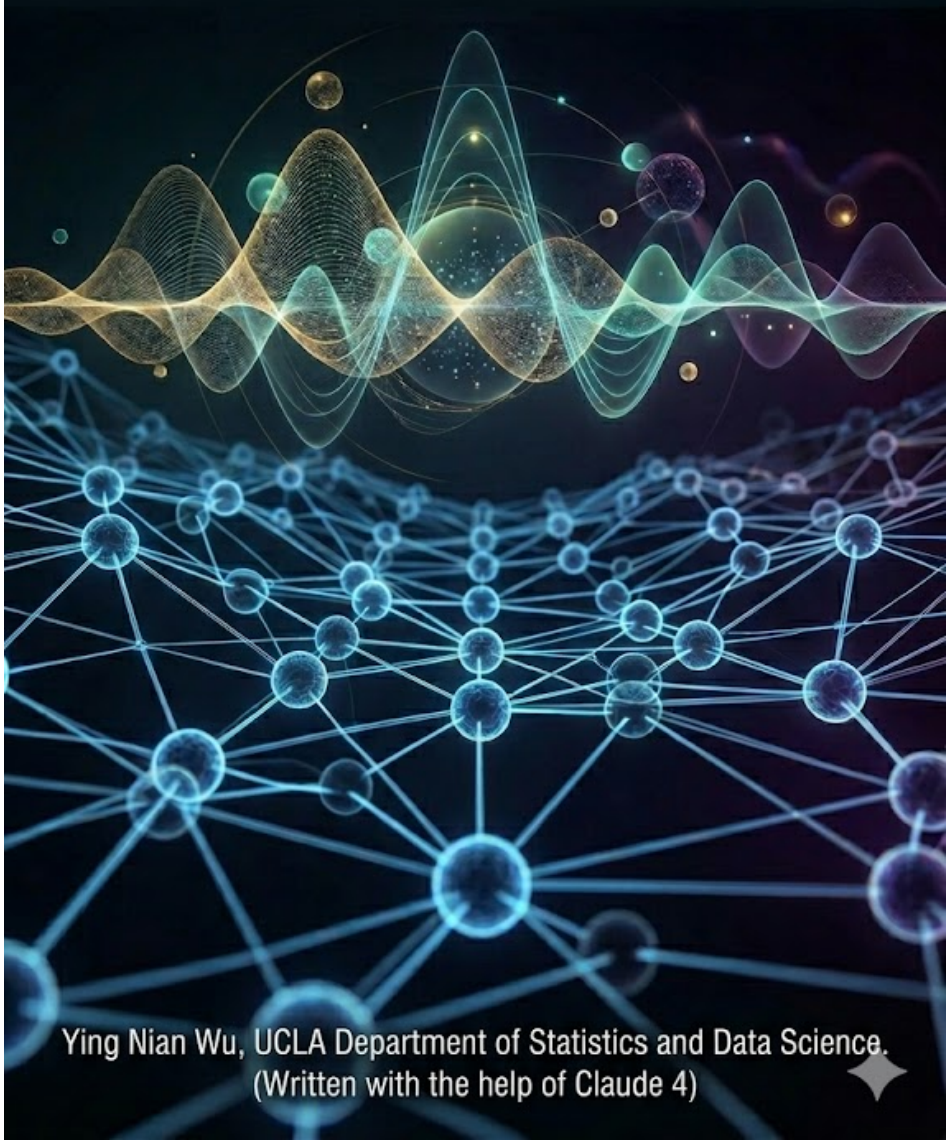
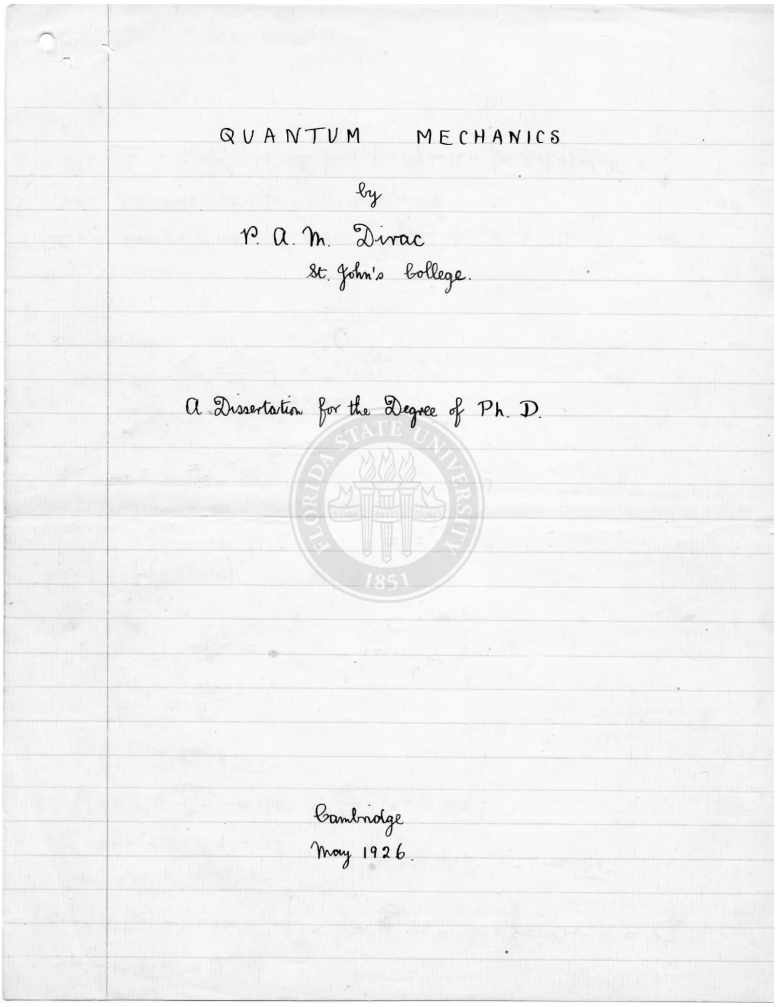
Quantum Physics as Recurrent Network

An Explanation for ML/AI Practitioners



Ying Nian Wu, UCLA Department of Statistics and Data Science.
(Written with the help of Claude 4)



QUANTUM MECHANICS

by

P. A. M. Dirac

St. John's College.

A Dissertation for the Degree of Ph. D.

Cambridge

May 1926.

Left: E. Schrodinger. Right: V. Fock.
The universe runs a Schrodinger RNN in the Fock space.
Feynman path integral and Feynman diagrams are for theoretical calculation.

Quantum Physics as Recurrent Network

A Self-Contained Introduction for Machine Learning Practitioners

Ying Nian Wu
UCLA Department of Statistics and Data Science

Version April 2026, written with the help of Claude 4

Abstract

Quantum mechanics is a linear recurrent neural network (RNN) with complex-valued weights, operating on a hidden state vector that lives in a complex vector space. Every piece of quantum mechanics notation maps directly onto machine learning concepts: wave functions are hidden states, observables are weight matrices, the Born rule is a normalised readout, and wavefunction collapse is an embedding lookup. Quantum field theory extends the hidden state to Fock space—a concatenation of fixed-particle-number subspaces—leaving the RNN equations unchanged. The Feynman path integral is repeated matrix multiplication. The Standard Model is a 19-parameter specification of the weight matrix and Hilbert space. We develop this correspondence carefully, with explicit matrix examples at each step, covering the harmonic oscillator, spinors, quantum electrodynamics, the electroweak theory, quantum chromodynamics, and renormalisation as parameter learning. An ML practitioner comfortable with linear RNNs, embedding tables, and equivariant networks has the complete mathematical toolkit to understand all of fundamental particle physics.

Contents

Preface

This book was written for machine learning practitioners—people who are comfortable with recurrent neural networks, transformer architectures, and the geometry of high-dimensional embedding spaces—who want to understand quantum physics without first spending years on classical mechanics, Lagrangian dynamics, and the historical scaffolding through which physicists have traditionally approached the subject. The central claim is simple enough to state in a single sentence:

If you understand a linear RNN, you already understand quantum mechanics. If you understand GPT, you already understand quantum field theory.

This is not a metaphor or a loose analogy. It is a precise structural correspondence. Every piece of quantum physics notation maps directly onto an ML concept, and the mapping is exact in both directions. We develop it carefully, with explicit matrix examples at each step, from the single qubit through the harmonic oscillator, the Dirac equation, quantum electrodynamics, the electroweak theory, and quantum chromodynamics, arriving at the complete Standard Model of particle physics.

The seven main messages of this book are as follows.

1. Quantum physics is a linear RNN

The Schrödinger equation is:

$$\frac{d\,h_t}{dt} = -i\,H\,h_t.$$

This is a continuous-time linear RNN with hidden state h_t and recurrent weight matrix $W = -iH$. The constraint that H is Hermitian ($H = H^\dagger$) makes W skew-Hermitian, which is the unique condition that preserves the norm of h_t: the hidden state rotates on the unit sphere in a complex Hilbert space, never shrinking or growing. Mamba's state-space model is $\dot{h} = Ah + Bx$; quantum mechanics is this model with $A = -iH$, $B = 0$ (no external input for an isolated system), and h_t a unit vector in $\mathbb{C}^n$.

The readout is the *Born layer*: the probability of outcome o upon measurement is $p(o) = |\langle q_o, h_t\rangle|^2$, where q_o is the eigenvector of the measured observable with eigenvalue o. This is a squared inner product—the "square-softmax" of the projection onto the measurement basis—and it is automatically normalised because the Schrödinger evolution is unitary and therefore preserves $\sum_o |\langle q_o, h_t\rangle|^2 = \|h_t\|^2 = 1$. No partition function needs to be computed.

After measurement, the state is reset: $h_t \leftarrow q_o$. This is the *Bohr layer*—a hard embedding lookup, identical to the re-embedding step in autoregressive GPT. The sampled outcome o is looked up in the eigenstate table and the corresponding eigenvector q_o is written back into the hidden state. The cycle is then: hidden state

$\rightarrow$ Born-rule sample $\rightarrow$ Bohr re-embed $\rightarrow$ hidden state. This is the autoregressive loop of quantum mechanics.

The three components—Schrödinger update, Born readout, Bohr re-embed—are exactly the three components of an autoregressive sequence model. The Schrödinger layer is the recurrent body. The Born layer is the unembedding layer that converts the hidden state to a probability distribution over outcomes. The Bohr layer is the embedding layer that converts the sampled token back into the hidden state. All of quantum mechanics is contained in this architecture.

2. Quantum mechanics and quantum field theory are the same, with a richer index in QFT

In quantum mechanics, the hidden state h_t lives in a Hilbert space whose basis vectors are labelled by a quantum number—a momentum k, a position x, an energy level n. The state is a superposition:

$$h_t = \sum_k c(k)\ |k\rangle\,.$$

Each basis vector $|k\rangle$ is monosemantic in exactly the sense of sparse autoencoders: it encodes a single definite value of the quantum number, nothing else. The coefficient $c(k)$ is the amplitude for that configuration.

Quantum field theory makes one change: it enriches the index on the basis vector. Instead of labelling a basis vector by a single momentum, it labels it by a particle number n *and* a collection of momenta $(k_1, \ldots, k_n)$, one per particle:

$$h_t = \sum_{n=0}^{\infty} \sum_{k_1,\ldots,k_n} c^{(n)}(k_1, \ldots, k_n)\ |n;\, k_1, \ldots, k_n\rangle\,.$$

This is the *Fock space* hidden state. It is a superposition over basis vectors that are monosemantic about both particle number and particle momenta simultaneously. The Hilbert space is larger—its basis has more indices—but it is still a Hilbert space, still spanned by a countable (or continuously labelled) orthonormal basis, and the hidden state is still a unit vector in it.

In GPT, the residual stream $h \in \mathbb{R}^d$ is a superposition of high-dimensional directions that encode rich semantic content: syntactic role, factual associations, discourse context. No explicit monosemantic decomposition is needed; the vector is just a vector, and its content is implicit in the coefficients over the basis. The Fock hidden state is the same: no block structure, no zero-padding, no tensor product notation is required. It is a vector in a Hilbert space with a richly-indexed basis.

The Schrödinger equation is identical: $\dot{h}_t = -iHh_t$. The Born rule is identical: $p(o) = |\langle q_o, h_t\rangle|^2$. When the Hamiltonian H contains creation and annihilation operators—blocks that shift the particle-number index—the rotation of h_t tells a story of particles appearing, disappearing, and interacting. Particle creation is a

rotation that transfers amplitude from the n-particle block to the $(n+1)$-particle block. Annihilation is the reverse. Scattering is a rotation that reshuffles amplitude among different momentum configurations at fixed particle number.

The move from quantum mechanics to quantum field theory is therefore the same kind of move as adding a second particle's momentum as an additional index, or adding spin as an additional index, or adding colour charge as an additional index. It is a change in what the basis vectors are labelled by—nothing more. The framework is unchanged.

3. Multi-step calculation: path integral as matrix multiplication

The RNN runs one step at a time: $h_{t+dt} = e^{-iH\,dt}\,h_t$. For a longer time $T = N \cdot dt$, the hidden state after N steps is:

$$h_T = \underbrace{e^{-iH\,dt} \cdots e^{-iH\,dt}}_{N}\, h_0 = e^{-iHT}\, h_0.$$

This is matrix exponentiation: repeated application of the single-step rotation matrix. To compute the matrix element $\langle q_f|e^{-iHT}|q_i\rangle$ (the amplitude to go from initial state q_i to final state q_f), insert a complete basis $I = \int dq\ |q\rangle\,\langle q|$ between every consecutive pair of steps. The result is:

$$\langle q_f|e^{-iHT}|q_i\rangle = \int \mathcal{D}q\; e^{iS[q]},$$

where $S[q] = \int_0^T \mathcal{L}(q,\dot{q})\,dt$ is the classical action and the integral sums over all paths $q(t)$ connecting q_i to q_f. This is the *Feynman path integral*.

It is not a new postulate. It is the same matrix multiplication, written in a basis. The Lagrangian $\mathcal{L}$ appears in the exponent not because it is fundamental but because it is what the short-time matrix element $\langle q_{k+1}|e^{-iH\,dt}|q_k\rangle$ evaluates to via the Trotter approximation. The classical path (saddle point of S) is the most probable trajectory—greedy decoding. Quantum corrections are all the off-saddle paths—the full sum. Feynman diagrams are the perturbative expansion of this sum around the saddle, organised by powers of the coupling constants.

The path integral is our analytical tool. It is not how the universe runs. The universe applies $e^{-iH\,dt}$ exactly, at every Planck step, without time-slicing or saddle-point approximation. We time-slice because we cannot compute e^{-iHT} directly; the universe has no such limitation.

4. The recipe: Lagrangian to Hamiltonian to path integral

The practical workflow of quantum field theory follows a clean recipe.

Step 1: write a Lagrangian. The Lagrangian density $\mathcal{L}(\phi, \partial_\mu\phi)$ is a compact, manifestly Lorentz-invariant specification of the dynamics. For the scalar field:

$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4$. Every term is a Lorentz scalar (no free indices), ensuring that the action $S = \int d^4x\,\mathcal{L}$ is the same in every inertial frame.

Step 2: derive the Hamiltonian. The Legendre transform $H = \int d^3x\,(\pi\dot{\phi} - \mathcal{L})$, where $\pi = \partial\mathcal{L}/\partial\dot{\phi}$ is the conjugate momentum, converts the Lagrangian into the Hamiltonian. Quantisation promotes ϕ and π to operators on Fock space satisfying canonical commutation relations. The result is the recurrent weight matrix $-iH$ of the Schrödinger RNN.

Step 3: run the RNN. The Schrödinger equation $\dot{h}_t = -iHh_t$ evolves the Fock hidden state. The Born rule reads out probabilities. The Bohr layer resets the state after measurement. This is the physical content of the theory.

Step 4: derive the path integral for analysis. Time-slice e^{-iHT}, insert resolutions of identity, evaluate short-time matrix elements by Trotter splitting. The Lagrangian reappears in the exponent. The perturbative expansion of the path integral generates Feynman diagrams, which are the computational tools for extracting predictions from the theory.

The Lagrangian thus plays two roles: it is the starting point for deriving the Hamiltonian (going up to the RNN), and it reappears at the end in the path integral (going down to perturbation theory). The Hamiltonian—the recurrent weight matrix—is the central object. The Lagrangian and the path integral are, respectively, the specification language and the analysis language for that matrix.

5. Internal structure of the embedding: symmetry forces the coupling

The hidden state vector h_t carries a rich internal structure. In quantum field theory, the basis vectors of the Fock space are labelled not only by momentum k and particle number n but by additional attributes: *spin* $s \in \{-\frac{1}{2}, +\frac{1}{2}\}$ (intrinsic angular momentum), *flavour* $f \in \{u, d, s, c, b, t\}$ (quark species), *colour* $\alpha \in \{r, g, b\}$ (strong charge), *helicity* $\lambda \in \{+1, -1\}$ (photon polarisation). Each attribute is an additional index on the basis vector.

These attributes are not chosen freely. They are determined by the *representation theory of the Lorentz group and the gauge group*. Diagram 3 in Chapter 13 makes this precise: a field $\Phi(x)$ is an embedding table assigning a vector to each spacetime point x, and the Lorentz transformation Λ acts on both the spacetime point (top of the diagram) and the field value (bottom of the diagram) via a matrix $M(\Lambda)$. Different choices of $M(\Lambda)$ give different particle species:

- $M(\Lambda) = 1$: scalar field, spin 0 (Higgs boson).
- $M(\Lambda) = L \in SL(2, \mathbb{C})$: left Weyl spinor, spin $\frac{1}{2}$ (electron, quarks).
- $M(\Lambda) = L \cdot (\text{right}) \cdot L^\dagger$: 4-vector field, spin 1 (photon, $W^\pm$, Z^0, gluons).

The two-sided action $M_L(L)\,\Phi\,M_R(L^\dagger)$ encodes left- and right-handed spinor structure simultaneously. Every particle in the Standard Model is a choice of left and

right representation matrices.

The gauge symmetries—$SU(3)_c \times SU(2)_L \times U(1)_Y$—act on additional internal indices (colour, isospin, hypercharge) independently of the Lorentz indices. The requirement of *local* gauge invariance—that the symmetry can be applied independently at each spacetime point—forces the introduction of gauge fields (photon, $W^{\pm}$, Z^0, gluons) and uniquely determines how matter and force fields couple. In QED, the coupling of the electron to the photon is forced by demanding that the Dirac Lagrangian be invariant under the local phase rotation $\Psi \to e^{i\alpha(x)}\Psi$. The photon is not added by hand; it is the mathematical necessity of making the electron's phase unobservable. In QCD, demanding local $SU(3)_c$ invariance forces the gluon field and uniquely determines the quark-gluon coupling.

The symmetry structure—global Lorentz invariance for the spin structure, local gauge invariance for the force-matter coupling—is encoded entirely in the index structure of the basis vectors and the transformation properties of the Hamiltonian. It is the physics analogue of equivariant network design in machine learning: the architecture (the RNN) is constrained to commute with a group of transformations, and this constraint determines the weight matrix up to a small number of free parameters.

