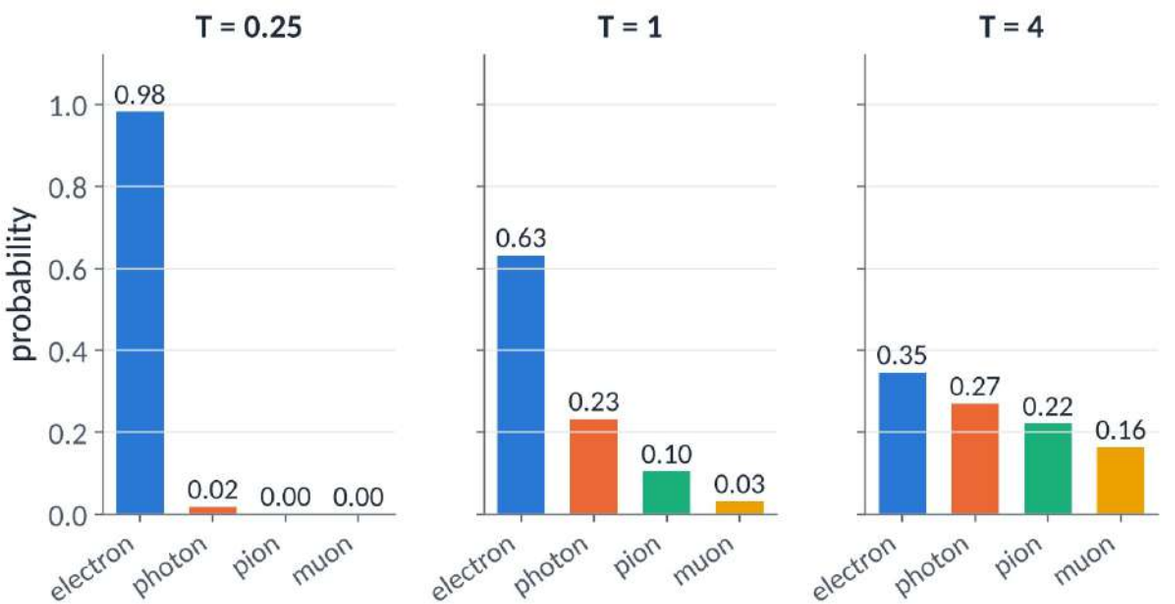
temperature: sharp ($T \to 0$, arg-max) or flat ($T \to \infty$, uniform)

$Z = \sum_j e^{z_j/T}$

normalisation: the partition function Z

$p(k \mid x)$

probability of class k; for two classes $\sigma(z_1 - z_0)$: logistic regression



Boltzmann (1877): $p_i = \frac{e^{-E_i/k_BT}}{Z}, \qquad Z = \sum_j e^{-E_j/k_BT}$

$$\mathbb{E}_{p_{\text{data}}}\left[-\log p_\theta\right] = H(p_{\text{data}}) + \mathrm{KL}\left(p_{\text{data}} \,\|\, p_\theta\right)$$

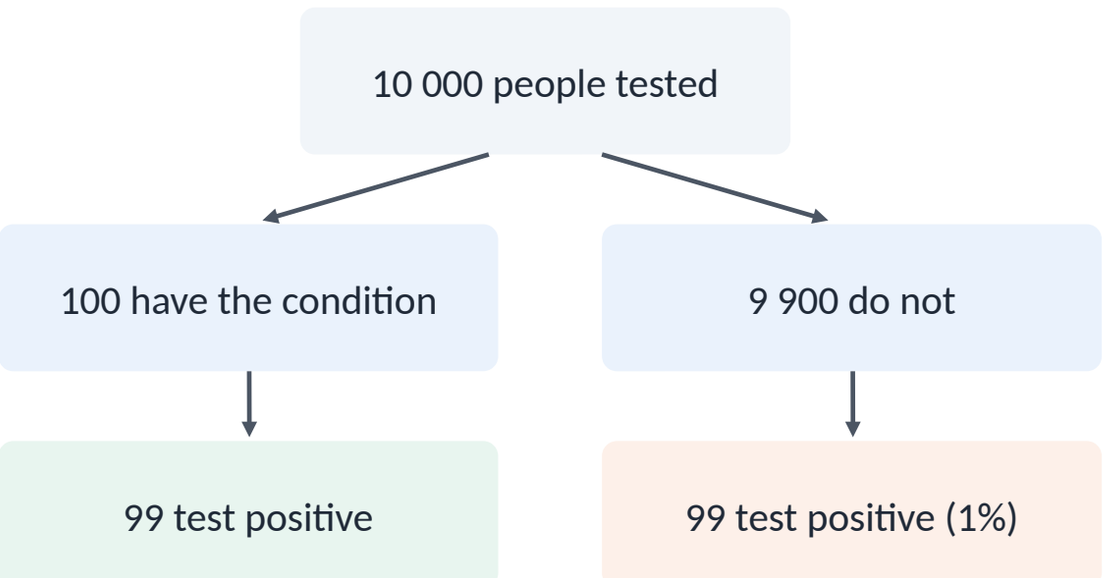
cross-entropy = entropy + KL divergence

Accuracy is not enough: confusion matrix, precision, recall — and the base rate

	predicted positive	predicted negative
actually positive	TP true positive	FN missed
actually negative	FP false alarm	TN true negative

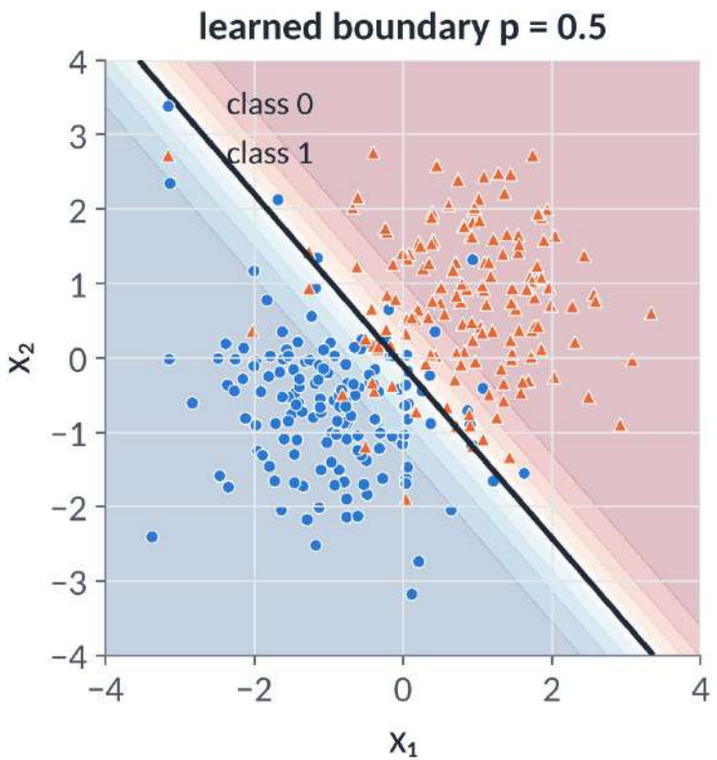
- **recall** (sensitivity, signal efficiency) $= \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$
- **precision** (purity) $= \mathrm{TP}/(\mathrm{TP}+\mathrm{FP})$
- **false-positive rate** $= \mathrm{FP}/(\mathrm{FP}+\mathrm{TN})$
- **ROC curve:** recall vs false-positive rate as the threshold moves; AUC = chance that a random positive scores above a random negative.

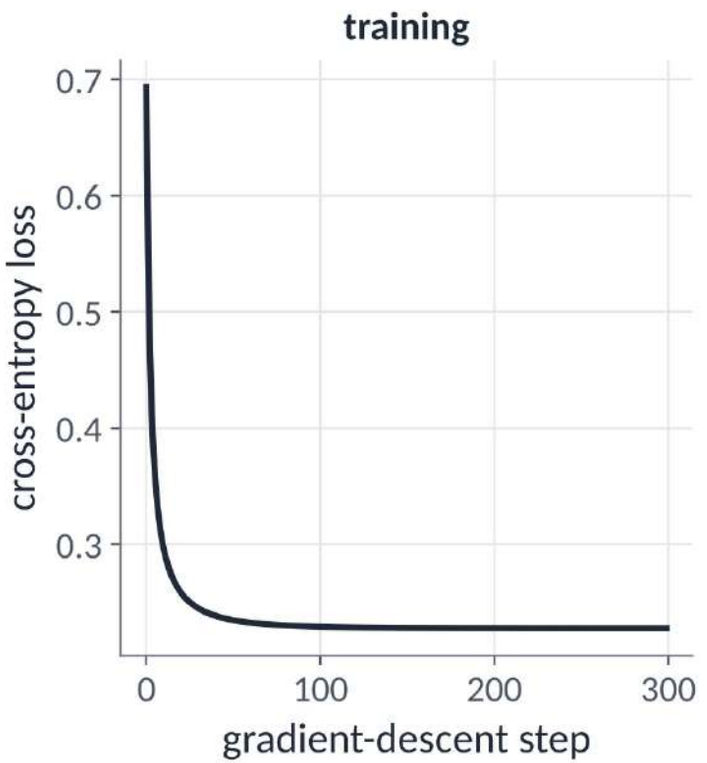
99% sensitivity, 99% specificity, 1% prevalence

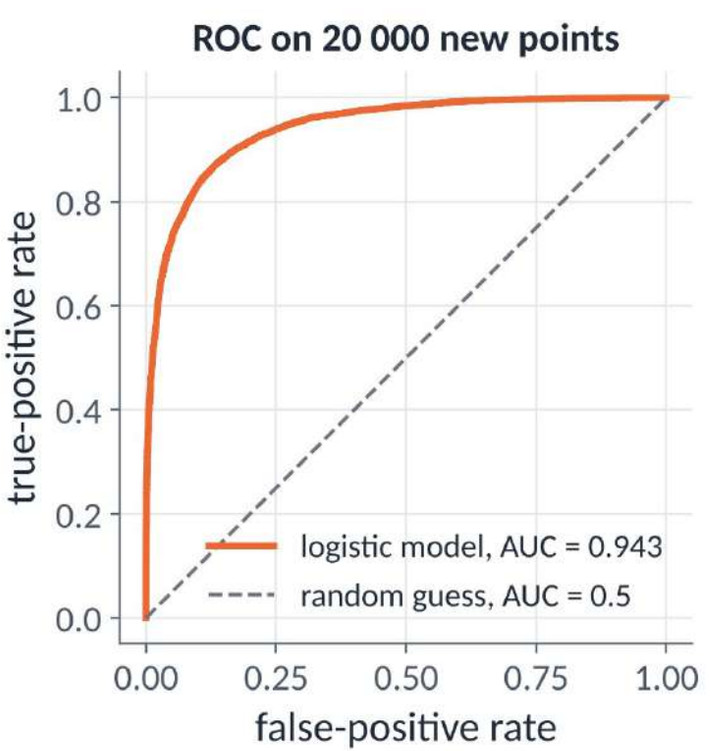


198 positive tests, but only 99 people have the condition: precision = 50%. Rare signals need very low false-positive rates — just as in particle physics.

Logistic regression trained by gradient descent: boundary, learning curve and ROC



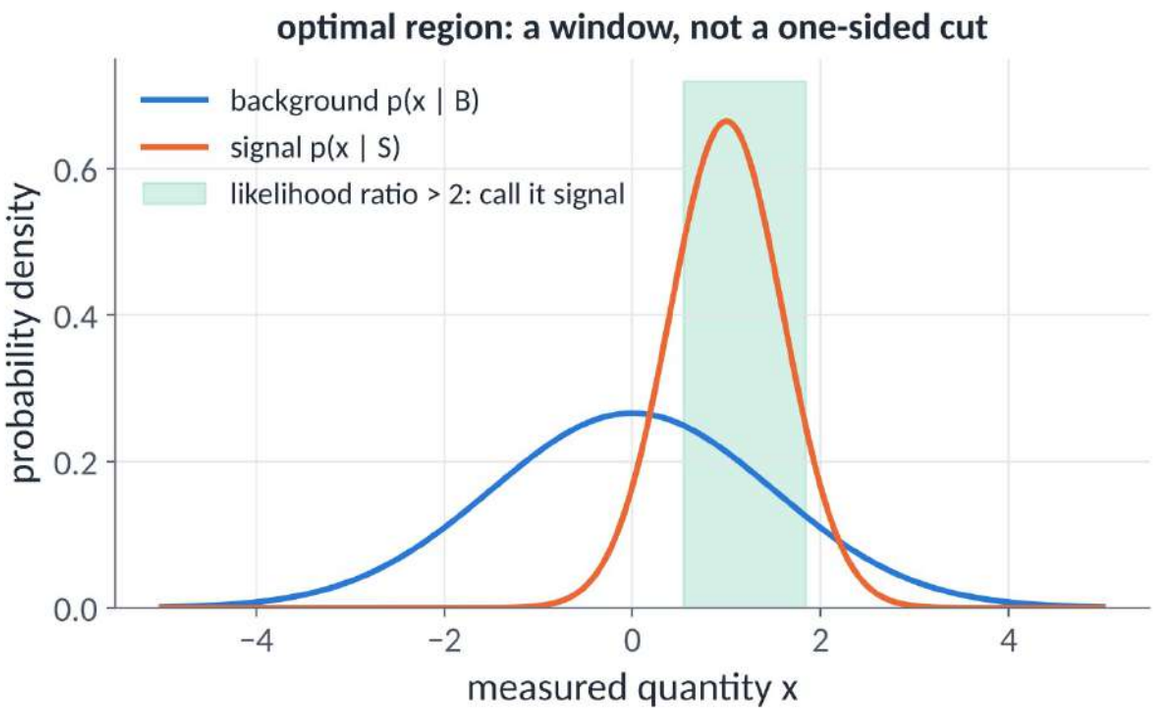


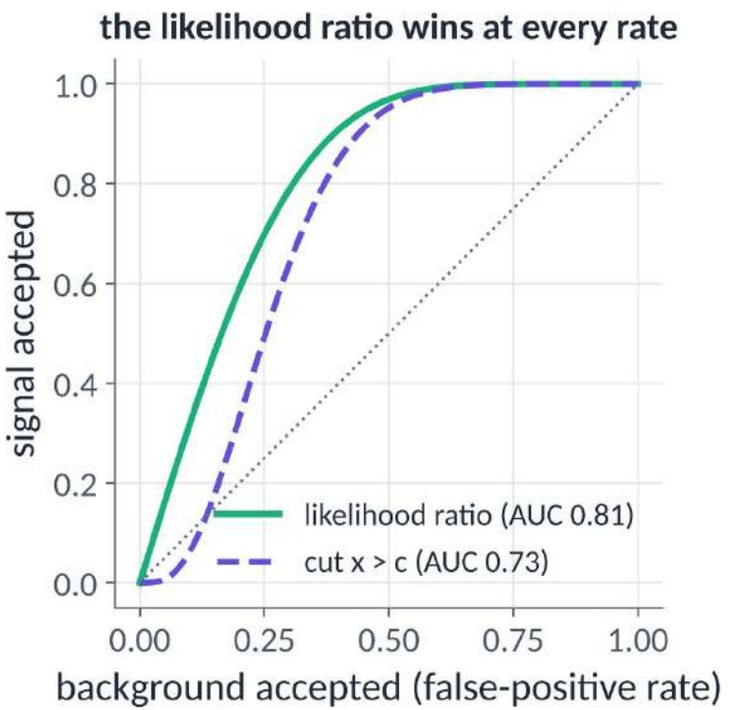


Two overlapping Gaussian classes, 300 training points, 300 gradient steps: the loss falls from $\ln 2 = 0.693$ to 0.228. On 20 000 new points: accuracy 86.5%, precision 0.84, recall 0.90, AUC 0.943. The best possible classifier for these classes reaches 87.0% and AUC 0.944: the remaining errors come from the overlap itself.

The best possible classifier is the likelihood ratio (Neyman–Pearson lemma)

$$\Lambda(x) = \frac{p(x \mid S)}{p(x \mid B)} > c \qquad d^*(x) = \frac{p(x \mid S)}{p(x \mid S) + p(x \mid B)} \Rightarrow \Lambda(x) = \frac{d^*(x)}{1 - d^*(x)}$$

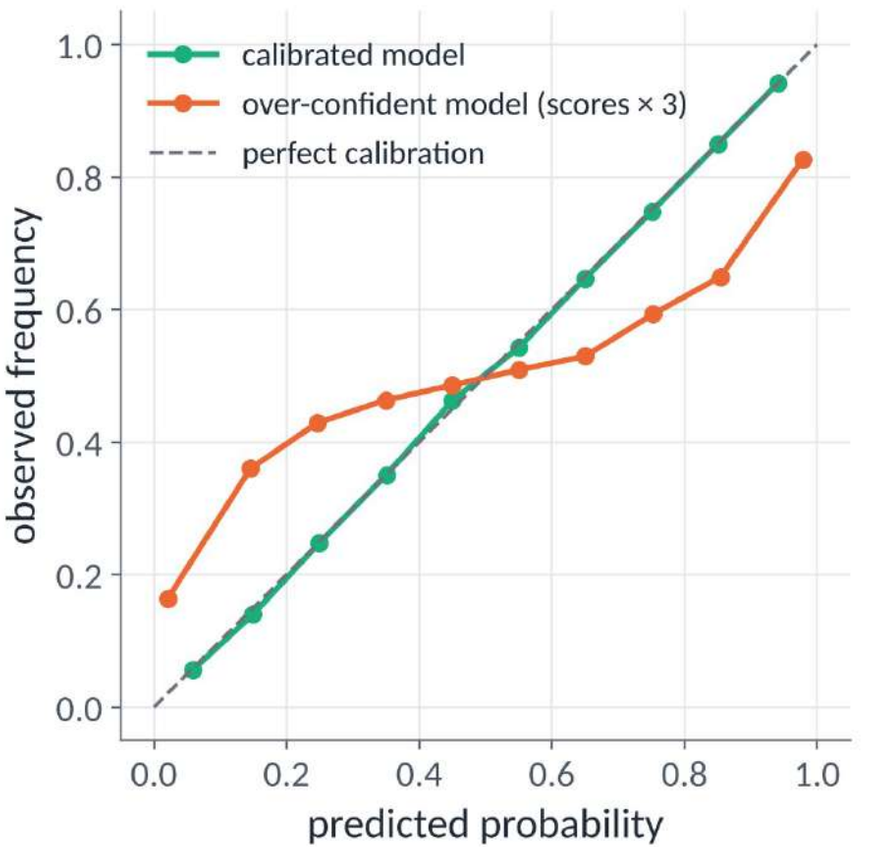




At a fixed false-positive rate no test accepts more signal than $\Lambda(x) > c$ (Neyman & Pearson, 1933). A classifier trained with cross-entropy learns $d^*(x)$ (shown for equal class sizes), a monotonic function of Λ — so it is optimal too. Here AUC 0.81 vs 0.73 for a simple cut.

Optional: are the predicted probabilities trustworthy? Calibration

- A model is calibrated if, among all cases where it predicts 70%, about 70% are positive.
- Over-confident models push probabilities towards 0 and 1 — common for large neural networks (Guo et al., 2017).
- Fix: temperature scaling — divide the logits by a $T > 1$ fitted on validation data (the softmax temperature of Key equation 4).
- Proper scoring rules (log-loss, Brier score) are minimised only by the true probabilities.



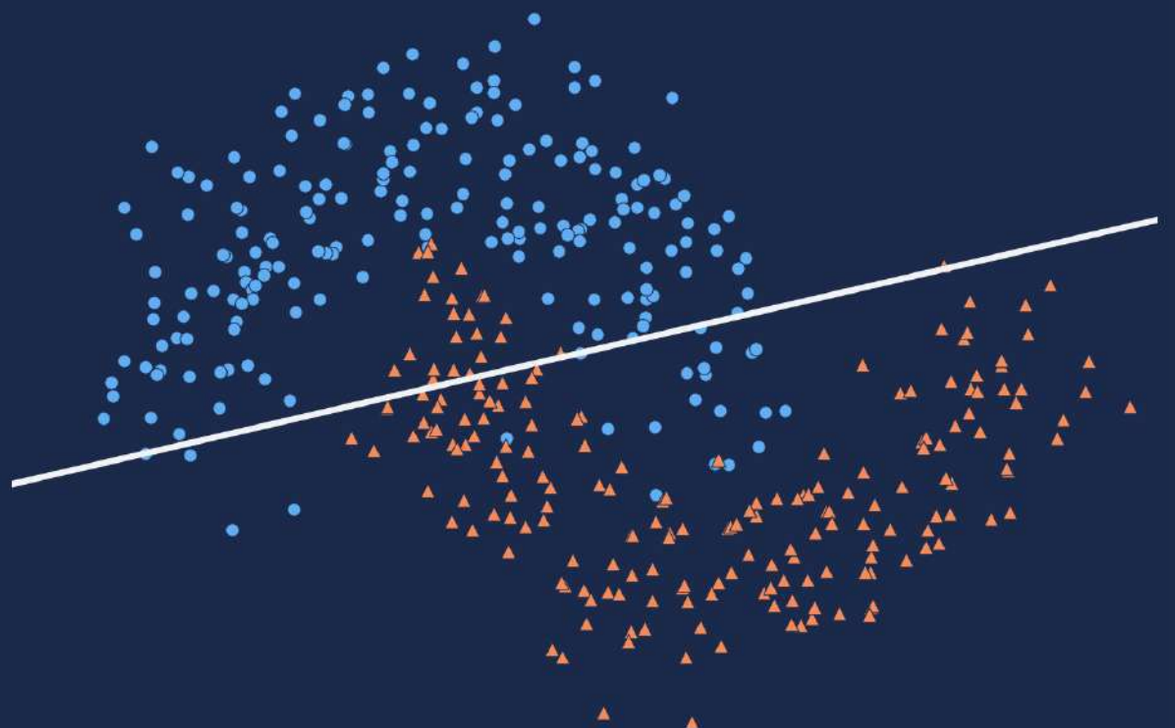
Three things to remember from Block 4

1. A loss function is a noise model: least squares $\Leftrightarrow$ Gaussian noise (χ^2), Poisson deviance for counts, cross-entropy for classes.

2. A regulariser is a prior (maximum a posteriori), and the softmax classifier is a Boltzmann distribution over the classes.

3. Judge classifiers by ROC, precision and recall, with the base rate in mind. The optimal test is the likelihood ratio.

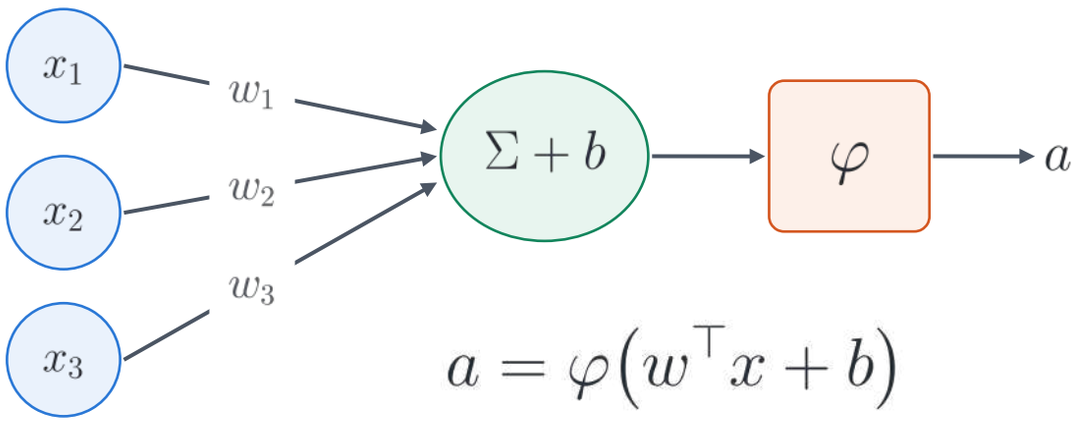
A straight line cannot separate these two classes. What can?



best straight line: 88% correct

- Logistic regression draws a straight boundary: at best 88% correct.
- We could invent nonlinear features by hand (x^2, xy, ...) — but which ones, in 10 or 10^6 dimensions?
- Neural networks learn the features themselves.