6. Renormalisation is parameter learning

The Standard Model has 19 free parameters: 3 gauge couplings, 9 fermion masses, 4 mixing angles, the Higgs mass and vacuum expectation value, and the QCD vacuum angle. These 19 numbers are not derivable from the symmetry structure—they must be measured from experiment. Given them, the theory predicts every other observable in particle physics.

Computing quantum corrections (loop diagrams) produces integrals over all momenta:

$$I = \int_0^\infty \frac{d^4k}{(2\pi)^4}\, f(k, p, m),$$

which diverge at large k (ultraviolet divergences). These divergences are not physical; they reflect the fact that the theory is being asked to extrapolate to arbitrarily short distances, beyond the domain where it has been tested.

Renormalisation resolves this by writing each bare parameter as a renormalised parameter plus a counterterm:

$$m_0 = m_R + \delta m, \qquad \lambda_0 = \lambda_R + \delta\lambda,$$

and choosing the counterterms to absorb the UV divergences. The renormalised parameters m_R, λ_R are then fixed by matching to a small set of experimental measurements. This is precisely *fine-tuning*: the bare parameters are pre-training initialisations; quantum corrections are gradient signals from the vacuum; the

counterterms are update steps; and the renormalised parameters are the trained weights that reproduce the experimental data.

A theory is *renormalisable* if all UV divergences at any loop order can be absorbed into finitely many counterterms—the parameters already present in the Lagrangian. Power counting determines whether a theory is renormalisable: an interaction term of mass dimension ≤ 4 is renormalisable in four spacetime dimensions. The Standard Model contains only dimension-≤ 4 operators, so it is renormalisable: a finite-parameter model that can be fully specified by 19 experimental measurements and then predicts everything else.

The Wilsonian picture deepens this: integrating out high-momentum modes above scale Λ generates an *effective action* at scale Λ/b, with renormalised parameters that flow according to the *renormalisation group equation* $\mu\, d\lambda/d\mu = \beta(\lambda)$. This is multi-scale representation learning: marginalising over fine-grained degrees of freedom produces a compressed model at coarser scales. The RG fixed point—where $\beta(\lambda^*) = 0$—is a scale-invariant representation, the attractor of the parameter flow. Relevant operators grow in importance at long scales (like high-level semantic features); irrelevant operators are suppressed (like fine-grained texture).

The key point for an ML practitioner: *renormalisability is the condition that the model has a finite number of parameters.* A non-renormalisable theory requires new counterterms at every loop order—its parameter count grows without bound as precision increases. This is not a theory in the predictive sense. A renormalisable theory has a fixed, finite parameter count, and every quantum correction at every loop order is absorbed into the same 19 numbers. The Standard Model is a 19-parameter model of all known particle physics.

A note on approach

We teach quantum physics *top-down*. The standard curriculum runs bottom-up: classical mechanics, then electromagnetism, then special relativity, then quantum mechanics, then quantum field theory. We invert this. The top—the Schrödinger RNN—is simpler than the bottom. Classical mechanics is not the foundation; it is an emergent approximation, derived from quantum mechanics in the limit where the Born-rule fluctuations are negligible and the hidden state is well approximated by a coherent state.

Quantisation—the procedure of converting a classical theory into a quantum one—is guesswork, not derivation. Both canonical quantisation (promoting Poisson brackets to commutators) and field quantisation (promoting classical fields to Fock-space operators) are recipes that work, but they work backwards from the quantum answer using a classical scaffolding that is not physically real. The classical Dirac field $\Psi(x)$, for instance, is not a physical object; it is a notational device for writing down the Hamiltonian.

In the top-down approach, one starts with the Fock space and the Hamiltonian, determined by symmetry requirements (diagram 3), and derives classical mechanics as the coherent-state limit. Classical physics is what quantum physics looks like from inside the rendering layer—from the perspective of an observer who is making coarse-grained measurements on a high-dimensional quantum hidden state.

The rendering layer itself—the Born rule and the Bohr reset—is an independent postulate, not derivable from the Schrödinger equation. This is the "measurement problem" in its correct framing: not a paradox, but the observation that a complete physical theory requires both a dynamics (Schrödinger) and an output interface (Born + Bohr), and the two are logically independent. The observer is left undefined in the Copenhagen axioms for the same reason that "point" is left undefined in Euclid: because the output interface is not the business of the physics engine, and its concrete implementation does not affect the predictions.

How to read this book

Chapter 1 develops quantum mechanics as an RNN, using explicit 2×2 and 3×3 matrix examples. The qubit, the harmonic oscillator, the Bloch sphere, and Feynman's path integral are all presented as special cases of the Schrödinger RNN on progressively richer Hilbert spaces.

Chapter 2 introduces quantum field theory by the single move of promoting the Hilbert space to Fock space. Scalar fields, the harmonic oscillator, coherent states, and the ladder operator algebra are developed in the Schrödinger, Heisenberg, and Feynman pictures, all shown to be the same RNN in different coordinates.

Chapter 3 develops the representation theory of the Lorentz group through diagram 3: the commutative diagram that makes every particle species a choice of two-sided matrix action on the field value. Spinors, Dirac spinors, and 4-vector fields are all derived as cases of this diagram.

Chapters 4–6 build the Standard Model sector by sector: the Dirac field (electrons and positrons), QED (gauge invariance forcing the coupling), the electroweak theory (Higgs mechanism and mass generation), and QCD (quarks, gluons, confinement, and asymptotic freedom).

Chapter 7 addresses the technical tools: the path integral as matrix multiplication (one section), Faddeev–Popov gauge fixing as integration over gauge orbits (one section), and renormalisation as parameter learning (one section).

The final sections step back: a recap of the whole picture through the ML correspondence table; Feynman diagrams as cumulants of a log-partition function (the EBM connection); the universe's source code and the rendering layer; gravity as an emergent equation of state; the observer and the measurement problem; the real mystery (the exponential of the astronomical); and a closing meditation on whether the universe is thought or stuff.

A reader who finishes the book will have covered, in order: the axioms of quantum mechanics, the harmonic oscillator, Dirac's equation, the full Standard Model, the Higgs mechanism, QCD confinement, the Feynman path integral, renormalisation group flow, and the conceptual foundations of quantum gravity—all in the language of linear algebra, matrix exponentials, and embedding spaces. The physics is not simplified. It is translated.

1 Quantum Mechanics as an RNN: The Core Model

1.1 The three-layer architecture

Quantum mechanics has a clean three-layer architecture that maps onto a recurrent neural network:

Schrödinger layer — hidden state update:

$$h_{t+dt} \;=\; h_t \;+\; W h_t \, dt, \qquad W = -iH. \tag{1}$$

Born layer — output / unembed upon measurement:

$$p_t(o) \;=\; |\langle h_t,\, q_o \rangle|^2, \qquad \text{(automatically normalised; see below)} \tag{2}$$

Bohr layer — input / embed upon measurement:

$$h_t \;\leftarrow\; q_o. \tag{3}$$

The hidden state $h_t \in \mathbb{C}^n$ is a *complex unit vector*. The recurrent weight matrix is $W = -iH$, where H is *Hermitian* ($H = H^\dagger$, i.e. $H^\dagger_{jk} = \overline{H_{kj}}$). The vectors (q_o) are orthonormal eigenvectors of an observable matrix; they serve as readout vectors in the Born layer and as embedding vectors in the Bohr layer.

1.2 Why the hidden state must be complex-valued

The Schrödinger update $(I - iH\,dt)$ is a rotation (unitary transformation) because:

$$\begin{aligned}(I - iH\,dt)(I - iH\,dt)^\dagger &= (I - iH\,dt)(I + iH^\dagger\,dt) \\ &= I + i(H^\dagger - H)\,dt + O(dt^2) \\ &= I + O(dt^2),\end{aligned} \tag{4}$$

where the last step uses $H = H^\dagger$. So $\|h_t\|^2$ is preserved under every update—the hidden state rotates on the unit sphere in $\mathbb{C}^n$. This is why the Born probabilities always sum to 1:

$$\sum_o p_t(o) = \sum_o |\langle h_t, q_o \rangle|^2 = \|h_t\|^2 = 1. \tag{5}$$

The $i = \sqrt{-1}$ in $W = -iH$ is not a notational trick; it is what turns a real symmetric matrix into a generator of rotations rather than a generator of shearing. Without it, the norm would drift.

1.3 Why the Born rule uses $|\cdot|^2$ instead of $\exp(\cdot)$

In a standard softmax output layer you compute $\exp(\langle h, q_o\rangle)$ and then divide by the partition function. In the Born layer you compute $|\langle h_t, q_o\rangle|^2$ and divide by nothing, because the normalisation constant is already identically 1. This works because the Schrödinger layer is unitary: it preserves $\sum_o |\langle h_t, q_o\rangle|^2 = \|h_t\|^2$. The Born rule is the "right" softmax for a unitary hidden layer—one that is automatically normalised without ever recomputing a partition function.

1.4 ML–QM correspondence table

ML concept	**QM concept**
Hidden state $h_t \in \mathbb{C}^n$	Quantum state / wave function
Recurrent weight $W = -iH$	Hamiltonian operator
Unitary step $(I - iH\,dt)$	Time-evolution operator
Readout inner product $\langle h_t, q_o\rangle$	Probability amplitude
$\lvert\langle h_t, q_o\rangle\rvert^2$ readout	Born rule (measurement prob.)
Eigenvector q_o of observable O	Basis state / outcome embedding
Hard reset $h_t \leftarrow q_o$	Wavefunction collapse
Distribution over hidden states ρ	Density matrix (mixed state)
Environment noise corrupting ρ	Decoherence
Tensor product $h_A \otimes h_B$	Two-qubit joint state

2 A Concrete Example: The Qubit

The simplest quantum system is a *qubit*, where $h_t \in \mathbb{C}^2$. It is the quantum analogue of a single binary variable. Working through a qubit makes all three layers fully concrete before we tackle infinite-dimensional systems.

2.1 The qubit state space

A qubit state is any unit vector in $\mathbb{C}^2$:

$$h = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \qquad |\alpha|^2 + |\beta|^2 = 1. \tag{6}$$

Two distinguished basis states are:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{7}$$

(Dirac bra-ket notation: $|v\rangle$ is a column vector; $\langle v| = |v\rangle^\dagger$ is its conjugate row.) A general state $h = \alpha\,|0\rangle + \beta\,|1\rangle$ is called a *superposition*—not "both at once" but an information state that assigns complex amplitudes to both outcomes.

2.2 Observables as Hermitian 2×2 matrices

The natural observables for a qubit are the *Pauli matrices*:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{8}$$

Each is Hermitian. Their eigenvalues are always ± 1.

Eigen decomposition of σ_z:

$$\sigma_z = QDQ^\dagger, \quad Q = I, \quad D = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{9}$$

Eigenvectors: $q_+ = |0\rangle$ (eigenvalue $+1$), $q_- = |1\rangle$ (eigenvalue -1).

Born rule for σ_z. Let $h = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}$.

$$p(+1) = \left|\langle h|q_+\rangle\right|^2 = \left|\tfrac{1}{\sqrt{2}}(1,\,-i)\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right|^2 = \left|\tfrac{1}{\sqrt{2}}\right|^2 = \tfrac{1}{2}.$$

$$p(-1) = \left|\langle h|q_-\rangle\right|^2 = \left|\tfrac{1}{\sqrt{2}}(1,\,-i)\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right|^2 = \left|\tfrac{-i}{\sqrt{2}}\right|^2 = \tfrac{1}{2}.$$

Equal probability. If the outcome is $+1$, the Bohr layer fires: $h \leftarrow |0\rangle$.

2.3 The Schrödinger layer on a qubit: explicit time evolution

Let $H = \sigma_z$. The continuous-time Schrödinger equation is

$$\frac{dh_t}{dt} = -i\sigma_z\, h_t, \qquad h_0 = \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix}. \tag{10}$$

Since σ_z is already diagonal, the matrix exponential is:

$$h_t = e^{-i\sigma_z t}\, h_0 = \begin{pmatrix} e^{-it} & 0 \\ 0 & e^{it} \end{pmatrix}\begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} = \begin{pmatrix} e^{-it}\alpha_0 \\ e^{it}\beta_0 \end{pmatrix}. \tag{11}$$

Each component rotates at unit angular speed in the complex plane—the hidden state precesses.

Rotating qubit: measuring a different observable. Start at $h_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$, so $h_t = \frac{1}{\sqrt{2}}\begin{pmatrix}e^{-it}\\e^{it}\end{pmatrix}$. Now measure σ_x. Its eigenvectors are $q_\pm = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\\pm 1\end{pmatrix}$ (eigenvalues ± 1).

$$p_t(+1) = |\langle h_t|q_+\rangle|^2 = \left|\tfrac{1}{2}(e^{it} + e^{-it})\right|^2 = \cos^2 t,$$
$$p_t(-1) = |\langle h_t|q_-\rangle|^2 = \left|\tfrac{1}{2}(e^{it} - e^{-it})\right|^2 = \sin^2 t.$$

The measurement probabilities oscillate sinusoidally in time. This is *quantum interference*, and it is a direct consequence of the hidden state being complex-valued and rotating.

2.4 Wavefunction collapse as a hard embedding lookup

After the measurement the Bohr layer resets $h_t \leftarrow q_o$. This is not a soft Bayesian update—it is a hard assignment, exactly like an embedding table lookup in NLP: given a discrete token o, replace the current representation with the pre-stored embedding q_o. All information encoded in the previous h_t that is orthogonal to q_o is permanently discarded. The universe has updated its records.

2.5 The Bloch sphere: a geometric view of the qubit

Every qubit state (up to an irrelevant global phase) can be written as

$$|h\rangle = \cos\tfrac{\theta}{2}\,|0\rangle + e^{i\phi}\sin\tfrac{\theta}{2}\,|1\rangle\,, \qquad \theta \in [0,\pi],\ \phi \in [0, 2\pi). \tag{12}$$

The pair (θ, ϕ) defines a point on the unit sphere in $\mathbb{R}^3$ via the *Bloch vector*

$$\vec{r} = \Big(\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta\Big). \tag{13}$$

The north pole ($\theta = 0$) is $|0\rangle$; the south pole ($\theta = \pi$) is $|1\rangle$; the equator contains all equal-amplitude superpositions.

Observables as axes. The expected value of each Pauli matrix equals the corresponding Bloch-vector component:

$$\langle\sigma_x\rangle = \sin\theta\cos\phi, \qquad \langle\sigma_y\rangle = \sin\theta\sin\phi, \qquad \langle\sigma_z\rangle = \cos\theta. \tag{14}$$

Measuring σ_z projects onto the north or south pole; measuring σ_x projects onto the x-axis poles $q_\pm$. Incompatible observables correspond to orthogonal axes—knowing the north/south coordinate exactly forces the east/west coordinate to be maximally uncertain.

Evolution as rotation. With $H = \frac{\omega}{2}\sigma_z$ the Schrödinger update is $e^{-i\omega t\,\sigma_z/2}$, which rotates the Bloch vector around the z-axis at angular speed ω:

$$\phi \;\longmapsto\; \phi + \omega t. \tag{15}$$

Every unitary 2×2 matrix is a rotation of the Bloch sphere—no more and no less. The RNN hidden state literally precesses like a compass needle in a magnetic field.

2.6 Schrödinger vs. Heisenberg picture

There are two equivalent ways to run the QM-RNN:

	Schrödinger picture	**Heisenberg picture**
Hidden state h_t	Rotates: $h_t = e^{-iHt}h_0$	Fixed at h_0
Observable O	Fixed	Rotates: $O_t = e^{iHt}Oe^{-iHt}$
ML analogy	Evolve the activations	Evolve the weights

Both pictures give identical Born-rule probabilities. The Schrödinger picture maps directly onto the RNN analogy; the Heisenberg picture is often more convenient for analytic calculations.

3 The Classical Hamiltonian

Before building matrix Hamiltonians for general systems we need to understand what H means classically: it is simply the *total energy*.

3.1 Free particle

For a particle of mass $m = 1$ moving in 1D with momentum $p = mv$:

$$H = \tfrac{1}{2}p^2. \tag{16}$$

3.2 Harmonic oscillator

Adding a restoring spring with unit spring constant gives potential energy $V(x) = \frac{1}{2}x^2$, so:

$$H = \tfrac{1}{2}(p^2 + x^2). \tag{17}$$

The harmonic oscillator is the most important system in physics because any smooth potential near a local minimum looks quadratic by Taylor expansion. It models vibrations in molecules, photons in a cavity, and forms the building block of quantum field theory. Conceptually, it plays the same role as the linear layer in neural networks: the foundational unit from which everything is built.

4 Promoting Scalars to Matrices (Quantisation)

The quantisation step is simple to state: replace every scalar observable $f(x, p)$ of classical mechanics with the corresponding matrix $f(X, P)$, subject to one constraint.

4.1 Position operator X

In the position basis, X is the operator that multiplies a wave function $v(x)$ by x:

$$(Xv)(x) = x \cdot v(x). \tag{18}$$

Its eigenvectors are delta functions $\delta_{x_0}(x)$ (one-hot vectors at each position x_0), with eigenvalue x_0.

4.2 Momentum operator P

In the position basis, momentum is a differential operator:

$$P = -i\,\frac{d}{dx}, \qquad (Pv)(x) = -iv'(x). \tag{19}$$

Discrete approximation. Discretise x with spacing s. Then $v'(x) \approx (v(x) - v(x-s))/s$, and P becomes the banded matrix:

$$P \approx \frac{-i}{s}\begin{pmatrix} 1 & 0 & 0 & \cdots \\ -1 & 1 & 0 & \cdots \\ 0 & -1 & 1 & \cdots \\ \vdots & & \ddots & \ddots \end{pmatrix}.$$

This is just a first-difference matrix scaled by $-i/s$. As $s \to 0$ it converges to $-i\,d/dx$.

Eigenvectors of P. Plane waves diagonalise P:

$$q_p(x) = e^{ipx}, \qquad Pq_p = -i\frac{d}{dx}e^{ipx} = p\,e^{ipx} = p\,q_p. \tag{20}$$

A particle with definite momentum p has its hidden state spread over all positions as a sinusoidal wave. If you measure position, the uncertainty is maximal.

Remark. The Fourier transform is exactly the change-of-basis matrix from the position eigenbasis $\{\delta_{x_0}\}$ to the momentum eigenbasis $\{e^{ipx}\}$. If you have ever used an FFT, you have been computing a quantum mechanical basis change.

4.3 The canonical commutation relation

Classical mechanics treats x and p as independent real numbers, so $xp = px$. Quantum mechanics replaces them with operators that satisfy:

$$[X, P] \;\equiv\; XP - PX \;=\; iI. \tag{21}$$

This single equation *is* the entire content of quantisation. It replaces Bohr's ad hoc rule about quantised orbits.

Verification with $X = x\cdot$ and $P = -i\,d/dx$.
Apply $(XP - PX)$ to any test function $v(x)$:

$$(XPv)(x) = x \cdot (-iv'(x)) = -ix\,v'(x),$$
$$(PXv)(x) = -i\frac{d}{dx}\Big(x\,v(x)\Big) = -iv(x) - ix\,v'(x).$$

Subtracting:

$$\Big[(XP - PX)v\Big](x) = -ix\,v'(x) - (-iv(x) - ix\,v'(x)) = iv(x).$$

So $[X, P]v = iIv$ for every v, confirming (21). ✓

Non-commutativity in 2×2: a toy model.
Take $X = \sigma_x$ and $P = \sigma_y$:

$$XP = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$
$$PX = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix},$$
$$[X, P] = \begin{pmatrix} 2i & 0 \\ 0 & -2i \end{pmatrix}.$$

This does not equal iI exactly (the exact relation requires infinite dimensions), but it shows concretely that position and momentum matrices genuinely fail to commute.

5 Observables, Measurement, and the Uncertainty Principle

5.1 Observables are Hermitian matrices

Any physical quantity that can be measured—position, momentum, energy, spin—is represented by a Hermitian matrix $O = O^\dagger$. The eigen decomposition

$$O = QDQ^\dagger, \qquad D = \mathrm{diag}(d_1, d_2, \ldots), \qquad Q = (q_1 \mid q_2 \mid \cdots) \tag{22}$$

provides:

- **Eigenvalues** (d_i): the only values that can appear as measurement outcomes.
- **Eigenvectors** (q_i): the embedding/readout vectors for the Born and Bohr layers.

If you observe energy, you use the eigen decomposition of H. If you observe position, you use that of X, and so on.

5.2 The uncertainty principle from incompatible eigenbases

Two observables A and B are *compatible* if $[A, B] = 0$, meaning they share a common eigenbasis and can be observed simultaneously with certainty. If $[A, B] \neq 0$ they are *incompatible*: their eigenbases differ, and knowing one exactly forces the other to be uncertain.

The logic in RNN terms: if the hidden state h_t is equal to eigenvector q_i of A, then A is observed with probability 1. Now express q_i in the eigenbasis of B:

$$q_i = \sum_j \underbrace{\langle q'_j, q_i \rangle}_{c_{ij}} \, q'_j. \tag{23}$$

If $[A, B] \neq 0$, no single $|c_{ij}|$ can equal 1; the amplitude is spread across multiple j, so the Born layer gives a distribution over B's outcomes. This is Heisenberg's uncertainty principle, viewed as a change-of- basis effect.

σ_z–σ_x uncertainty on a qubit.
Start with $h = |0\rangle$ (the $+1$ eigenvector of σ_z: we know $\sigma_z = +1$ with certainty). Now measure σ_x with eigenvectors $q_\pm = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}$:

$$p(+1) = |\langle 0|q_+\rangle|^2 = \left|\tfrac{1}{\sqrt{2}}\right|^2 = \tfrac{1}{2},$$
$$p(-1) = |\langle 0|q_-\rangle|^2 = \left|\tfrac{1}{\sqrt{2}}\right|^2 = \tfrac{1}{2}.$$

Knowing σ_z exactly forces σ_x to be maximally random. This is the uncertainty principle, expressed as a 2×2 matrix calculation.

6 Ladder Operators and the Harmonic Oscillator Spectrum

The harmonic oscillator $H = \frac{1}{2}(X^2 + P^2)$ has a beautiful algebraic structure discovered by Dirac that lets us find all eigenvalues *without ever writing down an infinite matrix.*

6.1 Defining the ladder operators

$$a = \tfrac{1}{\sqrt{2}}(X + iP), \qquad a^\dagger = \tfrac{1}{\sqrt{2}}(X - iP). \tag{24}$$

Expressing H in terms of a and $a^\dagger$.

$$\begin{aligned} a^\dagger a &= \tfrac{1}{2}(X - iP)(X + iP) = \tfrac{1}{2}\Big(X^2 + iXP - iPX + P^2\Big) \\ &= \tfrac{1}{2}\Big(X^2 + P^2\Big) + \tfrac{i}{2}[X, P] \\ &= H + \tfrac{i}{2}(iI) = H - \tfrac{1}{2}. \end{aligned}$$

Therefore $H = a^\dagger a + \frac{1}{2}$. ✓
Similarly $aa^\dagger = H + \frac{1}{2}$, so $[a, a^\dagger] = aa^\dagger - a^\dagger a = I$.

6.2 Climbing and descending the energy ladder

Suppose $Hq_n = d_n\, q_n$. We claim:

$$H(a^\dagger q_n) = (d_n + 1)\,(a^\dagger q_n), \qquad H(a\, q_n) = (d_n - 1)\,(a\, q_n). \tag{25}$$

Proof for $a^\dagger$:

$$\begin{aligned} H(a^\dagger q_n) &= \Big(a^\dagger a + \tfrac{1}{2}\Big)(a^\dagger q_n) = a^\dagger\Big(aa^\dagger\Big)q_n + \tfrac{1}{2}a^\dagger q_n \\ &= a^\dagger\Big(a^\dagger a + I\Big)q_n + \tfrac{1}{2}a^\dagger q_n = a^\dagger\Big(d_n + \tfrac{1}{2}\Big)q_n + a^\dagger q_n = (d_n + 1)(a^\dagger q_n). \quad \checkmark \end{aligned}$$

So $a^\dagger$ raises the energy by 1 and a lowers it by 1: they are the *rising* and *lowering* operators.

6.3 Ground state and the full spectrum

Because H is positive semi-definite ($\langle v, Hv\rangle \geq 0$), the sequence $\ldots, a^2 q, aq, q$ of lowered eigenvectors must terminate. The ground state q_0 satisfies $a\, q_0 = 0$, so:

$$Hq_0 = \Big(a^\dagger a + \tfrac{1}{2}\Big)q_0 = \tfrac{1}{2}\, q_0. \tag{26}$$

The ground-state energy is $\frac{1}{2}$, not 0. This *zero-point energy* has no classical analogue—the system cannot "rest" at the bottom of the potential well.

All eigenstates and eigenvalues follow by repeated application of $a^\dagger$:

$$q_n = \frac{(a^\dagger)^n}{\sqrt{n!}}\, q_0, \qquad d_n = n + \tfrac{1}{2}, \quad n = 0, 1, 2, \ldots \tag{27}$$

6.4 Truncated matrix representation

Although H is infinite-dimensional, we can truncate to a finite basis $\{q_0, q_1, \ldots, q_{N-1}\}$. In this eigenbasis the matrix elements of a are $\langle q_n, a\, q_m\rangle = \sqrt{m}\,\delta_{n,m-1}$. The truncated 5×5 versions are:

$$a \approx \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{3} & 0 \\ 0 & 0 & 0 & 0 & \sqrt{4} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad H \approx \begin{pmatrix} \frac{1}{2} & & & & \\ & \frac{3}{2} & & & \\ & & \frac{5}{2} & & \\ & & & \frac{7}{2} & \\ & & & & \frac{9}{2} \end{pmatrix}. \tag{28}$$

H is diagonal in its own eigenbasis (expected), and a is a *superdiagonal shift operator*—it moves probability amplitude one rung up the energy ladder.

6.5 Coherent states and the classical limit

A *coherent state* $|\alpha\rangle$ ($\alpha \in \mathbb{C}$) is defined as an eigenstate of the annihilation operator:

$$a\,|\alpha\rangle = \alpha\,|\alpha\rangle\,. \tag{29}$$

Expanding in the energy eigenbasis $\{q_n\}$:

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\,|n\rangle\,. \tag{30}$$

The Born probabilities over energy levels are Poisson distributed with mean $|\alpha|^2$:

$$p(n) = |\langle n|\alpha\rangle|^2 = e^{-|\alpha|^2} \frac{|\alpha|^{2n}}{n!}. \tag{31}$$

Classical oscillation from the Born layer. Under the Schrödinger evolution e^{-iHt}, a coherent state remains coherent with time-dependent amplitude $\alpha(t) = \alpha e^{-it}$ (a rotation in the complex plane). The expected position is:

$$\begin{aligned} \langle X\rangle_t &= \langle \alpha(t)|\, X\, |\alpha(t)\rangle = \langle \alpha(t)|\, \tfrac{1}{\sqrt{2}}(a + a^\dagger)\, |\alpha(t)\rangle \\ &= \tfrac{1}{\sqrt{2}}\Big(\alpha(t) + \overline{\alpha(t)}\Big) = \sqrt{2}\,|\alpha|\cos(t + \phi), \end{aligned} \tag{32}$$

where $\phi = \arg\alpha$. This is *classical sinusoidal motion* emerging exactly from the quantum RNN. Coherent states are the quantum states that behave most like classical oscillators—they are the bridge between the two theories.

Why $a\,|\alpha\rangle = \alpha\,|\alpha\rangle$ is special. For a general energy eigenstate $|n\rangle$, the expected position is $\langle X\rangle = 0$ for all time: the state is stationary and centred. For a coherent state, $\langle X\rangle_t$ oscillates sinusoidally. The difference is that coherent states are *superpositions* of many energy levels, and the relative phases between levels

advance at different rates $e^{-id_n t}$—their interference produces the oscillating envelope. Remove the imaginary unit from the Schrödinger update and the phases stop rotating; the interference pattern freezes; the oscillation disappears.

7 The Feynman Path Integral

The Schrödinger equation tells us *how* the hidden state evolves. The Feynman path integral tells us *why* it is the only consistent choice. It also reframes the entire theory in terms of matrix multiplications, making the RNN analogy maximally explicit.

7.1 The propagator as a matrix element

The central object is the *propagator*

$$K(x_f, t;\, x_i, 0) \;=\; \langle x_f|\; e^{-iHt}\; |x_i\rangle\,, \tag{33}$$

the (x_f, x_i) matrix element of the time-evolution operator $U(t) = e^{-iHt}$. In ML terms, $U(t)$ is the recurrent weight matrix exponentiated; K is one entry of that matrix. The Born-rule amplitude for "start at x_i, measure position at time t, find x_f" is exactly K.

If we discretise position onto N_x grid points, $U(t)$ is a literal $N_x \times N_x$ complex matrix, and $K(x_f, t; x_i, 0)$ is the entry in row x_f, column x_i. The probability density is $|K|^2$.

7.2 Time slicing: the propagator as a product of matrices

The key identity is:

$$e^{-iHT} = \underbrace{e^{-iH\epsilon} \cdots e^{-iH\epsilon}}_{N \text{ factors}}, \qquad \epsilon = \frac{T}{N}. \tag{34}$$

This is exact for any N. It says: running the RNN for time T is identical to running it N steps of size ϵ.

Now insert a *resolution of identity* $I = \int dx\; |x\rangle\,\langle x|$ between every consecutive pair of factors:

$$\begin{aligned} K(x_f, T;\, x_i, 0) &= \langle x_f|\; e^{-iHT}\; |x_i\rangle \\ &= \int dx_1 \cdots dx_{N-1}\; \langle x_f|\, e^{-iH\epsilon}\, |x_{N-1}\rangle \cdots \langle x_1|\, e^{-iH\epsilon}\, |x_i\rangle\,. \end{aligned} \tag{35}$$

Each factor $\langle x_{k+1}|\, e^{-iH\epsilon}\, |x_k\rangle$ is a single entry of the short-time evolution matrix; integrating over all intermediate positions $(x_1, \ldots, x_{N-1})$ is a matrix multiplication:

$$\boxed{K = U^N, \qquad U_{x'x} = \langle x'|\, e^{-iH\epsilon}\, |x\rangle\,.} \tag{36}$$

A path $(x_i = x_0 \to x_1 \to \cdots \to x_{N-1} \to x_N = x_f)$ is one term in the matrix product; summing over all paths is matrix multiplication.

7.3 The short-time matrix element

For the Hamiltonian $H = P^2/2m + V(X)$ (kinetic plus potential), the short-time matrix element can be evaluated by inserting a momentum-basis resolution of identity $I = \int \frac{dp}{2\pi} |p\rangle \langle p|$:

$$\begin{aligned} U_{x'x} &= \langle x' | \, e^{-i(P^2/2m + V(X))\epsilon} \, |x\rangle \\ &\approx \langle x' | \, e^{-iP^2\epsilon/2m} e^{-iV(X)\epsilon} \, |x\rangle \qquad \text{(Trotter splitting, valid for small } \epsilon) \\ &= e^{-iV(x)\epsilon} \int \frac{dp}{2\pi} \, e^{ip(x'-x)} \, e^{-ip^2\epsilon/2m}. \end{aligned} \tag{37}$$

The p-integral is a Gaussian: $\int \frac{dp}{2\pi} e^{ip\Delta x - ip^2\epsilon/2m} = \sqrt{\frac{m}{2\pi i\epsilon}} \, e^{im(\Delta x)^2/2\epsilon}$, so:

$$U_{x'x} = \sqrt{\frac{m}{2\pi i\epsilon}} \, \exp\!\left(i\epsilon \left[\frac{m}{2} \left(\frac{x'-x}{\epsilon} \right)^2 - V(x) \right] \right). \tag{38}$$

The exponent is $i\epsilon \times \mathcal{L}(x, \dot{x})$, where $\mathcal{L} = \frac{1}{2}m\dot{x}^2 - V(x)$ is the classical *Lagrangian* and $\dot{x} \approx (x' - x)/\epsilon$ is the finite-difference velocity.

7.4 The path integral formula

Substituting (38) into (35) and taking $N \to \infty$ ($\epsilon \to 0$) gives the Feynman path integral:

$$K(x_f, T;\, x_i, 0) = \int_{x(0)=x_i}^{x(T)=x_f} \mathcal{D}x(t) \; \exp\!\left(i \int_0^T \mathcal{L}(x, \dot{x}) \, dt \right) = \int \mathcal{D}x \; e^{iS[x]}, \tag{39}$$

where $S[x] = \int_0^T \mathcal{L} \, dt$ is the classical *action* of the path $x(t)$, and $\int \mathcal{D}x$ sums over all continuous paths connecting x_i to x_f.

Physical interpretation. Every path from x_i to x_f contributes a complex amplitude $e^{iS[x]}$ of modulus 1. Paths whose action differs by $\sim 2\pi$ contribute with opposite phase and cancel; paths near the classical trajectory (where $\delta S = 0$ by Hamilton's principle) have stationary phase and reinforce. The classical path is the one that survives destructive interference; quantum fluctuations are all the other paths—they are suppressed but not zero, and their residual contribution is the origin of quantum corrections.

Remark. The saddle-point approximation to $\int \mathcal{D}x \, e^{iS[x]}$ recovers classical mechanics exactly. Perturbative corrections around the saddle are Feynman diagrams.

7.5 Concrete matrix multiplication: a three-slice example

Discretise position to three points $\{-1, 0, +1\}$ and time to $N = 3$ equal slices of width ϵ. The propagator from x_i to x_f in time $T = 3\epsilon$ is:

$$K(x_f, 3\epsilon;\, x_i, 0) = \left[U^3 \right]_{x_f,\, x_i}, \tag{40}$$

where U is the 3×3 short-time evolution matrix with entries $U_{x'x} \propto e^{iS_{\text{step}}(x',x)}$.

Free particle ($V = 0$) on a three-point grid.
Set $m = 1$, $\epsilon = 0.5$, grid $\{-1, 0, +1\}$. The normalisation constant $C = \sqrt{m/2\pi i\epsilon} = 1/\sqrt{i\pi}$ is common to all entries. The phase for each step is $e^{im(x'-x)^2/2\epsilon} = e^{i(x'-x)^2}$. The 3×3 matrix U (rows = destination, columns = source) is:

$$U = C \begin{pmatrix} e^{i\cdot 0} & e^{i\cdot 1} & e^{i\cdot 4} \\ e^{i\cdot 1} & e^{i\cdot 0} & e^{i\cdot 1} \\ e^{i\cdot 4} & e^{i\cdot 1} & e^{i\cdot 0} \end{pmatrix} = C \begin{pmatrix} 1 & e^{i} & e^{4i} \\ e^{i} & 1 & e^{i} \\ e^{4i} & e^{i} & 1 \end{pmatrix}.$$

The ($x' = 0$, $x = 0$) entry at $t = \epsilon$ is $C \cdot 1$: the particle stays at the origin with amplitude C. The ($x' = +1$, $x = -1$) entry is $C\, e^{4i}$: jumping two lattice sites in one step acquires a large phase e^{4i} and is strongly suppressed by the destructive interference from neighbouring paths.
Propagating for three steps:

$$K(\,\cdot\,, 3\epsilon;\,\cdot\,, 0) = U^3.$$

U^3 is computed as two ordinary 3×3 complex matrix multiplications. Each of the $3^3 = 27$ paths $(-1/0/+1) \to (\cdots) \to (\cdots) \to (-1/0/+1)$ contributes one term to this product; the matrix multiplication sums them all simultaneously. The Born probability to find the particle at x_f given it started at x_i is $|[U^3]_{x_f,x_i}|^2$, renormalised over x_f.

7.6 The path integral as a sum over RNN trajectories

The ML reading of the path integral is now immediate. The short-time matrix U is the RNN's one-step transition matrix in the position *eigenbasis*. The full propagator U^N is the RNN unrolled for N steps. The path $(x_0, x_1, \ldots, x_N)$ is a single hidden-state trajectory; the amplitude $\prod_k U_{x_{k+1},x_k}$ is the product of transition amplitudes along that trajectory; and the path integral sums (integrates) over all trajectories—exactly as a matrix power does.

Path integral	**RNN / ML**
Short-time propagator $U_{x'x}$	One-step transition matrix
Path $(x_0 \to x_1 \to \cdots \to x_N)$	One hidden-state trajectory
Amplitude $\prod_k U_{x_{k+1},x_k}$	Product of transition amplitudes
Sum over all paths $= U^N$	Matrix power = unrolled RNN
Stationary-phase path	Most probable trajectory
Action $S[x]$ in the exponent	Log-amplitude of the trajectory
Classical limit $(\hbar \to 0)$	Greedy / Viterbi decoding

The classical limit $\hbar \to 0$ (which we have set to 1 throughout, but can be restored by replacing S with $S/\hbar$) makes the phase $e^{iS/\hbar}$ oscillate infinitely rapidly for all paths except the one where $\delta S = 0$. Only that path—the classical trajectory—survives. This is the quantum analogue of greedy decoding in a sequence model: as the temperature goes to zero, the most probable path dominates. Quantum mechanics is classical mechanics plus all the off-path fluctuations whose amplitude is $e^{iS/\hbar}$.

8 Spinors as Structured Embeddings: the $\infty \times 2$ Hidden State

A real electron has *both* position and spin: at every point x in space it carries two simultaneous probability amplitudes, one for spin-up and one for spin-down. This section shows how to incorporate that structure by promoting the hidden state from a vector to a *matrix*. The result is a clean algebraic story—orbital mechanics and spin mechanics are two commuting operator families acting on opposite indices of the same matrix—and the various spinor types of relativistic physics (Weyl, Dirac, Majorana) differ only in how many columns that matrix has and what constraints tie them together.