A neuron computes a weighted sum and passes it through a nonlinearity

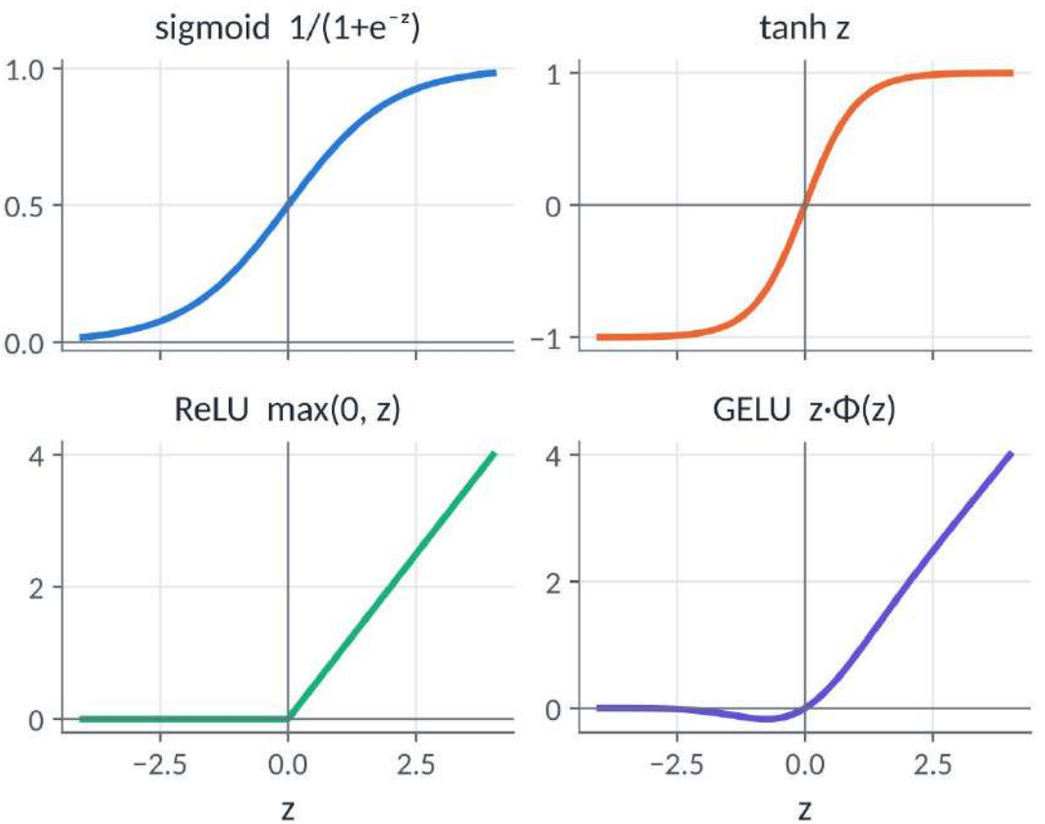


$$a = \varphi\left(w^\top x + b\right)$$

Without φ, stacked layers collapse into a single linear map:

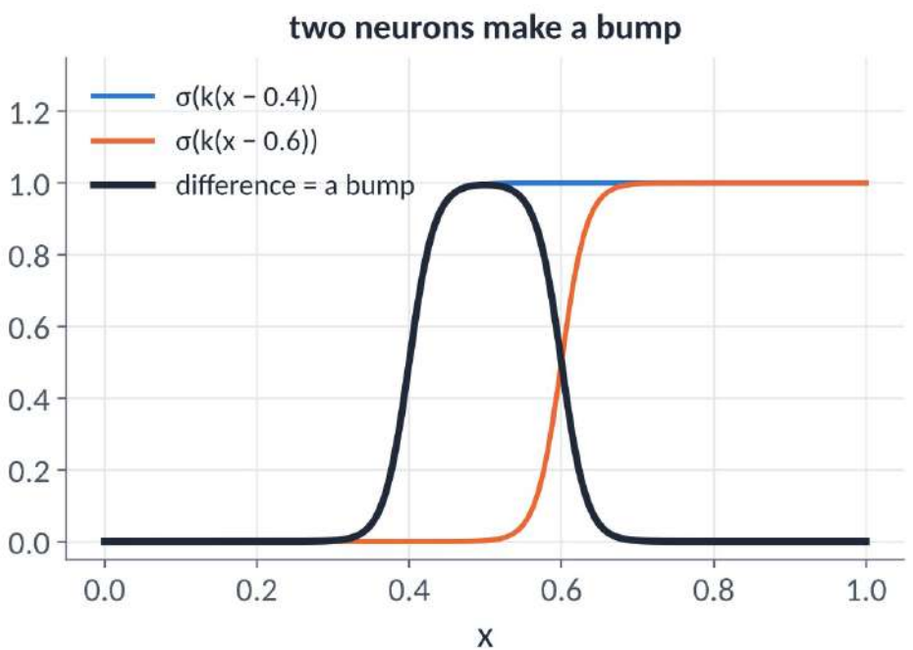
$$W_2(W_1 x + b_1) + b_2 = (W_2 W_1)\, x + (W_2 b_1 + b_2)$$

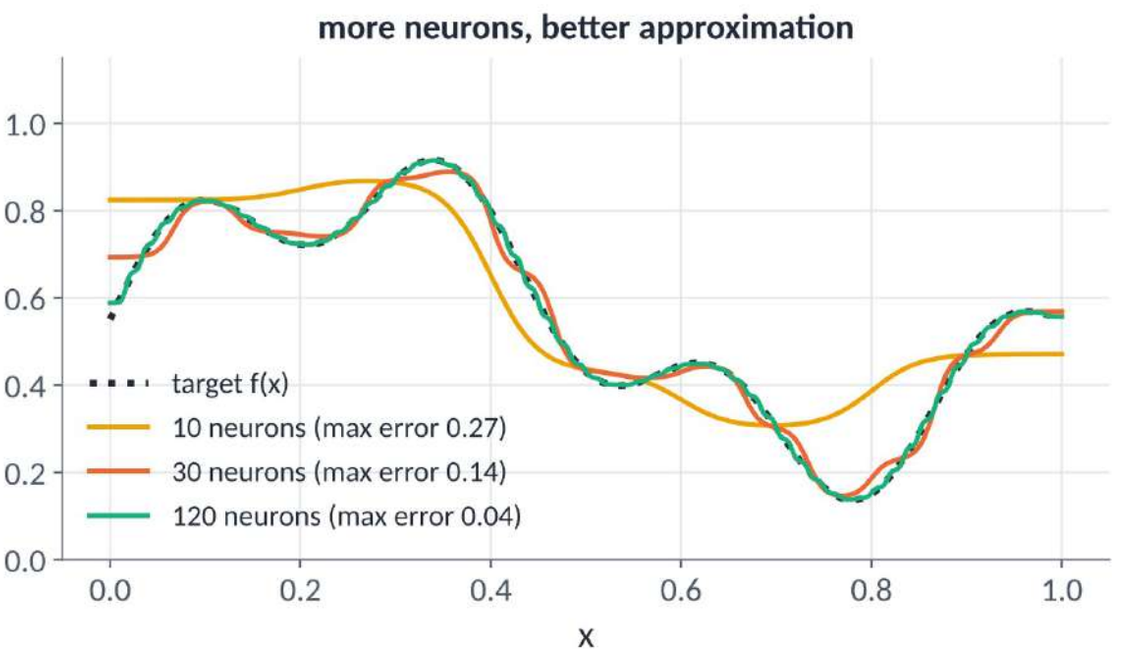
- ReLU is cheap and does not saturate for $z > 0$: the default in deep networks; GELU is used in transformers.
- Loosely inspired by biology (McCulloch & Pitts 1943; Rosenblatt's perceptron 1958).



Universal approximation: one hidden layer with enough neurons can approximate any continuous function arbitrarily well

$$F(x) = \sum_{j=1}^{M} v_j \, \varphi\big(w_j^\top x + b_j\big), \qquad \sup_{x \in K} |F(x) - f(x)| < \varepsilon$$





Cybenko (1989), Hornik (1991). The fine print: the theorem does not say how many neurons are needed (possibly exponentially many), how to find the weights, or how well the network will generalise.

Key equation 5: backpropagation — the chain rule, run backwards through the network

forward $$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}, \qquad a^{(l)} = \varphi\big(z^{(l)}\big)$$

backward $$\delta^{(l)} = \big(W^{(l+1)}\big)^{\top} \delta^{(l+1)} \odot \varphi'\big(z^{(l)}\big), \qquad \frac{\partial L}{\partial W^{(l)}} = \delta^{(l)}\, a^{(l-1)\top}$$

$\delta^{(l)}$

error signal: how the loss changes with the input of each neuron in layer l

$\big(W^{(l+1)}\big)^{\top}$

errors flow back along the same weights, transposed

$\varphi'\big(z^{(l)}\big)$

the local slope of the activation gates the signal

$a^{(l-1)}$

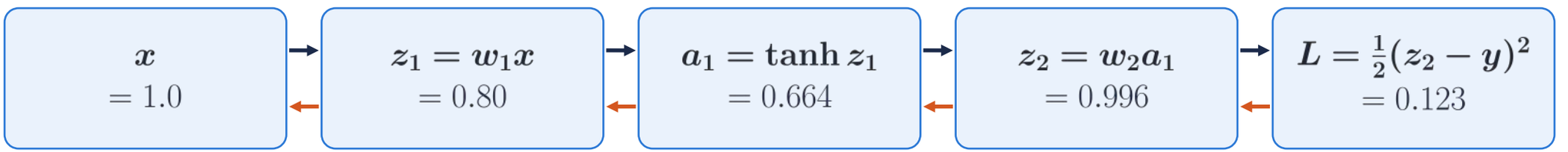
activity of the layer below: gradient = error × input

start at the output: $\delta^{(L)} = \dfrac{\partial L}{\partial z^{(L)}} = \hat{y} - y$ (squared error, or softmax with cross-entropy); $\dfrac{\partial L}{\partial b^{(l)}} = \delta^{(l)}$

Cost: one forward plus one backward pass ≈ 2–3 forward passes, whatever the number of parameters (reverse-mode automatic differentiation). Made popular for neural networks by Rumelhart, Hinton and Williams (1986).

Backpropagation by hand: every node needs only its local derivative

$w_1 = 0.8, \quad w_2 = 1.5, \quad$ target $y = 0.5$



forward pass → (values)

← backward pass (gradients)

$\partial L/\partial z_2 = z_2 - y \;=\; \mathbf{0.496}$

$\partial L/\partial w_2 = \partial L/\partial z_2 \cdot a_1 \;=\; \mathbf{0.329}$

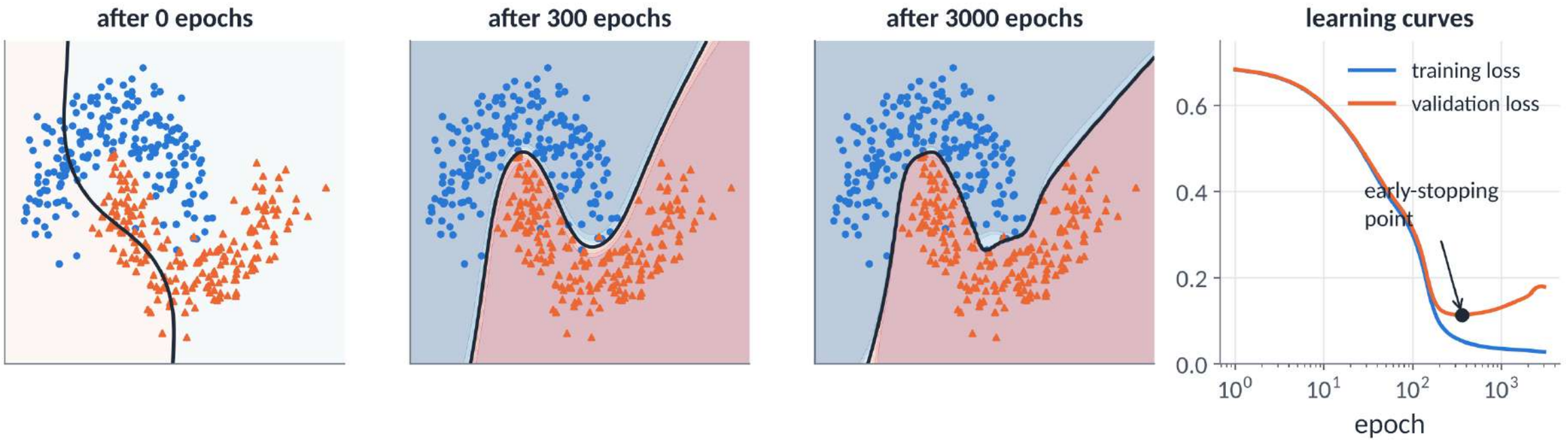
$\partial L/\partial a_1 = \partial L/\partial z_2 \cdot w_2 \;=\; \mathbf{0.744}$

$\partial L/\partial z_1 = \partial L/\partial a_1 \cdot (1 - a_1^2) \;=\; \mathbf{0.416}$

$\partial L/\partial w_1 = \partial L/\partial z_1 \cdot x \;=\; \mathbf{0.416}$

Each gradient = (gradient arriving from above) × (local derivative), reusing the stored forward values. A finite-difference check gives the same numbers. Then every weight takes the step $w \leftarrow w - \eta\, \partial L/\partial w$.

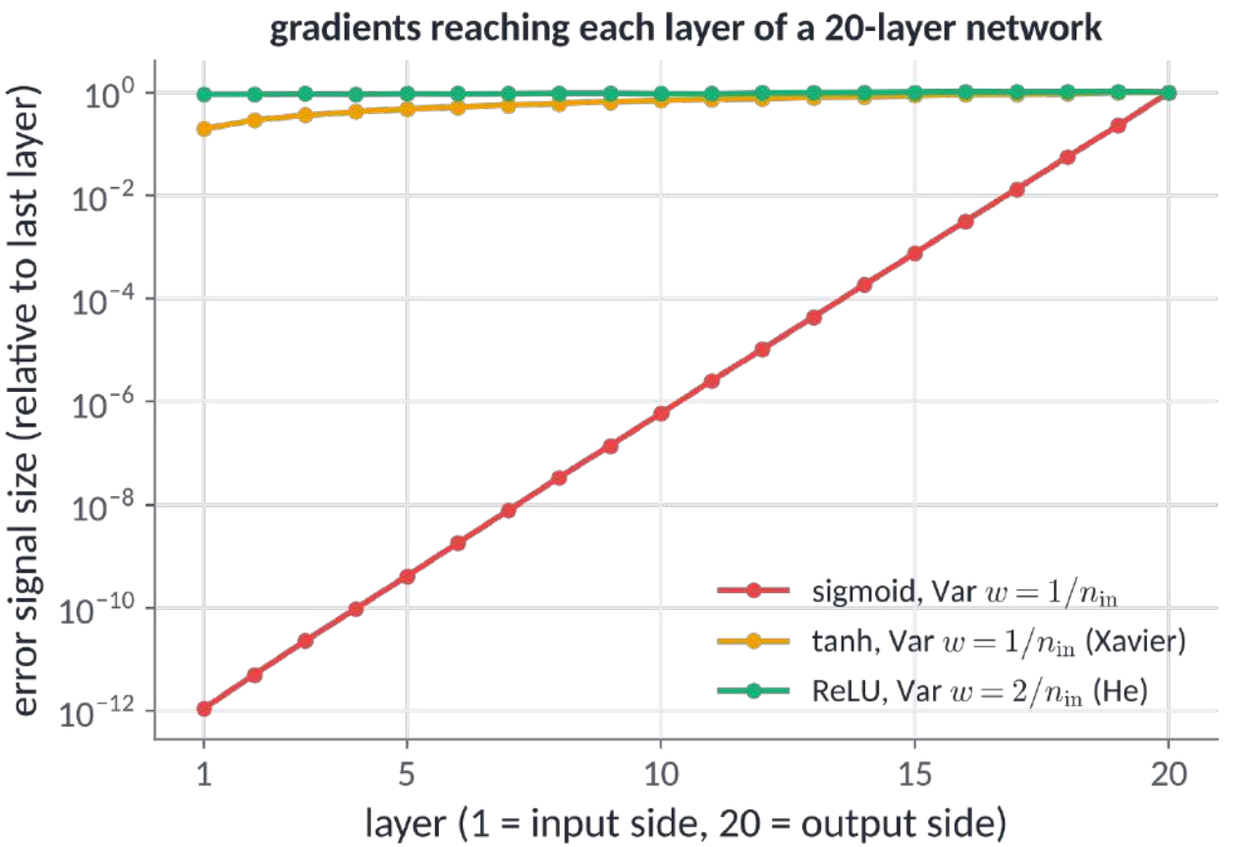
A small network learns the curved boundary — and starts to overfit if trained too long



- Network 2–16–2–1 (85 parameters, $\tanh$ units), trained with Adam on 400 points: validation accuracy 96% (a straight line: 88%).
- The hidden layers transform the inputs until a straight line in their space separates the classes.
- Validation loss is lowest after about 355 epochs (0.114) and then rises (0.179 at 3000) while the training loss keeps falling: the network becomes over-confident. Early stopping is a regulariser.

Training deep networks became practical when gradients could flow through many layers

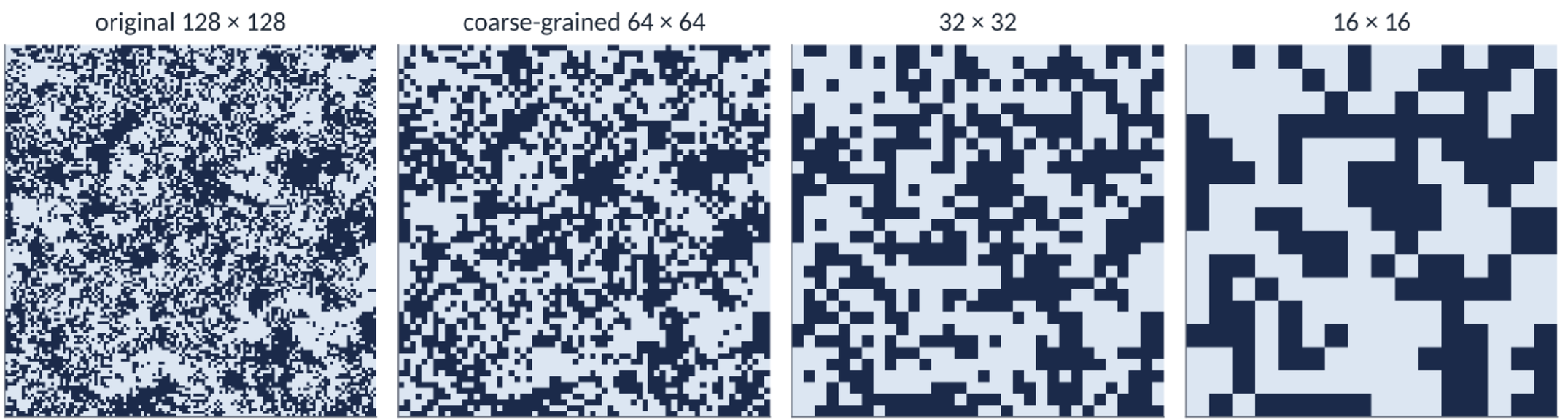
- **Backpropagation multiplies one factor per layer.** Sigmoid slopes are at most $1/4$, so with $\mathrm{Var}\, w = 1/n_{\mathrm{in}}$ the error signal shrinks exponentially: vanishing gradients.
- **ReLU with He initialisation** ($\mathrm{Var}\, w = 2/n_{\mathrm{in}}$) keeps the signal size constant through depth.
- **Normalisation layers and residual connections** (Day 2) stabilise very deep networks.
- **Adam, learning-rate schedules, early stopping** on validation data.



After 20 layers the sigmoid signal is down to 10^{-12}; $\tanh$ keeps 0.20; ReLU with He initialisation keeps 0.92.

What depth buys: features built from features — like coarse-graining in statistical physics

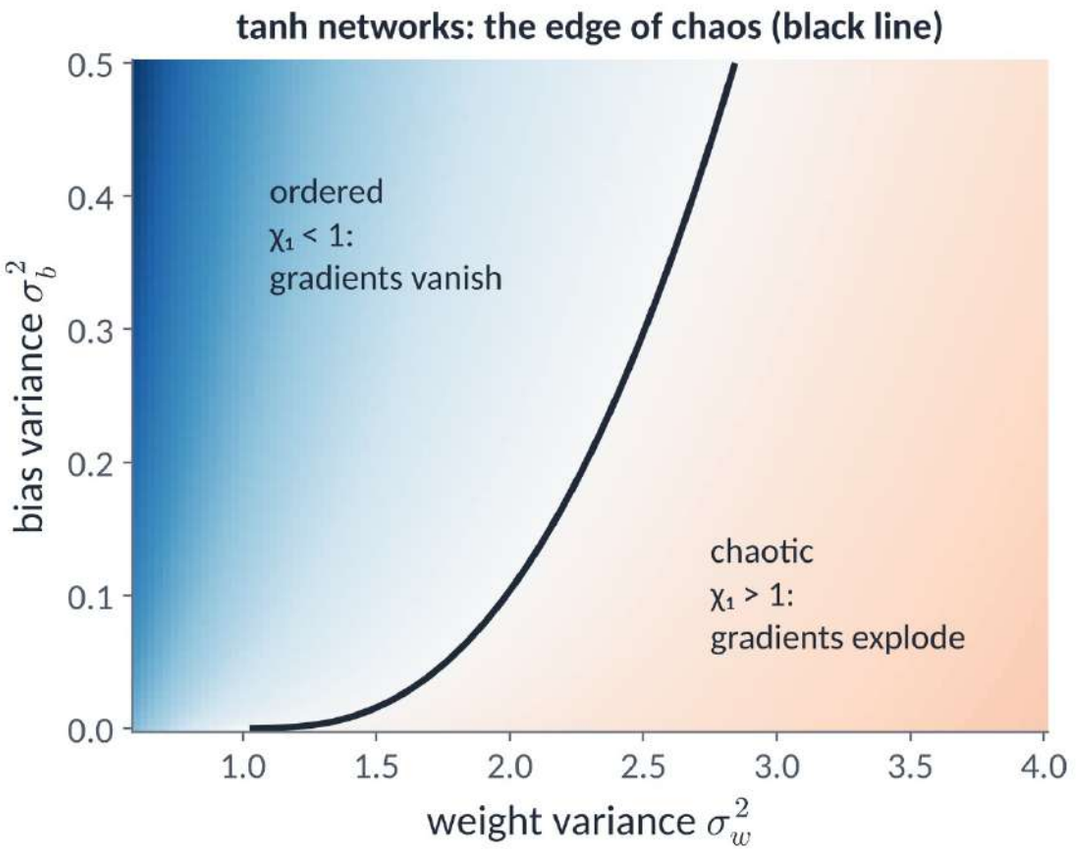
- Early layers detect local patterns; deeper layers combine them into larger, more abstract features (edges → textures → parts → objects in image networks).
- Compositional functions can be represented with exponentially fewer parameters by deep networks than by shallow ones (Lin, Tegmark & Rolnick, 2017).
- Renormalisation group: each block-spin step keeps the large-scale information and discards detail. Mehta & Schwab (2014) mapped variational RG exactly onto stacked restricted Boltzmann machines — an analogy, not a general theorem.



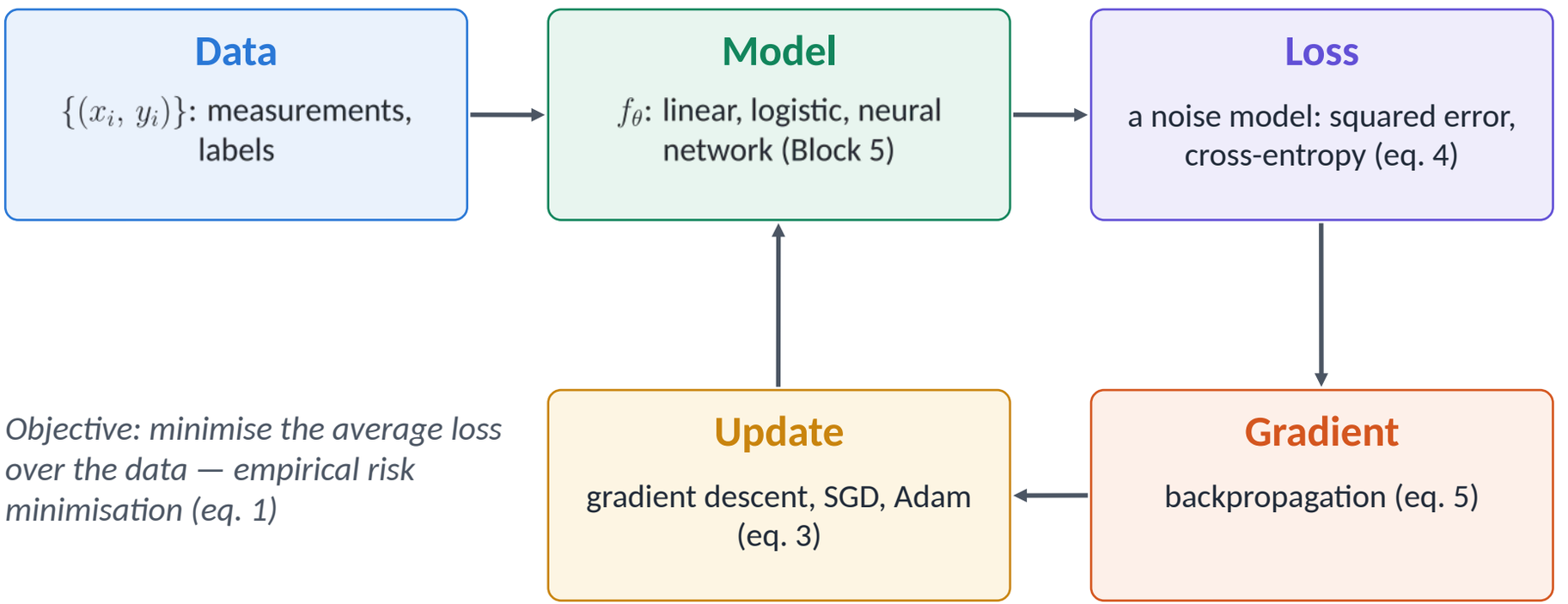
Optional · expert corner: deep random networks have an order-to-chaos transition

$$q^{(l)} = \sigma_w^2 \, \mathbb{E}_{z\sim\mathcal{N}(0,\,q^{(l-1)})}\left[\varphi(z)^2\right] + \sigma_b^2 \quad \longrightarrow \quad \text{fixed point } q^*$$
$$\chi_1 = \sigma_w^2 \, \mathbb{E}_{z\sim\mathcal{N}(0,\,q^*)}\left[\varphi'(z)^2\right]$$

- A signal propagating through a random network obeys a mean-field recursion for its variance q.
- $\chi_1 < 1$: nearby inputs merge and gradients vanish (ordered phase).
- $\chi_1 > 1$: inputs decorrelate and gradients explode (chaotic phase).
- Deep networks are trainable only near the critical line $\chi_1 = 1$ (Poole et al., 2016; Schoenholz et al., 2017).
- The same mean-field theory describes chaos in random recurrent networks (Sompolinsky, Crisanti & Sommers, 1988).



Day 1 in one loop: data, model, loss, gradient, update — and validation



Objective: minimise the average loss over the data — empirical risk minimisation (eq. 1)

Validation

held-out data, bias–variance trade-off, regularisation, cross-validation (eq. 2)

A one-parameter Hooke's-law fit and a trillion-parameter language model run exactly this loop. Next time: which architectures and generative models make modern AI work — and why they look like physics.

Five questions to check your understanding

1 **Training error keeps falling as parameters are added, while validation error rises. This is...**

A underfitting B an optimiser bug C overfitting D the noise floor

2 **Gradient descent on $L = \frac{1}{2}\lambda\theta^2$ with $\lambda = 4$. Which learning rates converge?**

A any $\eta > 0$ B $0 < \eta < 2$ C $0 < \eta < 1$ D $0 < \eta < 0.5$

3 **Least-squares fitting is maximum likelihood for which noise?**

A Gaussian B Poisson C Laplace D uniform

4 **Softmax at temperature $T \to 0$ puts...**

A equal probability on all classes B all probability on the highest score C probability $\propto$ score

5 **Sensitivity 99%, specificity 99%, prevalence 0.1%: what fraction of positive tests is real?**

A 99% B 50% C 9% D 0.1%

Answers — and the reason behind each

1. **C overfitting.** Training error is optimistic; held-out error measures generalisation (Block 2).

2. **D $0 < \eta < 0.5$.** Stability requires $\eta < 2/\lambda = 0.5$; beyond it the iterates oscillate with growing amplitude (Block 3).

3. **A Gaussian.** $-\log$ of a Gaussian is $(y - f)^2/2\sigma^2$ plus a constant (Block 4).

4. **B all probability on the highest score.** $T \to 0$ is the 'ground state' ($\arg\max$); $T \to \infty$ gives the uniform distribution (Key equation 4).

5. **C about 9%.** True positives 0.099% of people, false positives 0.999%: precision $= 0.099/(0.099 + 0.999) \approx 9\%$ — the base rate dominates.