8.1 The Pauli spinor as an $\infty \times 2$ matrix

A spin-$\frac{1}{2}$ particle requires two simultaneous wave functions—one per spin component—which we arrange as the two columns of a matrix:

$$\Psi \;=\; \Big(\psi_\uparrow \,\Big|\, \psi_\downarrow\Big), \qquad \Psi \;\in\; \mathbb{C}^{\infty\times 2}. \tag{41}$$

Row x and column $s \in \{\uparrow, \downarrow\}$ give the amplitude $\Psi_{xs} = \psi_s(x)$ to find the particle at position x with spin s. The norm condition $\|\Psi\|_F^2 = \sum_{x,s} |\Psi_{xs}|^2 = 1$ uses the Frobenius norm of the matrix, and is equivalent to requiring that the total probability (summed over positions and both spin states) equals 1.

ML reading. Ψ is an embedding table with two tokens. Each row $\Psi_{x,:} \in \mathbb{C}^2$ is the spinor at position x—the two-dimensional hidden state of the spin subsystem,

parametrised continuously by x. The "vocabulary" has only two elements (spin-up and spin-down), but their amplitudes vary continuously with position.

8.2 The Pauli matrices

The 2×2 spin operators acting on the column index of Ψ are the *Pauli matrices*:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{42}$$

They are worth understanding completely, because they are the atomic unit of all spinor physics: every 2×2 Hermitian matrix is a linear combination of $\{I, \sigma_x, \sigma_y, \sigma_z\}$, and the Pauli matrices alone carry the entire spin-$\frac{1}{2}$ algebra.

Basic properties. Each Pauli matrix is simultaneously:

- **Hermitian** ($\sigma_i^\dagger = \sigma_i$): a valid observable with real eigenvalues.
- **Traceless** ($\operatorname{Tr}\sigma_i = 0$): its two eigenvalues sum to zero.
- **Unitary** ($\sigma_i^\dagger \sigma_i = I$): applying it twice returns to the identity.
- **Involutory** ($\sigma_i^2 = I$): eigenvalues are ±1.

The eigenvalues ±1 are the two outcomes of a spin measurement along axis i: spin-up ($+1$) or spin-down (-1).

Commutation relations. The Pauli matrices do not commute; their commutator encodes the non-commutativity of spin measurements along different axes:

$$[\sigma_i,\, \sigma_j] = 2i\, \epsilon_{ijk}\, \sigma_k, \tag{43}$$

where ϵ_{ijk} is the Levi-Civita symbol. The spin operators $S_i = \frac{1}{2}\sigma_i$ therefore satisfy $[S_i, S_j] = i\epsilon_{ijk}S_k$—the angular momentum algebra $\mathfrak{su}(2)$. The factor of $\frac{1}{2}$ is what makes spin-$\frac{1}{2}$ half-integer.

Anticommutation and the multiplication rule. The Pauli matrices also anticommute:

$$\{\sigma_i,\, \sigma_j\} = 2\,\delta_{ij}\, I, \tag{44}$$

which combined with the commutation relation gives the complete multiplication rule:

$$\sigma_i\, \sigma_j = \delta_{ij}\, I + i\, \epsilon_{ijk}\, \sigma_k. \tag{45}$$

This single identity encodes the full algebra. For example: $\sigma_x\sigma_y = i\sigma_z$ and $\sigma_x^2 = I$.

Completeness. Any 2×2 Hermitian matrix M expands uniquely in the Pauli basis:

$$M = a_0\, I + a_x\, \sigma_x + a_y\, \sigma_y + a_z\, \sigma_z, \qquad a_\mu \in \mathbb{R}, \tag{46}$$

with $a_0 = \frac{1}{2}\operatorname{Tr} M$ and $a_i = \frac{1}{2}\operatorname{Tr}(\sigma_i M)$. Every spin Hamiltonian, every qubit gate, every 2×2 observable lives in this four-dimensional real space. The spacetime matrix $X = x^\mu \sigma_\mu$ of diagram 3 is precisely this expansion applied to a spacetime 4-vector: the Pauli matrices are the bridge between spinor algebra and spacetime geometry.

Rotations from the exponential. Because $(\hat{n} \cdot \vec{\sigma})^2 = I$ for any unit vector $\hat{n}$ (immediate from (44)), the Taylor series of the matrix exponential collapses into:

$$e^{-i\frac{\theta}{2}\hat{n}\cdot\vec{\sigma}} = I\cos\tfrac{\theta}{2} - i(\hat{n}\cdot\vec{\sigma})\sin\tfrac{\theta}{2}. \tag{47}$$

This is the $SU(2)$ element L that appears in diagram 3. It rotates the spinor column by angle θ around axis $\hat{n}$. The factor of $\frac{1}{2}$ means a full 360 degree rotation gives $e^{-i\pi\hat{n}\cdot\vec{\sigma}} = -I$: spinors change sign under a full rotation and require 720 degree to return to themselves. This sign is physically observable in neutron interferometry experiments, and is ultimately the origin of Fermi–Dirac statistics and the Pauli exclusion principle.

180 **degree rotation about x; verifying the sign.**
Set $\hat{n} = \hat{x}$, $\theta = \pi$:

$$e^{-i\frac{\pi}{2}\sigma_x} = I\cos\tfrac{\pi}{2} - i\sigma_x\sin\tfrac{\pi}{2} = -i\sigma_x = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}.$$

Acting on spin-up $|\uparrow\rangle = (1,0)^T$:

$$-i\sigma_x\,|\uparrow\rangle = -i\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = -i\begin{pmatrix} 0 \\ 1 \end{pmatrix} = -i\,|\downarrow\rangle\,.$$

A 180 degree rotation about x sends spin-up to spin-down with phase $-i$. Applying the rotation twice: $(-i\sigma_x)^2 = (-i)^2\sigma_x^2 = (-1)I = -I$. A 360ř rotation gives $-|\uparrow\rangle$, not $+|\uparrow\rangle$, confirming that spinors acquire a sign under full rotation.

ML reading. The Pauli matrices are the weight matrices of the 2×2 spin RNN. When the Hamiltonian contains $\vec{B}\cdot\vec{\sigma}/2$, the spin column of the hidden state Ψ rotates in the direction of $\vec{B}$ at the Larmor frequency. The three Pauli matrices span all possible spin rotations; the exponential (47) is the one-parameter family of spin weight matrices along any axis; and the completeness relation (46) says that the Pauli matrices form a basis for the weight-matrix space of any 2×2 quantum system.

8.3 Two-sided operator algebra: the key structural fact

The $\infty \times 2$ shape endows the hidden state with **two independent, commuting families of operators**:

$$\underbrace{A\,\Psi}_{\substack{\text{left action}\\ (\infty\times\infty)\text{ on row index}}} \qquad \text{and} \qquad \underbrace{\Psi\,B}_{\substack{\text{right action}\\ (2\times2)\text{ on column index}}} \quad . \tag{48}$$

They commute because matrix multiplication associates: $A(\Psi B) = (A\Psi)B$. This simple observation organises all of quantum mechanics for spin-$\frac{1}{2}$ particles.

Left action = orbital mechanics. Any operator that acts on position is a left action; it is applied independently and identically to each column:

$$[X\,\Psi]_{xs} = x\,\Psi_{xs}, \tag{49}$$

$$[P\,\Psi]_{xs} = -i\,\partial_x\,\Psi_{xs}, \tag{50}$$

$$[L_i\,\Psi]_{xs} = [L_i\psi_s](x). \tag{51}$$

Position, momentum, and orbital angular momentum are *blind to spin*: they see the same operator in each column.

Right action = spin mechanics. Any operator that acts on the spinor index is a right action; it mixes the two columns at each position without touching x:

$$[\Psi\,\sigma_i^T]_{xs} = \sum_{s'} \Psi_{xs'}\,(\sigma_i)_{ss'} = [\sigma_i\,\psi(x)]_s. \tag{52}$$

Right-multiplying Ψ by σ_i^T is exactly the Pauli spin operator σ_i applied to the spinor $\psi(x)$ at each point x. The orbital wave function is unchanged.

Mixed action = spin–orbit coupling. An operator of the form $A_i\,\Psi\,\sigma_i^T$ acts on both indices simultaneously and cannot be factored into a pure left or pure right term. The canonical example is:

$$[\vec{L}\cdot\vec{S}]\,\Psi \;=\; \sum_i L_i\,\Psi\,\frac{\sigma_i^T}{2}, \tag{53}$$

where L_i acts on the row (orbital) and $\frac{\sigma_i}{2}$ acts on the column (spin).

Term	**Matrix form**	**Acts on**	**Type**
Kinetic $P^2/2m$	$\dfrac{P^2}{2m}\Psi$	Row only	Left
Potential $V(X)$	$V(X)\,\Psi$	Row only	Left
Zeeman $\vec{\sigma}\cdot\vec{B}$	$\sum_i B_i(X)\,\Psi\,\sigma_i^T$	Both	Mixed
Spin–orbit $\vec{L}\cdot\vec{S}$	$\sum_i L_i\,\Psi\,\dfrac{\sigma_i^T}{2}$	Both	Mixed

8.4 The Pauli Hamiltonian in two-sided form

For an electron (charge $-e$, mass m) in an electromagnetic field $(\vec{A}, \phi)$ with magnetic field $\vec{B} = \nabla \times \vec{A}$, the Pauli Hamiltonian is:

$$H_{\text{Pauli}} \;=\; \frac{1}{2m}\Big(\vec{P} + e\vec{A}(X)\Big)^2 \;-\; \frac{e}{2m}\,\vec{B}(X)\cdot\vec{\sigma} \;-\; e\phi(X). \tag{54}$$

Acting on $\Psi \in \mathbb{C}^{\infty\times 2}$, this separates cleanly:

$$H_{\text{Pauli}}\,\Psi \;=\; \underbrace{\left[\frac{\big(\vec{P} + e\vec{A}\big)^2}{2m} - e\phi\right]\Psi}_{\text{pure left action: same on both columns}} \;-\; \underbrace{\frac{e}{2m}\sum_i B_i(X)\,\Psi\,\sigma_i^T}_{\text{mixed: position-dependent spin}} \quad . \tag{55}$$

The kinetic and scalar potential terms are left-acting: they update each spin column independently, as though spin did not exist. The magnetic term couples both indices: $B_i(X)$ is an orbital operator (acting on row) and σ_i^T is a spin operator (acting on column), multiplied together. Spin–orbit coupling $\propto \vec{L}\cdot\vec{S}$ has the same mixed structure.

Larmor precession: pure right-acting Hamiltonian.
Set $\vec{A} = 0$, $\phi = 0$, $\vec{B} = B\hat{z}$ uniform, and suppress spatial motion. The Hamiltonian reduces to a pure right action:

$$H = \omega_L\,\frac{\sigma_z}{2}, \qquad \omega_L = -\frac{eB}{m},$$

which is exactly the 2×2 RNN of qubit with $H = \omega_L\sigma_z/2$. Acting on the full matrix Ψ, the Schrödinger equation $i\partial_t\Psi = \Psi\,(\omega_L\sigma_z/2)$ (right action) yields:

$$\Psi(t) = \Psi(0)\,e^{-i\omega_L\sigma_z t/2} = \Psi(0)\begin{pmatrix} e^{-i\omega_L t/2} & 0 \\ 0 & e^{i\omega_L t/2}\end{pmatrix}.$$

Each row's spinor precesses around the z-axis at the Larmor frequency ω_L. The position wave function is completely unaffected—the Hamiltonian never touched the row index.

Stern–Gerlach: Born rule on the matrix hidden state.
Prepare the particle in a Gaussian wave packet with spin pointing at polar angle θ from $\hat{z}$:

$$\Psi_{xs} = \phi(x)\cdot u_s, \qquad \phi(x) = \pi^{-1/4}e^{-x^2/2}, \qquad u = \begin{pmatrix}\cos\frac{\theta}{2} \\ \sin\frac{\theta}{2}\end{pmatrix}.$$

A Stern–Gerlach device measures σ_z. The eigenvectors of the right-acting spin operator are $q_+ = e_1 = (1,0)^T$ and $q_- = e_2 = (0,1)^T$. The Born probabilities

are:

$$p\left(+\tfrac{1}{2}\right) = \int |\Psi_{x,\uparrow}|^2\, dx = \int |\phi(x)|^2 |\cos\tfrac{\theta}{2}|^2\, dx = \cos^2\tfrac{\theta}{2},$$
$$p\left(-\tfrac{1}{2}\right) = \sin^2\tfrac{\theta}{2}.$$

The *position* wave function $\phi(x)$ factored out completely: this measurement cares only about the column structure of Ψ. After measurement (Born layer), say outcome $+\frac{1}{2}$, the hidden state resets to $\Psi \leftarrow (\phi(x) \mid 0)$—the second column is zeroed, the first column is unchanged. The measurement was a hard assignment on the column index; the row index was unaffected.

8.5 $SU(2)$ as the symmetry of the column index

A global spin rotation by $U \in SU(2)$ maps each spinor $\psi(x) \mapsto U\psi(x)$. In matrix notation, this is a right action on Ψ:

$$\psi(x) \mapsto U\psi(x) \quad \Longleftrightarrow \quad \Psi \;\mapsto\; \Psi\, U^T. \tag{56}$$

(Right multiplication by U^T rotates each row-spinor by U; the transpose arises from the row-vs-column convention and reduces to U itself for the physically important cases of z- and x-axis rotations, since $\sigma_z^T = \sigma_z$ and $\sigma_x^T = \sigma_x$.)

This is a symmetry of the free Hamiltonian ($\vec{B} = 0$):

$$H_0\,(\Psi U^T) = (H_0\Psi)\,U^T, \tag{57}$$

since H_0 acts purely on the left. A uniform magnetic field $\vec{B}$ picks a preferred direction and breaks $SU(2)$ down to $U(1)$ rotations about $\hat{B}$.

The spin-$\frac{1}{2}$ phase. A full 2π rotation $U = e^{-i\pi\hat{n}\cdot\vec{\sigma}} = -I$ acts as $\Psi \mapsto \Psi(-I) = -\Psi$. The hidden state acquires a global minus sign. Since all observables are quadratic in Ψ, this sign is invisible in any single measurement—but it produces measurable interference effects when two paths with different rotation histories are recombined (as in neutron interferometry). In the RNN language: the Bohr layer computes $|\langle h_t, q_o\rangle|^2$, which is invariant under $h_t \mapsto -h_t$. The sign lives in the hidden state but cannot be read out by a single Born-layer query.

8.6 Total angular momentum: the two-sided generator

A physical rotation of space must act on *both* indices of Ψ simultaneously: the spatial wave function rotates (row index), and the spinor rotates (column index), both by the same angle. A rotation by ϕ about $\hat{z}$ is therefore a *two-sided* action:

$$\Psi \;\longmapsto\; e^{i\phi L_z}\,\Psi\, e^{i\phi\sigma_z/2}. \tag{58}$$

Differentiating with respect to ϕ at $\phi = 0$ gives the generator:

$$\boxed{J_z\,\Psi \;=\; L_z\,\Psi \;+\; \Psi\,\frac{\sigma_z}{2}.} \tag{59}$$

The two terms are the orbital angular momentum L_z (left action, on rows) and the spin angular momentum $S_z = \sigma_z/2$ (right action, on columns). They commute trivially—L_z and $\sigma_z/2$ act on different indices—and their sum is the total angular momentum $J_z = L_z + S_z$.

Consistency check.

$$\frac{d}{d\phi}\bigg|_{\phi=0} e^{i\phi L_z}\Psi e^{i\phi\sigma_z/2} = iL_z\Psi + \Psi\,\frac{i\sigma_z}{2} = i\Big(L_z\Psi + \Psi\,\frac{\sigma_z}{2}\Big) = iJ_z\Psi. \quad \checkmark$$

The Lie algebra element J_z acts as a two-sided derivation: L_z on the left and $\sigma_z/2$ on the right.

Eigenstates and half-integer angular momentum. Under the rotation (58), the state Ψ is an eigenstate of J_z with eigenvalue m_j if and only if it is simultaneously an eigenstate of L_z with eigenvalue m_ℓ and of S_z with eigenvalue $m_s = \pm\frac{1}{2}$, with $m_j = m_\ell + m_s$. Since m_s is a half-integer, J_z can have half-integer eigenvalues even though the *left* factor $e^{i\phi L_z}$ alone would only produce integer eigenvalues. The half-integer angular momenta of spin-$\frac{1}{2}$ particles are a direct consequence of the two-sided matrix structure—they cannot arise from any operator acting only on the row index.

The $j = \frac{1}{2}$ eigenstates.
Take a hydrogen s-wave: $\psi_{\ell=0,m=0}(x) = R(r)Y_0^0(\hat{r})$ (spherically symmetric, L_z eigenvalue 0). Pair it with each spin eigenstate:

$$\Psi_+ = R(r)\Big(\,e_1\,\Big|\,0\Big), \qquad \Psi_- = R(r)\Big(\,0\,\Big|\,e_2\Big),$$

where $e_1 = (1,0)^T$, $e_2 = (0,1)^T$.
$J_z\Psi_+ = 0\cdot\Psi_+ + R(r)(e_1|0)\,\frac{\sigma_z}{2} = R(r)(\frac{1}{2}e_1|0) = \frac{1}{2}\Psi_+$.
$J_z\Psi_- = 0\cdot\Psi_- + R(r)(0|e_2)\,\frac{\sigma_z}{2} = R(r)(0|-\frac{1}{2}e_2) = -\frac{1}{2}\Psi_-$.
The two states are $j = \frac{1}{2}$, $m_j = \pm\frac{1}{2}$ eigenstates. The half-integer eigenvalue came entirely from the right (column) action.

9 Quantum Field Theory: Fock Space Is the Only Change

In a GPT language model, the residual stream $h_t \in \mathbb{R}^d$ encodes extraordinarily rich semantics: syntactic role, factual associations, discourse context, and sentiment all coexist in a single vector. The vector is not a list of labelled slots; it is a superposition over directions in a high-dimensional space, each direction carrying a single interpretable concept. Sparse autoencoder research calls these directions *monosemantic*: each one encodes exactly one thing.

The Fock-space hidden state is built on exactly the same idea.

9.1 The basis vectors of Fock space

A basis vector of Fock space is labelled by a complete specification of a particle configuration: a definite number n of particles, each with a definite momentum. We write:

$$|n;\, \mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_n\rangle\,, \tag{60}$$

meaning: exactly n particles present, with momenta $\mathbf{k}_1, \ldots, \mathbf{k}_n$. For $n = 0$ there is a unique basis vector, the vacuum $|0\rangle$. These basis vectors are:

- **Mutually orthogonal**: $\Big\langle n;\, \mathbf{k}_1, \ldots \Big|\, m;\, \mathbf{q}_1, \ldots \Big\rangle = 0$ whenever $n \neq m$ or any momenta differ.
- **Monosemantic**: each one encodes exactly one particle configuration—a definite number of particles with definite momenta, nothing else. This is the quantum analogue of a one-hot vector: the basis vector $|n; \mathbf{k}_1, \ldots, \mathbf{k}_n\rangle$ is "hot" at the single configuration $(n, \mathbf{k}_1, \ldots, \mathbf{k}_n)$ and zero everywhere else.
- **Complete**: every state in Fock space can be expressed as a superposition of these basis vectors.

9.2 The Fock hidden state as an explicit superposition

The general Fock hidden state is a superposition over all basis vectors (60), summing over all particle numbers and integrating over all momentum configurations:

$$\begin{aligned} h_t = {} & c_0(t)\ |0\rangle \\ & + \int \frac{d^3\mathbf{k}_1}{(2\pi)^3}\ c_1(\mathbf{k}_1;\, t)\ |1;\, \mathbf{k}_1\rangle \\ & + \int \frac{d^3\mathbf{k}_1}{(2\pi)^3}\frac{d^3\mathbf{k}_2}{(2\pi)^3}\ c_2(\mathbf{k}_1, \mathbf{k}_2;\, t)\ |2;\, \mathbf{k}_1, \mathbf{k}_2\rangle \\ & + \cdots \\ & + \int \prod_{j=1}^{n} \frac{d^3\mathbf{k}_j}{(2\pi)^3}\ c_n(\mathbf{k}_1, \ldots, \mathbf{k}_n;\, t)\ |n;\, \mathbf{k}_1, \ldots, \mathbf{k}_n\rangle \\ & + \cdots \end{aligned} \tag{61}$$

The coefficient $c_n(\mathbf{k}_1, \ldots, \mathbf{k}_n;\, t)$ is the *probability amplitude* for finding exactly n particles with momenta $\mathbf{k}_1, \ldots, \mathbf{k}_n$ at time t.

The integral is a continuous sum. The integral $\int \frac{d^3\mathbf{k}}{(2\pi)^3}$ is simply a sum over all momentum values $\mathbf{k}$, where the momentum index is continuous rather than discrete. The factor $(2\pi)^{-3}$ is a conventional normalisation, the same as the normalisation of the Fourier transform: it ensures that completeness and orthogonality relations take their standard form. For a finite volume V (a box), momenta are discrete, and the integral becomes an ordinary sum $\frac{1}{V}\sum_{\mathbf{k}}$. In the infinite-volume limit $V \to \infty$, the sum becomes the integral. There is no conceptual difference between a discrete

sum and a continuous integral in this context; both are linear superpositions over basis vectors.

Normalisation. The unit-norm condition $\|h_t\|^2 = 1$ becomes:

$$\|h_t\|^2 = |c_0|^2 + \int \frac{d^3\mathbf{k}_1}{(2\pi)^3}\, |c_1(\mathbf{k}_1)|^2 + \int \frac{d^3\mathbf{k}_1}{(2\pi)^3}\frac{d^3\mathbf{k}_2}{(2\pi)^3}\, |c_2(\mathbf{k}_1, \mathbf{k}_2)|^2 + \cdots = 1. \tag{62}$$

The Born rule gives the probability of each outcome: the probability of finding exactly n particles with momenta in a small cell around $(\mathbf{k}_1, \ldots, \mathbf{k}_n)$ is $|c_n(\mathbf{k}_1, \ldots, \mathbf{k}_n)|^2 \prod_j \frac{d^3\mathbf{k}_j}{(2\pi)^3}$.

9.3 Each basis vector is monosemantic

In sparse autoencoder interpretations of GPT, a monosemantic direction f_k in the residual stream encodes exactly one concept: "France", "past tense", "toxic language", and so on. The embedding $h = \sum_k \alpha_k f_k$ is a superposition of such directions, each weighted by its relevance to the current token.

The Fock basis vector $|n; \mathbf{k}_1, \ldots, \mathbf{k}_n\rangle$ is monosemantic in exactly this sense: it encodes exactly one particle configuration, with no ambiguity. The coefficient $c_n(\mathbf{k}_1, \ldots, \mathbf{k}_n;\, t)$ plays the role of α_k: it is the amplitude (complex-valued weight) with which the hidden state h_t "attends to" that particular configuration. The superposition (61) is the quantum analogue of the sparse superposition of monosemantic features.

There are two differences from the GPT case. First, the weights $c_n(\mathbf{k}_1, \ldots, \mathbf{k}_n;\, t)$ are complex rather than real: they carry phase information that determines how different basis vectors interfere under the Schrödinger evolution. Second, the superposition is exact (not approximate or sparse): the hidden state is generically a superposition over all particle configurations simultaneously, with the interaction Hamiltonian continuously redistributing amplitude among them.

9.4 Why this superposition belongs to Fock space

Fock space is defined as the direct sum of fixed-particle-number Hilbert spaces:

$$\mathcal{F} = \mathcal{H}^{(0)} \oplus \mathcal{H}^{(1)} \oplus \mathcal{H}^{(2)} \oplus \cdots \tag{63}$$

Here $\mathcal{H}^{(n)}$ is the Hilbert space of exactly n-particle states: it is spanned by all basis vectors $|n; \mathbf{k}_1, \ldots, \mathbf{k}_n\rangle$ for all momentum values. The direct sum (63) means that any element of $\mathcal{F}$ can be written as a sum of elements, one from each $\mathcal{H}^{(n)}$.

The superposition (61) is exactly of this form. The n-particle piece:

$$h_t^{(n)} = \int \prod_{j=1}^{n} \frac{d^3\mathbf{k}_j}{(2\pi)^3}\; c_n(\mathbf{k}_1, \ldots, \mathbf{k}_n;\, t)\; |n;\, \mathbf{k}_1, \ldots, \mathbf{k}_n\rangle \;\in\; \mathcal{H}^{(n)} \tag{64}$$

lives entirely within $\mathcal{H}^{(n)}$—it is a superposition only of n-particle basis vectors. The full Fock hidden state is the direct sum of these pieces:

$$h_t = h_t^{(0)} \oplus h_t^{(1)} \oplus h_t^{(2)} \oplus \cdots \in \mathcal{H}^{(0)} \oplus \mathcal{H}^{(1)} \oplus \mathcal{H}^{(2)} \oplus \cdots = \mathcal{F}. \tag{65}$$

The direct sum is the mathematical structure that accommodates superpositions across sectors with different particle numbers. Each sector $\mathcal{H}^{(n)}$ is an orthogonal subspace of $\mathcal{F}$; the direct sum is the union of all of them, allowing the hidden state to have components in multiple sectors simultaneously.

There is nothing more to Fock space than this. A Fock hidden state is a superposition of monosemantic basis vectors, each labelled by a particle configuration $(n, \mathbf{k}_1, \ldots, \mathbf{k}_n)$, with complex amplitudes $c_n(\mathbf{k}_1, \ldots, \mathbf{k}_n;\, t)$. The direct sum structure simply organises these basis vectors by particle number. There is no block structure, no concatenation to maintain, no special data structure required. The Fock hidden state is a vector in a Hilbert space, exactly as the GPT residual stream is a vector in $\mathbb{R}^d$. The only difference is the richness of the index that labels the basis vectors.

9.5 The Schrödinger equation on Fock space

The Schrödinger equation:

$$\frac{dh_t}{dt} = -iH\, h_t \tag{66}$$

acts on $h_t \in \mathcal{F}$ in exactly the same way as on any Hilbert space. The Hamiltonian H is a linear operator on $\mathcal{F}$; it maps each basis vector to a linear combination of basis vectors. The free Hamiltonian H_{free} maps each basis vector $|n; \mathbf{k}_1, \ldots, \mathbf{k}_n\rangle$ to a scalar multiple of itself (phase rotation at the energy of those particles), preserving particle number. The interaction Hamiltonian H_{int} maps $|n; \mathbf{k}_1, \ldots, \mathbf{k}_n\rangle$ to a linear combination of basis vectors with different particle numbers—this is particle creation and annihilation.

When H_{int} acts on h_t, it rotates amplitude from one set of basis vectors to another. The coefficients $c_n(\mathbf{k}_1, \ldots, \mathbf{k}_n;\, t)$ evolve continuously, with amplitude flowing between different particle configurations. The story of particle creation, propagation, scattering, and annihilation is the story of how the Schrödinger equation redistributes amplitude among monosemantic basis vectors over time.

QM (fixed particle number)	**QFT (variable particle number)**
Basis: $\|\mathbf{k}\rangle$, one index	Basis: $\|n; \mathbf{k}_1, \ldots, \mathbf{k}_n\rangle$, richer index
$h_t = \int \frac{d^3k}{(2\pi)^3} c(\mathbf{k}; t) \|\mathbf{k}\rangle$	$h_t = \sum_n \int \prod_j \frac{d^3k_j}{(2\pi)^3} c_n(\mathbf{k}_{1\cdots n}; t) \|n; \mathbf{k}_{1\cdots n}\rangle$
$\dot{h}_t = -iHh_t$	$\dot{h}_t = -iHh_t$ (unchanged)
H rotates amplitude among $\|\mathbf{k}\rangle$	H also rotates amplitude among different n
Born rule: $\|c(\mathbf{k})\|^2$	Born rule: $\|c_n(\mathbf{k}_1, \ldots, \mathbf{k}_n)\|^2$

The right column is obtained from the left by enriching the basis index. The RNN equations are identical.

10 Lorentz transformations, rotations, boosts, and Einstein notation

Quantum field satisfies special relativity and is invariant under Lorentz transformations.

10.1 Spacetime and the invariant interval

In special relativity, spacetime is $\mathbb{R}^4$ with coordinates $x^\mu = (x^0, x^1, x^2, x^3) = (ct, x, y, z)$ (setting $c = 1$ throughout). The fundamental invariant is the *spacetime interval*:

$$s^2 = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = \eta_{\mu\nu}\, x^\mu x^\nu, \tag{67}$$

where $\eta_{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)$ is the *Minkowski metric*. The interval s^2 is the same in every inertial frame. Positive s^2 is timelike (causally connected), zero is lightlike (on the light cone), negative is spacelike (causally disconnected).

ML reading. The Minkowski metric $\eta_{\mu\nu}$ is the inner product matrix of spacetime. Ordinary Euclidean space has inner product matrix $\delta_{\mu\nu} = \mathrm{diag}(+1, +1, +1)$; Minkowski space has $\eta_{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)$. The sign flip on the spatial components is what distinguishes time from space, and it is the source of all the counterintuitive features of special relativity.

10.2 Einstein index notation: upper and lower indices

Physics uses a compact notation that makes Lorentz invariance manifest. Every repeated index—one upper, one lower—implies a sum. This is the *Einstein summation convention*:

$$\eta_{\mu\nu}\, x^\mu x^\nu \;\equiv\; \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} \eta_{\mu\nu}\, x^\mu x^\nu. \tag{68}$$

The rule is: a free index (appearing once) labels a component; a contracted index (appearing twice, once up and once down) is summed over.

Upper indices (contravariant components). x^μ is a *contravariant* vector—it transforms by the Lorentz matrix Λ:

$$x^\mu \longmapsto x'^\mu = \Lambda^\mu{}_\nu \, x^\nu. \tag{69}$$

Here $\Lambda^\mu{}_\nu$ is the Lorentz matrix with the first index (row) up and the second index (column) down.

Lower indices (covariant components). The metric $\eta_{\mu\nu}$ is used to *lower* an index:

$$x_\mu = \eta_{\mu\nu} \, x^\nu = (x^0, \, -x^1, \, -x^2, \, -x^3). \tag{70}$$

The covariant vector x_μ transforms by the inverse-transpose of Λ:

$$x_\mu \longmapsto x'_\mu = (\Lambda^{-T})_\mu{}^\nu \, x_\nu = \Lambda_\mu{}^\nu \, x_\nu. \tag{71}$$

A contracted upper-lower pair $x^\mu x_\mu = x^\mu \eta_{\mu\nu} x^\nu = s^2$ is Lorentz invariant: it has no free indices.

The inverse metric. The inverse of $\eta_{\mu\nu}$ is $\eta^{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)$ (coincidentally the same matrix for Minkowski), used to *raise* an index:

$$x^\mu = \eta^{\mu\nu} \, x_\nu. \tag{72}$$

Derivatives. The partial derivative $\partial/\partial x^\mu$ with respect to a contravariant coordinate is itself *covariant*—it carries a lower index:

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \left(\frac{\partial}{\partial t}, \, \frac{\partial}{\partial x}, \, \frac{\partial}{\partial y}, \, \frac{\partial}{\partial z} \right). \tag{73}$$

The d'Alembertian (wave operator) is the Lorentz-invariant contraction:

$$\partial^\mu \partial_\mu = \eta^{\mu\nu} \partial_\mu \partial_\nu = \partial_t^2 - \nabla^2 = \Box. \tag{74}$$

The Klein–Gordon equation $(\Box + m^2)\phi = 0$ is Lorentz invariant because $\Box$ is a contracted pair with no free indices.

Index counting in $\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2} m^2 \phi^2$.
The kinetic term: $\partial_\mu \phi$ has one lower free index μ; $\partial^\mu \phi = \eta^{\mu\nu} \partial_\nu \phi$ has one upper free index μ. Contracting: $(\partial_\mu \phi)(\partial^\mu \phi)$ has no free indices—it is a Lorentz scalar. The mass term $m^2 \phi^2$ has no indices. Every term in $\mathcal{L}$ is a scalar. Since d^4x is also a scalar (the Lorentz Jacobian $|\det \Lambda| = 1$), the action $S = \int d^4x \, \mathcal{L}$ is invariant. ✓

10.3 Lorentz transformations: definition

A *Lorentz transformation* is a 4×4 real matrix $\Lambda^\mu{}_\nu$ that preserves the spacetime interval (67):

$$\eta_{\mu\nu}\,\Lambda^\mu{}_\rho\,\Lambda^\nu{}_\sigma = \eta_{\rho\sigma}, \qquad \text{i.e.} \qquad \Lambda^T \eta\, \Lambda = \eta. \tag{75}$$

In matrix notation: $\Lambda^T \eta \Lambda = \eta$. This is the defining equation of the *Lorentz group* $O(3,1)$. Taking the determinant: $(\det \Lambda)^2 = 1$, so $\det \Lambda = \pm 1$.

The *proper orthochronous* Lorentz group $SO^+(3,1)$ consists of transformations with $\det \Lambda = +1$ and $\Lambda^0{}_0 \geq 1$ (time-orientation preserving). This is the physically relevant subgroup for continuous transformations connected to the identity. It has six parameters: three rotations and three boosts.

10.4 Rotations

A *rotation* is a Lorentz transformation that acts only on the spatial components and leaves $x^0 = t$ unchanged. For a rotation by angle θ about the z-axis:

$$\Lambda_{\text{rot},z}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{76}$$

Rotations form a subgroup $SO(3) \subset SO^+(3,1)$. They preserve both the time component x^0 and the spatial length $|\mathbf{x}|^2$ separately, as well as the full interval s^2.

The generators of infinitesimal rotations are the angular momentum matrices J_i. For rotation about the z-axis:

$$J_z = -i \frac{d}{d\theta} \Lambda_{\text{rot},z}\bigg|_{\theta=0} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{77}$$

so $\Lambda_{\text{rot},z}(\theta) = e^{-i\theta J_z}$. The three rotation generators satisfy $[J_i, J_j] = i\epsilon_{ijk} J_k$—the $\mathfrak{su}(2) \cong \mathfrak{so}(3)$ algebra, which is the same algebra as the Pauli spin matrices (but now acting on spacetime rather than on spinors).

10.5 Boosts

A *boost* is a Lorentz transformation that mixes the time component x^0 with one spatial component. For a boost with velocity v along the x-axis (setting $c = 1$):

$$\Lambda_{\text{boost},x}(v) = \begin{pmatrix} \gamma & -\gamma v & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \gamma = \frac{1}{\sqrt{1 - v^2}}. \tag{78}$$

It is conventional and algebraically cleaner to parametrise by the *rapidity* $\xi = \tanh^{-1} v$, so that $\gamma = \cosh\xi$ and $\gamma v = \sinh\xi$:

$$\Lambda_{\text{boost},x}(\xi) = \begin{pmatrix} \cosh\xi & -\sinh\xi & 0 & 0 \\ -\sinh\xi & \cosh\xi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = e^{-i\xi K_x}, \tag{79}$$

where K_x is the boost generator along x:

$$K_x = -i\frac{d}{d\xi}\Lambda_{\text{boost},x}\Big|_{\xi=0} = \begin{pmatrix} 0 & -i & 0 & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{80}$$

Rapidity adds linearly for collinear boosts: two boosts of rapidity ξ_1 and ξ_2 along the same axis compose to rapidity $\xi_1 + \xi_2$. This is the Lorentz analogue of angle addition for rotations.

Key difference from rotations. Rotations involve sin and cos; boosts involve sinh and cosh. The rotation matrix is orthogonal ($\Lambda^T\Lambda = I$); the boost matrix is not orthogonal in the Euclidean sense, but satisfies the Lorentz condition $\Lambda^T\eta\Lambda = \eta$. Rotations act on a compact parameter space ($\theta \in [0, 2\pi)$); boosts act on a non-compact parameter space ($\xi \in (-\infty, \infty)$). This non-compactness means that finite-dimensional representations of the Lorentz group cannot be unitary—a fact that forces spinors to transform non-unitarily under boosts.

10.6 The Lie algebra of the Lorentz group

The full Lorentz group $SO^+(3,1)$ has six generators: three rotations J_i and three boosts K_i ($i = x, y, z$). Their commutation relations are:

$$[J_i, J_j] = i\,\epsilon_{ijk}\,J_k, \tag{81}$$

$$[K_i, K_j] = -i\,\epsilon_{ijk}\,J_k, \tag{82}$$

$$[J_i, K_j] = i\,\epsilon_{ijk}\,K_k. \tag{83}$$

The first relation (81) is the $\mathfrak{su}(2)$ rotation algebra. The second (82) says that two boosts in different directions produce a rotation (the *Wigner rotation*—the physical origin of Thomas precession). The third (83) says that boosts transform as vectors under rotations.

The algebra (81)–(83) is $\mathfrak{so}(3,1)$. Defining the complexified combinations:

$$A_i = \tfrac{1}{2}(J_i + iK_i), \qquad B_i = \tfrac{1}{2}(J_i - iK_i), \tag{84}$$

one finds $[A_i, A_j] = i\epsilon_{ijk}A_k$, $[B_i, B_j] = i\epsilon_{ijk}B_k$, $[A_i, B_j] = 0$. The Lorentz algebra *factorises* into two independent $\mathfrak{su}(2)$ algebras. This is the algebraic origin of the (j_L, j_R) classification of representations introduced in diagram 3.