Backup: normal equations and ridge regression in closed form

Setting: N data points; the rows of X ($N \times p$) are the features $\phi(x_i)^\top$, y holds the targets; ridge loss $L_\lambda(\theta) = \frac{1}{N}\|y - X\theta\|^2 + \lambda\|\theta\|^2$.

1. **Least squares: set the gradient to zero.** For $\lambda = 0$ the loss L_0 is quadratic in θ; its minimum is where the gradient vanishes:
$$\nabla_\theta L_0 = -\tfrac{2}{N} X^\top (y - X\theta) = 0 \quad \Rightarrow \quad X^\top X\,\theta = X^\top y$$
These are the normal equations. If $X^\top X$ is singular ($p > N$, or collinear features), the solution is not unique.

2. **Add the penalty.** The gradient of $\lambda\|\theta\|^2$ is $2\lambda\theta$; add it and multiply by $N/2$:
$$(X^\top X + N\lambda I)\,\theta = X^\top y$$
For $\lambda > 0$ it is positive definite, $v^\top (X^\top X + N\lambda I)\, v = \|Xv\|^2 + N\lambda\|v\|^2 > 0$, hence always invertible.

3. **Diagonalise with the SVD.** Write $X = USV^\top$ with orthonormal u_k (data space), v_k (parameter space) and singular values $s_k \geq 0$. Then $X^\top X + N\lambda I = V(S^2 + N\lambda I)V^\top$ and $X^\top y = VSU^\top y$, so
$$\hat\theta_{\text{ridge}} = V(S^2 + N\lambda I)^{-1} S U^\top y = \sum_k \frac{s_k}{s_k^2 + N\lambda}\,(u_k^\top y)\, v_k$$

4. **Compare with least squares.** For $\lambda = 0$ the factor is $1/s_k$; ridge multiplies it by a shrinkage factor between 0 and 1:
$$\frac{s_k}{s_k^2 + N\lambda} = \underbrace{\frac{1}{s_k}}_{\text{least squares}} \times \underbrace{\frac{s_k^2}{s_k^2 + N\lambda}}_{\text{shrinkage}}$$

$$\hat\theta_{\text{ridge}} = \sum_k \frac{s_k^2}{s_k^2 + N\lambda}\,\frac{u_k^\top y}{s_k}\, v_k$$

Noise in y gives the least-squares coefficient along v_k a variance σ^2/s_k^2. Directions with $s_k^2 \gg N\lambda$ are kept; poorly determined ones ($s_k^2 \ll N\lambda$) are suppressed. The intercept is usually not penalised.

Backup: derivation of the bias–variance decomposition

Setting: $y = f(x) + \varepsilon$ with $\mathbb{E}\varepsilon = 0$, $\mathrm{Var}\,\varepsilon = \sigma^2$; the fit $\hat{f}$ depends on the random training set, the new noise ε is independent of it. All averages $\mathbb{E}$ are over training sets and noise, at a fixed input x.

1. **Insert the data model.** Error of the prediction $\hat{f}(x)$ for a new measurement y at the same x:
$$\begin{aligned}\mathbb{E}\big[(y-\hat{f})^2\big] &= \mathbb{E}\big[(f-\hat{f}+\varepsilon)^2\big] \\ &= \mathbb{E}\big[(f-\hat{f})^2\big] + 2\,\mathbb{E}\big[(f-\hat{f})\,\varepsilon\big] + \mathbb{E}\,\varepsilon^2\end{aligned}$$
2. **Split off the noise.** $\hat{f}$ is built from the training set, and ε is independent of it with zero mean, so $\mathbb{E}[(f-\hat{f})\,\varepsilon] = \mathbb{E}[f-\hat{f}]\,\mathbb{E}\varepsilon = 0$. With $\mathbb{E}\varepsilon^2 = \sigma^2$:
$$\mathbb{E}\big[(y-\hat{f})^2\big] = \mathbb{E}\big[(f-\hat{f})^2\big] + \sigma^2$$
3. **Add and subtract the average fit** $\bar{f}(x) \equiv \mathbb{E}\hat{f}(x)$. $f - \bar{f}$ is not random, so it comes out of the average:
$$\begin{aligned}\mathbb{E}\big[(f-\hat{f})^2\big] &= \mathbb{E}\big[(f-\bar{f}+\bar{f}-\hat{f})^2\big] \\ &= (f-\bar{f})^2 + \mathbb{E}\big[(\hat{f}-\bar{f})^2\big] \\ &\quad + 2\,(f-\bar{f})\,\mathbb{E}\big[\bar{f}-\hat{f}\big]\end{aligned}$$
4. **The cross term vanishes,** because $\mathbb{E}\big[\bar{f}-\hat{f}\big] = \bar{f} - \mathbb{E}\hat{f} = 0$ by the definition of $\bar{f}$.

$$\mathbb{E}\big[(y-\hat{f})^2\big] = \underbrace{(f-\bar{f})^2}_{\text{bias}^2} + \underbrace{\mathbb{E}\big[(\hat{f}-\bar{f})^2\big]}_{\text{variance}} + \underbrace{\sigma^2}_{\text{noise}}$$

Bias: systematic error of the average fit. Variance: scatter of the fit from one training set to the next. σ^2: a floor no model can beat. Averaging over x gives the curves of Block 2.

Backup: convergence of gradient descent on a quadratic

Setting: near a minimum, $L(\theta) \approx L^* + \frac{1}{2}(\theta-\theta^*)^\top H(\theta-\theta^*)$, Hessian $H = \sum_k \lambda_k\, u_k u_k^\top$ (eigenvalues $\lambda_k > 0$, orthonormal u_k).

1. **Error recursion.** The gradient is $\nabla L = H(\theta-\theta^*)$. For the error $e_t = \theta_t - \theta^*$, one step $\theta_{t+1} = \theta_t - \eta\nabla L(\theta_t)$ gives
$$e_{t+1} = (I - \eta H)\, e_t$$
2. **The modes decouple.** Expand $e_t = \sum_k c_k(t)\, u_k$; since $Hu_k = \lambda_k u_k$, each coefficient evolves on its own:
$$c_k(t+1) = (1-\eta\lambda_k)\, c_k(t) \;\Rightarrow\; c_k(t) = (1-\eta\lambda_k)^t\, c_k(0)$$
3. **Stability.** All modes shrink if and only if $|1-\eta\lambda_k| < 1$ for every k, i.e. $0 < \eta < 2/\lambda_{\max}$. The slowest mode sets the rate per step, $\rho = \max\left(|1-\eta\lambda_{\min}|,\ |1-\eta\lambda_{\max}|\right)$.
4. **Best learning rate.** ρ is smallest when the two extreme modes shrink equally fast, $1-\eta\lambda_{\min} = -(1-\eta\lambda_{\max})$:
$$\eta_{\text{opt}} = \frac{2}{\lambda_{\max}+\lambda_{\min}}, \qquad \rho = \frac{\kappa-1}{\kappa+1}, \qquad \kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$$
Steps per factor e: $1/\ln\frac{\kappa+1}{\kappa-1} \approx \kappa/2$ for $\kappa \gg 1$.
5. **Heavy-ball momentum,** $v_{t+1} = \beta v_t - \eta\nabla L$, $\theta_{t+1} = \theta_t + v_{t+1}$. Per mode, $c_{t+1} = (1+\beta-\eta\lambda_k)\, c_t - \beta\, c_{t-1}$. For $\eta = 4/(\sqrt{\lambda_{\max}}+\sqrt{\lambda_{\min}})^2$, $\beta = \rho_{\text{HB}}^2$ all roots have modulus $\sqrt{\beta}$:
$$\rho_{\text{HB}} = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \;\Rightarrow\; \approx \sqrt{\kappa}/2 \text{ steps per factor } e$$

$\kappa = 100$: gradient descent: $\rho = \frac{99}{101}$, about 50 steps per factor e
heavy ball: $\rho = \frac{9}{11}$, about 5 steps per factor e

The condition number κ, not the number of parameters, makes gradient descent slow; momentum replaces κ by $\sqrt{\kappa}$.

Backup: SGD as Langevin dynamics

Setting: a mini-batch of B examples gives a noisy gradient, $\nabla L_B = \nabla L + \xi$ with $\mathbb{E}\xi = 0$ and $\mathrm{Var}\,\xi = \sigma_g^2/B$ (σ_g^2: variance of the gradient of a single example; the batch averages B independent ones).

1. **SGD step = drift + random kick.**
$$\theta_{t+1} = \theta_t - \eta\nabla L(\theta_t) - \eta\,\xi_t, \qquad \mathrm{Var}(\eta\,\xi_t) = \eta^2\sigma_g^2/B$$
2. **Langevin dynamics.** Overdamped motion in the potential L at temperature T, $d\theta = -\nabla L\,dt + \sqrt{2T}\,dW$. Its Euler step with $dt = \eta$ has the same drift and a kick of variance $2T\eta$:
$$\theta_{t+1} = \theta_t - \eta\nabla L(\theta_t) + \sqrt{2T\eta}\,\zeta_t, \qquad \zeta_t \sim \mathcal{N}(0,1)$$
3. **Match the kicks.** This fixes the temperature; for isotropic noise the stationary state is then Boltzmann:
$$\frac{\eta^2\sigma_g^2}{B} = 2T\eta \;\Rightarrow\; T_{\text{eff}} = \frac{\eta\,\sigma_g^2}{2B}, \qquad p(\theta) \propto e^{-L(\theta)/T_{\text{eff}}}$$
4. **Exact check for** $L = \frac{1}{2}\lambda\theta^2$ (θ measured from the minimum): $\theta_{t+1} = (1-\eta\lambda)\,\theta_t - \eta\,\xi_t$. Since ξ_t is independent of θ_t, the variance v_t of θ_t obeys
$$v_{t+1} = (1-\eta\lambda)^2\,v_t + \eta^2\sigma_g^2/B$$
In the stationary state $v_{t+1} = v_t = v$:
$$v = \frac{\eta\,\sigma_g^2}{B\lambda\,(2-\eta\lambda)} = \frac{T_{\text{eff}}}{\lambda}\left(1 - \frac{\eta\lambda}{2}\right)^{-1}$$
Boltzmann in a harmonic well gives T/λ; the last factor is the finite-step correction (5% for $\eta\lambda = 0.1$).

$\eta = 0.1$, $\sigma_g^2 \approx 1$, $\lambda \approx 1$: $B = 2$: predicted width 0.161, measured 0.165
$B = 32$: predicted width 0.040, measured 0.041

Small batches and large learning rates are ‘hot’: SGD explores a region of width $\propto \sqrt{\eta/B}$ around the minimum.

Backup: gradient of the softmax cross-entropy

Setting: scores (logits) $z_1, \dots, z_K$ at temperature $T = 1$, softmax $p_k = e^{z_k} / \sum_j e^{z_j}$; for one example of true class y the loss is $L = -\log p_y$; e_y is the one-hot vector of class y.

1. **Write the loss in terms of the scores.**
$$L = -\log p_y = -z_y + \log \sum_j e^{z_j}$$

2. **Differentiate.** $\partial z_y / \partial z_k = \delta_{ky}$, and the derivative of the logarithm of the sum is
$$\frac{\partial}{\partial z_k} \log \sum_j e^{z_j} = \frac{e^{z_k}}{\sum_j e^{z_j}} = p_k$$
so the gradient is 'prediction − target':
$$\frac{\partial L}{\partial z_k} = p_k - \delta_{ky}, \qquad \nabla_z L = p - e_y$$

3. **Second derivative.** With $\partial p_k / \partial z_m = p_k \delta_{km} - p_k p_m$:
$$H_{km} = \frac{\partial^2 L}{\partial z_k \, \partial z_m} = p_k \delta_{km} - p_k p_m$$
$$v^\top H v = \sum_k p_k v_k^2 - \Big(\sum_k p_k v_k \Big)^2 = \mathrm{Var}_p(v) \;\geq\; 0$$
H is the covariance matrix of the one-hot vector under p: positive semi-definite, so L is convex in z.

4. **Linear scores.** For $z = Wx + b$, L is a convex function of a linear map, hence convex in (W, b); by the chain rule
$$\frac{\partial L}{\partial W} = (p - e_y)\, x^\top, \qquad \frac{\partial L}{\partial b} = p - e_y$$

$$\frac{\partial L}{\partial z_k} = p_k - \delta_{ky}, \qquad H = \mathrm{diag}(p) - pp^\top \succeq 0$$

Gradient descent reaches the global minimum when one exists. Separable data have none (L falls forever as W is scaled up): add a small L2 penalty $\lambda \|W\|^2$.

Backup: backpropagation for a two-layer network

Setting: forward pass $z_1 = W_1 x + b_1$, $h = \varphi(z_1)$ (element-wise), $\hat{y} = W_2 h + b_2$, $L = \frac{1}{2}\|\hat{y} - y\|^2$. Error signals: $\delta_2 = \partial L/\partial \hat{y}$ and $\delta_1 = \partial L/\partial z_1$.

1. **Output layer.** Differentiating $\frac{1}{2}\|\hat{y} - y\|^2$ gives $\delta_2 = \hat{y} - y$. Since $\hat{y}_i = \sum_j (W_2)_{ij} h_j + (b_2)_i$, we have $\partial \hat{y}_i / \partial (W_2)_{ij} = h_j$, hence
$$\frac{\partial L}{\partial W_2} = \delta_2 \, h^\top, \qquad \frac{\partial L}{\partial b_2} = \delta_2$$
2. **Send the error back through W_2.** h affects L only through $\hat{y}$; the chain rule gives $\partial L/\partial h_j = \sum_i (\delta_2)_i \, (W_2)_{ij}$, i.e.
$$\frac{\partial L}{\partial h} = W_2^\top \delta_2$$
3. **Through the activation.** h_j depends only on $z_{1,j}$, with slope $\varphi'(z_{1,j})$: the Jacobian is diagonal, so the product is element-wise:
$$\delta_1 = \left(W_2^\top \delta_2\right) \odot \varphi'(z_1)$$
4. **First layer.** Exactly as in step 1, with the input x in place of h:
$$\frac{\partial L}{\partial W_1} = \delta_1 \, x^\top, \qquad \frac{\partial L}{\partial b_1} = \delta_1$$
5. **Deeper networks.** Repeat steps 2 and 3 layer by layer, from the output down: each δ is computed once and reused for all weights of its layer. With the activities $a^{(0)} = x$, $a^{(1)} = h$ one gets the general rule below.

any depth: $\delta^{(l)} = \left(W^{(l+1)}\right)^\top \delta^{(l+1)} \odot \varphi'\left(z^{(l)}\right), \qquad \frac{\partial L}{\partial W^{(l)}} = \delta^{(l)} \, a^{(l-1)\top}$

For a mini-batch, stack the examples as columns: every line becomes a matrix product. All gradients together cost about 2–3 forward passes.

Backup: proof of the Neyman–Pearson lemma

Setting: a test accepts x as signal if $x \in R$; false-positive rate $P_B(R) = \int_R p_B\,dx$, power $P_S(R) = \int_R p_S\,dx$, with $p_S(x) = p(x \mid S)$ and $p_B(x) = p(x \mid B)$.

1. **The candidate region.** Cut on $\Lambda = p_S/p_B$, the likelihood ratio, at the threshold $c \geq 0$ that gives false-positive rate α:
$$R^* = \{x :\ p_S(x) > c\,p_B(x)\}, \qquad P_B(R^*) = \alpha$$
Compare it with any region R with $P_B(R) \leq \alpha$.

2. **A non-negative integral.**
$$I = \int \big(\mathbf{1}_{R^*}(x) - \mathbf{1}_R(x)\big)\big(p_S(x) - c\,p_B(x)\big)\,dx \ \geq\ 0$$
because the integrand is ≥ 0 at every x. If $x \in R^*$: second factor > 0, first factor $1 - \mathbf{1}_R \geq 0$. If $x \notin R^*$: second factor ≤ 0, first factor $-\mathbf{1}_R \leq 0$.

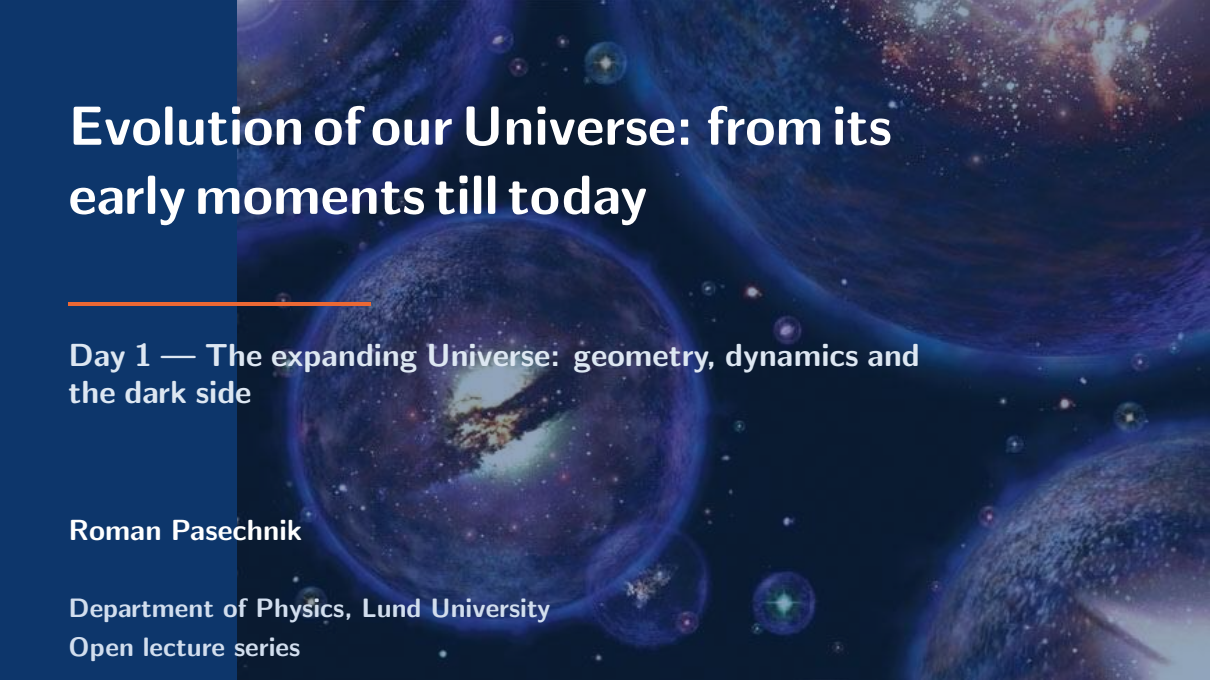
3. **Expand the integral.** $\int \mathbf{1}_R\,p_S\,dx = P_S(R)$, and similarly for the other three terms:
$$I = P_S(R^*) - P_S(R) - c\,\big[P_B(R^*) - P_B(R)\big] \ \geq\ 0$$
$$\Rightarrow \quad P_S(R^*) - P_S(R) \ \geq\ c\,\big[\alpha - P_B(R)\big] \ \geq\ 0$$

4. **Any monotonic function of Λ is optimal too.** If $d = g(\Lambda)$ with g strictly increasing, then $\{d > g(c)\} = \{\Lambda > c\}$: the same regions, the same ROC curve. The classifier output $d^* = \Lambda/(1+\Lambda)$ (equal class sizes) is such a function.

$P_S(R^*) \geq P_S(R)$ for every region R with $P_B(R) \leq \alpha$

At any fixed false-positive rate, cutting on the likelihood ratio (or on d^*) accepts the most signal (Neyman & Pearson, 1933).

Evolution of our Universe: The expanding Universe



Evolution of our Universe: from its early moments till today

Day 1 — The expanding Universe: geometry, dynamics and the dark side

Roman Pasechnik

Department of Physics, Lund University
Open lecture series

Hour 1

The Universe at a glance

scales, units and the cosmological principle

- How old and how big is the Universe, and how do we measure it?
- The ladder of cosmic structures and the cosmological principle
- Why is the night sky dark?
- Natural units and the hierarchy of scales; why gravity rules the cosmos



Image: NASA, ESA, Hubble Heritage Team (STScI/AURA)

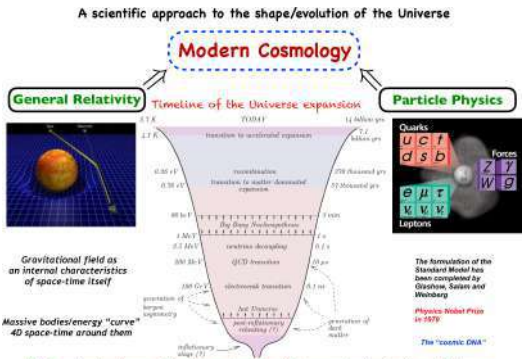
Two days, ten hours: the plan

"My goal is simple. It is a complete understanding of the universe, why it is as it is and why it exists at all." — Stephen Hawking

	Day 1 — the expanding Universe		Day 2 — the hot Big Bang
H1	The Universe at a glance: scales, units, the cosmological principle	H1	The hot early Universe: thermodynamics in an expanding world
H2	The expanding Universe: Hubble's law, redshift and geometry	H2	Particle physics meets cosmology: forces, the vacuum, phase transitions
H3	Cosmic dynamics: the Friedmann equations	H3	The first three minutes: neutrino decoupling and nucleosynthesis
H4	Measuring the Universe: distances, standard candles, acceleration	H4	Recombination and the cosmic microwave background
H5	The dark side: dark matter and dark energy	H5	Beyond the hot Big Bang: inflation, structure, dark-matter genesis, open questions

Day 1 builds the expanding Universe and its inventory today; Day 2 runs the film backwards.

Modern cosmology = General Relativity + Particle Physics

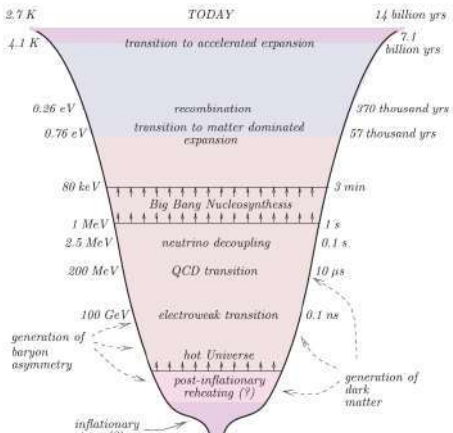


From the author's talk "Shedding light on the dark sides of the Universe"

- **Macro-world:** gravity = geometry of space-time (GR) ⇒ expansion, horizons, fate
- **Micro-world:** particles and forces (Standard Model) ⇒ content, thermal history
- The early Universe: the ultimate high-energy laboratory
- Gaps: dark matter, dark energy, the matter excess

In cosmology the largest and the smallest scales of Nature meet.

The cosmic history in one picture



Gorbunov & Rubakov (2011), Fig. 1.9

- Bottom to top: from hot and dense to cold and dilute
- Left: temperature; right: time after the Big Bang

Updated numbers (used in this course)

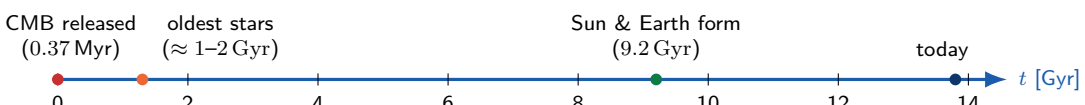
EW crossover	$160\,\mathrm{GeV}$	$10^{-11}\,\mathrm{s}$
QCD crossover	$155\,\mathrm{MeV}$	$2\times10^{-5}\,\mathrm{s}$
matter–radiation equality	$0.80\,\mathrm{eV}$	$51\,\mathrm{kyr}$
last scattering (CMB)	$0.26\,\mathrm{eV}$	$3.7\times10^{5}\,\mathrm{yr}$
acceleration begins	$4.4\,\mathrm{K}$	$7.7\,\mathrm{Gyr}$

By the end of Day 2 every line will be explained.

The Universe is old: 13.8 billion years

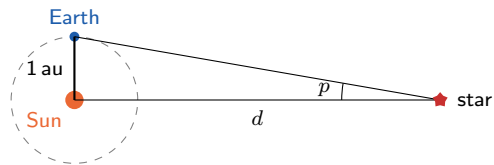
- CMB + ΛCDM: $t_0 = 13.80\,\mathrm{Gyr}$
  (Planck 2018: $13.797 \pm 0.023\,\mathrm{Gyr}$)
- Every artefact must be *younger*:
  - Earth, meteorites: $4.54\,\mathrm{Gyr}$ (radioactive dating)
  - Sun: $4.6\,\mathrm{Gyr}$
  - oldest stars, globular clusters: 12–13 Gyr
  - CMB photons: travelling for $\approx 13.8\,\mathrm{Gyr}$

Age is a consistency test: a cosmological model that is younger than its oldest stars is wrong. (We will use this in Hour 3.)