Lorentz invariant from index contraction.
The electromagnetic field-strength tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ has two lower indices. Under a Lorentz transformation:

$$F_{\mu\nu} \longmapsto F'_{\mu\nu} = \Lambda_\mu{}^\rho \, \Lambda_\nu{}^\sigma \, F_{\rho\sigma}.$$

To form a scalar, raise both indices with $\eta^{\mu\rho}\eta^{\nu\sigma}$ to get $F^{\mu\nu}$, then contract:

$$F^{\mu\nu}F_{\mu\nu} = \eta^{\mu\rho}\eta^{\nu\sigma}F_{\rho\sigma}F_{\mu\nu}.$$

All four indices are contracted: no free indices remain, so this is a Lorentz scalar. Evaluating explicitly: $F^{\mu\nu}F_{\mu\nu} = 2(|\mathbf{B}|^2 - |\mathbf{E}|^2)$. The Maxwell Lagrangian $\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$ is therefore Lorentz invariant by index counting alone.

10.7 Lorentz invariance: the practical test

An expression is Lorentz invariant if and only if it has *no free Lorentz indices*. The practical test:

Expression	**Free indices**	**Lorentz status**
$x^\mu x_\mu = \eta_{\mu\nu}x^\mu x^\nu$	none	Scalar (invariant)
$\partial_\mu\phi\,\partial^\mu\phi$	none	Scalar (invariant)
$F_{\mu\nu}F^{\mu\nu}$	none	Scalar (invariant)
x^μ	μ	4-vector (covariant)
$F_{\mu\nu}$	μ, ν	rank-2 tensor
$\partial_\mu F^{\mu\nu}$	ν	4-vector (Maxwell eqn lhs)
$g_{\mu\nu}R^{\mu\nu} = R$	none	Scalar (Ricci scalar)

The Lagrangian density $\mathcal{L}$ must be a scalar (no free indices) for the action $S = \int d^4x\,\mathcal{L}$ to be Lorentz invariant. The spacetime measure d^4x transforms with $|\det\Lambda| = 1$, so it contributes no index structure. Writing a Lorentz-invariant Lagrangian is therefore the same as writing an expression with all indices contracted.

Summary: upper vs. lower at a glance.

Object	Index position	Transforms as
Position $x^\mu = (t, \mathbf{x})$	upper	$x'^\mu = \Lambda^\mu{}_\nu x^\nu$
Gradient $\partial_\mu = (\partial_t, \nabla)$	lower	$\partial'_\mu = \Lambda_\mu{}^\nu \partial_\nu$
Metric $\eta_{\mu\nu}$	two lower	invariant tensor
Inverse metric $\eta^{\mu\nu}$	two upper	invariant tensor
Field A_μ (photon potential)	lower	$A'_\mu = \Lambda_\mu{}^\nu A_\nu$
Field strength $F^{\mu\nu}$	two upper	$F'^{\mu\nu} = \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma F^{\rho\sigma}$
Contracted pair $V^\mu W_\mu$	none	Lorentz scalar

11 The Scalar Field (Spin 0)

The simplest quantum field is the *scalar field* $\phi(x)$: a real-valued function of spacetime, transforming trivially under Lorentz transformations (spin 0). It is the ideal starting point because all the essential structure of QFT—Fock space, the three RNN pictures, the path integral, and Lorentz invariance—appears without the additional index structure of spinors or gauge fields.

11.1 Classical scalar field: infinitely many coupled oscillators

A classical scalar field $\phi(\mathbf{x}, t)$ assigns a real number to each point in space at each moment of time. Decomposing into Fourier modes:

$$\phi(\mathbf{x}, t) = \int \frac{d^3k}{(2\pi)^3}\, \tilde{\phi}(\mathbf{k}, t)\, e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{85}$$

each Fourier coefficient $\tilde{\phi}(\mathbf{k}, t)$ satisfies a harmonic oscillator equation:

$$\ddot{\tilde{\phi}}(\mathbf{k}, t) + \omega_{\mathbf{k}}^2\, \tilde{\phi}(\mathbf{k}, t) = 0, \qquad \omega_{\mathbf{k}} = \sqrt{|\mathbf{k}|^2 + m^2}. \tag{86}$$

A classical scalar field is therefore *an infinite collection of independent harmonic oscillators*, one per momentum mode $\mathbf{k}$. The mass m sets the minimum frequency: even the zero-momentum mode oscillates at frequency m.

11.2 The classical Lagrangian and the Klein–Gordon equation

The equation of motion (86) follows from $\delta S[\phi] = 0$, where the *Lagrangian density* is:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}m^2\phi^2 = \frac{1}{2}\dot{\phi}^2 - \frac{1}{2}|\nabla\phi|^2 - \frac{1}{2}m^2\phi^2. \tag{87}$$

The action integrates over all spacetime: $S[\phi] = \int d^4x\, \mathcal{L}$. The Euler–Lagrange equation gives the *Klein–Gordon equation*:

$$(\partial_\mu\partial^\mu + m^2)\,\phi = 0, \qquad \partial_\mu\partial^\mu \equiv \partial_t^2 - \nabla^2. \tag{88}$$

11.3 Lorentz invariance of the scalar Lagrangian

A Lorentz transformation Λ maps spacetime points as $x \mapsto \Lambda x$. For a scalar field, the value at the transformed point is unchanged:

$$\phi'(x) = \phi(\Lambda^{-1}x). \tag{89}$$

This is the defining property of spin-0: *only the argument transforms; the value carries no Lorentz index.* The Lagrangian density (87) is Lorentz invariant because $(\partial_\mu\phi)(\partial^\mu\phi)$ contracts two Lorentz indices with the metric, and d^4x is Lorentz invariant. Therefore $S[\phi]$ is frame-independent.

11.4 Quantisation: promoting modes to ladder operators

Quantisation promotes the Fourier coefficients to operators satisfying:

$$\Big[a(\mathbf{k}),\, a^\dagger(\mathbf{k}')\Big] = (2\pi)^3\,\delta^{(3)}(\mathbf{k}-\mathbf{k}'), \tag{90}$$

one copy of the harmonic oscillator algebra per mode. The field operator is:

$$\hat\phi(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3\sqrt{2\omega_{\mathbf{k}}}}\Big[a(\mathbf{k})\,e^{i\mathbf{k}\cdot\mathbf{x}} + a^\dagger(\mathbf{k})\,e^{-i\mathbf{k}\cdot\mathbf{x}}\Big]. \tag{91}$$

The general Fock hidden state is:

$$h_t = \underbrace{c_0(t)\,|0\rangle}_{\mathcal{H}^{(0)}} \oplus \underbrace{\int \frac{d^3k}{(2\pi)^3}\,c_1(\mathbf{k};t)\;|\mathbf{k}\rangle}_{\mathcal{H}^{(1)}} \oplus \underbrace{\int \frac{d^3k_1 d^3k_2}{(2\pi)^6}\,c_2(\mathbf{k}_1,\mathbf{k}_2;t)\;|\mathbf{k}_1,\mathbf{k}_2\rangle}_{\mathcal{H}^{(2)}} \oplus \cdots, \tag{92}$$

with $|c_0|^2 + \int|c_1|^2 + \int|c_2|^2 + \cdots = 1$. In ML terms, this is the concatenated hidden state: each block is a "particle-sector embedding," and the full Fock vector is their concatenation. The free Hamiltonian:

$$H_{\text{free}} = \int \frac{d^3k}{(2\pi)^3}\,\omega_{\mathbf{k}}\,a^\dagger(\mathbf{k})\,a(\mathbf{k}), \tag{93}$$

is block-diagonal in this concatenation: it phase-rotates each coefficient c_n independently, preserving particle number.

11.5 The three RNN pictures in Fock space

All three pictures carry over without modification.

Schrödinger picture. $h_t \in \mathcal{F}$ evolves; field operators $\hat\phi(\mathbf{x})$ are fixed matrices on $\mathcal{F}$:

$$\frac{dh_t}{dt} = -iH\,h_t, \qquad h_t = e^{-iHt}\,h_0. \tag{94}$$

Heisenberg picture. h_0 is fixed; field operator evolves:

$$\phi(\mathbf{x},t) = e^{iHt}\,\hat{\phi}(\mathbf{x})\,e^{-iHt} = \int \frac{d^3k}{(2\pi)^3\sqrt{2\omega_{\mathbf{k}}}}\Big[a(\mathbf{k})\,e^{i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}t)} + a^\dagger(\mathbf{k})\,e^{-i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}t)}\Big]. \quad (95)$$

The four-vector $(\omega_{\mathbf{k}}, \mathbf{k})$ satisfies $\omega_{\mathbf{k}}^2 - |\mathbf{k}|^2 = m^2$, the relativistic dispersion relation. Lorentz covariance is built into the Heisenberg-picture field automatically.

Born rule. $p = |\langle q_o, h_t\rangle|^2$ as always; eigenvectors q_o are Fock basis states; the Born rule gives the probability of finding n particles with specified momenta.

11.6 Interactions: coupling particle-number sectors

Adding an interaction term to the Hamiltonian:

$$H = H_{\text{free}} + \lambda \int d^3x\ \hat{\phi}(\mathbf{x})^4, \quad (96)$$

the $\hat{\phi}^4$ term, expanded in ladder operators, generates products like $a^\dagger a^\dagger aa$, $a^\dagger a^\dagger a^\dagger a$, and so on. These shift particle number; the Hamiltonian is no longer block-diagonal in the concatenated Fock vector. Different blocks are coupled; the RNN mixes them; the full hidden state rotates across particle-number sectors as t advances. This is particle creation and annihilation, encoded faithfully in a single linear, unitary RNN.

12 The Feynman Path Integral for the Scalar Field

The Feynman path integral for the scalar field follows from the same time-slicing argument as in Section 7. The identity $e^{-iHT} = (e^{-iH\epsilon})^N$ holds for any H on any Hilbert space, including the scalar field Hamiltonian on Fock space. The only difference from the single-particle case is that the resolution of identity we insert between time slices uses field-configuration eigenstates of $\hat{\phi}(\mathbf{x})$ rather than position eigenstates of $\hat{X}$.

12.1 Field-configuration eigenstates are Fock-space basis vectors

The field operator $\hat{\phi}(\mathbf{x})$ has simultaneous eigenstates $|\phi(\cdot)\rangle$:

$$\hat{\phi}(\mathbf{x})\ |\phi(\cdot)\rangle = \phi(\mathbf{x})\ |\phi(\cdot)\rangle\,, \qquad \text{for all } \mathbf{x}. \quad (97)$$

Here $\phi(\cdot)$ is a *classical field configuration*—a specific function from $\mathbb{R}^3$ to $\mathbb{R}$. These states live in Fock space $\mathcal{F}$; they are coherent-state superpositions across all particle-number sectors. The eigenvalue $\phi(\mathbf{x})$ is the *classical value* that the quantum field would take with certainty if measured at $\mathbf{x}$ in this particular Fock state.

This is the critical point to absorb: the $\phi(\mathbf{x})$ appearing in the path integral is not an independent classical field living outside quantum mechanics. It is a *label*

for a Fock-space basis state—a real number that parametrises which coherent superposition of $0, 1, 2, \ldots$ particles we are projecting onto. The resolution of identity is:

$$I_{\mathcal{F}} = \int \mathcal{D}\phi(\cdot)\ |\phi(\cdot)\rangle \langle\phi(\cdot)| \,. \tag{98}$$

Inserting this between time slices is exactly inserting a complete basis of $\mathcal{F}$—it is matrix multiplication on the Fock hidden state.

12.2 Time-slicing: matrix multiplication over field space

Divide $[0, T]$ into N steps of width $\epsilon = T/N$:

$$e^{-iHT} = \underbrace{e^{-iH\epsilon} \cdots e^{-iH\epsilon}}_{N}. \tag{99}$$

Insert $I_{\mathcal{F}}$ between every consecutive pair of factors:

$$K[\phi_f, T;\, \phi_i, 0] = \langle\phi_f|\ e^{-iHT}\ |\phi_i\rangle = \int \mathcal{D}\phi_1 \cdots \mathcal{D}\phi_{N-1} \prod_{k=0}^{N-1} \underbrace{\langle\phi_{k+1}|\ e^{-iH\epsilon}\ |\phi_k\rangle}_{\text{one entry of the short-time transition matrix}} . \tag{100}$$

A "path" in field space $\phi_0 \to \phi_1 \to \cdots \to \phi_N$ is a sequence of Fock-basis labels, one per time step. The amplitude along that path is the product of transition matrix entries. The path integral sums over all sequences—this is exactly repeated matrix multiplication on the Fock space.

What the matrix entries encode. Each transition amplitude $\langle\phi_{k+1}|\, e^{-iH\epsilon}\, |\phi_k\rangle$ is controlled by the full quantum Hamiltonian H, including the interaction terms. The interaction vertices—the $a^\dagger a^\dagger a a$ terms that create and annihilate particles—contribute to the off-diagonal structure of the transition matrix. The particle-creation story is not lost when we write $e^{iS[\phi]}$; it is encoded in how the matrix entry depends on ϕ_{k+1} and ϕ_k.

12.3 The short-time amplitude and the Lagrangian

For small ϵ, Trotter splitting and a Gaussian integration over the conjugate momentum $\pi = \dot{\phi}$ (exactly as in Section 7) gives:

$$\langle\phi_{k+1}|\ e^{-iH\epsilon}\ |\phi_k\rangle\ \propto\ \exp\!\Big(i\epsilon \int d^3x\ \mathcal{L}(\phi_k, \dot{\phi}_k)\Big), \qquad \dot{\phi}_k \approx \frac{\phi_{k+1} - \phi_k}{\epsilon}. \tag{101}$$

The classical Lagrangian $\mathcal{L}$ re-emerges in the exponent—not because we are summing over classical field configurations, but because the short-time matrix element of any Hamiltonian of the form kinetic + potential always factors into $e^{i\epsilon\mathcal{L}}$ via a Gaussian integral. This is a property of the Trotter expansion, not an independent assumption.

12.4 The field path integral

Taking $N \to \infty$:

$$\boxed{K[\phi_f, T;\, \phi_i, 0] = \int_{\phi(\cdot,0)=\phi_i}^{\phi(\cdot,T)=\phi_f} \mathcal{D}\phi \, \exp\!\Big(i \int_0^T dt \int d^3x \, \mathcal{L}(\phi, \dot{\phi})\Big) = \int \mathcal{D}\phi \, e^{iS[\phi]}.} \quad (102)$$

Concept	**Particle QM**	**Scalar QFT**
Basis labels	Position eigenvalue $x \in \mathbb{R}$	Field eigenvalue $\phi(\cdot) \in L^2(\mathbb{R}^3)$
Basis states	$\lvert x\rangle$ in Hilbert space	$\lvert\phi(\cdot)\rangle$ in Fock space
Path	Trajectory $x(t)$	Field-eigenvalue sequence $\phi(\mathbf{x}, t)$
Path weight	$e^{iS[x]}$	$e^{iS[\phi]}$
Matrix multiplication	U^N, finite dim	$(U_{\phi'\phi})^N$, infinite dim
Classical limit	Euler–Lagrange for $x(t)$	Klein–Gordon for $\phi(\mathbf{x}, t)$
Quantum corrections	Off-classical-path trajectories	Off-saddle field configurations

Stationary phase recovers the Klein–Gordon equation.
In the limit $\hbar \to 0$, the path integral is dominated by the field history where $\delta S[\phi] = 0$:

$$\delta S = \int d^4x \, \delta\phi \left[-\partial_\mu \partial^\mu \phi - m^2 \phi\right] = 0 \implies (\partial_\mu \partial^\mu + m^2)\phi = 0.$$

Classical field theory is the greedy-decoding limit of the QFT RNN: as $\hbar \to 0$, only the saddle-point field history survives. Quantum fluctuations are all the off-saddle field-eigenvalue sequences—suppressed but never zero, and their interference is the origin of every quantum correction (i.e., every Feynman diagram).

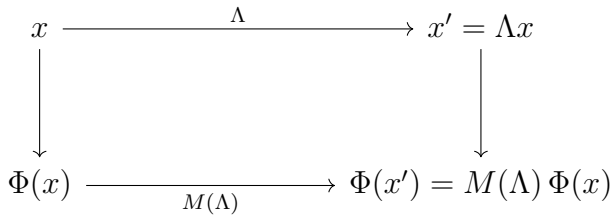
13 Representations of the Lorentz Group

We have treated the scalar field $\phi(x)$, whose value at each spacetime point is a single number. The full spectrum of particles—electrons, photons, quarks—requires fields whose values carry *additional structure* that transforms consistently when we change inertial frame. The right language is representation theory, and the cleanest entry point is a sequence of three commutative diagrams, each adding one layer of structure to the one before.

13.1 Diagram 1: the abstract transformation law

A field $\Phi(x)$ lives at each spacetime point x and takes values in some vector space V. Under a Lorentz transformation Λ, the point and the value transform together.

This is the fundamental diagram:



The vertical arrows are evaluation of the field at a point. The top arrow is the Lorentz transformation on spacetime. The bottom arrow is the induced action on the field value. The diagram commutes: transform the point and then evaluate, or evaluate and then transform the value—you get the same answer.

The assignment $\Lambda \mapsto M(\Lambda)$ must be a group homomorphism: $M(\Lambda_1\Lambda_2) = M(\Lambda_1)M(\Lambda_2)$. This is the *representation condition*. Different choices of M give different species of field.

ML reading. The field $\Phi(x)$ is an equivariant embedding table: each row is the "embedding" of spacetime point x in V. A Lorentz transformation acts on the row index via Λ and on the row content via $M(\Lambda)$ simultaneously. Classifying all possible fields is classifying all equivariant representations of the Lorentz group—the same problem as finding all equivariant maps in geometric deep learning.

13.2 Diagram 2: spacetime as 2×2 matrices

To make the representation $M(\Lambda)$ explicit, we first rewrite the spacetime point $x^\mu = (x^0, x^1, x^2, x^3)$ as a 2×2 Hermitian matrix by contracting with the Pauli matrices:

$$X = x^\mu \sigma_\mu = x^0 I + x^1\sigma_1 + x^2\sigma_2 + x^3\sigma_3 = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix}. \tag{103}$$

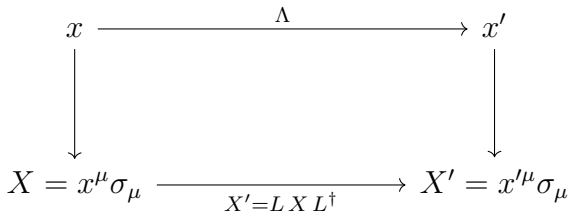
Two facts make this the right thing to do. First, every 2×2 Hermitian matrix arises this way: the map $x^\mu \mapsto X$ is a vector-space isomorphism between $\mathbb{R}^4$ and the space of 2×2 Hermitian matrices. Second, the determinant of X is the spacetime interval:

$$\det X = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = x^\mu x_\mu. \tag{104}$$

Now let $L \in SL(2, \mathbb{C})$ (a complex 2×2 matrix with $\det L = 1$) act on X by *conjugation*:

$$X' = L\,X\,L^\dagger. \tag{105}$$

Then X' is still Hermitian ($X'^\dagger = (LXL^\dagger)^\dagger = LXL^\dagger = X'$), and its determinant is preserved: $\det X' = \det L \cdot \det X \cdot \det L^\dagger = \det X$, since $\det L = 1$ and $\det L^\dagger = \overline{\det L} = 1$. Therefore X' corresponds to a new spacetime point x'^μ with the same spacetime interval as x^μ: the map $X \mapsto LXL^\dagger$ *is* a Lorentz transformation $\Lambda(L)$.