Measuring distances: light travel time and parallax

- Light: $c = 3.0 \times 10^5\,\mathrm{km\,s^{-1}}$
- Sun: $1\,\mathrm{au} = 1.496 \times 10^{11}\,\mathrm{m} = 8.3$ light-minutes
- Light year: $1\,\mathrm{ly} = c \times 1\,\mathrm{yr} = 9.46 \times 10^{15}\,\mathrm{m}$
- Parsec: the distance at which $1\,\mathrm{au}$ subtends $1''$



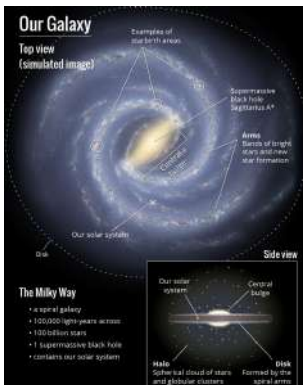
The parsec in one line

$$1'' = \frac{\pi}{180 \cdot 3600} = 4.848 \times 10^{-6}\,\mathrm{rad}$$

$$1\,\mathrm{pc} = \frac{1\,\mathrm{au}}{\tan 1''} = 206\,265\,\mathrm{au}$$
$$= 3.086 \times 10^{16}\,\mathrm{m} = 3.26\,\mathrm{ly}$$

- kpc: galaxies; Mpc: clusters, cosmic web
- Gpc: the observable Universe
- $1\,\mathrm{Mpc} = 3.26 \times 10^6\,\mathrm{ly}$
  $= 3.1 \times 10^{22}\,\mathrm{m}$

The ladder of cosmic structures



NASA and STScI / WebbTelescope.org

Solar System (planets, Kuiper belt)	$\sim 100\,\mathrm{au}$
Nearest stars (α Centauri)	$1.3\,\mathrm{pc}$
Centre of the Milky Way	$8\,\mathrm{kpc}$
Milky Way stellar disc ($\sim 10^{11}$ stars)	$\approx 30\,\mathrm{kpc}$ across
Magellanic Clouds (satellite galaxies)	50–60 kpc
Andromeda galaxy (M31)	$0.78\,\mathrm{Mpc}$
Clusters of galaxies (size)	1–3 Mpc
Virgo cluster (nearest large cluster)	$16.5\,\mathrm{Mpc}$
Superclusters, filaments, voids	$\sim 100\,\mathrm{Mpc}$
Hubble radius c/H_0	$4.4\,\mathrm{Gpc}$
Particle horizon (observable Universe)	$14\,\mathrm{Gpc}$

Galaxies and clusters of galaxies



NASA, ESA, Hubble

Barred spiral NGC 1672:
$\sim 10^{10}$–10^{11} stars

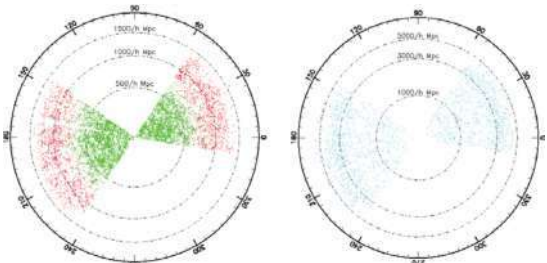


Dan Watt

The Coma cluster: thousands of galaxies, $\sim 100\,\mathrm{Mpc}$ away

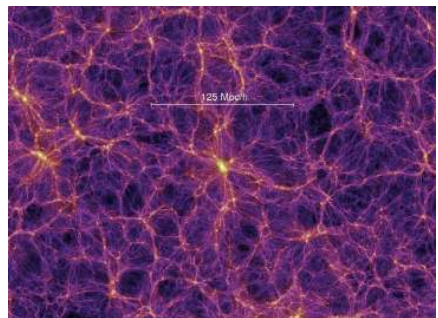
- The observable Universe contains $\sim 10^{11}$–10^{12} galaxies
- Clusters are the **largest gravitationally bound systems**: a few Mpc, 10^{14}–$10^{15} M_\odot$, hundreds to thousands of galaxies

The cosmic web



SDSS galaxy maps; Gorbunov & Rubakov (2011), Fig. 13.1

Observed (SDSS): left, galaxies (green) and luminous red galaxies (red); right, quasars; distance grows outwards



Springel et al. / MPA (Millennium simulation)

Simulated: dark matter in a thin slice; scale bar $125\,h^{-1}\,\mathrm{Mpc} \approx 185\,\mathrm{Mpc}$

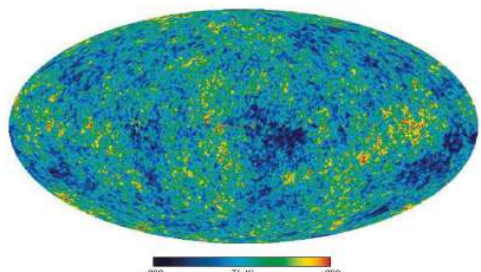
- Filaments, walls and voids, up to $\sim 100\,\mathrm{Mpc}$ across
- On larger scales the distribution becomes **uniform**

The cosmological principle

Cosmological principle

On scales $\gtrsim 100\,\mathrm{Mpc}$ the Universe is **homogeneous** (the same everywhere) and **isotropic** (the same in all directions).

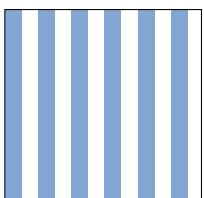
- Galaxy counts in spheres of radius $\gtrsim 200\,\mathrm{Mpc}$ agree to a few per cent
- CMB: the same temperature in all directions to $\delta T/T \sim 10^{-5}$ (after removing the 10^{-3} dipole due to our motion)
- Copernican principle: we are not at a special place



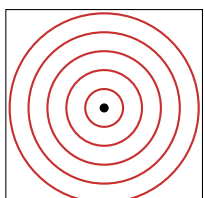
NASA/WMAP

Temperature map of the whole sky: fluctuations of $\pm 200\,\mu\mathrm{K}$ around $T_0 = 2.7255\,\mathrm{K}$

Homogeneity versus isotropy



homogeneous,
not isotropic

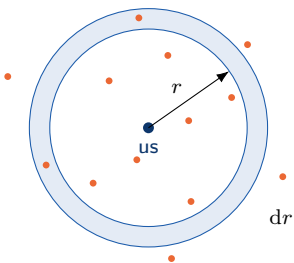


isotropic about one point,
not homogeneous

- Stripes: every point is equivalent, but one direction is special
- Bull's eye: all directions equivalent, but only from the centre
- **Isotropy about every point** $\Rightarrow$ homogeneity

We see isotropy (CMB) and assume we are typical (Copernicus) $\Rightarrow$ the Universe is homogeneous and isotropic.

Why is the night sky dark? Olbers' paradox



Assume an infinite, static, eternal Universe, uniformly filled with stars (n per volume, luminosity L):

Shell by shell

stars in the shell: $n\,4\pi r^2\,\mathrm{d}r$;
flux of each: $L/(4\pi r^2)$
$\mathrm{d}F = n\,L\,\mathrm{d}r$ — **the same for every shell**

$$F = nL\int_0^\infty \mathrm{d}r \to \infty$$

Stars block each other first: every line of sight ends on a stellar surface ⇒ the whole sky as bright as the Sun's surface!

Resolving Olbers' paradox: the Universe had a beginning

- **Finite age:** we only receive light emitted within the last $13.8\,\mathrm{Gyr}$
- **Expansion:** light from far away is redshifted — each photon arrives with less energy and less often
- Stars do not live long enough to fill space with light anyway

How far must a line of sight go to hit a star?

Mean free path $\lambda = 1/(n_* \pi R_*^2)$ with $n_* \sim 10^9\,\mathrm{Mpc}^{-3}$ (10^{11} stars per galaxy, $\sim 10^{-2}$ galaxies per Mpc^3), $R_* \sim R_\odot = 7 \times 10^8$ m:
$\lambda \sim 10^{18}\,\mathrm{Mpc} \sim 10^{24}\,\mathrm{ly} \approx 10^{14} \times (c\,t_0)$

A dark night sky is evidence for a beginning.

Natural units: $\hbar = c = k_B = 1$

- Energy, mass, temperature, inverse length and inverse time all measured in **GeV**
- $E = mc^2 \to m$; $E = k_B T \to T$; $\hbar c = 0.197\,\mathrm{GeV\,fm}$
- To return to SI: insert $\hbar$, c, k_B to fix the units

One unit for everything — the natural language of particle cosmology.

Conversion table

$$
\begin{aligned}
1\,\mathrm{GeV} &= 1.16 \times 10^{13}\,\mathrm{K} \\
1\,\mathrm{GeV} &= 1.78 \times 10^{-27}\,\mathrm{kg} \\
1\,\mathrm{GeV}^{-1} &= 1.97 \times 10^{-16}\,\mathrm{m} \\
1\,\mathrm{GeV}^{-1} &= 6.58 \times 10^{-25}\,\mathrm{s} \\
1\,\mathrm{eV} &= 11\,605\,\mathrm{K} \\
1\,\mathrm{cm}^{-3} &= 7.7 \times 10^{-42}\,\mathrm{GeV}^3
\end{aligned}
$$

Natural units at work: the photons of the CMB

- $T_0 = 2.7255\,\mathrm{K} = 2.35 \times 10^{-4}\,\mathrm{eV}$
- Mean photon energy of a blackbody: $\langle E \rangle \approx 2.70\,T$
- $\langle E \rangle \approx 6.3 \times 10^{-4}\,\mathrm{eV}$
- $\lambda = \dfrac{2\pi\hbar c}{\langle E \rangle} = \dfrac{1.24 \times 10^{-6}\,\mathrm{eV\,m}}{6.3 \times 10^{-4}\,\mathrm{eV}} \approx 2\,\mathrm{mm}$

The relic light of the Big Bang is **microwave** radiation (411 photons per cm^3).

Two more conversions

Proton:
$m_p = 0.938\,\mathrm{GeV} \leftrightarrow 1.1 \times 10^{13}\,\mathrm{K}$.

QCD crossover:
$155\,\mathrm{MeV} \leftrightarrow 1.8 \times 10^{12}\,\mathrm{K}$; above it protons dissolve into a quark–gluon plasma (Day 2).

Gravity is weak — the Planck scale

$$G = \frac{1}{M_{\mathrm{Pl}}^2}, \qquad M_{\mathrm{Pl}} = 1.22 \times 10^{19}\,\mathrm{GeV}$$

- Planck length $l_{\mathrm{Pl}} = 1.6 \times 10^{-35}\,\mathrm{m}$
- Planck time $t_{\mathrm{Pl}} = 5.4 \times 10^{-44}\,\mathrm{s}$
- Planck mass $M_{\mathrm{Pl}} = 2.2 \times 10^{-8}\,\mathrm{kg}$
- Quantum gravity becomes important at these scales

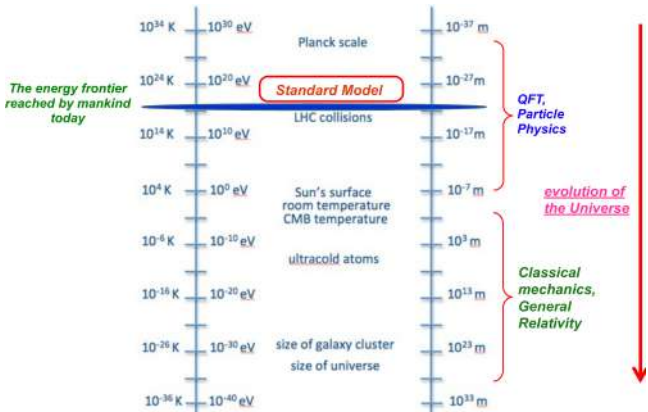
Two protons

$$\frac{F_{\mathrm{grav}}}{F_{\mathrm{el}}} = \frac{G m_p^2}{e^2/4\pi\varepsilon_0} = \frac{(m_p/M_{\mathrm{Pl}})^2}{\alpha}$$

$$\approx 8 \times 10^{-37}$$

Gravity rules the cosmos only because it always attracts and matter is electrically neutral on large scales.

Energy, length and temperature scales of fundamental physics



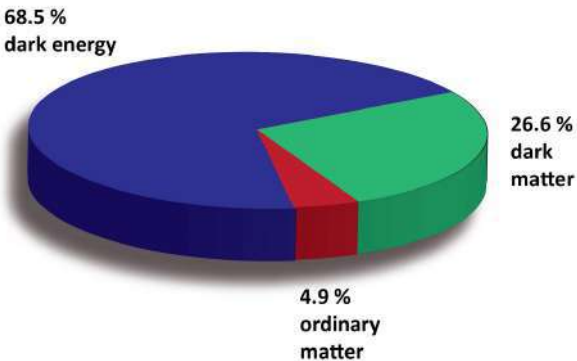
From the author's talk "Shedding light on the dark sides of the Universe"

- Going back in time = going **up** in energy and temperature
- LHC: $E \sim 10^4\,\mathrm{GeV} \leftrightarrow 10^{17}\,\mathrm{K}$
- Thermal particles had such energies ($\langle E \rangle \approx 3T$) at $t \sim 10^{-14}$ s

Radiation-era clock (Day 2)

$$t \approx 0.3\, g_*^{-1/2}\, M_{\mathrm{Pl}}/T^2$$

What is the Universe made of? A teaser

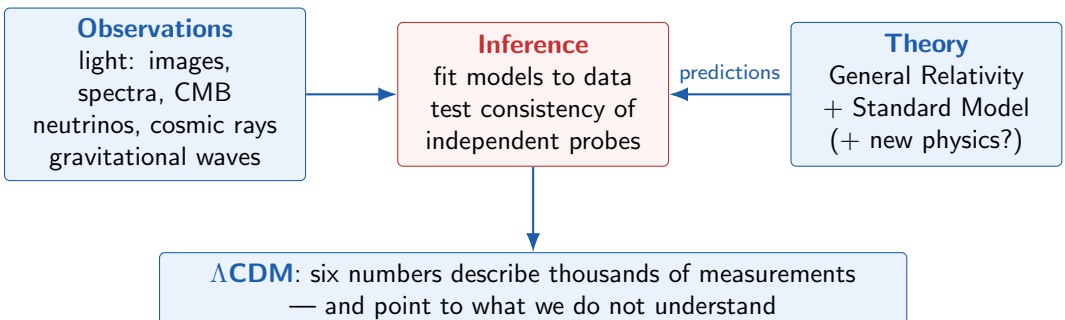


ESA/Planck

- **68.5 %** dark energy
- **26.6 %** dark matter
- **4.9 %** ordinary matter (atoms)
- Light, neutrinos: $< 0.3\,\%$

"Our job is to figure out what the rest is!" — and to understand *how we know* each number (Hours 3–5).

How cosmology works



- We cannot experiment with the Universe — but it has done the experiments for us

Key points of Hour 1

- The Universe is $13.8\,\mathrm{Gyr}$ old; its observable part has a radius of $\approx 14\,\mathrm{Gpc}$
- Structures form a hierarchy up to $\sim 100\,\mathrm{Mpc}$ (cosmic web); beyond that the Universe is **homogeneous and isotropic** (cosmological principle; CMB isotropic to 10^{-5})
- A dark night sky $\Rightarrow$ the Universe has a finite age and expands
- Natural units $\hbar = c = k_B = 1$: everything in GeV; $1\,\mathrm{eV} \leftrightarrow 11\,605\,\mathrm{K}$
- Gravity is extremely weak ($M_{\mathrm{Pl}} = 1.22 \times 10^{19}\,\mathrm{GeV}$) but dominates on large scales
- Today: $68.5\,\%$ dark energy, $26.6\,\%$ dark matter, $4.9\,\%$ ordinary matter

Hour 2

The expanding Universe

Hubble's law, redshift and geometry

- From Einstein's static Universe to Hubble's expansion
- Why a homogeneous expansion gives $v = H_0 d$ — and has no centre
- Redshift as the stretching of space: $1 + z = a_0/a$
- The three possible geometries and the FLRW metric
- How far can we see? The particle horizon

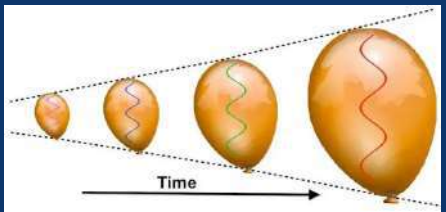


Image: from the author's talk

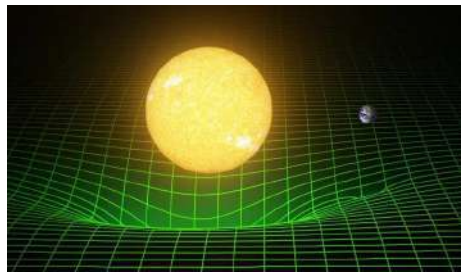
General Relativity in one slide

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}$$

Law of an expanding universe

All matter and energy in the universe

From the author's talk "Shedding light on the dark sides of the Universe"

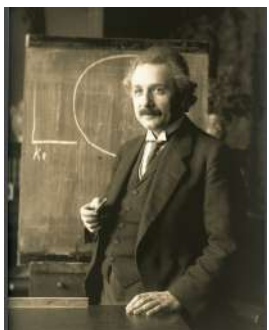


LIGO/T. Pyle

- Gravity = curvature of space-time
- Metric: $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$ — 10 functions of space and time
- Left: curvature built from $g_{\mu\nu}$; right: energy, momentum, pressure
- Tested: light bending (1919), GPS clocks, gravitational waves (2015)

"Matter tells space-time how to curve; space-time tells matter how to move." (J. A. Wheeler)

Static or expanding? Theory ahead of observation



public domain

- **1917** Einstein applies GR to the whole Universe; to keep it *static* he adds the cosmological constant Λ
- **1922, 1924** Friedmann: homogeneous solutions that expand or contract
- **1927** Lemaître: an expanding Universe explains the redshifts of galaxies
- **1929** Hubble: redshift grows with distance

History

Doppler shifts of spiral nebulae were measured by V. Slipher from 1912 on (the first, M31, a blueshift; most later ones redshifts). Einstein later abandoned the static model: the theory had been clever, but observations were way behind.

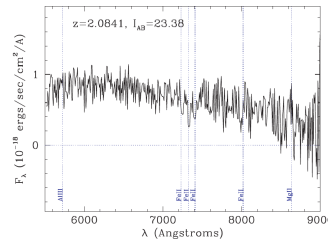
Redshift

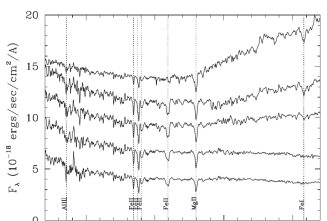
$$z \equiv \frac{\lambda_{\rm obs} - \lambda_{\rm em}}{\lambda_{\rm em}}, \qquad 1 + z = \frac{\lambda_{\rm obs}}{\lambda_{\rm em}}$$

- Atoms emit and absorb at fixed wavelengths: a fingerprint
- In distant galaxies the whole pattern of lines is shifted to the **red**, by the same factor $1 + z$
- Small z: Doppler effect, $z \simeq v/c$

Quick estimate

Coma cluster: $z = 0.023 \Rightarrow v \approx 6900\,\mathrm{km\,s^{-1}}$





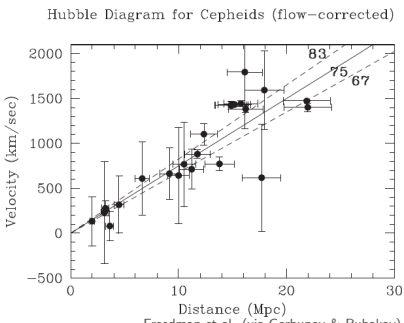
Gorbunov & Rubakov (2011), Fig. 1.1

Hubble's law



Huntington Library / public domain

E. Hubble, 1929: redshift $\propto$ distance



Modern Hubble diagram, distances from Cepheids

$$v = H_0\, d$$

$$H_0 = 100\, h\, \mathrm{km\,s^{-1}\,Mpc^{-1}}$$

- Planck 2018: $h = 0.674$
- H_0 has units of $1/\text{time}$

Homogeneous expansion implies Hubble's law

Forward

Physical position $\vec{r}(t) = a(t)\,\vec{x}$, comoving $\vec{x}$ fixed:

$$\vec{v} = \dot{\vec{r}} = \dot{a}\,\vec{x} = \frac{\dot{a}}{a}\,\vec{r}$$

$$\vec{v} = H(t)\,\vec{r}, \qquad H(t) \equiv \frac{\dot{a}}{a}$$

$H_0 \equiv H(t_0)$: the Hubble constant is the value of the Hubble *parameter* today.

Converse

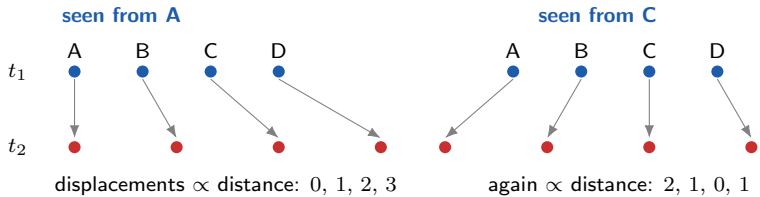
Homogeneity: velocity of galaxy 1 seen from galaxy 2 depends only on their separation:

$$\vec{v}(\vec{r}_1) - \vec{v}(\vec{r}_2) = \vec{v}(\vec{r}_1 - \vec{r}_2)$$

$\Rightarrow$ $\vec{v}$ is linear in $\vec{r}$; isotropy $\Rightarrow \vec{v} = H\vec{r}$

The linear law is the **only** expansion law compatible with the cosmological principle.

No centre: every galaxy sees the same law



- All distances grow by the same factor $a(t_2)/a(t_1)$: each galaxy sees all others recede with $v \propto d$
- Like raisins in a rising bread, or dots on an inflating balloon: **no galaxy is the centre**
- Galaxies do not move *through* space — space itself expands (plus small peculiar motions)

The Big Bang happened everywhere



From the author's talk "Shedding light on the dark sides of the Universe"

- Going back in time, *all* distances shrink $\propto a(t) \to 0$
- The Big Bang is a **moment in time**, not a point in space
- An infinite Universe stays infinite all the way back — but ever denser and hotter
- There is no "outside" into which the Universe expands

The Big Bang happened at every point of the Universe at once.

Hubble time and Hubble radius

$$t_H = \frac{1}{H_0} = 14.5\,\mathrm{Gyr}, \qquad \frac{c}{H_0} = 4.45\,\mathrm{Gpc}$$

- t_H: the time to reach today's distances at today's speeds — a first estimate of the age
- c/H_0: distance at which $v = H_0 d$ reaches c
- Beyond c/H_0 galaxies recede *faster than light*

Handy numbers

$$1/H_0 = 9.78\,h^{-1}\,\mathrm{Gyr}$$
$$c/H_0 = 2998\,h^{-1}\,\mathrm{Mpc}$$
$$H_0 = 2.18 \times 10^{-18}\,\mathrm{s}^{-1} = 1.44 \times 10^{-42}\,\mathrm{GeV}$$

No contradiction with relativity: nothing overtakes a light signal *locally*; it is space that expands.

Cosmological redshift = stretching of space



Chain of nearby observers

Each step is a tiny Doppler shift:

$$\frac{\mathrm{d}\lambda}{\lambda} = \frac{\mathrm{d}v}{c} = \frac{H\,\mathrm{d}r}{c} = H\,\mathrm{d}t = \frac{\dot{a}}{a}\mathrm{d}t = \frac{\mathrm{d}a}{a}$$

$$\Rightarrow\ \lambda \propto a(t)$$

$$1 + z = \frac{a(t_0)}{a(t_{\mathrm{em}})} = \frac{1}{a(t_{\mathrm{em}})}$$

With $a_0 = 1$. The Doppler formula $z \simeq v/c$ is only the local, infinitesimal version.

What else redshifts?

- Photon energy: $E = \dfrac{2\pi\hbar c}{\lambda} \propto \dfrac{1}{a}$
- Momenta of *all* free particles: $p \propto \dfrac{1}{a}$ (the same chain-of-observers argument)
- Peculiar velocities of freely moving bodies die out, $v \propto 1/a$ (unless sustained by gravity)
- Number densities: $n \propto a^{-3}$ (volume $\propto a^3$)

Radiation

$$\rho_\gamma = n_\gamma \langle E \rangle \propto a^{-3} \cdot a^{-1} = a^{-4}$$

Non-relativistic matter

$$\rho_m = m\, n \propto a^{-3}$$

Radiation dilutes faster than matter: in the past it must have dominated (Hour 3, Day 2).

A blackbody stays a blackbody

Photons in an expanding box

Free photons: number conserved, each $E \propto 1/a$.

Planck spectrum at t_1, $\mathrm{d}n \propto \dfrac{E^2 \mathrm{d}E}{\mathrm{e}^{E/T_1} - 1}$

later: $E \to E\,\dfrac{a_1}{a}$, $\mathrm{d}n \to \mathrm{d}n \left(\dfrac{a_1}{a}\right)^3$

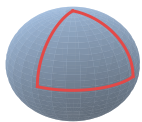
$\Rightarrow$ again Planck, with $T = T_1\, a_1/a$

$$T \propto \frac{1}{a}, \qquad T(z) = T_0\,(1+z)$$

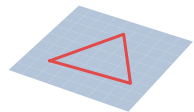
- Measured at high z: excitation of atoms and molecules in distant gas clouds; Sunyaev–Zel'dovich effect in clusters
- The early Universe was **hot**

Three possible geometries of homogeneous, isotropic space

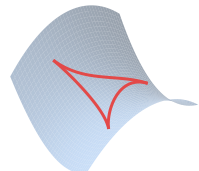
closed: $k > 0$
(angles sum > 180°)



flat: $k = 0$
(angles sum = 180°)



open: $k < 0$
(angles sum < 180°)



- Closed (sphere), flat (plane), open (saddle)
- Angle sum of a triangle: $> 180°$, $= 180°$, $< 180°$

Circle of radius ρ (curvature radius R)

$$C = 2\pi R \sin\frac{\rho}{R}, \quad 2\pi\rho, \quad 2\pi R \sinh\frac{\rho}{R}$$

The Friedmann–Lemaître–Robertson–Walker (FLRW) metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1-k\,r^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)\right]$$

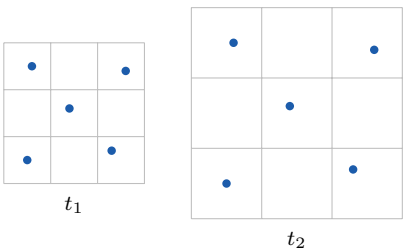
- $k > 0$ closed, $k = 0$ flat, $k < 0$ open; $|k| = 1/R_0^2$ (R_0: curvature radius today, $a_0 = 1$)
- The whole Universe: **one function** $a(t)$ and **one number** k
- t: cosmic time, measured by comoving clocks

Flat case

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2)$$

Course book: a_0 kept general, $k = \pm 1, 0$; identical with $a_0 \leftrightarrow R_0$.

Comoving and physical distances



- Galaxies sit (nearly) at fixed **comoving** coordinates; the grid expands
- Physical distance $= a(t) \times$ comoving distance
- With $a_0 = 1$: comoving distance $=$ physical distance *today*
- Bound systems do *not* expand: atoms (electromagnetic forces), the Solar System, galaxies, clusters (gravity)

$$D(t) = a(t)\,\chi, \qquad \dot{D} = \dot{a}\,\chi = H\,D \quad \text{(Hubble's law again)}$$

How light travels in an expanding Universe

- Light moves on null paths: $\mathrm{d}s^2 = 0$
- Radial ray (flat space): $\mathrm{d}t^2 = a^2(t)\,\mathrm{d}\chi^2$

$$\chi = \int_{t_{\mathrm{em}}}^{t_0} \frac{c\,\mathrm{d}t}{a(t)}$$

comoving distance covered by light between emission and observation

Conformal time

$$\eta = \int \frac{\mathrm{d}t}{a(t)}$$

$$\mathrm{d}s^2 = a^2(\eta)\left(\mathrm{d}\eta^2 - \mathrm{d}\vec{x}^{\,2}\right)$$

In $(\eta, \vec{x})$ light rays are straight 45° lines, as in special relativity.

Redshift from the metric — and cosmic time dilation

Two wave crests

Crests emitted at t_e and $t_e + \delta t_e$ arrive at t_0 and $t_0 + \delta t_0$; both cover the same χ:

$$\int_{t_e}^{t_0} \frac{\mathrm{d}t}{a} = \int_{t_e+\delta t_e}^{t_0+\delta t_0} \frac{\mathrm{d}t}{a} \;\Rightarrow\; \frac{\delta t_e}{a(t_e)} = \frac{\delta t_0}{a(t_0)}$$

$$\frac{\delta t_0}{\delta t_e} = \frac{\lambda_0}{\lambda_e} = 1 + z$$

- Not only wavelengths: **every** process far away appears slowed down by $1 + z$
- Observed: light curves of distant supernovae last $(1 + z)$ times longer
  (Leibundgut et al. 1996; Goldhaber et al. 2001)

How far can we see? The particle horizon

- Light emitted at $t = 0$ has travelled the comoving distance $\int_0^t \mathrm{d}t'/a(t')$
- Regions further away are causally disconnected from us — still today

$$d_{\mathrm{hor}}(t) = a(t) \int_0^t \frac{c\,\mathrm{d}t'}{a(t')}$$

Power law $a \propto t^n$ ($n < 1$, $c = 1$)

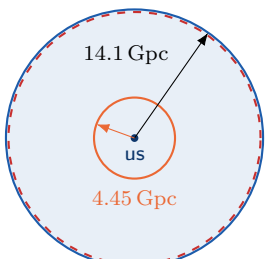
$$d_{\mathrm{hor}} = t^n \int_0^t t'^{-n} \mathrm{d}t' = \frac{t}{1-n}$$

radiation ($n = \frac{1}{2}$): $d_{\mathrm{hor}} = 2t$

matter ($n = \frac{2}{3}$): $d_{\mathrm{hor}} = 3t$

The horizon is larger than $c\,t$: space expanded while the light was on its way.

The observable Universe today

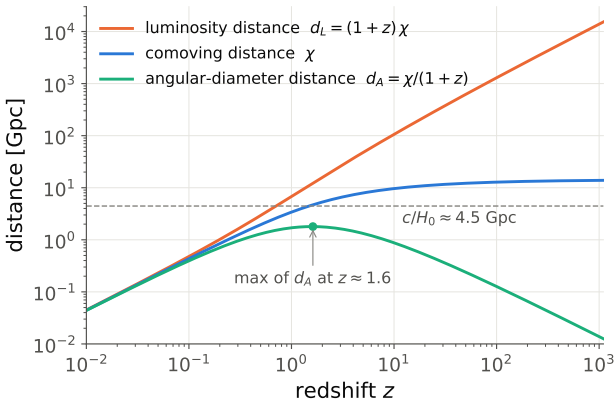


Full ΛCDM, today

particle horizon $= 14.1\,\mathrm{Gpc} = 46 \times 10^9\,\mathrm{ly}$
$> \; c\,t_0 = 13.8 \times 10^9\,\mathrm{ly}$

- Hubble radius $c/H_0 = 4.45\,\mathrm{Gpc}$ (orange)
- CMB last-scattering surface: comoving distance $13.9\,\mathrm{Gpc}$ — but only $12.7\,\mathrm{Mpc}$ (42 million ly) away when the light set off

Distances in an expanding Universe: a preview



- "Distance" is no longer unique
- **comoving** χ: on today's grid
- **luminosity** d_L: from brightness
- **angular-diameter** d_A: from apparent size
- All agree for $z \ll 1$ ($\approx cz/H_0$)

Made precise in Hour 4 — the basis of modern precision cosmology.

Key points of Hour 2

- GR: geometry of space-time $\leftrightarrow$ matter and energy; the Universe is described by the **FLRW metric**: one function $a(t)$, one curvature k
- A homogeneous, isotropic expansion gives **Hubble's law** $v = H_0 d$, with no centre; $H_0 = 67.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $1/H_0 = 14.5\,\mathrm{Gyr}$, $c/H_0 = 4.45\,\mathrm{Gpc}$
- The Big Bang is a moment in time, not a point in space
- Redshift is the stretching of space: $1 + z = 1/a$; all momenta $\propto 1/a$, $T \propto 1/a$, $\rho_\gamma \propto a^{-4}$, $\rho_m \propto a^{-3}$
- Distant clocks run slow by $1 + z$
- The observable Universe has a radius of $14.1\,\mathrm{Gpc} = 46 \times 10^9\,\mathrm{ly}$

Hour 3

Cosmic dynamics

the Friedmann equations

- The Friedmann equation from Newton's laws — and what GR adds
- How matter, radiation and vacuum energy dilute
- Solving for $a(t)$: eras, the age of the Universe, acceleration
- Einstein's static Universe and the fate of the Universe

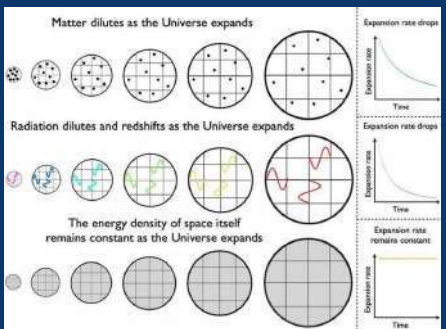
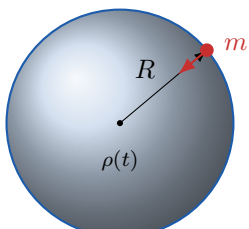


Image: from the author's talk

The Friedmann equation from Newton (1): the set-up



$R(t) = a(t)\,x$, x fixed

- Uniform ball of density $\rho(t)$, radius $R(t) = a(t)\,x$
- Matter moves with the expansion: mass inside is conserved,
$$M = \tfrac{4\pi}{3}\rho R^3 = \text{const}$$
- Shell theorem: only the mass *inside* acts on a galaxy at the surface

$$\ddot{R} = -\frac{GM}{R^2}$$

Legitimate for a small ball ($v \ll c$); homogeneity: any point can be the centre.

The Friedmann equation from Newton (2): energy conservation

First integral of the equation of motion

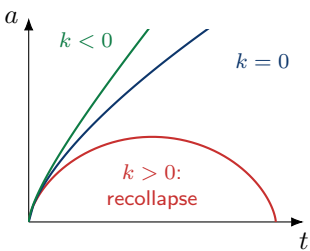
$$\ddot{R} = -\frac{GM}{R^2} \;\Big|\cdot \dot{R} \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\dot{R}^2 - \frac{GM}{R}\right) = 0 \quad \Rightarrow \quad \frac{1}{2}\dot{R}^2 - \frac{GM}{R} = E$$

$$R = a\,x, \quad M = \frac{4\pi}{3}\rho\, a^3 x^3 : \qquad \frac{1}{2}\,\dot{a}^2 x^2 - \frac{4\pi G}{3}\,\rho\, a^2 x^2 = E \qquad \Big|\cdot \frac{2}{a^2 x^2}$$

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}$$

- $k = -2E/x^2$ (E: energy per unit mass): the same for every shell (homogeneity)
- GR gives exactly the same equation; k becomes the **curvature** of space

Like a ball thrown upwards: the critical density



matter only; same early history

- $E < 0$ ($k > 0$): bound — expansion stops, recollapse
- $E = 0$ ($k = 0$): marginal — expands forever, $\dot{a} \to 0$
- $E > 0$ ($k < 0$): unbound — expands forever

Critical density ($k = 0$)

$$\rho_c = \frac{3H^2}{8\pi G}: \qquad \rho > \rho_c \;\Leftrightarrow\; k > 0 \text{ (closed)}$$

Geometry and destiny are linked — *if* the Universe contains only matter.

The critical density today

Critical density

$$\rho_{c,0} = \frac{3H_0^2}{8\pi G} = 1.05 \times 10^{-5}\, h^2\,\mathrm{GeV\,cm^{-3}} = 4.8 \times 10^{-6}\,\mathrm{GeV\,cm^{-3}} = 8.5 \times 10^{-27}\,\mathrm{kg\,m^{-3}}$$

$\approx$ **5 proton masses per cubic metre**

- Air: $\sim 10^{25}$ molecules per m^3
- Interstellar gas: ~ 1 atom per $\mathrm{cm}^3 = 10^6$ per m^3
- Best laboratory vacua: still billions of molecules per m^3

The Universe on average is emptier than any vacuum we can make — yet this tiny density decides its geometry and history.

What General Relativity adds: pressure gravitates

- Matter as a perfect fluid, at rest on the comoving grid:

$$T^{\mu}{}_{\nu} = \mathrm{diag}(\rho, -p, -p, -p)$$

- Einstein's equations for FLRW give two equations:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}$$
$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p)$$

- $\rho =$ *total* energy density (incl. mc^2)
- Newton: $\ddot{a}/a = -\frac{4\pi G}{3}\rho$; GR: source of gravity is $\rho + 3p$
- Pressure is a form of energy and **gravitates**: radiation decelerates the expansion twice as strongly as dust of the same ρ

Accelerated expansion requires $\rho + 3p < 0$, i.e. $p < -\rho/3$.

Energy conservation in an expanding Universe

First law for a comoving volume $V = a^3$

No heat flows (homogeneity):

$$\mathrm{d}E = -p\,\mathrm{d}V, \quad E = \rho\, a^3$$

$$\mathrm{d}(\rho a^3) = -p\,\mathrm{d}(a^3)$$

$$\dot{\rho}\, a^3 + 3\rho a^2 \dot{a} = -3pa^2\dot{a}$$

$$\dot{\rho} + 3H(\rho + p) = 0$$

- The **continuity equation**: expansion is adiabatic
- It also follows from the two Friedmann equations (they are consistent)
- Two equations + an equation of state $p(\rho)$ determine $a(t)$

Equation of state and dilution

$p = w\rho$, w constant

$$\dot{\rho} = -3(1+w)\frac{\dot{a}}{a}\rho$$

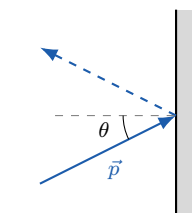
$$\Rightarrow \quad \frac{\mathrm{d}\rho}{\rho} = -3(1+w)\,\frac{\mathrm{d}a}{a}$$

$$\rho \propto a^{-3(1+w)}$$

component	w	$\rho \propto$
matter ("dust")	0	a^{-3}
radiation, relativistic particles	$\frac{1}{3}$	a^{-4}
vacuum energy (Λ)	-1	const
curvature term $-k/a^2$	"$-\frac{1}{3}$"	a^{-2}

Going back in time: radiation wins; going forward: vacuum energy wins.

Why $p = \rho/3$ for radiation



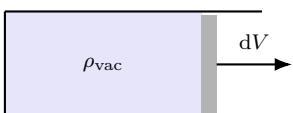
Kinetic theory

particles hitting unit wall area per unit time at angle θ: $n\,v\cos\theta\,\dfrac{\mathrm{d}\Omega}{4\pi}$

$$P = \int_{\text{half}} \frac{\mathrm{d}\Omega}{4\pi}\, n\,v\cos\theta \cdot 2p\cos\theta = \frac{1}{3}\, n\, \langle p\,v\rangle$$

- Photons: $v = c$, $pc = E \Rightarrow P = \frac{1}{3}\, n\,\langle E\rangle = \rho/3$
- Non-relativistic gas: $P = nT \ll \rho = nm \quad (w \approx 0)$

Why $p = -\rho$ for vacuum energy



pulling the piston *adds* energy $\rho_{\rm vac}\mathrm{d}V$

Piston

Vacuum energy density is the same everywhere: $\mathrm{d}E = \rho_{\rm vac}\,\mathrm{d}V$. First law: $\mathrm{d}E = -p\,\mathrm{d}V$

$\Rightarrow\ p_{\rm vac} = -\rho_{\rm vac}$ (a tension)

- Lorentz invariance: the vacuum looks the same to all observers $\Rightarrow T_{\mu\nu} = \rho_{\rm vac}\, g_{\mu\nu} \Rightarrow p = -\rho$
- $\rho + 3p = -2\rho < 0$: **gravity becomes repulsive**

Density parameters and the practical Friedmann equation

- Measure every component in units of the critical density today:

$$\Omega_i = \frac{\rho_{i,0}}{\rho_{c,0}}, \qquad \Omega_k = -\frac{k}{H_0^2}$$

- Friedmann equation today ($a = 1$):

$$\Omega_r + \Omega_m + \Omega_k + \Omega_\Lambda = 1$$

Divide by H_0^2, use $\rho_i \propto a^{-3(1+w_i)}$, $a = \frac{1}{1+z}$

$$\frac{H^2}{H_0^2} = \sum_i \frac{\rho_i}{\rho_{c,0}} - \frac{k}{H_0^2 a^2}$$

$$H^2(z) = H_0^2 \left[\Omega_r (1+z)^4 + \Omega_m (1+z)^3 + \Omega_k (1+z)^2 + \Omega_\Lambda\right]$$

The cosmic inventory today

component	Ω_i
dark energy Ω_Λ	0.685
cold dark matter Ω_c	0.265
baryons Ω_b (stars: $\lesssim 0.005$)	0.049
neutrinos Ω_ν	0.001–0.003
photons Ω_γ (CMB)	5.4×10^{-5}
curvature Ω_k	0.001 ± 0.002

- $\Omega_m = \Omega_c + \Omega_b + \Omega_\nu = 0.265 + 0.049 + 0.001 = 0.315$
- Radiation in the early Universe (γ + three light ν): $\Omega_r = 9.2 \times 10^{-5}$
- Space is flat to $\lesssim 0.5\,\%$

95 % of today's energy density is of unknown nature.

Evolution of our Universe: From the expansion history to the hot Big Bang and inflation



Evolution of our Universe: from its early moments till today

Day 2 — From the expansion history to the hot Big Bang and inflation

Roman Pasechnik

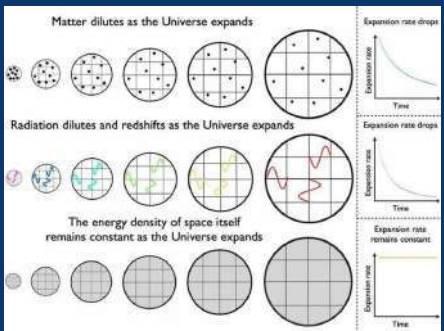
Department of Physics, Lund University

Open lecture series

Hour 1

How the Universe expands — and how we measure it

- Solving the Friedmann equation: eras, equality, the age of the Universe
- Deceleration or acceleration?
- Distances in an expanding Universe
- Supernovae reveal the accelerating Universe



From the author's talk "Shedding light on the dark sides of the Universe"

Where we stopped

the Friedmann equation and the cosmic inventory

- FLRW ($c = 1$): $\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\,\mathrm{d}\vec{x}^{\,2}$, $1 + z = 1/a$ ($a_0 = 1$), $H = \dot{a}/a$
- Friedmann: $H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}$, $\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p)$
- Continuity: $\dot{\rho} + 3H(\rho + p) = 0 \Rightarrow \rho \propto a^{-3(1+w)}$ ($p = w\rho$)

the cosmic inventory today

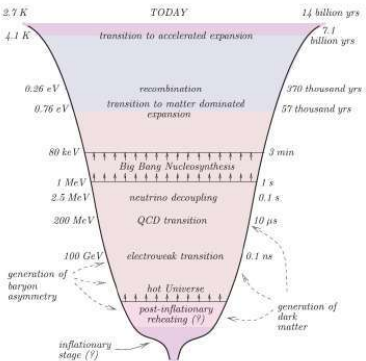
dark energy Ω_Λ	0.685
cold dark matter Ω_c	0.265
baryons Ω_b	0.049
massive neutrinos Ω_ν	≈ 0.001
radiation Ω_r (γ, ν)	9.2×10^{-5}
curvature Ω_k	0.001 ± 0.002

$$H^2(z) = H_0^2\left[\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda\right]$$
$$\Omega_i = \rho_{i,0}/\rho_{c,0}, \quad \rho_{c,0} = 3H_0^2/8\pi G, \quad \Omega_k = -k/H_0^2, \quad H_0 = 67.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$$

▶ Today: **solve** it and **measure** the Universe (Hours 1–2), then run the film backwards (Hours 3–5).

Today's plan

from today's expansion back to the first moments



Gorbunov & Rubakov (2011), Fig. 1.9 — 2011 values; current values on Slide 68

H1 How the Universe expands: eras, age, acceleration, distances, supernovae

H2 What it is made of: rulers, tensions, dark matter, dark energy

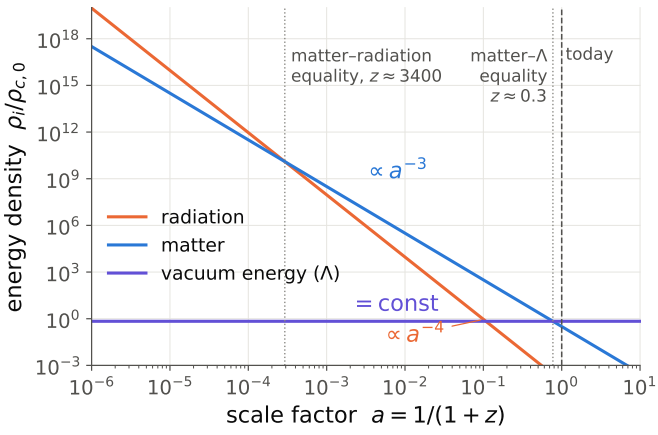
H3 The hot early Universe: thermodynamics, the clock, phase transitions, antimatter

H4 The first three minutes and the cosmic microwave background

H5 Beyond the hot Big Bang: inflation, structure, the origin of dark matter

▶ Goal: by the end of today every line of this diagram is explained.

The three eras of cosmic history



- $\rho_r \propto a^{-4}$, $\rho_m \propto a^{-3}$, $\rho_\Lambda =$ const: each dominates in turn
- **Radiation era**: until $z_{\rm eq} \approx 3400$
- **Matter era**: structures grow
- **Dark-energy era**: since $z \approx 0.3$

Each era has its own simple law $a(t)$ — next slide.

One component at a time: solving for $a(t)$

Flat, one component, $w \neq -1$

$$H^2 = H_0^2\,\Omega\,a^{-3(1+w)}$$
$$\Rightarrow\ \dot{a} = H_0\sqrt{\Omega}\,a^{1-\frac{3}{2}(1+w)}$$
$$\Rightarrow\ a^{\frac{3}{2}(1+w)-1}\,\mathrm{d}a = H_0\sqrt{\Omega}\,\mathrm{d}t$$
$$\Rightarrow\ a^{\frac{3}{2}(1+w)} \propto t \quad \text{(choosing } a = 0 \text{ at } t = 0\text{)}$$

$$a \propto t^{\frac{2}{3(1+w)}}, \qquad H = \frac{\dot{a}}{a} = \frac{2}{3(1+w)\,t}$$

	w	$\rho(a)$	$a(t)$	$H(t)$
radiation	$\frac{1}{3}$	a^{-4}	$t^{1/2}$	$\frac{1}{2t}$
matter	0	a^{-3}	$t^{2/3}$	$\frac{2}{3t}$
vacuum	-1	const	e^{Ht}	const

- Radiation, matter: **decelerating**, $a \to 0$ at $t = 0$
- Vacuum (H const): **accelerating**, de Sitter

- The real Universe passes through these regimes one after another.

Matter–radiation equality

- $\rho_r \propto (1+z)^4$, $\rho_m \propto (1+z)^3$: equal when

$$1 + z_{\rm eq} = \frac{\Omega_m}{\Omega_r} = \frac{0.315}{9.2 \times 10^{-5}} \approx 3400$$

- $T_{\rm eq} = T_0(1+z_{\rm eq}) \approx 0.80\,{\rm eV} \approx 9300\,{\rm K}$
- $t_{\rm eq} \approx 51\,{\rm kyr}$ (radiation only: $\approx 65\,{\rm kyr}$)

Exact two-component solution

$x \equiv a/a_{\rm eq}$: $\quad H^2 = \frac{1}{2}H_{\rm eq}^2\left(x^{-4} + x^{-3}\right)$

$$\mathrm{d}t = \frac{\mathrm{d}a}{aH} = \frac{\mathrm{d}x}{xH} = \frac{\sqrt{2}}{H_{\rm eq}}\frac{x\,\mathrm{d}x}{\sqrt{1+x}}$$

$$t_{\rm eq} = \frac{\sqrt{2}}{H_{\rm eq}}\int_0^1 \frac{x\,\mathrm{d}x}{\sqrt{1+x}} = \frac{4(\sqrt{2}-1)}{3H_{\rm eq}}$$

$$H_{\rm eq} = H_0\sqrt{2\Omega_r}\,(1+z_{\rm eq})^2 \;\Rightarrow\; t_{\rm eq} \approx 51\,{\rm kyr}$$

- After equality, dark matter can clump: the seeds of galaxies start to grow (Hour 5).

Matter + Λ: an exact solution

- Late Universe, flat, radiation neglected:

$$H^2 = H_0^2 \left(\Omega_m a^{-3} + \Omega_\Lambda\right)$$

$$a(t) = \left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{1/3} \sinh^{2/3}\left(\tfrac{3}{2}\sqrt{\Omega_\Lambda}\, H_0 t\right)$$

Check by substitution, $\tau \equiv \frac{3}{2}\sqrt{\Omega_\Lambda} H_0 t$

$$H = \frac{\dot a}{a} = \frac{2}{3}\frac{\cosh\tau}{\sinh\tau}\,\dot\tau = \sqrt{\Omega_\Lambda}\, H_0 \coth\tau$$

$$H^2 = \Omega_\Lambda H_0^2 \left(1 + \frac{1}{\sinh^2\tau}\right) = H_0^2\left(\Omega_\Lambda + \frac{\Omega_m}{a^3}\right) \quad \text{since} \quad \frac{\Omega_\Lambda}{\sinh^2\tau} = \frac{\Omega_m}{a^3}\ \checkmark$$

- Early times: $\sinh\tau \approx \tau \Rightarrow a \propto t^{2/3}$ (matter era); late times: $a \propto \mathrm{e}^{\sqrt{\Omega_\Lambda} H_0 t}$

The age of the Universe

- Today $a(t_0) = 1 \Rightarrow \sinh \tau_0 = \sqrt{\Omega_\Lambda / \Omega_m}$:

$$t_0 = \frac{2}{3\sqrt{\Omega_\Lambda}\, H_0} \operatorname{arsinh} \sqrt{\frac{\Omega_\Lambda}{\Omega_m}} = 13.8\,\mathrm{Gyr}$$

- Matter only (Einstein–de Sitter, $\Omega_\Lambda \to 0$):
  $t_0 = \dfrac{2}{3H_0} = 9.7\,\mathrm{Gyr}$
- Radiation changes t_0 by only $\approx 5\,\mathrm{Myr}$

Numbers

$1/H_0 = 14.5\,\mathrm{Gyr}$
$\sqrt{\Omega_\Lambda/\Omega_m} = 1.475, \quad \operatorname{arsinh} 1.475 = 1.181$
$t_0 = 0.806 \times 1.181 \times 14.5\,\mathrm{Gyr} = 13.8\,\mathrm{Gyr}$

Einstein–de Sitter is younger than the oldest stars (12–13 Gyr): the **age problem**, solved by Λ.

Deceleration or acceleration?

$$q \equiv -\frac{\ddot{a}\,a}{\dot{a}^2} = \frac{1}{2}\sum_i \Omega_i(z)\,(1+3w_i)$$

$$\Omega_i(z) \equiv \frac{\rho_i(z)}{\rho_c(z)}, \qquad \rho_c(z) = \frac{3H^2(z)}{8\pi G}$$

(divide the acceleration equation by H^2)

- Today: $q_0 = \frac{1}{2}\Omega_m - \Omega_\Lambda = -0.53$: **accelerating**
- $\ddot{a} = 0$ when $\Omega_m(1+z)^3 = 2\Omega_\Lambda$: $1 + z_{\rm acc} = \left(2\Omega_\Lambda/\Omega_m\right)^{1/3} = 1.63$

Milestones of the late Universe

Acceleration starts: $z = 0.63$, $t = 7.7\,\mathrm{Gyr}$ ($6.1\,\mathrm{Gyr}$ ago)

$\rho_\Lambda = \rho_m$: $z = 0.30$, $t = 10.3\,\mathrm{Gyr}$ ($3.5\,\mathrm{Gyr}$ ago)

The Sun formed $4.6\,\mathrm{Gyr}$ ago — in an already accelerating Universe.

Comoving distance to redshift z

From $\mathrm{d}s^2 = 0$ to an integral over z

incoming light ray: $a\,\mathrm{d}\chi = -c\,\mathrm{d}t$

$$a = \frac{1}{1+z}\colon\quad \mathrm{d}t = \frac{\mathrm{d}a}{aH} = -\frac{\mathrm{d}z}{(1+z)H}$$

$$\Rightarrow\quad \mathrm{d}\chi = -\frac{c\,\mathrm{d}t}{a} = \frac{c\,\mathrm{d}z}{H(z)}$$

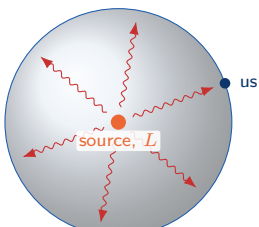
ΛCDM

$\chi(z=1) = 3.4\,\mathrm{Gpc}$; $\chi(z=1090) = 13.9\,\mathrm{Gpc}$: the CMB comes from near the particle horizon, $14.1\,\mathrm{Gpc}$

$$\chi(z) = \int_0^z \frac{c\,\mathrm{d}z'}{H(z')}$$

- χ grows into the past ($\mathrm{d}t < 0$ for $\mathrm{d}z > 0$)
- $H(z)$ from the Friedmann equation $\Rightarrow$ $\chi(z)$ depends on $\Omega_m, \Omega_\Lambda, \Omega_k, \ldots$
- Small z: $\chi \simeq cz/H_0$ — Hubble's law

Luminosity distance: how bright is a distant candle?



area today: $4\pi\chi^2$

- Photons emitted with luminosity L spread over a sphere of area $4\pi\chi^2$ (flat space)
- Each photon arrives with energy reduced by $1+z$
- Photons arrive at a rate reduced by $1+z$ (time dilation)

$$F = \frac{L}{4\pi\chi^2(1+z)^2} \equiv \frac{L}{4\pi d_L^2} \;\Rightarrow\; d_L = (1+z)\,\chi$$

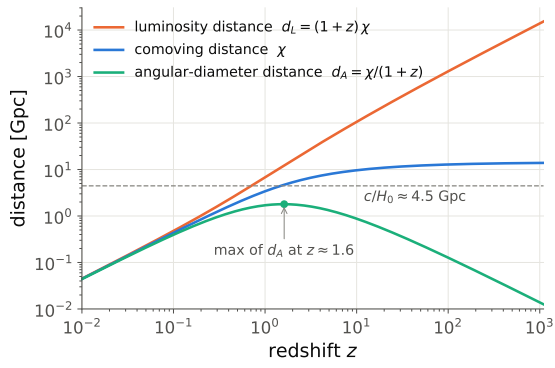
Curved space (R_0: curvature radius, Day 1): $\chi \to R_0 \sin(\chi/R_0)$ or $R_0 \sinh(\chi/R_0)$ — brightness measures curvature too.

Angular-diameter distance: how big does it look?