The map $L \mapsto \Lambda(L)$ is a 2-to-1 group homomorphism $SL(2,\mathbb{C}) \to SO^+(3,1)$: the two matrices L and $-L$ give the same Λ. This is why the Lorentz group has "double covers" and why half-integer spin is possible.

L and $L^\dagger$ together represent Λ. The matrix L acts on the left index of X and $L^\dagger$ acts on the right index. These are the two independent $SL(2,\mathbb{C})$ factors that arise from the factorisation of the Lorentz algebra $\mathfrak{so}(3,1)_\mathbb{C} \cong \mathfrak{sl}(2,\mathbb{C}) \oplus \mathfrak{sl}(2,\mathbb{C})$, labelled (j_L, j_R) in the previous section. For real spacetime vectors, L and $L^\dagger$ are conjugates of each other (not independent), so they jointly encode the single Lorentz transformation Λ. Different choices of how to split the representation between left and right give the different particle types.

Rotation and boost in $SL(2,\mathbb{C})$.
For a rotation by θ about $\hat{z}$:

$$L = e^{-i\theta\sigma_3/2} = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix}, \qquad L^\dagger = \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix}.$$

One can verify that $X' = LXL^\dagger$ rotates (x^1, x^2) by angle θ while leaving x^0 and x^3 unchanged—exactly the Lorentz rotation Λ.
For a boost with rapidity ξ along $\hat{z}$:

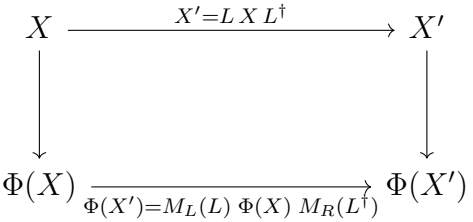
$$L = e^{-\xi\sigma_3/2} = \begin{pmatrix} e^{-\xi/2} & 0 \\ 0 & e^{\xi/2} \end{pmatrix}, \qquad L^\dagger = \begin{pmatrix} e^{-\xi/2} & 0 \\ 0 & e^{\xi/2} \end{pmatrix}.$$

(Both real here since σ_3 is real.) $X' = LXL^\dagger$ mixes x^0 and x^3 as a hyperbolic rotation—exactly the Lorentz boost.

13.3 Diagram 3: fields on matrix spacetime

Now we promote X from a label for a spacetime point to the *argument* of the field. The field value $\Phi(X)$ is itself a matrix, and the Lorentz group acts on both its left and right indices:

$$\Phi(X) \;\longmapsto\; \Phi(X') = M_L(L)\,\Phi(X)\,M_R(L^\dagger). \tag{106}$$



This is the master transformation law. Every particle species corresponds to a specific choice of the left representation M_L and the right representation M_R:

Field	$\Phi(X)$ **lives in**	$M_L(L)$	$M_R(L^\dagger)$	**Particle**
Scalar	$\mathbb{R}$ (number)	1	1	Higgs
Left Weyl	$\mathbb{C}^2$ (column)	L	1	ν_L, e_L
Right Weyl	$\mathbb{C}^2$ (row)	1	$L^\dagger$	e_R
4-vector	$\mathbb{C}^{2\times 2}$ (matrix)	L	$L^\dagger$	photon, $W^\pm$, Z

Reading across the rows:

Scalar. $M_L = M_R = 1$; the field value is a number, blind to both the left and right actions of L. Diagram 3 reduces to diagram 1 with $M(\Lambda) = 1$.

Left Weyl spinor. $M_L = L$, $M_R = 1$; the field value is a 2-component column vector and only the left $SL(2,\mathbb{C})$ factor acts:

$$\psi_L(X) \;\longmapsto\; L\,\psi_L(X). \tag{107}$$

The left index of X and the index of ψ_L transform by the same L. This is the $(\frac{1}{2}, 0)$ representation.

Right Weyl spinor. $M_L = 1$, $M_R = L^\dagger$; only the right factor acts:

$$\psi_R(X) \;\longmapsto\; \psi_R(X)\,L^\dagger. \tag{108}$$

This is the $(0, \frac{1}{2})$ representation. The right index of X and the index of ψ_R transform by the same $L^\dagger$.

4-vector. $M_L = L$, $M_R = L^\dagger$; the field value is a 2×2 matrix and both factors act:

$$A(X) \;\longmapsto\; L\,A(X)\,L^\dagger. \tag{109}$$

This is the $(\frac{1}{2}, \frac{1}{2})$ representation. The field value transforms *in exactly the same way as the spacetime point X itself*: the bottom arrow in diagram 3 is identical to the bottom arrow in diagram 2. The photon field A_μ, written as the 2×2 matrix $A = A_\mu \sigma^\mu$, is a second copy of spacetime living at each point in spacetime.

Dirac spinor. A Dirac spinor stacks both chiralities: $\Psi = (\psi_L, \psi_R)^T \in \mathbb{C}^4$, with

$$M_L(L) = \begin{pmatrix} L & 0 \\ 0 & 1 \end{pmatrix}, \qquad M_R(L^\dagger) = \begin{pmatrix} 1 & 0 \\ 0 & L^\dagger \end{pmatrix}, \tag{110}$$

acting on the left and right halves of Ψ respectively. The mass term couples the two halves; a massless particle can be purely left- or right-handed, but mass requires both.

13.4 General representations: (j_L, j_R) as matrix dimensions

The pattern is now clear. For a general (j_L, j_R) representation, the field value $\Phi(X)$ lives in a $(2j_L + 1) \times (2j_R + 1)$ matrix, and:

$$\Phi(X) \longmapsto D^{(j_L)}(L)\, \Phi(X)\, D^{(j_R)}(L^\dagger), \tag{111}$$

where $D^{(j)}(L)$ is the $(2j+1)$-dimensional spin-j representation of $SL(2, \mathbb{C})$. The left and right matrix dimensions count the number of "left-handed" and "right-handed" spinor indices. Scalars have no indices (1×1); Weyl spinors have one index (2×1 or 1×2); 4-vectors have one index on each side (2×2); higher-spin particles have more.

Remark. This two-sided action is the relativistic generalisation of the non-relativistic $\infty \times 2$ spinor matrix of Section 8, where orbital operators acted on the left (row) index and spin operators acted on the right (column) index. There the two families commuted because matrix multiplication associates: $A(\Psi B) = (A\Psi)B$. The same principle operates here, now with L and $L^\dagger$ as the two non-commuting-with-each-other but independently-commuting families.

13.5 Quantisation: Fock space for every representation

Once the classical field $\Phi(X)$ and its Lagrangian are specified by the choice of (j_L, j_R), the quantisation procedure is the same as for the scalar field. The Fourier modes become ladder operators; the Fock space is the direct sum over particle-number sectors; the Schrödinger equation $\dot{h}_t = -iHh_t$ governs the Fock hidden state; the Born rule reads out probabilities; and the path integral is repeated matrix multiplication.

The representation (j_L, j_R) determines the structure of the ladder operators—how they transform under L and $L^\dagger$—but does not alter the RNN architecture.

Field	(j_L, j_R)	$\Phi(X)$ **shape**	**Statistics**	**d.o.f.**	**Fock algebra**
Scalar	$(0,0)$	1×1	Boson	1	$[a, a^\dagger] = 1$
Left Weyl	$(\frac{1}{2}, 0)$	2×1	Fermion	2	$\{b, b^\dagger\} = 1$
Right Weyl	$(0, \frac{1}{2})$	1×2	Fermion	2	$\{b, b^\dagger\} = 1$
Dirac	$(\frac{1}{2}, 0)\oplus(0, \frac{1}{2})$	4×1	Fermion	4	$\{b, b^\dagger\} = 1$
4-vector	$(\frac{1}{2}, \frac{1}{2})$	2×2	Boson	2	$[a, a^\dagger] = 1$

In every case, the Fock hidden state h_t is a concatenation of particle-number-sector vectors, the Schrödinger equation rotates it unitarily, the Born rule reads out probabilities, and the Feynman path integral is the same time-sliced matrix multiplication. *The only input that varies across particle species is the shape of the field value matrix $\Phi(X)$ and the choice of (M_L, M_R) in diagram 3. The RNN framework is universal.*

13.6 Diagram 3 in spin-index notation

In diagram 3, the field $\Phi(X)$ at each spacetime point X is a matrix transforming as $\Phi(X) \to M_L(L)\,\Phi(X)\,M_R(L^\dagger)$. The left action M_L and right action M_R act on the row and column indices of Φ respectively. Writing these indices explicitly: let $s \in \{1, \ldots, 2j_L+1\}$ label the row (left spinor index) and $\dot{s} \in \{1, \ldots, 2j_R+1\}$ label the column (right spinor index). Then $\Phi^s{}_{\dot{s}}(X)$ is the $(s, \dot{s})$ component of the field matrix, and the transformation law is:

$$\Phi^s{}_{\dot{s}}(X) \;\longmapsto\; M_L(L)^s{}_{s'}\;\Phi^{s'}{}_{\dot{s}'}(\Lambda^{-1}X)\;M_R(L^\dagger)^{\dot{s}'}{}_{\dot{s}}. \tag{112}$$

Here s' and $\dot{s}'$ are contracted (summed) indices; s and $\dot{s}$ are free indices labelling the component of the transformed field. The upper left index s transforms with M_L and the lower right index $\dot{s}$ transforms with M_R—consistent with the convention that upper indices transform with the matrix and lower indices transform with its inverse or conjugate.

For the special cases:

$$\text{Scalar:}\quad \Phi(X),\quad s, \dot{s}\text{ absent},\quad M_L = M_R = 1. \tag{113}$$

$$\text{Left Weyl spinor:}\quad \psi^s(X),\quad s \in \{1,2\},\quad \dot{s}\text{ absent},\quad \psi^s \to L^s{}_{s'}\psi^{s'}. \tag{114}$$

$$\text{Right Weyl spinor:}\quad \bar{\chi}_{\dot{s}}(X),\quad \dot{s} \in \{1,2\},\quad s\text{ absent},\quad \bar{\chi}_{\dot{s}} \to \bar{\chi}_{\dot{s}'}(L^\dagger)^{\dot{s}'}{}_{\dot{s}}. \tag{115}$$

$$\text{4-vector:}\quad A^s{}_{\dot{s}}(X) \;=\; A_\mu(X)\,(\sigma^\mu)^s{}_{\dot{s}},\quad A^s{}_{\dot{s}} \to L^s{}_{s'}A^{s'}{}_{\dot{s}'}(L^\dagger)^{\dot{s}'}{}_{\dot{s}}. \tag{116}$$

The spin indices s and $\dot{s}$ are the Lorentz-group indices of diagram 3, now written explicitly. Each undotted index s transforms with $L \in SL(2,\mathbb{C})$; each dotted index $\dot{s}$ transforms with $\bar{L} = (L^\dagger)^{-T}$.

Fock reading. A basis vector of $\mathcal{F}_{\mathrm{SM}}$ for a single left Weyl particle carries the index $s \in \{1, 2\}$ as one of its labels: $|k, s\rangle$ is a state with momentum k and spin s. The two values of s correspond to the two spin states (helicities) of the particle. The full hidden state h_t is a superposition over all values of k and s:

$$h_t^{(1,\mathrm{Weyl})} = \sum_{s\in\{1,2\}} \int \frac{d^3k}{(2\pi)^3}\, c(k,s)\ |k,s\rangle\,. \tag{117}$$

Adding a new internal index a (colour, isospin) just adds another label to $|k, s, a\rangle$ and another sum to the expansion.

14 Additional Indices: Gauge Structure and the Standard Model

Diagram 3 classifies all particle species by how their field value transforms under Lorentz transformations. But the fields of the Standard Model carry more structure than Lorentz indices alone. Each field also carries *internal* indices—colour, isospin, hypercharge—that label its transformation properties under the gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$. This section introduces these additional indices systematically, shows how they sit alongside the Lorentz indices of diagram 3, and previews the full index structure of the Standard Model fields.

Throughout, we keep the Fock hidden state in view. The basis vectors of $\mathcal{F}_{\mathrm{SM}}$ are labelled by all quantum numbers simultaneously—Lorentz spin, colour, isospin, flavour, momentum. Adding a new index to a field adds a new label to the Fock basis vectors. The Schrödinger RNN $\dot{h}_t = -iH_{\mathrm{SM}}\, h_t$ runs on this larger labelled space. The index structure is the structure of the labels; the RNN is unchanged.

14.1 Internal indices: local gauge transformations

For each fixed spacetime point x, one can equip the field with an additional *internal* index a that has nothing to do with Lorentz structure. The index a labels a degree of freedom in an internal vector space V_{int}, and the field becomes:

$$\Phi^{s}{}_{\dot{s},\,a}(x), \tag{118}$$

where $s, \dot{s}$ are Lorentz spin indices (as before) and $a \in \{1, \ldots, N\}$ is the internal index. The value of N determines which gauge group acts on a:

N	Gauge group	Generator matrices	Field representation	Example
1	$U(1)$	1×1: phase $e^{i\alpha}$	Singlet (no a index)	Electron e_R
2	$SU(2)_L$	2×2: $\tau^a = \sigma^a/2$	Doublet	(ν_L, e_L)
3	$SU(3)_c$	3×3: $T^a = \lambda^a/2$	Triplet (fundamental)	Quarks
8	$SU(3)_c$	8×8: adjoint representation	Octet	Gluons

A *local gauge transformation* at spacetime point x is a group element $g(x)$ applied to the internal index:

$$\Phi^{s}{}_{\dot{s},\,a}(x) \;\longmapsto\; U(g(x))^{b}{}_{a}\;\Phi^{s}{}_{\dot{s},\,b}(x), \tag{119}$$

where $U(g(x))^{b}{}_{a}$ is the (b, a) matrix element of the group element in the chosen representation. The key feature is *locality*: $g(x)$ can depend on x, and different points can be transformed by different group elements. This is in contrast to the global Lorentz transformation of diagram 3, which applies the same Λ (or L) to every spacetime point simultaneously.

Why local gauge invariance is the more natural condition. A global symmetry (g independent of x) is a fine constraint, but asking the Lagrangian to be invariant under *local* symmetry transformations is a stronger and more surprising requirement. It turns out to be exactly what forces the gauge fields to exist: demanding local $U(1)$ invariance of the Dirac Lagrangian introduces the photon; demanding local $SU(3)_c$ invariance introduces the gluons. The gauge bosons are not added by hand—they are the mathematical consequence of requiring that the internal phase at each point be unobservable. We return to this in the sections on QED and QCD.

14.2 Two ways to write the gauge transformation

The gauge transformation (119) on the internal index a can be written in two equivalent ways, each with its own advantages.

Way 1: tensor notation with explicit index a. Keep the index a written out. The transformation is:

$$\Phi_a(x) \;\longmapsto\; U^{b}{}_{a}(x)\,\Phi_b(x), \tag{120}$$

where a (free, lower) is the transformed index and b (contracted, upper-lower pair) is summed. The gauge group element $U^{b}{}_{a}$ has upper index b (the output) and lower index a (the input), consistent with the Einstein convention that the transformation raises the output index. This notation makes Lorentz invariance and gauge covariance visible simultaneously: one scans the free indices to check covariance. The covariant derivative in tensor notation is:

$$(D_\mu)^{b}{}_{a} = \partial_\mu\,\delta^{b}_{a} - ig\,A^{b}_{\mu\,a}, \tag{121}$$

where $A^{b}_{\mu\,a}$ is the gauge field with both a spacetime index μ and two internal indices b, a. The kinetic term $(D_\mu\Phi)^b(D^\mu\Phi)_b = \sum_b |(D_\mu\Phi)^b|^2$ contracts all indices.

Way 2: matrix notation without explicit index a. Suppress the internal index a and treat $\Phi(x)$ as a vector in the internal space V_{int}, and $U(x)$ as a matrix acting on that vector:

$$\Phi(x) \;\longmapsto\; U(x)\,\Phi(x), \tag{122}$$

where $\Phi(x) \in \mathbb{C}^N$ (a column vector) and $U(x) \in U(N)$ or $SU(N)$ (an $N \times N$ unitary matrix). This is simpler to write and emphasises the group-theory structure. The gauge field becomes a matrix:

$$A_\mu(x) = A^a_\mu(x)\,T^a \;\in\; \mathfrak{g}\ (\text{Lie algebra, } N \times N \text{ matrix}), \tag{123}$$

and the covariant derivative is:

$$D_\mu\Phi = \partial_\mu\Phi - ig\,A_\mu\,\Phi, \tag{124}$$

a matrix acting on a vector at each point.

Which to use? Both notations are equally valid and encode the same physics. Tensor notation with explicit indices a is useful when checking Lorentz and gauge covariance simultaneously, when writing Feynman rules, or when treating the field as part of a larger multi-index object. Matrix notation is more compact for group-theoretic arguments, for writing Lagrangians, and for emphasising the covariant derivative as a connection. In this book we use both, switching freely depending on which makes the structure more transparent.

Fock reading. In either notation, the internal index a adds a label to each Fock basis vector. A quark basis state carries three labels for its colour: $|k, s, a\rangle$ with $a \in \{r, g, b\}$. The colour-singlet condition (only gauge-invariant states are physical) projects the Fock hidden state onto the subspace where all colour indices are contracted to zero. A meson basis state is $\sum_a c^a\,|k_q, s_q, a\rangle \otimes |k_{\bar{q}}, s_{\bar{q}}, a\rangle$, summing over colour with equal amplitude—a colour-singlet combination.

14.3 The full index structure of the Standard Model

Combining the Lorentz spin index s (from diagram 3) with the internal indices (from the gauge group), each Standard Model field carries a specific collection of indices. The following table gives the complete picture.