- Object of proper size D at redshift z, seen under the angle θ
- At emission the comoving separation $\chi\theta$ corresponded to the physical size $D = a(t_e)\,\chi\theta = \chi\theta/(1+z)$

$$d_A \equiv \frac{D}{\theta} = \frac{\chi}{1+z}, \qquad d_L = (1+z)^2\, d_A$$

- d_A has a **maximum** at $z \approx 1.6$: more distant objects look *larger* again



Standard candles: type Ia supernovae

- **Standard candle**: known $L \Rightarrow$ $d_L = \sqrt{L/4\pi F}$ from the measured flux F
- Magnitudes: $m = -2.5 \log_{10} F + \text{const}$ (5 mag = factor 100 fainter); M = the value of m at $10\,\mathrm{pc}$

$$\mu \equiv m - M = 5 \log_{10} \frac{d_L}{10\,\mathrm{pc}}$$

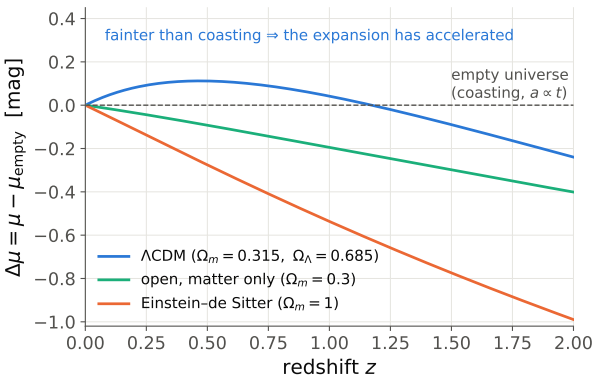
Type Ia supernovae

Thermonuclear explosion of a carbon–oxygen **white dwarf** near the Chandrasekhar mass $1.4 M_\odot$; peak $\sim 10^{10} L_\odot$, seen beyond $z \sim 1$.

Phillips relation: brighter ones fade more slowly $\Rightarrow$ after correction, distances to 5–7 % each.

parallaxes (Gaia): stars in the Milky Way → **Cepheids**, tip of the red-giant branch (**TRGB**) in nearby galaxies → **SNe Ia** in the same galaxies: calibrate M → **SNe Ia in the Hubble flow** $z \sim 0.02$–1

1998: the accelerating Universe



computed for this course (Planck 2018 ΛCDM)

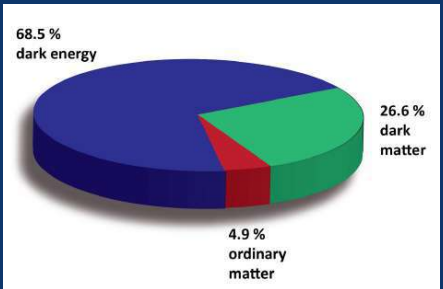
- Riess et al. 1998, Perlmutter et al. 1999: distant SNe Ia are **fainter** than expected
- Plotted: μ minus that of an empty, coasting Universe ($a \propto t$); at $z = 0.5$ ΛCDM is 0.1 mag fainter, EdS 0.3 mag brighter
- Best fit $\Omega_m \approx 0.3$, $\Omega_\Lambda \approx 0.7$ (Nobel 2011)

Acceleration: expansion **slower in the past** ⇒ more time, longer light paths, fainter sources.

Hour 2

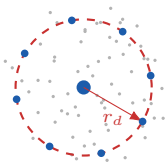
What is the Universe made of?

- Standard rulers: independent probes agree
- Two tensions: H_0 and evolving dark energy?
- Weighing galaxies and clusters: dark matter
- What dark matter must be
- Dark energy as vacuum energy — and the worst prediction in physics



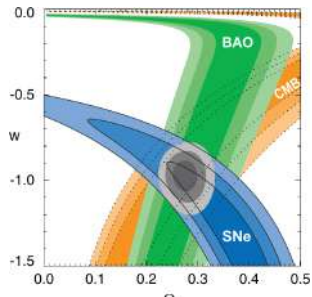
ESA/Planck

Standard rulers and concordance



Baryon acoustic oscillations (BAO): excess of galaxy pairs at the separation r_d

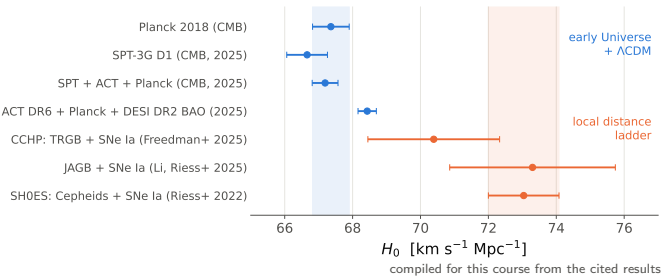
- Sound waves in the early plasma imprint a comoving scale $r_d \approx 147\,\mathrm{Mpc}$ (Hour 4)
- At redshift z it subtends the angle $r_d/D_M(z)$ ($D_M = \chi$ if flat) and spans $\Delta z = r_d H(z)/c$ along the line of sight
- DESI (2025): 1–2 % per redshift bin up to $z \approx 2.3$
- CMB: the same ruler at last scattering (r_s at $z_* \approx 1090$) subtends $\theta_* = r_s/D_M(z_*) = 0.596^\circ \Rightarrow$ flat (Hour 4)



Kowalski et al. 2008, via Gorbunov & Rubakov, Fig. 13.6

▶ Candles and rulers constrain *different* parameter combinations: they meet at $\Omega_m \approx 0.3$, $\Omega_\Lambda \approx 0.7$, $w \approx -1$.

Two tensions: H_0 and evolving dark energy



compiled for this course from the cited results

Hubble tension

Early Universe (ruler r_d) + ΛCDM: 67–$68.5\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$
Distance ladder: SH0ES (Cepheids) 73.0 ± 1.0, CCHP (TRGB) 70.4 ± 1.9
Planck vs SH0ES: $\approx 5\sigma$

Is dark energy constant?

Fit $w(a) = w_0 + w_a(1-a)$: DESI DR2 (2025) prefers $(w_0, w_a) \neq (-1, 0)$ at 3.1σ (BAO+CMB), 2.8–4.2σ with supernovae; recalibrated DES SNe + DESI + CMB (2025): 3.2σ, $w_0 = -0.80 \pm 0.05$, $w_a = -0.72 \pm 0.21$

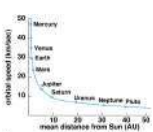
- Systematics — or physics beyond ΛCDM? **Unresolved (2026)**.

Weighing with orbits: flat rotation curves

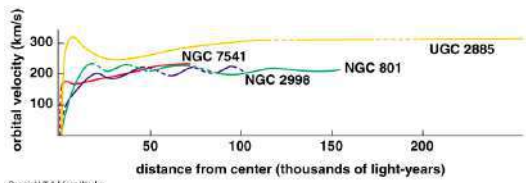
Circular orbit of a star or a gas cloud

$$\frac{mv^2}{r} = \frac{GM(r)\,m}{r^2}$$

$$v = \sqrt{\frac{GM(r)}{r}}, \qquad M(r) = \frac{v^2 r}{G}$$



Solar System:
$v \propto r^{-1/2}$
("Keplerian")

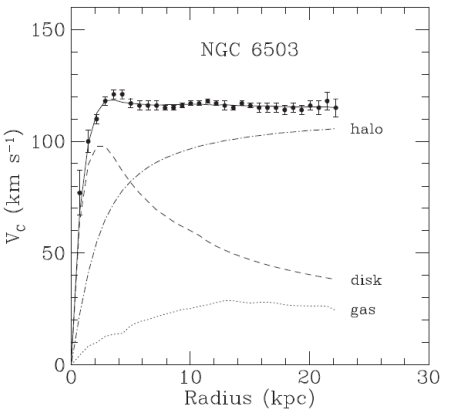


© Addison Wesley

- Orbital speeds weigh the mass *inside* the orbit — visible or not
- Rubin & Ford (1970s, Doppler shifts across the disc): v **flat** out to the edge of the visible disc; 21-cm radio data: flat far beyond

Expected $v \propto r^{-1/2}$ beyond the stars; observed $v \approx$ const.

Dark-matter haloes



Begeman, Broeils & Sanders 1991

NGC 6503: disc and gas cannot give the flat part — a dark halo dominates at large r

Flat curve, $v = \text{const}$

$$M(r) = \frac{v^2 r}{G} \propto r, \qquad \frac{\mathrm{d}M}{\mathrm{d}r} = 4\pi r^2 \rho$$

$$\Rightarrow \quad \rho(r) = \frac{v^2}{4\pi G\, r^2}$$

The Milky Way

$v \approx 230\,\mathrm{km\,s^{-1}}$ at the Sun's orbit, $r_\odot \approx 8.2\,\mathrm{kpc}$:
$M(< r_\odot) = v^2 r_\odot / G \approx 1 \times 10^{11} M_\odot$;
whole halo ($\gtrsim 100\,\mathrm{kpc}$): $\sim 10^{12} M_\odot$, but stars + gas only a few $\times 10^{10} M_\odot$;
near the Sun $\rho_{\mathrm{DM}} \approx 0.3$–$0.4\,\mathrm{GeV\,cm^{-3}}$

Weighing clusters: the virial theorem

Bound, stationary system of galaxies

$\mathcal{G} \equiv \sum_i m_i \vec{r}_i \cdot \vec{v}_i$

$\frac{\mathrm{d}\mathcal{G}}{\mathrm{d}t} = \sum_i m_i v_i^2 + \sum_i \vec{r}_i \cdot \vec{F}_i = 2K + U$

gravity ($U \propto 1/r$): $\sum_i \vec{r}_i \cdot \vec{F}_i = U$

bound, stationary: time average $\langle \mathrm{d}\mathcal{G}/\mathrm{d}t \rangle = 0$

$$2\langle K \rangle + \langle U \rangle = 0$$

Virial mass

Uniform sphere: $U = -\frac{3}{5} G M^2/R$;
line-of-sight dispersion σ: $K = \frac{3}{2} M \sigma^2$

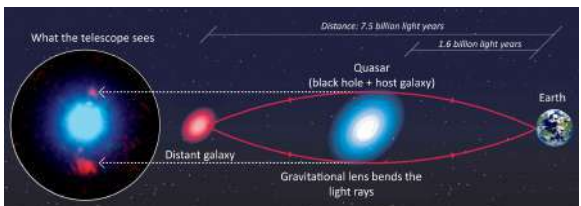
$$\Rightarrow \; M \simeq \frac{5\,\sigma^2 R}{G}$$

Coma cluster

$\sigma \approx 1000\,\mathrm{km\,s^{-1}}$, $R \approx 2$–$3\,\mathrm{Mpc}$:
$M \approx (2.3\text{–}3.5) \times 10^{15} M_\odot$,
mass/light ~ 300

- Zwicky 1933: **"dunkle Materie"**; Lundmark (Lund) 1930: galaxies weigh far more than their light.

Mass bends light: gravitational lensing



From the author's talk "Shedding light on the dark sides of the Universe"

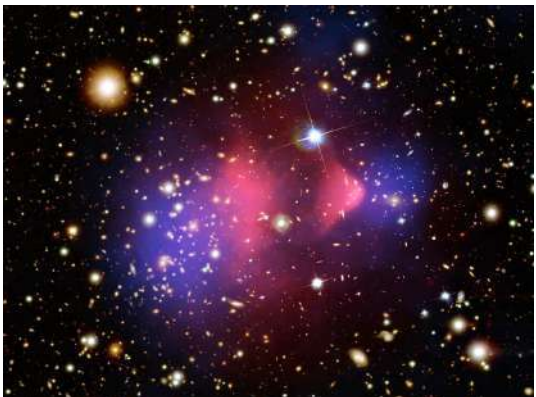
$$\alpha = \frac{4GM}{c^2\, b} \quad \text{(GR)}$$

Newton (a particle with $v = c$ passing at distance b): $\Delta p_\perp / p = 2GM/(c^2 b)$ — GR doubles it.

- Sun: $\alpha = 1.75''$ (Eddington 1919)
- **Strong**: arcs, rings, multiple images; **weak**: $\sim 1\,\%$ distortions of many galaxies

Lensing weighs the **total** mass, whatever it is made of.

The Bullet cluster: dark matter caught in the act



NASA/CXC/CfA/M. Markevitch et al.

- Two clusters collided (1E 0657–558)
- Pink: X-ray gas at $\sim 10^8\,\mathrm{K}$ — most of the ordinary matter — slowed down by the collision
- Blue: total mass from lensing — passed through, with the galaxies
- Offset significant at 8σ (Clowe et al. 2006)

Most mass is **collisionless** and not where most of the baryons are: $\sigma/m \lesssim 1\,\mathrm{cm}^2/\mathrm{g}$.

What dark matter must be

- Light elements, CMB: $\Omega_b h^2 = 0.0224$; CMB, BAO: $\Omega_m h^2 = 0.143$ (Hour 4)

$$1 - \frac{\Omega_b}{\Omega_m} \approx 0.84$$ is non-baryonic

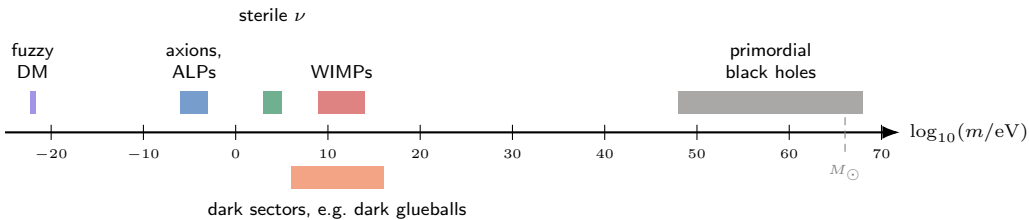
Required properties

neutral, (nearly) collisionless; **cold** (slow when structures began to grow); stable over $\gg 13.8\,\mathrm{Gyr}$; $\Omega_c h^2 = 0.120$

- **Not neutrinos**: $\Omega_\nu h^2 = \sum m_\nu / 93\,\mathrm{eV} < 10^{-3}$, and they are *hot*: they stream over $\sim 0.5\,\mathrm{Gpc}\,(1\,\mathrm{eV}/m_\nu)^{1/2}$, erasing galaxies
- **Not dark baryons** (dead stars, planets): the baryon budget is fixed by BBN and CMB; microlensing finds too few compact objects

No Standard-Model particle fits: dark matter is **new physics**.

Candidates and searches



Direct detection

Halo particles hit nuclei underground. LZ (2024): $\sigma_{\mathrm{SI}} < 2.2 \times 10^{-48}\,\mathrm{cm}^2$ at $40\,\mathrm{GeV}$

Indirect detection

Annihilation or decay products: γ rays, e^+, $\bar{p}$, ν

Colliders

Produced at the LHC: missing transverse energy

- WIMPs are thermal relics (Hour 5). So far **no detection**.

Dark energy = vacuum energy?

- The vacuum must look the same to all inertial observers:

$$T^{\mathrm{vac}}_{\mu\nu} = \rho_{\mathrm{vac}}\, g_{\mu\nu} \;\Rightarrow\; p_{\mathrm{vac}} = -\rho_{\mathrm{vac}}$$

- Exactly Einstein's cosmological constant:

$$G_{\mu\nu} - \Lambda\, g_{\mu\nu} = 8\pi G\, T_{\mu\nu}, \qquad \Lambda = 8\pi G \rho_{\mathrm{vac}}$$

Observed

$\rho_\Lambda = \Omega_\Lambda\, \rho_{c,0} \approx 2.5 \times 10^{-47}\,\mathrm{GeV}^4$
$\approx (2.24\,\mathrm{meV})^4$
$\approx 3.3 \times 10^{-6}\,\mathrm{GeV\,cm^{-3}}$:
about 3.5 proton masses per m^3

Signature $(+,-,-,-)$; with $(-,+,+,+)$ the Λ term appears with the opposite sign.

The vacuum catastrophe

Zero-point energy of a field ($\hbar = c = 1$)

Each mode $\vec{k}$ is an oscillator with $\omega_k = |\vec{k}|$ and ground-state energy $\frac{1}{2}\omega_k$ (uncertainty principle); $\mathrm{d}^3k/(2\pi)^3$ modes per volume:

$$\rho_{\mathrm{vac}} = \int^{k_{\mathrm{max}}} \frac{\mathrm{d}^3k}{(2\pi)^3}\,\frac{k}{2} = \frac{1}{4\pi^2}\int_0^{k_{\mathrm{max}}} k^3\,\mathrm{d}k$$

$$\rho_{\mathrm{vac}} = \frac{k_{\mathrm{max}}^4}{16\pi^2}$$

source	ρ_{vac} [GeV^4]	$\rho_{\mathrm{vac}}/\rho_\Lambda$
cutoff at M_{Pl}	10^{74}	10^{121}
electroweak: $m_H^2 v^2/8$	10^{8}	10^{55}
QCD condensates	10^{-3}–10^{-2}	10^{44}
observed ρ_Λ	2.5×10^{-47}	1

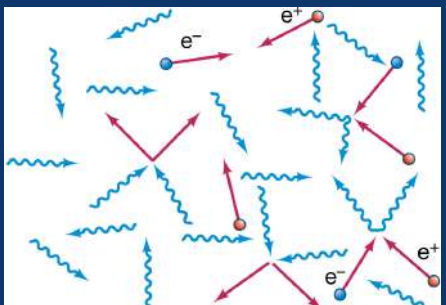
$m_H = 125\,\mathrm{GeV}$, $v = 246\,\mathrm{GeV}$: Higgs mass and vacuum value (Hour 3)

"The worst prediction in physics": nothing known cancels 50 or more orders of magnitude.

Hour 3

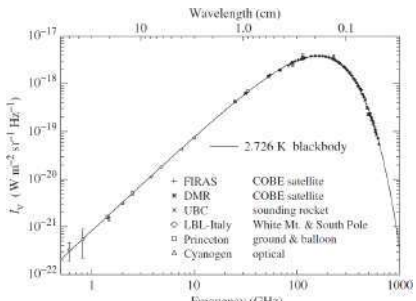
The hot early Universe

- Evidence for a hot, thermal early Universe
- Hot gases: $\rho \propto T^4$, degrees of freedom, entropy
- The clock: temperature $\leftrightarrow$ time; freeze-out
- The vacuum changes: the Higgs field, phase transitions, gravitational waves
- Why is there more matter than antimatter?



From the author's talk "Shedding light on the dark sides of the Universe"

The CMB: a perfect blackbody, preserved by the expansion



COBE/FIRAS and earlier data (via Gorbunov & Rubakov 2011, Fig. 1.3)

Penzias & Wilson 1965; COBE/FIRAS 1990:
$T_0 = 2.7255\,\mathrm{K}$

From here on: $\hbar = c = k_B = 1$

Free particles in an expanding Universe

momenta redshift, $p \propto 1/a$, and the number of particles in a comoving volume is conserved:

$$f(p,t) = f_i\Big(p\,\frac{a(t)}{a_i}\Big)$$

photons: $f_i = 1/(\mathrm{e}^{p/T_i} - 1) \Rightarrow$
$f = 1/(\mathrm{e}^{p/T} - 1)$ with $T = T_i\,a_i/a$

A Planck spectrum stays a Planck spectrum, with $T \propto 1/a$.

- A perfect blackbody needs **thermal equilibrium**: the early Universe was a hot, opaque plasma.

Equilibrium distributions and the relativistic gas

$$f(p) = \frac{1}{\mathrm{e}^{(E-\mu)/T} \mp 1}$$

$E = \sqrt{p^2 + m^2}$; $-$: bosons, $+$: fermions

$$n = g\int \frac{\mathrm{d}^3p}{(2\pi)^3}\, f, \qquad \rho = g\int \frac{\mathrm{d}^3p}{(2\pi)^3}\, E\, f,$$

$$p = g\int \frac{\mathrm{d}^3p}{(2\pi)^3}\, \frac{|\vec{p}|^2}{3E}\, f \quad \text{(pressure)}$$

- g = internal states: γ 2, e^- 2, ν 1, quark 2×3, gluon 2×8
- $\mu \simeq 0$: $\mu/T \sim 10^{-9}$ (baryons, charged leptons); $|\mu_\nu|/T \lesssim 0.02$

$T \gg m$: $E = p$, $\mathrm{d}^3p = 4\pi p^2 \mathrm{d}p$, $x = p/T$

$$\rho = \frac{g}{2\pi^2}\int_0^\infty \frac{p^3 \mathrm{d}p}{\mathrm{e}^{p/T} - 1} = \frac{gT^4}{2\pi^2} \underbrace{\int_0^\infty \frac{x^3 \mathrm{d}x}{\mathrm{e}^x - 1}}_{\pi^4/15}$$

$$\rho = g\frac{\pi^2}{30}T^4 \left(\times \tfrac{7}{8}\right)$$

$$n = g\frac{\zeta(3)}{\pi^2}T^3 \left(\times \tfrac{3}{4}\right), \qquad p = \frac{\rho}{3}$$

brackets: fermions; $\zeta(3) = 1.202$; mean energy $\rho/n = 2.70\,T$ (bosons), $3.15\,T$ (fermions)

Non-relativistic gas ($T \ll m$): Boltzmann suppression

- $E \simeq m + \dfrac{p^2}{2m}$, $f \simeq \mathrm{e}^{-(E-\mu)/T}$ (classical)
- Gaussian integral over $\vec{p}$:

$$n = g\left(\frac{mT}{2\pi}\right)^{3/2} \mathrm{e}^{-(m-\mu)/T}$$

- $\rho \simeq m\,n$, $p = nT \ll \rho$ ("dust", $w \simeq 0$)

Electrons without an asymmetry

$$\frac{n_{e^-}}{n_\gamma} = \frac{\pi^2}{\zeta(3)}\left(\frac{m_e}{2\pi T}\right)^{3/2} \mathrm{e}^{-m_e/T}$$

$T = m_e/10$: $\approx 10^{-3}$
$T = m_e/30$: $\approx 10^{-11}$

▶ Once $T < m$, particle–antiparticle pairs annihilate away **exponentially fast** — unless an asymmetry protects them.

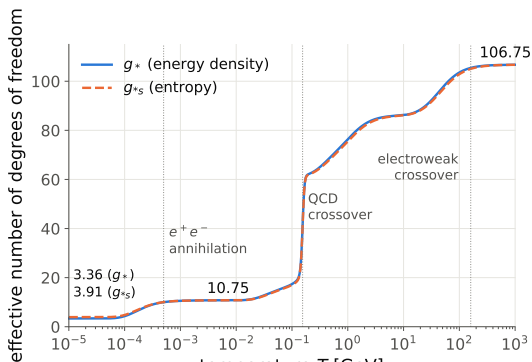
Counting degrees of freedom: $g_*(T)$

$$\rho = \frac{\pi^2}{30}\, g_*(T)\, T^4$$

$$g_* = \sum_{\text{bosons}} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{\text{fermions}} g_i \left(\frac{T_i}{T}\right)^4$$

sum over *relativistic* species ($m_i \ll T$); entropy version g_{*s}: powers 3

bosons: γ 2, $W^\pm, Z$ 9, gluons 16, H 1	28
fermions: quarks 72, e, μ, τ 12, ν 6	90
Standard Model: $28 + \frac{7}{8} \cdot 90$	**106.75**



$1\,\text{MeV}$ (γ, $e^\pm$, 3ν): $g_* = 2 + \frac{7}{8}(4+6) = 10.75$;
today: $g_* = 3.36$, $g_{*s} = 3.91$ ($T_\nu < T_\gamma$, Hour 4)

Entropy is conserved in the expanding Universe

Comoving volume $V = a^3$, $\mu = 0$

Expansion is adiabatic (continuity equation, Day 1):
$\mathrm{d}(\rho V) = -p\,\mathrm{d}V$

First law: $T\,\mathrm{d}S = \mathrm{d}(\rho V) + p\,\mathrm{d}V = 0$

Euler relation: $\rho V = TS - pV$

$\Rightarrow$ entropy density $s \equiv \dfrac{S}{V} = \dfrac{\rho + p}{T}$

- Relativistic gas: $s = \frac{4}{3}\rho/T$; holds while all species stay in equilibrium

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(s\,a^3\right) = 0, \qquad s = \frac{2\pi^2}{45}\,g_{*s}\,T^3$$

Consequence

$$T \propto g_{*s}^{-1/3}\,a^{-1}$$

When a species annihilates, g_{*s} drops and its entropy goes to the rest of the plasma: T falls *slower* than $1/a$ — the plasma is **heated** (e.g. $e^+e^- \to \gamma\gamma$, Hour 4)

Photons, baryons and entropy today

the most important small number: η

Today: $T_0 = 2.7255\,\mathrm{K} = 2.35 \times 10^{-4}\,\mathrm{eV}$

$$n_\gamma = \frac{2\zeta(3)}{\pi^2} T_0^3 = 411\,\mathrm{cm}^{-3}$$

$$\rho_\gamma = \frac{\pi^2}{15} T_0^4 = 0.26\,\mathrm{eV\,cm}^{-3}$$

$$s_0 = \frac{2\pi^2}{45}\, 3.91\, T_0^3 = 2.9 \times 10^3\,\mathrm{cm}^{-3}$$

Baryons

$n_b = \Omega_b\, \rho_{c,0}/m_p = 2.5 \times 10^{-7}\,\mathrm{cm}^{-3}$ (about 0.25 per m^3)

$$\eta \equiv \frac{n_b}{n_\gamma} = 6.1 \times 10^{-10}$$

$$\frac{n_b}{s} = 8.7 \times 10^{-11}$$

n_b/s stays constant even when species annihilate

▶ **1.6 billion photons per baryon** — this ratio controls nucleosynthesis and recombination (Hour 4).

The radiation-era clock

how old was the Universe at temperature T?

Derivation

$$H^2 = \frac{8\pi}{3}\frac{\rho}{M_{\rm Pl}^2}, \quad \rho = \frac{\pi^2}{30} g_* T^4 \quad \Rightarrow \quad H = \sqrt{\frac{8\pi^3 g_*}{90}}\frac{T^2}{M_{\rm Pl}} = 1.66\sqrt{g_*}\,\frac{T^2}{M_{\rm Pl}}$$

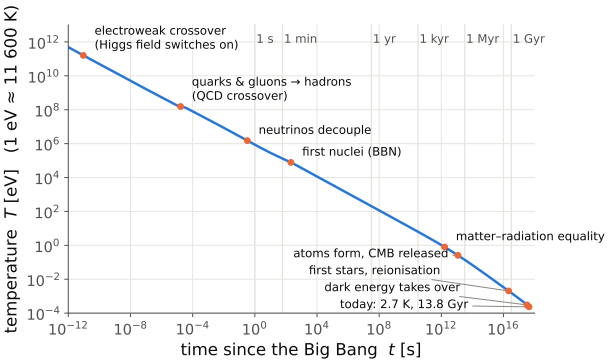
Radiation era: $a \propto t^{1/2}$, $H = 1/(2t) \;\Rightarrow\; t = \dfrac{1}{2H} = \dfrac{0.30\, M_{\rm Pl}}{\sqrt{g_*}\, T^2}$

$$t \approx \frac{2.4\,\mathrm{s}}{\sqrt{g_*}} \left(\frac{1\,\mathrm{MeV}}{T}\right)^2$$

$M_{\rm Pl} = G^{-1/2} = 1.22 \times 10^{19}\,\mathrm{GeV}$;
$\hbar M_{\rm Pl}/(1\,\mathrm{MeV})^2 = 8.0\,\mathrm{s}$

Rule of thumb: $t/(1\,\mathrm{s}) \sim (1\,\mathrm{MeV}/T)^2$

The thermal history of the Universe



event	T	t
electroweak crossover	$160\,\mathrm{GeV}$	$10^{-11}\,\mathrm{s}$
QCD crossover	$155\,\mathrm{MeV}$	$2\times10^{-5}\,\mathrm{s}$
ν decoupling	1–$2\,\mathrm{MeV}$	0.2–$0.7\,\mathrm{s}$
nucleosynthesis	$80\,\mathrm{keV}$	$3.5\,\mathrm{min}$
recombination	$0.3\,\mathrm{eV}$	$0.3\,\mathrm{Myr}$

$T \leftrightarrow z$: $1+z=(g_{*s}/3.91)^{1/3}\,T/T_0$

$z\approx1100$: $3000\,\mathrm{K}$, atoms form

$z\approx6\times10^9$: $1\,\mathrm{MeV}\approx10^{10}\,\mathrm{K}$, no nuclei

Equilibrium versus expansion: freeze-out

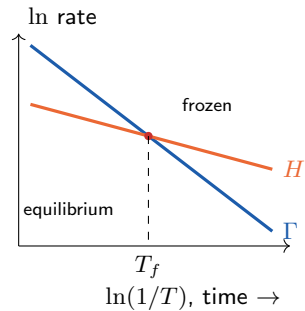
- Rate per particle $\Gamma = n\langle\sigma v\rangle$ (σ: cross section, v: speed); collisions per Hubble time $\sim \Gamma/H$
- $\Gamma \gg H$: **equilibrium**; $\Gamma \ll H$: **decoupled**, abundances frozen

freeze-out: $\Gamma(T_f) \simeq H(T_f)$

Why the early Universe was a thermal soup

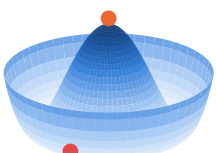
QED at $T \gg m_e$: $\sigma \sim \alpha^2/T^2$, $n \sim T^3 \Rightarrow \Gamma \sim \alpha^2 T$

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_{\rm Pl}}{1.66\sqrt{g_*}\,T} \gg 1 \quad \text{for} \quad T \ll 10^{14}\,\text{GeV}$$



Relics: decoupled while relativistic (CMB photons, neutrinos) or frozen out non-relativistic (n/p, free electrons, dark matter)

The Higgs vacuum: the origin of mass



computed for this course

$$V(\phi) = -\mu^2|\phi|^2 + \lambda|\phi|^4$$

($\mu^2 > 0$: the Higgs mass parameter, not a chemical potential)

Minimum and the Higgs boson

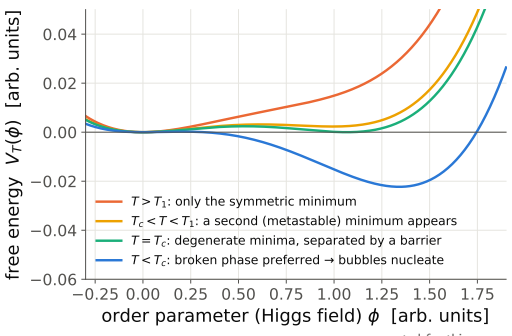
$$\frac{\partial V}{\partial|\phi|^2} = -\mu^2 + 2\lambda|\phi|^2 = 0 \;\Rightarrow\; |\langle\phi\rangle|^2 = \frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}$$

oscillations about it, $|\phi|^2 = \frac{1}{2}(v+h)^2$: $m_H^2 = -\mu^2 + 3\lambda v^2 = 2\lambda v^2$

The vacuum is filled with a **Higgs condensate**, $v = 246\,\mathrm{GeV}$; the Higgs boson has $m_H = 125\,\mathrm{GeV}$.

Masses $\propto v$: $m_W = \frac{1}{2} g_W v$, $m_f = y_f v/\sqrt{2}$ (weak coupling $g_W \approx 0.65$, Yukawa couplings y_f); photon and gluons stay massless.

Symmetry restoration at high temperature



- The hot plasma gives the Higgs a thermal mass^2 $\propto T^2$:

$$V_T \simeq D(T^2 - \bar{T}^2)\phi^2 - ET\phi^3 + \frac{\lambda}{4}\phi^4$$

($\bar{T}$: below it $\phi = 0$ is unstable; D, E: one-loop thermal corrections)

- $T \gg \bar{T}$: minimum at $\phi = 0$ — W, Z, quarks and leptons were **massless**
- **First order**: $\langle\phi\rangle$ jumps (bubbles); **crossover**: smooth change

SM (lattice): electroweak **crossover** at $160\,\mathrm{GeV}$ (10^{-11} s); QCD crossover at $155\,\mathrm{MeV}$ ($20\,\mu$s): hadrons form. First order needs new physics.

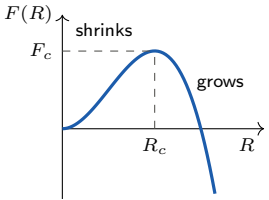
First-order transitions: bubble nucleation

Free energy of a bubble of radius R

$$F(R) = \underbrace{4\pi R^2 \sigma_{\rm w}}_{\text{wall}} - \underbrace{\tfrac{4\pi}{3} R^3 \Delta V}_{\text{volume}}$$

$$F'(R) = 8\pi R \sigma_{\rm w} - 4\pi R^2 \Delta V = 0 \;\Rightarrow\; R_c = \frac{2\sigma_{\rm w}}{\Delta V}$$

$$F_c = \frac{16\pi\sigma_{\rm w}^3}{\Delta V^2} - \frac{32\pi\sigma_{\rm w}^3}{3\Delta V^2} = \frac{16\pi}{3}\,\frac{\sigma_{\rm w}^3}{\Delta V^2}$$



$\sigma_{\rm w}$: wall tension; ΔV: free-energy difference of the phases ($\to 0$ at the critical temperature T_c: supercooling)

Nucleation rate per volume $\Gamma_{\rm nuc} \simeq T^4 \mathrm{e}^{-F_c/T}$; done when $\Gamma_{\rm nuc} \sim H^4$ (one bubble per Hubble volume per Hubble time):

$$\frac{F_c}{T} \simeq 4 \ln \frac{M_{\rm Pl}}{1.66\sqrt{g_*}\,T} \approx 150 \quad \text{at } T \sim 100\,\mathrm{GeV}$$

Gravitational waves from a first-order transition

- Bubble collisions, sound waves, turbulence $\Rightarrow$ a **stochastic GW background**
- Emitted at the transition ($T = T_*$) with $f_* \sim \beta$, β^{-1} being its duration ($\beta/H_* \sim 10$–10^3); redshifted: $f_0 = f_*\, a_*/a_0$

a_*/a_0 from entropy conservation

$$H_* = 1.66\sqrt{g_*}\,T_*^2/M_{\rm Pl}, \quad \frac{a_*}{a_0} = \left(\frac{g_{*s,0}}{g_{*s}}\right)^{1/3}\frac{T_0}{T_*}$$

$$\Rightarrow \; H_*\frac{a_*}{a_0} = 1.66\cdot 3.91^{1/3}\, g_*^{1/6}\,\frac{T_* T_0}{M_{\rm Pl}}$$

$$f_0 \simeq \frac{\beta}{H_*}\, H_*\frac{a_*}{a_0} \approx 1.65\times 10^{-5}\,{\rm Hz}\,\frac{\beta}{H_*}\left(\frac{T_*}{100\,{\rm GeV}}\right)\left(\frac{g_*}{100}\right)^{1/6}$$

▶ A **thermometer**: QCD-scale transitions $\rightarrow$ nHz (pulsar timing arrays; a background was reported in 2023); electroweak scale $\rightarrow$ mHz (LISA, $\sim$2035).

The matter–antimatter asymmetry

- No antimatter domains seen; $\eta = 6.1 \times 10^{-10}$: a few extra quarks per **billion** quark–antiquark pairs in the hot plasma

Baryon-symmetric Universe?

$B\bar{B}$ annihilation ($\langle \sigma v \rangle \sim 1/m_\pi^2$) freezes out at $T_f \approx 20\,\mathrm{MeV}$: $\Omega_B = \Omega_{\bar{B}} \sim 10^{-11}$ — 10^9 times too little

Sakharov's conditions (1967)

(1) B violation; (2) C and CP violation; (3) departure from equilibrium

Standard Model: (1) ✓ sphalerons violate $B + L$ for $T \gtrsim 100\,\mathrm{GeV}$; (2) × CKM CP violation far too small; (3) × a crossover stays in equilibrium

New physics needed

Leptogenesis: CP-violating decays of heavy right-handed neutrinos (they also explain tiny ν masses); sphalerons convert $L \to B$

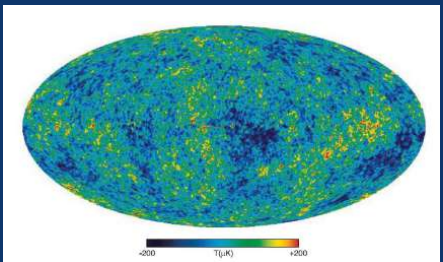
Electroweak baryogenesis: strong first-order transition + new CP violation at the bubble walls — testable with GWs

▶ The very existence of matter is evidence for physics **beyond the Standard Model**.

Hour 4

The first three minutes and the cosmic microwave background

- Neutrinos decouple; relic neutrinos today
- How the light elements were made
- When and why atoms formed; the CMB is released
- Reading the CMB sky: acoustic peaks, a ruler, six numbers



NASA/WMAP

Neutrino decoupling

Weak rates by dimensional analysis

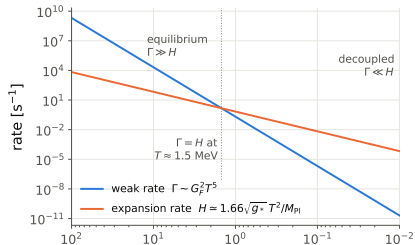
$E \ll m_W$: Fermi contact interaction,
$G_F/\sqrt{2} = g_W^2/8m_W^2$, $G_F = 1.17 \times 10^{-5}\,\mathrm{GeV}^{-2}$

$[\sigma] = \mathrm{GeV}^{-2} = [G_F]$, the only other scale is E:
$\sigma \sim G_F^2 E^2$

plasma: $E \sim T$, $n \sim T^3$, $v \simeq 1$: $\Gamma = n\sigma v \sim G_F^2 T^5$

$$\frac{\Gamma}{H} \simeq \frac{G_F^2 T^5 M_{\mathrm{Pl}}}{1.66\sqrt{g_*}\,T^2} \simeq \left(\frac{T}{1.5\,\mathrm{MeV}}\right)^3$$

($g_* = 10.75$; $\Gamma \propto T^5$ falls much faster than $H \propto T^2$)



computed for this course

Neutrinos decouple

at $T \approx 1$–$2\,\mathrm{MeV}$, $t \approx 0.2$–$0.7\,\mathrm{s}$;
then stream freely with $T_\nu \propto 1/a$

An order-of-magnitude estimate; detailed calculations: $\approx 2\,\mathrm{MeV}$ (ν_e), $\approx 3\,\mathrm{MeV}$ ($\nu_{\mu,\tau}$) (Dolgov, Phys. Rept. 370 (2002) 333)

e^+e^- annihilation heats the photons; relic neutrinos

Two sealed boxes: $(\gamma, e^\pm)$ and (ν)

$T < m_e$: $e^+e^- \to \gamma\gamma$; the entropy goes to the photons only

$g_{*s}(\gamma, e^\pm) = 2 + \frac{7}{8} \cdot 4 = \frac{11}{2} \;\to\; 2$

$g_{*s}(aT_\gamma)^3 = \text{const} \;\Rightarrow\; aT_\gamma$ grows by $\left(\frac{11}{4}\right)^{1/3}$; neutrinos: $aT_\nu = \text{const}$

$$\frac{T_\nu}{T_\gamma} = \left(\frac{4}{11}\right)^{1/3} = 0.714$$

Relic neutrinos today ($C\nu B$)

$T_{\nu,0} = 0.714 \times 2.7255\,\mathrm{K} = 1.95\,\mathrm{K}$

$n_\nu = \frac{3}{4} \cdot 2 \cdot \frac{\zeta(3)}{\pi^2} T_\nu^3 = \frac{3}{11}\, n_\gamma = 112\,\mathrm{cm}^{-3}$ per flavour

radiation: $\rho_r = \rho_\gamma \left[1 + \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} N_{\mathrm{eff}}\right]$,

$N_{\mathrm{eff}}^{\mathrm{SM}} = 3.044$ (Planck: 2.99 ± 0.17)

$\Rightarrow \Omega_r = 9.2 \times 10^{-5}$

massive today: $\Omega_\nu h^2 = \sum m_\nu / 93.1\,\mathrm{eV}$

Oscillations: $\sum m_\nu \geq 0.059\,\mathrm{eV}$; DESI+CMB: $< 0.064\,\mathrm{eV}$; KATRIN (2025): $m_\nu < 0.45\,\mathrm{eV}$

Big Bang nucleosynthesis: neutrons and protons in equilibrium

- For a few minutes the whole Universe fused nuclei: ^{4}He ($\approx 25\,\%$ by mass), D, ^{3}He, ^{7}Li
- Known physics; **one free parameter**: $\eta = n_b/n_\gamma$
- Gamow, Alpher, Herman (1948): a hot Universe makes helium — and leaves relic radiation of a few kelvin

- $n + \nu_e \leftrightarrow p + e^-$, $n + e^+ \leftrightarrow p + \bar{\nu}_e$: $\mu_n + \mu_\nu = \mu_p + \mu_e$, with $\mu_e, \mu_\nu \simeq 0 \Rightarrow \mu_n \simeq \mu_p$

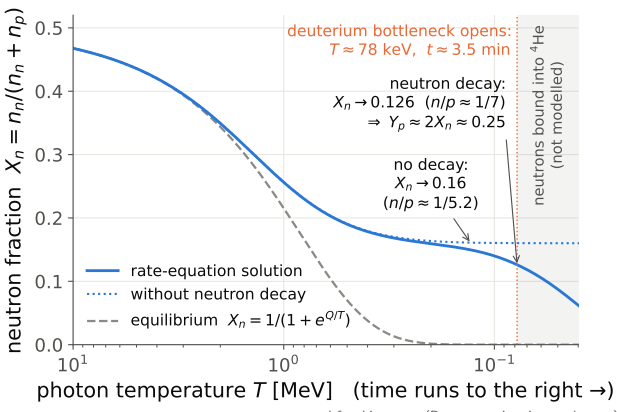
Non-relativistic densities

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{3/2} \mathrm{e}^{-(m_n - m_p)/T}\, \mathrm{e}^{(\mu_n - \mu_p)/T}$$

$$\frac{n}{p} = \mathrm{e}^{-Q/T}, \qquad Q = m_n - m_p = 1.293\,\mathrm{MeV}$$

▶ At $T \gg Q$ as many neutrons as protons; as T drops neutrons become rarer — while the weak reactions keep up.

Neutron freeze-out and decay



computed for this course (Born-approximation weak rates)

T	$\Gamma_{n\to p}$	H
$1\,\mathrm{MeV}$	$1.6\,\mathrm{s}^{-1}$	$0.68\,\mathrm{s}^{-1}$
$0.7\,\mathrm{MeV}$	$0.33\,\mathrm{s}^{-1}$	$0.33\,\mathrm{s}^{-1}$

- $\Gamma \simeq H$ at $T_f \approx 0.7$–$0.8\,\mathrm{MeV}$, $t \approx 1\,\mathrm{s}$: $n/p \simeq \mathrm{e}^{-1.293/0.75} \approx \frac{1}{5.5}$ (full: $\frac{1}{5.2}$, $X_n \approx 0.16$)
- Then free decay, $\tau_n = 878.4\,\mathrm{s}$, until D survives ($78\,\mathrm{keV}$, $3.5\,\mathrm{min}$): $X_n \to 0.16 \times 0.79 \approx 0.126$, $n/p \approx 1/7$

The deuterium bottleneck

- First step: $p + n \leftrightarrow \mathrm{D} + \gamma$, $B_\mathrm{D} = 2.22\,\mathrm{MeV}$
- D survives only when the photons able to break it, $\eta^{-1}\mathrm{e}^{-B_\mathrm{D}/T}$ per baryon, become fewer than one:

$$T \sim \frac{B_\mathrm{D}}{\ln \eta^{-1}} = \frac{2.22\,\mathrm{MeV}}{21} \approx 0.1\,\mathrm{MeV}$$

Saha equation for deuterium

$n_i = g_i (m_i T/2\pi)^{3/2} \mathrm{e}^{(\mu_i - m_i)/T}$, $\mu_\mathrm{D} = \mu_n + \mu_p$, $m_n + m_p - m_\mathrm{D} = B_\mathrm{D}$, $g_\mathrm{D} = 3$, $g_{n,p} = 2$, $m_\mathrm{D} \simeq 2m_N$ (m_N: nucleon mass):

$$\frac{n_\mathrm{D}}{n_n n_p} = \frac{g_\mathrm{D}}{g_n g_p} \Big(\frac{4\pi}{m_N T}\Big)^{3/2} \mathrm{e}^{B_\mathrm{D}/T}, \quad n_b = \eta\,\frac{2\zeta(3)}{\pi^2} T^3$$

$$\frac{X_\mathrm{D}}{X_n X_p} \simeq 8.1\,\eta \Big(\frac{T}{m_N}\Big)^{3/2} \mathrm{e}^{B_\mathrm{D}/T} \qquad (X_i = n_i/n_b)$$

▶ $= 1$ at $66\,\mathrm{keV}$; full kinetics: 70–80 keV, $t \approx 3.5$–4.5 min — **too many photons**.

Helium-4: a quarter of all baryons

nucleus	B [MeV]	B/A [MeV]
D	2.22	1.11
^{3}H	8.48	2.83
^{3}He	7.72	2.57
^{4}He	28.30	**7.07**

- Once D survives, the chain D $\to$ ^{3}H, ^{3}He $\to$ ^{4}He runs within minutes
- No stable nuclei with $A = 5$ or 8: BBN stops; traces of ^{7}Li

All surviving neutrons end in ^{4}He ($2p + 2n$)

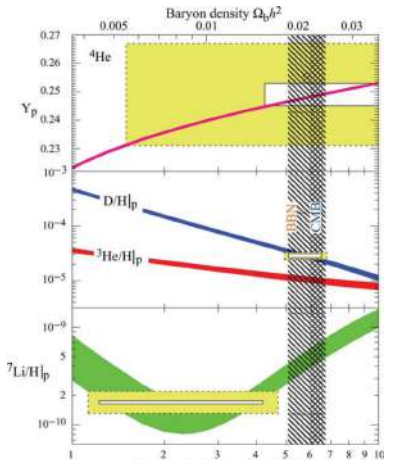
$n_{\rm He} = \frac{1}{2} n_n$, $m_{\rm He} \simeq 4 m_N$:

$$Y_p \equiv \frac{\rho_{\rm He}}{\rho_b} = \frac{4\, n_{\rm He}}{n_n + n_p} = \frac{2 n_n}{n_n + n_p}$$

$$Y_p \simeq \frac{2\,(n/p)}{1 + n/p} = 2 X_n \approx \frac{2/7}{8/7} = 0.25$$

full calculation 0.247; observed 0.245 ± 0.003 (metal-poor H II regions)

The light elements: a baryometer



Gorbunov & Rubakov (2011), Fig. 8.3 (PDG)

- Y_p depends on η only logarithmically
- D/H is a sensitive **baryometer**: $\mathrm{D/H} \propto \eta^{-1.6}$ — more baryons burn D more efficiently
- $\mathrm{D/H} = (2.55 \pm 0.03) \times 10^{-5}$ from pristine gas clouds in front of quasars (stars only destroy D)

Two independent measurements of $\Omega_b h^2$

BBN (D/H), $t \approx 3$ min: 0.0221 ± 0.0005
CMB acoustic peaks, $t \approx 3.7 \times 10^5$ yr: 0.02237 ± 0.00015

Open: the lithium problem — predicted ^{7}Li/H is ~ 3 times the observed value.

Recombination: the Saha equation

$e + p \leftrightarrow \mathrm{H} + \gamma$, $B = 13.6\,\mathrm{eV}$

$n_i = g_i (m_i T/2\pi)^{3/2} \mathrm{e}^{(\mu_i - m_i)/T}$

equilibrium: $\mu_p + \mu_e = \mu_\mathrm{H}$; $m_p + m_e - m_\mathrm{H} = B$

$g_e = g_p = 2$, $g_\mathrm{H} = 4$, $m_\mathrm{H} \simeq m_p$

$$\frac{n_p\, n_e}{n_\mathrm{H}} = \left(\frac{m_e T}{2\pi}\right)^{3/2} \mathrm{e}^{-B/T}$$

Free-electron fraction $X_e = n_e/n_{\mathrm{H,tot}}$

neutrality: $n_e = n_p = X_e\, n_{\mathrm{H,tot}}$

$n_\mathrm{H} = (1 - X_e)\, n_{\mathrm{H,tot}}$, $n_{\mathrm{H,tot}} = \eta_\mathrm{H}\, n_\gamma$

$\eta_\mathrm{H} = (1 - Y_p)\, \eta$, $n_\gamma = \frac{2\zeta(3)}{\pi^2} T^3$

$$\frac{1 - X_e}{X_e^2} = 3.84\, \eta_\mathrm{H} \left(\frac{T}{m_e}\right)^{3/2} \mathrm{e}^{B/T}$$

$3.84 = 4\sqrt{2}\,\zeta(3)/\sqrt{\pi}$

Half of the hydrogen has recombined ($X_e = \frac{1}{2}$) at $T \approx 0.32\,\mathrm{eV}$, $z \approx 1370$ — far below $B = 13.6\,\mathrm{eV}$.

Why so late? A worked example

Set $X_e = \frac{1}{2}$ and $x = B/T$:

$$\mathrm{e}^x x^{-3/2} = A \equiv \frac{2}{3.84\,\eta_{\mathrm{H}}(B/m_e)^{3/2}} \approx 8 \times 10^{15}$$

Take logs and iterate:

$$x = \ln A + \tfrac{3}{2}\ln x$$

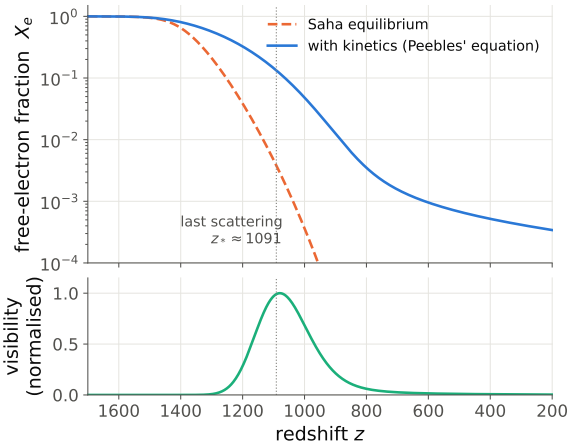
Iteration

$x_0 = \ln A = 36.7$
$x_1 = 36.7 + 1.5 \ln 36.7 = 42.1$
$x_2 = 42.3 \;\Rightarrow\; T = B/42 \approx 0.32\,\mathrm{eV}$

- $\ln A \approx \underbrace{21.5}_{\ln \eta_{\mathrm{H}}^{-1}} + \underbrace{15.8}_{\frac{3}{2}\ln(m_e/B)} - 0.7$: the 1.6×10^9 photons per baryon supply more than half of it
- The Wien tail of the Planck spectrum holds enough photons with $E > 13.6\,\mathrm{eV}$ to re-ionise every new atom until $T \approx B/40$

▶ $T \approx 0.32\,\mathrm{eV} \approx 3700\,\mathrm{K}$ — cooler than the surface of the Sun!

Last scattering: the CMB is released

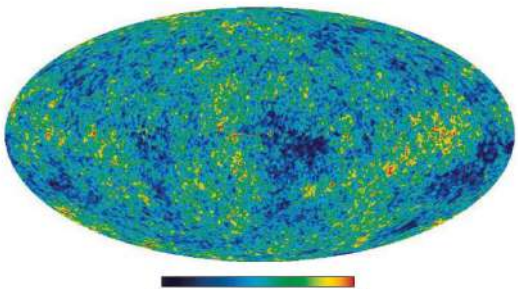


- Kinetics delays recombination (capture to $1s$ re-ionises a neighbour, Ly-α is re-absorbed, slow $2s \to 1s + 2\gamma$; Peebles 1968): $X_e = \frac{1}{2}$ at $z \approx 1270$, residual $X_e \sim 3 \times 10^{-4}$
- Thomson scattering, $\sigma_T = 6.65 \times 10^{-25}\mathrm{cm}^2$; optical depth $\tau = \int n_e \sigma_T c \, \mathrm{d}t$

last scattering ($\tau = 1$): $z_* \approx 1090$, $T_* \approx 0.26\,\mathrm{eV} \approx 3000\,\mathrm{K}$, $t_* \approx 3.7 \times 10^5\,\mathrm{yr}$

lower panel: visibility = probability of last scattering at z

The CMB sky and its statistics



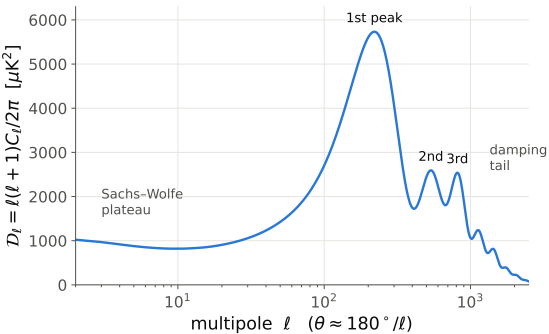
NASA/WMAP (colour scale $\pm 200\ \mu\mathrm{K}$)

- Dipole $3.36\,\mathrm{mK}$: our motion, $370\,\mathrm{km\,s^{-1}}$
- Without dipole and Galaxy: $\delta T/T \sim 10^{-5}$ — density and potential fluctuations at t_*

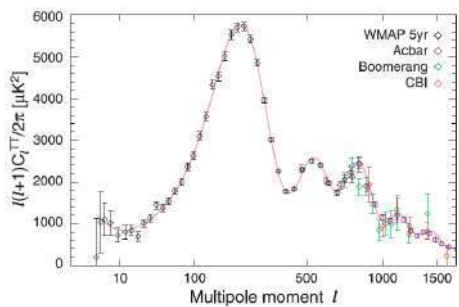
$$\frac{\delta T}{T}(\hat{n}) = \sum_{\ell m} a_{\ell m}\, Y_{\ell m}(\hat{n})$$
$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell\, \delta_{\ell\ell'} \delta_{mm'}$$

- Spherical harmonics: Fourier analysis on the sphere; multipole $\ell \leftrightarrow$ angle $\theta \approx 180°/\ell$
- Isotropy: C_ℓ does not depend on m; plot $\mathcal{D}_\ell = \ell(\ell+1)C_\ell/2\pi$
- Cosmic variance: $\Delta C_\ell / C_\ell = \sqrt{2/(2\ell+1)}$

The angular power spectrum



computed for this course (Planck 2018 best fit)



WMAP team (WMAP 5-yr, ACBAR, Boomerang, CBI)

- $\ell \lesssim 30$: Sachs–Wolfe plateau · $\ell \approx 220, 540, 810, \ldots$: acoustic peaks · $\ell \gtrsim 1000$: damping tail

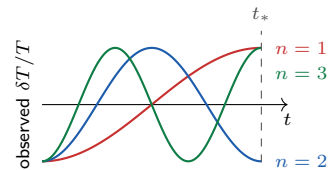
Acoustic oscillations: why there are peaks

- Before t_*: photons and baryons form **one fluid** (Thomson + Coulomb); photons give pressure, baryons inertia, dark matter the potential wells

$$c_s = \frac{1}{\sqrt{3(1+R)}}, \qquad R = \frac{3\rho_b}{4\rho_\gamma}$$

at z_*: $R \approx 0.6$, $c_s \approx 0.45$

- Each Fourier mode k (comoving): a driven oscillator, $\ddot{\delta} + c_s^2\,(k/a)^2\,\delta \simeq$ gravity
- It oscillates as $\cos\left(k\,r_s(t)\right)$, $r_s = \int c_s\,\mathrm{d}t/a$; at t_* each mode is frozen at its phase



mode in a potential well ($\Phi < 0$); up = compression

extremal modes: $k\,r_s(t_*) = n\pi$

odd n: compressions; even n: rarefactions

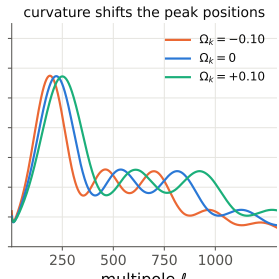
- A harmonic series — like the overtones of a musical instrument of size r_s.

The sound horizon: a ruler on the sky

Standard ruler

$$r_s = \int_0^{t_*} \frac{c_s\,\mathrm{d}t}{a} \approx 144\,\mathrm{Mpc}\ \text{(comoving)}$$

$$\theta_* = \frac{r_s}{D_M(z_*)} = \frac{144.4\,\mathrm{Mpc}}{13.87\,\mathrm{Gpc}} = 0.596^\circ$$



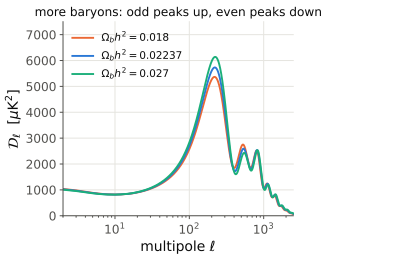
computed for this course

- Peak spacing $\Delta\ell \approx \pi/\theta_* \approx 300$; first peak at $\ell_1 \approx 220$
- Galaxies: the same waves ($r_d \approx 147\,\mathrm{Mpc}$) — the BAO ruler of Hour 2

Closed ($\Omega_k < 0$): the ruler looks larger, peaks move to lower ℓ; open: to higher ℓ

- The measured θ_* shows that space is **flat**: $\Omega_k = 0.001 \pm 0.002$ (with BAO).

Peak heights and the Sachs–Wolfe plateau



computed for this course

Baryons add inertia: odd peaks up, even down $\Rightarrow \Omega_b h^2 = 0.0224$ (as BBN). **Dark matter** deepens the wells $\Rightarrow \Omega_c h^2 = 0.120$. Damping tail: photon diffusion.

- Scales larger than r_s had no time to oscillate: they show the **primordial** potentials Φ

Two competing effects (matter era)

climbing out of a well: $\delta T/T = \Phi \quad (\Phi < 0)$

denser, hotter plasma in the well: $\delta T/T = -\frac{2}{3}\Phi$

$$\frac{\delta T}{T} = \frac{\Phi}{3} \qquad \Rightarrow \qquad |\Phi| \sim 3 \times 10^{-5}$$

Large-scale **cold spots** are potential wells; nearly equal power on all large scales: a nearly **scale-invariant** primordial spectrum

What the CMB tells us: six numbers

parameter	Planck 2018	meaning
$\Omega_b h^2$	0.0224	baryons — agrees with BBN
$\Omega_c h^2$	0.120	non-baryonic dark matter
$100\,\theta_*$	1.0411	acoustic scale $\Rightarrow$ flat space
n_s	0.965 ± 0.004	tilted spectrum: hint of inflation
A_s	2.1×10^{-9}	amplitude of primordial fluctuations
τ	0.054	reionisation at $z \approx 7.7$

Derived (ΛCDM)

$H_0 = 67.4 \pm 0.5\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$
$\Omega_\Lambda = 0.685,\ \ \Omega_m = 0.315$
$t_0 = 13.80\,\mathrm{Gyr}$

Assumed: flat, $w = -1$, three light neutrinos ($\sum m_\nu = 0.06\,\mathrm{eV}$)

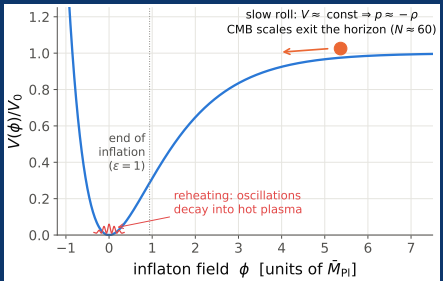
Six numbers describe thousands of measured multipoles — and agree with supernovae, BAO and BBN. (Only H_0 from the local ladder disagrees.)

Planck 2018 VI, A&A 641 (2020) A6

Hour 5

Beyond the hot Big Bang

- Puzzles of the hot Big Bang: horizon, flatness
- Inflation: how it works, how much is needed, what it predicts
- Quantum seeds and the growth of structure
- Where does dark matter come from?
- The complete history and the open questions



computed for this course

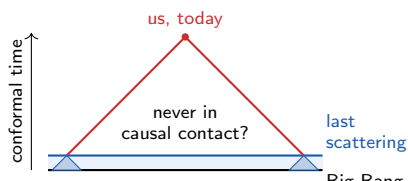
The horizon problem

Matter-era estimate: $H = H_0\sqrt{\Omega_m}(1+z)^{3/2}$, comoving horizon $\chi_{\rm hor} = 2/(aH)$

$$\chi_{\rm hor}(z_*) \simeq \frac{2}{H_0\sqrt{\Omega_m}}(1+z_*)^{-1/2} \approx 0.5\,{\rm Gpc}, \qquad \chi_{\rm hor,0} \simeq \frac{2}{H_0\sqrt{\Omega_m}} \approx 16\,{\rm Gpc}$$

$$\left(\frac{\chi_{\rm hor,0}}{\chi_{\rm hor}(z_*)}\right)^3 \simeq (1+z_*)^{3/2} \approx 3.6\times 10^4$$

(horizon volumes inside today's horizon)



ΛCDM: $\chi_{\rm hor,0} = 14.1\,{\rm Gpc}$ (Day 1), $\chi_{\rm hor}(z_*) = 280\,{\rm Mpc}$, subtending $\theta \approx 1.2^\circ \Rightarrow 4/\theta^2 \approx 10^4$ sky patches

- $\sim 10^4$ causally disconnected sky patches have the **same temperature to** 10^{-5}. How?

The flatness problem

Friedmann equation with $\Omega_{\rm tot} = \rho/\rho_c$

$$\Omega_{\rm tot} - 1 = \frac{k}{(aH)^2}$$

radiation: $aH = \dot{a} \propto t^{-1/2} \Rightarrow |\Omega_{\rm tot} - 1| \propto t$

matter: $aH \propto t^{-1/3} \Rightarrow |\Omega_{\rm tot} - 1| \propto t^{2/3}$

Backwards from $|\Omega_k| \lesssim 0.005$ (95%)

$$|\Omega_{\rm tot} - 1| = |\Omega_k| \left(\frac{a_0 H_0}{aH}\right)^2$$

$t \approx 1\,\text{s}$ ($T = 1\,\text{MeV}$): $\lesssim 2 \times 10^{-18}$

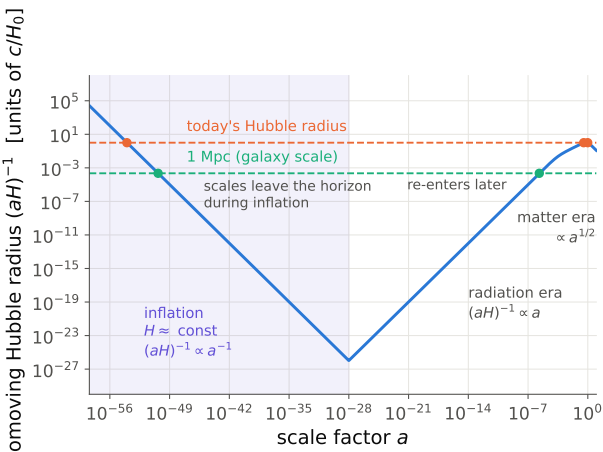
Planck time ($T \sim M_{\rm Pl}$): $\lesssim 10^{-62}$

$(a_0 H_0/aH)^2 = 3.7 \times 10^{-16},\ 1.2 \times 10^{-60}$

- Flatness is **unstable**: any deviation grows with time

Today's flatness needs a fine tuning of at least 1 part in 10^{60} at the Planck time. Also unexplained: the huge entropy ($S \sim 10^{90}$ in the observable Universe) and the origin of the 10^{-5} seeds.

The comoving Hubble radius: the idea of inflation



computed for this course

- Radiation and matter eras: $(aH)^{-1}$ grows — scales **enter** the horizon
- A **shrinking** $(aH)^{-1}$ needs acceleration:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\dot{a}} = -\frac{\ddot{a}}{\dot{a}^2} < 0 \;\Leftrightarrow\; \ddot{a} > 0$$

- Inflation ($H \approx$ const): scales **exit**, and re-enter much later

▶ CMB patches were in contact **before** inflation; and $|\Omega_{\mathrm{tot}} - 1| = |k|/(aH)^2 \to 0$.

How much inflation is needed?

- Today's Hubble scale must have been inside the Hubble radius when inflation began: $(a_i H_i)^{-1} \geq (a_0 H_0)^{-1}$
- $H \approx$ const during inflation ($H_i \approx H_e$):

$$\frac{a_e}{a_i} \gtrsim \frac{a_e H_e}{a_0 H_0} \sim 10^{26}$$

$$N = \ln \frac{a_e}{a_i} \gtrsim 60 \text{ e-folds}$$

Hot Universe starting at $T \approx 10^{15}\,\mathrm{GeV}$

$a_e \approx \left(\frac{3.91}{106.75}\right)^{1/3} \frac{T_0}{T} \approx 8 \times 10^{-29}$

$H_e = 1.66\sqrt{g_*}\, T^2/M_{\mathrm{Pl}} \approx 1.4 \times 10^{12}\,\mathrm{GeV}$

$a_0 H_0 = 1.4 \times 10^{-42}\,\mathrm{GeV}$

$\Rightarrow a_e H_e / a_0 H_0 \approx 10^{26}$

- Flatness solved too: $|\Omega_{\mathrm{tot}} - 1|$ shrinks by $\mathrm{e}^{-2N} \sim 10^{-52}$

Inflation with a scalar field

- A homogeneous scalar field $\phi(t)$, the **inflaton**, with potential $V(\phi)$:

$$\rho = \tfrac{1}{2}\dot{\phi}^2 + V(\phi), \qquad p = \tfrac{1}{2}\dot{\phi}^2 - V(\phi)$$
$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0$$

- A ball rolling down a hill, braked by **Hubble friction** $3H\dot{\phi}$ (the field equation follows from the continuity equation with this ρ, p)

Slow roll: $\dot{\phi}^2 \ll V$

$p \simeq -\rho \simeq -V$: vacuum-like

$H^2 \simeq \dfrac{8\pi}{3}\dfrac{V}{M_{\rm Pl}^2} \approx$ const

$a \propto \mathrm{e}^{Ht}$: quasi-de Sitter expansion

The same field theory describes dynamical dark energy ("quintessence") and, when oscillating, cold dark matter (axion-like fields).

Slow roll and the number of e-folds

Slow-roll equations, $\bar{M}_{\rm Pl} \equiv M_{\rm Pl}/\sqrt{8\pi}$

$$3H\dot{\phi} \simeq -V', \qquad 3\bar{M}_{\rm Pl}^2 H^2 \simeq V$$
$$N = \int H\,\mathrm{d}t = \int \frac{H}{\dot{\phi}}\,\mathrm{d}\phi = -\int \frac{V}{\bar{M}_{\rm Pl}^2 V'}\,\mathrm{d}\phi$$

$$N \simeq \frac{1}{\bar{M}_{\rm Pl}^2}\int_{\phi_e}^{\phi} \frac{V}{V'}\,\mathrm{d}\phi$$

Slow-roll conditions

$$\epsilon = \frac{\bar{M}_{\rm Pl}^2}{2}\left(\frac{V'}{V}\right)^2 \ll 1$$
$$|\eta_V| = \bar{M}_{\rm Pl}^2\,\frac{|V''|}{V} \ll 1$$

- $\epsilon \simeq -\dot{H}/H^2$: inflation ($\ddot{a} > 0$) ends at $\epsilon \simeq 1$

Reheating: the hot Big Bang begins

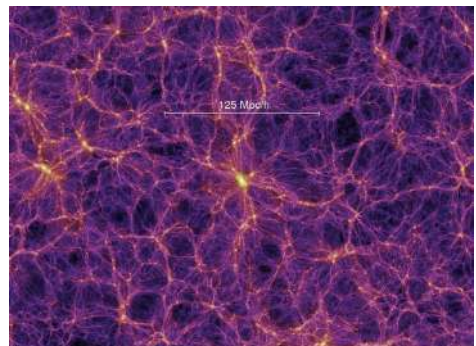
The inflaton then oscillates about the minimum ($\langle p \rangle = 0$: dust-like) and decays into SM particles, which thermalise at $T_{\rm reh}$ ($\gtrsim$ few MeV). The hot "Big Bang" is the **end of inflation**.

Quantum fluctuations: the seeds of structure

- The inflaton fluctuates quantum mechanically: $\delta\phi \simeq H/2\pi$ per Hubble time
- Stretched beyond the Hubble radius, the fluctuations freeze: inflation ends at slightly different times, $\delta t = \delta\phi/\dot{\phi}$

$$\zeta \simeq H\,\delta t \simeq \frac{H^2}{2\pi\dot{\phi}}, \qquad \sqrt{A_s} \approx 4.6 \times 10^{-5}$$

(ζ: primordial curvature perturbation; power $A_s \propto k^{n_s-1}$)



Springel et al. / MPA (Millennium simulation)

- H, $\dot{\phi}$ change slowly: nearly scale invariant, slightly **red**: $n_s - 1 = 2\eta_V - 6\epsilon$
- Primordial GWs (power A_t): $r \equiv A_t/A_s = 16\,\epsilon$

▶ The observable Universe: a **footprint of quantum fluctuations**.

Worked example: $V = \frac{1}{2}m^2\phi^2$

Derivation

$$\frac{V'}{V} = \frac{2}{\phi}, \quad \frac{V''}{V} = \frac{2}{\phi^2} \;\Rightarrow\; \epsilon = \eta_V = \frac{2\bar{M}_{\rm Pl}^2}{\phi^2}$$

end: $\epsilon = 1 \Rightarrow \phi_e = \sqrt{2}\,\bar{M}_{\rm Pl}$

$$N = \frac{1}{\bar{M}_{\rm Pl}^2}\int_{\phi_e}^{\phi} \frac{\phi}{2}\,\mathrm{d}\phi = \frac{\phi^2 - \phi_e^2}{4\bar{M}_{\rm Pl}^2}$$

$N = 60$: $\phi \approx 15.6\,\bar{M}_{\rm Pl}$ (super-Planckian)

Predictions (previous slide)

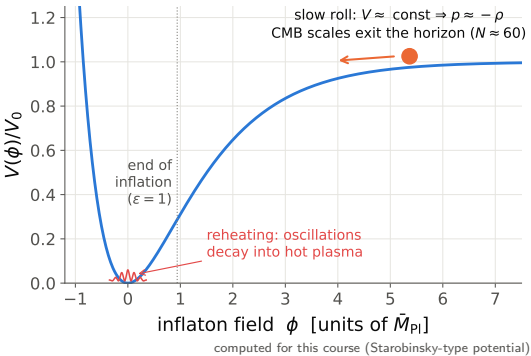
$$n_s - 1 = 2\eta_V - 6\epsilon = -\frac{2}{N} \approx -0.033$$

$$r = 16\epsilon = \frac{8}{N} \approx 0.13$$

- Data: $n_s = 0.965$ ✓, but $r < 0.036$ (BICEP/Keck 2021) ×

▶ The simplest model is **excluded** by the absence of primordial gravitational waves.

Confronting the data: n_s, r and primordial gravitational waves



- Plateau models — Starobinsky R^2 (1980), Higgs inflation: $n_s \simeq 1 - \frac{2}{N} \approx 0.967$, $r \simeq \frac{12}{N^2} \approx 0.003 \quad (N = 60)$

	n_s	r
$m^2\phi^2$, $N = 60$	0.967	0.13
plateau, $N = 60$	0.967	0.003
Planck, BICEP/Keck	0.965 ± 0.004	< 0.036

- $V^{1/4} \approx 1.0 \times 10^{16}\,\mathrm{GeV}\,(r/0.01)^{1/4}$: $r < 0.036 \Rightarrow V^{1/4} < 1.4 \times 10^{16}\,\mathrm{GeV}$
- Tensor modes leave **B-mode** polarisation in the CMB: Simons Observatory, LiteBIRD aim at $r \sim 10^{-3}$

▶ ACT DR6 + Planck + DESI (2025): $n_s = 0.974 \pm 0.003$ — mild pressure on the simplest plateau models.

Structure formation: gravitational instability

Linearised Newtonian fluid, $\delta = \delta\rho/\bar{\rho}$

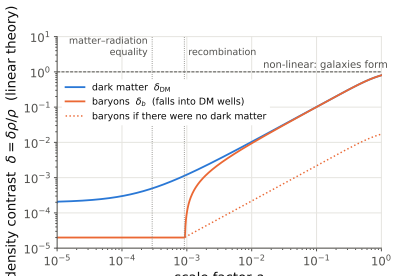
$$\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2}\nabla^2\delta = 4\pi G\bar{\rho}\,\delta$$

Dark matter ($c_s = 0$), matter era:
$H = \frac{2}{3t}$, $4\pi G\bar{\rho} = \frac{3}{2}H^2 = \frac{2}{3t^2}$
$\delta \propto t^n$: $n(n-1) + \frac{4}{3}n - \frac{2}{3} = 0 \;\Rightarrow\; n = \frac{2}{3}, -1$

$$\delta \propto t^{2/3} \propto a$$

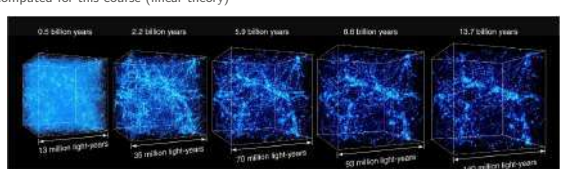
- Gravity amplifies, Hubble friction $2H\dot{\delta}$ slows: power-law growth, not exponential
- Radiation era: only $\delta \propto \ln a$ (Mészáros)
- Λ era: growth stops

Why we need dark matter



computed for this course (linear theory)

- Baryons are locked to photons until z_*: $\delta_b \sim 10^{-5}$
- Growth $\propto a$ gives only $\times 1100$: $\delta \sim 10^{-2}$ today — **no galaxies**
- Dark matter grows logarithmically before $z_{\rm eq}$, then $\propto a$: $\delta_{\rm DM} \sim 10^{-3}$ at z_* (galaxy scales); baryons then fall into its wells



N-body simulation of dark matter, 0.5–13.7 Gyr; from the author's talk "Shedding light on the dark sides of the Universe" (original source not identified)

Where does dark matter come from? Thermal freeze-out

- $X\bar{X} \leftrightarrow$ Standard-Model particles keeps X in equilibrium ($n = n_{\rm eq}$) while $T \gtrsim m$

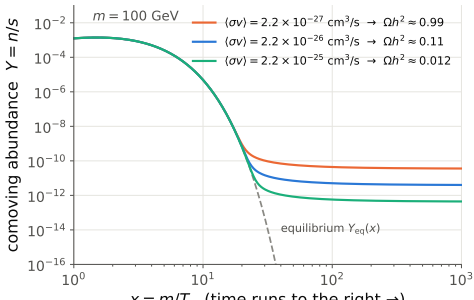
$$\frac{\mathrm{d}n}{\mathrm{d}t} + 3Hn = -\langle\sigma v\rangle\left(n^2 - n_{\rm eq}^2\right)$$

$Y = n/s$, $x = m/T$ (radiation era)

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = -\frac{\lambda}{x^2}\left(Y^2 - Y_{\rm eq}^2\right), \qquad \lambda = \frac{s\,\langle\sigma v\rangle}{H}\Big|_{T=m}$$

- $3Hn$: dilution; right-hand side: annihilation minus production
- Freeze-out when $n_{\rm eq}\langle\sigma v\rangle \simeq H$: $x_f \approx 20$–25 (logarithmic)
- Afterwards Y stays constant: a **relic abundance**, $Y_\infty \approx x_f/\lambda$

The “WIMP miracle” — and alternatives



computed for this course ($m = 100\,\mathrm{GeV}$, constant g_*; full calculation: $\Omega h^2 = 0.12$)

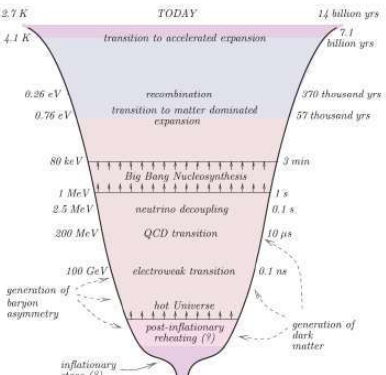
$$\Omega_c h^2 \approx 0.12 \times \frac{2.2\times10^{-26}\mathrm{cm}^3\mathrm{s}^{-1}}{\langle\sigma v\rangle}$$

- Stronger annihilation ⇒ later freeze-out ⇒ fewer survivors
- Weak-scale masses and couplings give $\langle\sigma v\rangle \sim 10^{-26}\mathrm{cm}^3\mathrm{s}^{-1}$: the **WIMP miracle**; no WIMP signal yet (LZ 2024)
- Must be **cold**: free streaming erases structure below the comoving horizon at $T \sim m$ — $\sim 0.1\,\mathrm{Mpc}\,(1\,\mathrm{keV}/m)$ for warm DM; eV neutrinos stream $\sim 0.5\,\mathrm{Gpc}\,(1\,\mathrm{eV}/m_\nu)^{1/2}$

Other mechanisms: freeze-in (feeble couplings); axion misalignment (oscillating field = cold matter); sterile neutrinos (keV); dark-sector glueballs; primordial black holes

The complete history

a tour from the first moments to today



Gorbunov & Rubakov (2011), Fig. 1.9 — 2011 values; current values in the table

	inflation → **reheating** → baryogenesis, dark matter
10^{-11} s	electroweak crossover (Higgs condensate)
$20\,\mu$s	QCD crossover (hadrons form)
~ 1 s	neutrino decoupling, n/p freeze-out
3.5 min	nucleosynthesis: 25% helium
51 kyr	matter–radiation equality
370 kyr	last scattering, CMB released
0.1–1 Gyr	end of the dark ages: first stars, reionisation ($z \approx 7.7$)
7.7 Gyr	expansion starts to accelerate
13.8 Gyr	today

Open questions — and the future

- What is **dark matter**?
- Is **dark energy** a cosmological constant? (DESI: 2.8–4.2σ hints of evolution)
- The **Hubble tension**: 67–68.5 vs $73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$
- Origin of the **baryon asymmetry**
- **Neutrino masses**: cosmology vs oscillations
- Did **inflation** happen, and which model (n_s, r)?
- First-order transitions, GW backgrounds, quantum gravity

The future, if Λ is constant

$a \propto \mathrm{e}^{\sqrt{\Omega_\Lambda} H_0 t}$ (doubling every $12\,\mathrm{Gyr}$): galaxies beyond the Local Group cross our **event horizon** ($5.1\,\mathrm{Gpc}$: light emitted beyond it now never reaches us); stars burn out — "heat death". Phantom energy ($w < -1$): Big Rip; $\rho_\Lambda < 0$: recollapse.

Coming soon

Euclid, Rubin LSST, DESI; Simons Observatory, LiteBIRD; LISA, Einstein Telescope; dark-matter and axion searches

▶ Cosmology = General Relativity + particle physics: where the macro- and micro-world meet.

Further reading

Books

- D. S. Gorbunov, V. A. Rubakov, *Introduction to the Theory of the Early Universe: Hot Big Bang Theory* (World Scientific, 2011) — the course book, Ch. 4–11
- B. Ryden, *Introduction to Cosmology*, 2nd ed. (CUP, 2017)
- D. Baumann, *Cosmology* (CUP, 2022)
- S. Weinberg, *Cosmology* (OUP, 2008); E. W. Kolb, M. S. Turner, *The Early Universe* (1990)
- S. Weinberg, *The First Three Minutes* (1977)
- Lecture notes for Days 1 and 2 (this course): all derivations in full, with exercises

Reviews and data

- Particle Data Group (2024): "Big-Bang cosmology", "Cosmological parameters", "Big-Bang nucleosynthesis", "Cosmic microwave background", "Neutrinos in cosmology", "Dark matter", "Inflation"
- Planck 2018 VI, A&A 641 (2020) A6; DESI DR2, Phys. Rev. D 112 (2025) 083515; ACT DR6, arXiv:2503.14452
- Riess et al., AJ 116 (1998) 1009; Perlmutter et al., ApJ 517 (1999) 565
- BICEP/Keck, PRL 127 (2021) 151301; KATRIN, Science 388 (2025) 180
- A. D. Dolgov, Phys. Rept. 370 (2002) 333