Field	**Symbol**	**Lorentz**	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$	**Indices**
Left quark	$Q_L^{i\alpha}$	$(\frac{1}{2},0)$, $s\in\{1,2\}$	**3**, $\alpha\in\{r,g,b\}$	**2**, $i\in\{1,2\}$	$+\frac{1}{6}$	s,i,α
Right up quark	u_R^{α}	$(\frac{1}{2},0)$, $s\in\{1,2\}$	**3**, $\alpha\in\{r,g,b\}$	**1** (singlet)	$+\frac{2}{3}$	s,α
Right down quark	d_R^{α}	$(\frac{1}{2},0)$, $s\in\{1,2\}$	**3**, $\alpha\in\{r,g,b\}$	**1**	$-\frac{1}{3}$	s,α
Left lepton	L_L^{i}	$(\frac{1}{2},0)$, $s\in\{1,2\}$	**1** (singlet)	**2**, $i\in\{1,2\}$	$-\frac{1}{2}$	s,i
Right lepton	e_R	$(\frac{1}{2},0)$, $s\in\{1,2\}$	**1**	**1**	-1	s only
Gluon	G_μ^a	$(\frac{1}{2},\frac{1}{2})$, $\mu\in\{0,1,2,3\}$	**8**, $a\in\{1,\dots,8\}$	**1**	0	μ,a
W boson	W_μ^i	$(\frac{1}{2},\frac{1}{2})$, $\mu\in\{0,1,2,3\}$	**1**	**3**, $i\in\{1,2,3\}$	0	μ,i
B boson	B_μ	$(\frac{1}{2},\frac{1}{2})$, $\mu\in\{0,1,2,3\}$	**1**	**1**	0	μ only
Higgs	ϕ^i	$(0,0)$ (scalar)	**1**	**2**, $i\in\{1,2\}$	$+\frac{1}{2}$	i only

Reading across the table: each Standard Model field is completely characterised by its index collection. The transformation law under any symmetry transformation is determined by which indices the transformation acts on.

Unpacking the left quark doublet $Q_L^{i\alpha}$.

The left quark doublet is the most index-rich field in the table. It carries:

- A Lorentz spinor index $s\in\{1,2\}$ (suppressed in the table for brevity, implicit in the "$(\frac{1}{2},0)$" label), transforming as $Q_L^{i\alpha}\to L^s{}_{s'}Q_L^{i\alpha}$ under $L\in SL(2,\mathbb{C})$.
- A colour index $\alpha\in\{r,g,b\}$, transforming as $Q_L^{i\alpha}\to U^\alpha{}_{\alpha'}Q_L^{i\alpha'}$ under $SU(3)_c$.
- An isospin index $i\in\{1,2\}$ (corresponding to up-type and down-type quarks), transforming as $Q_L^{i\alpha}\to V^i{}_{i'}Q_L^{i'\alpha}$ under $SU(2)_L$.
- A hypercharge $Y=+\frac{1}{6}$, contributing a phase $e^{i\alpha(x)/6}$ under $U(1)_Y$.

All four transformations are independent: Lorentz acts on s, colour on α, isospin on i, hypercharge as an overall phase. They commute because they act on different indices—the same two-sided commutativity of diagram 3, now generalised to four independent index slots.

In matrix notation (suppressing s):

$$Q_L(x)\ \in\ \mathbb{C}^2\otimes\mathbb{C}^3,\qquad Q_L\ \longmapsto\ V(x)\otimes U(x)\,Q_L, \tag{125}$$

where $V(x)\in SU(2)_L$ acts on the isospin factor and $U(x)\in SU(3)_c$ acts on the colour factor. The full field transforms as an element of the tensor product of the isospin and colour representation spaces.

Fock basis vectors for Standard Model fields. A single-particle basis state for the left quark doublet carries all indices simultaneously:

$$|\mathbf{p},\ s,\ f,\ i,\ \alpha\rangle\,, \tag{126}$$

where $\mathbf{p}$ is the 3-momentum, $s\in\{1,2\}$ is the Lorentz spinor index, $f\in\{u,d\}$ labels the flavour (up or down, the isospin-doublet components), $i\in\{1,2\}$ is the isospin index (same information as f in this context), and $\alpha\in\{r,g,b\}$ is the colour. The Fock hidden state for a single quark is a superposition over all values of all

these indices:

$$h_t^{(1,\text{quark})} = \sum_{s,f,\alpha} \int \frac{d^3p}{(2\pi)^3}\, c(\mathbf{p}, s, f, \alpha)\ |\mathbf{p}, s, f, \alpha\rangle\,. \tag{127}$$

Adding more particles adds more index-tuples to the basis label. The Standard Model Fock space $\mathcal{F}_{\text{SM}}$ has basis vectors labelled by all the occupation numbers for all combinations of $(\mathbf{p}, s, f, \alpha)$ for all species. The Schrödinger RNN rotates the hidden state h_t in this space; the Hamiltonian H_{SM} is the operator that couples different index configurations, creating the story of particle interactions, decays, and scatterings.

14.4 Gauge invariance as index contraction

A physical observable must be invariant under all gauge transformations. In index notation, this means that all internal indices must be contracted—no free gauge indices in any gauge-invariant quantity, exactly as no free Lorentz indices appear in a Lorentz scalar.

For $U(1)_Y$: the field Φ has charge Y and Φ^* has charge $-Y$. The product $\Phi^*\Phi$ has zero total charge and is gauge invariant.

For $SU(2)_L$: the doublet ϕ^i and its conjugate $\epsilon_{ij}\phi^{*j}$ (where ϵ_{ij} is the $SU(2)$ antisymmetric tensor) can be contracted: $\phi^i\epsilon_{ij}\phi^j = 0$ (antisymmetry), but the mass term $\bar{\Psi}\phi$ contracts the isospin index of the doublet ϕ^i with the isospin index of the lepton doublet L^i, giving a singlet.

For $SU(3)_c$: the quark field q^α (fundamental, index up) and the antiquark $\bar{q}_\alpha$ (anti-fundamental, index down) contract to a colour singlet $\bar{q}_\alpha q^\alpha = \sum_\alpha \bar{q}_\alpha q^\alpha$. Three quarks $\epsilon_{\alpha\beta\gamma}q^\alpha q^\beta q^\gamma$ also contract to a singlet via the totally antisymmetric tensor $\epsilon_{\alpha\beta\gamma}$.

Group	**Invariant contraction**	**Notation**	**Physical state**
$U(1)_Y$	$Y_1 + Y_2 = 0$	$\Phi^*\Phi$	Charge-neutral product
$SU(2)_L$	Doublet $\times$ doublet via ϵ_{ij}	$L^i\phi_i$	Yukawa coupling
$SU(3)_c$	$q\bar{q}$	$\bar{q}_\alpha q^\alpha$	Meson ($q\bar{q}$ singlet)
$SU(3)_c$	qqq	$\epsilon_{\alpha\beta\gamma}q^\alpha q^\beta q^\gamma$	Baryon

The Fock hidden state restricted to the physical (gauge-invariant) subspace is a superposition only over colour-singlet, isospin-singlet, charge-neutral basis vectors. The Hamiltonian H_{SM} preserves this subspace: every interaction vertex contracts all gauge indices, so the gauge-invariant subspace is invariant under the Schrödinger evolution. The rotation of the Fock hidden state stays within the physical Hilbert space at all times.

15 Direct Product of Fock Spaces

The previous section introduced Fock space as a direct sum: basis vectors are labelled by particle number and momenta, and the direct sum is the Hilbert space that accommodates superpositions across different particle-number sectors. This section introduces the *direct product* (tensor product) of Fock spaces, and shows that it is an even simpler operation: it just enlarges the set of indices that label each basis vector.

We have already seen this happen twice. When we added the spin index s (from diagram 3), each basis vector $|\mathbf{k}\rangle$ became $|\mathbf{k}, s\rangle$: the same particle, now labelled by an additional quantum number. When we added the colour index α (from $SU(3)_c$), $|\mathbf{k}, s\rangle$ became $|\mathbf{k}, s, \alpha\rangle$. Both of these were direct products. We now make the structure explicit.

15.1 From direct sum to direct product: adding index dimensions

Recall the direct sum. In the scalar Fock space, the basis vectors are:

$$|n;\, \mathbf{k}_1, \ldots, \mathbf{k}_n\rangle\,, \tag{128}$$

and the hidden state is a superposition:

$$h_t = \sum_n \int \prod_j \frac{d^3\mathbf{k}_j}{(2\pi)^3}\, c_n(\mathbf{k}_1, \ldots, \mathbf{k}_n;\, t)\;\, |n;\, \mathbf{k}_1, \ldots, \mathbf{k}_n\rangle\,. \tag{129}$$

Different values of n label orthogonal subspaces $\mathcal{H}^{(n)}$; the direct sum $\mathcal{F} = \bigoplus_n \mathcal{H}^{(n)}$ is the full Fock space. The direct sum organises the basis vectors by how many particles they contain.

Now suppose each particle additionally carries a spin index $s \in \{1, 2\}$ (from the $(\frac{1}{2}, 0)$ Lorentz representation of diagram 3). The basis vectors become:

$$|n;\, (\mathbf{k}_1, s_1), \ldots, (\mathbf{k}_n, s_n)\rangle\,. \tag{130}$$

Nothing has changed in the direct sum structure: the particle number n still organises the sectors. What has changed is that each particle now carries an extra label s_j. The Hilbert space is larger—each $\mathcal{H}^{(n)}$ is now spanned by basis vectors (130) rather than (128)—but the direct sum is the same construction.

This enlargement of the index set is the direct product. The single-particle space $\mathcal{H}^{(1)}_{\text{scalar}}$, spanned by $|\mathbf{k}\rangle$, becomes $\mathcal{H}^{(1)}_{\text{spin}} = \mathcal{H}^{(1)}_{\text{scalar}} \otimes \mathcal{H}_{\text{spin}}$, spanned by $|\mathbf{k}\rangle \otimes |s\rangle = |\mathbf{k}, s\rangle$. The direct product of two spaces *multiplies* their basis sets: if $\mathcal{H}_A$ has basis $\{e_i\}$ and $\mathcal{H}_B$ has basis $\{f_j\}$, then $\mathcal{H}_A \otimes \mathcal{H}_B$ has basis $\{e_i \otimes f_j\} = \{(i, j)\}$—all pairs. No new basis vectors are created by the direct sum (which concatenates); the direct product creates new joint basis vectors by taking all combinations of the individual bases.

The key distinction.

	Direct sum $\oplus$	**Direct product $\otimes$**
What it does	Combines spaces for different n	Combines index types per particle
New basis vectors	From each summand separately	All combinations of both bases
Dimension	$\dim A + \dim B$	$\dim A \times \dim B$
Example	$\mathcal{H}^{(0)} \oplus \mathcal{H}^{(1)} \oplus \cdots$	$\mathcal{H}_{\mathbf{k}} \otimes \mathcal{H}_s \otimes \mathcal{H}_\alpha$
Role in Fock space	Organises by particle number	Enlarges each particle's index set

15.2 Tensor product: explicit basis and general element

Let $\mathcal{H}_A$ have orthonormal basis $\{e_i\}_{i=1}^m$ and $\mathcal{H}_B$ have orthonormal basis $\{f_j\}_{j=1}^n$. The tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$ has orthonormal basis:

$$\{e_i \otimes f_j\}_{i=1,\ldots,m;\, j=1,\ldots,n}, \qquad \dim(\mathcal{H}_A \otimes \mathcal{H}_B) = mn. \tag{131}$$

A general element is a superposition over all joint basis vectors:

$$w = \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij}\, e_i \otimes f_j. \tag{132}$$

The coefficient array w_{ij} is a complex $m \times n$ matrix. In ML terms: it is the outer product space—all pairwise combinations of the two bases.

Product states vs. entangled states. A *product state* is one where the coefficient matrix factorises: $w_{ij} = u_i v_j$ for some vectors $u \in \mathcal{H}_A$ and $v \in \mathcal{H}_B$. Then $w = (\sum_i u_i e_i) \otimes (\sum_j v_j f_j) = u \otimes v$. In a product state, the two systems are independent: the A-part and B-part carry no correlations beyond what each individually knows. A *non-product* (entangled) state has w_{ij} that cannot be factored as $u_i v_j$—the two systems are correlated in a way that cannot be described by independent choices in each factor.

Two spin-$\frac{1}{2}$ particles.
Each has spin space $\mathbb{C}^2$ with basis $\{|\uparrow\rangle, |\downarrow\rangle\}$. The two-particle spin space is $\mathbb{C}^2 \otimes \mathbb{C}^2 = \mathbb{C}^4$ with basis:

$$\{|\uparrow\uparrow\rangle,\ |\uparrow\downarrow\rangle,\ |\downarrow\uparrow\rangle,\ |\downarrow\downarrow\rangle\}.$$

A product state: $|\uparrow\rangle \otimes |\downarrow\rangle = |\uparrow\downarrow\rangle$ (particle 1 is spin-up, particle 2 is spin-down, no correlations).
An entangled state: $\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$—the singlet. The coefficient matrix $w_{ij} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ has rank 2 and cannot be written as $u_i v_j$. Measuring particle 1

as spin-up forces particle 2 to be spin-down, regardless of the distance between them.

Operators on tensor product spaces. An operator that acts only on $\mathcal{H}_A$ is $A \otimes I_B$; one that acts only on $\mathcal{H}_B$ is $I_A \otimes B$. On the basis:

$$(A \otimes I_B)(e_i \otimes f_j) = (Ae_i) \otimes f_j, \tag{133}$$

$$(I_A \otimes B)(e_i \otimes f_j) = e_i \otimes (Bf_j). \tag{134}$$

They commute: $(A \otimes I_B)(I_A \otimes B) = A \otimes B = (I_A \otimes B)(A \otimes I_B)$. This commutativity is the mathematical foundation of the two-sided operator algebra of diagram 3: Lorentz operators act on the spin index ($A \otimes I$) and gauge operators act on the internal index ($I \otimes U$), and they commute because they act on different slots of the joint basis vector.

In terms of matrices: if A is $m \times m$ and B is $n \times n$, then $A \otimes B$ is the $mn \times mn$ *Kronecker product*:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots \\ a_{21}B & a_{22}B & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \tag{135}$$

15.3 Direct product of Fock spaces in the Standard Model

In the Standard Model, each particle species has its own Fock space. The full Standard Model Fock space is a direct product of all of them:

$$\mathcal{F}_{\text{SM}} = \mathcal{F}_{e^-} \otimes \mathcal{F}_{e^+} \otimes \mathcal{F}_{\nu_e} \otimes \mathcal{F}_{\gamma} \otimes \mathcal{F}_{u,r} \otimes \mathcal{F}_{u,g} \otimes \mathcal{F}_{u,b} \otimes \cdots \otimes \mathcal{F}_{h}, \tag{136}$$

where each factor is the Fock space for one species (electron, positron, electron-neutrino, photon, up quark in each colour, …, Higgs). A basis vector of $\mathcal{F}_{\text{SM}}$ specifies the occupation of every mode of every species simultaneously:

$$|n_{e^-}, \{(\mathbf{k}_j, s_j)\};\ n_{e^+}, \{(\mathbf{q}_j, s_j)\};\ n_{\gamma}, \{(\mathbf{p}_j, \lambda_j)\};\ \ldots\rangle\,. \tag{137}$$

This is a single monosemantic basis vector: it specifies exactly how many electrons are present and what their momenta and spins are, exactly how many positrons, exactly how many photons of each helicity, and so on for every species. The full Standard Model hidden state is a superposition of all such basis vectors:

$$h_t = \sum_{\{n_i\}} \int \prod_{\text{modes}} (\cdots)\; c\Big(\{n_i\}, \{\mathbf{k}_j^{(i)}, s_j^{(i)}\};\, t\Big)\; \Big|\{n_i\}, \{\mathbf{k}_j^{(i)}, s_j^{(i)}\}\Big\rangle\,, \tag{138}$$

with complex amplitudes $c(\cdots\,;t)$ that evolve under the Schrödinger equation $\dot{h}_t = -iH_{\text{SM}}\, h_t$.

Why direct product, not direct sum, for different species. The electron Fock space $\mathcal{F}_{e^-}$ and the photon Fock space $\mathcal{F}_{\gamma}$ are combined by direct *product*,

not direct sum. This is because a physical state can contain *both* electrons and photons simultaneously: $|1\,e^-, 1\,\gamma\rangle$ is a state with one electron and one photon, not a superposition of "one electron, no photons" and "no electrons, one photon." The direct product allows independent occupation numbers for each species at every momentum mode. The direct sum would combine them into a single particle-number count, losing the species distinction.

Within a single species, the Fock space is a direct *sum* over particle numbers for that species—the structure we developed in the previous section. So the full picture is: a direct product of Fock spaces (one per species), each of which is a direct sum over particle numbers for that species.

Enlarging the index set. An equivalent and often more transparent description avoids the species-by-species decomposition entirely. One simply specifies a larger index set for each particle: instead of separate spaces for electrons and quarks, one introduces a *species index* (or flavour and colour indices) that labels which type of particle a given quantum is. The basis vector:

$$|n;\,(\mathbf{k}_1, s_1, f_1, c_1), \ldots, (\mathbf{k}_n, s_n, f_n, c_n)\rangle\,, \tag{139}$$

where f is flavour ($e^-, \nu_e, u, d, \ldots$) and c is colour (r, g, b for quarks, trivial for leptons), directly encodes all the information of (137). This is the index-enrichment perspective: the direct product of species-specific Fock spaces is the same Hilbert space as the Fock space of a single particle type with a richer (flavour + colour) index.

Index	**Values**	**Origin**
Momentum $\mathbf{k}$	$\mathbb{R}^3$ (continuous)	Translation symmetry
Lorentz spin s	$\{1, \ldots, 2j_L + 1\}$	Diagram 3, $SL(2,\mathbb{C})$ rep
Flavour f	$\{e^-, \nu_e, u, d, s, c, b, t, \ldots\}$	Species label
Colour c	$\{r, g, b\}$ for quarks, trivial for leptons	$SU(3)_c$ rep
Isospin i	$\{1, 2\}$ for doublets	$SU(2)_L$ rep
Helicity λ	$\{\pm 1\}$ for photons/gluons	Gauge boson polarisation
Particle number n: organises the direct sum (not an index on each particle)		

A basis vector of $\mathcal{F}_{\text{SM}}$ is a joint specification of all of the above for each of the n particles present. The Schrödinger RNN rotates the hidden state among all such basis vectors. When H_{SM} contains a QED vertex $e\bar{\Psi}\gamma^\mu\Psi A_\mu$, it couples basis vectors that differ by one electron (or positron) and one photon: it is an off-diagonal block connecting $|n_{e^-}, n_\gamma\rangle$ to $|n_{e^-} \pm 1, n_\gamma \pm 1\rangle$. The story of QED—electrons emitting

and absorbing photons, pair creation and annihilation—is the story of how $H_{\rm SM}$ moves amplitude among the labelled basis vectors of $\mathcal{F}_{\rm SM}$.

15.4 The word "tensor": ML vs. physics

The word *tensor* has different meanings in machine learning and in physics, and the collision is worth resolving cleanly.

ML tensor: a multi-dimensional array. In PyTorch, TensorFlow, or JAX, a tensor is a multi-dimensional array of numbers—a generalisation of scalars (rank 0), vectors (rank 1), and matrices (rank 2) to arbitrary rank. A batch of images with shape (B, C, H, W) is a rank-4 tensor. No transformation law is implied; the word refers only to the data structure.

Physics tensor: an array with a transformation law. In physics, a tensor is an array whose components transform according to the chain rule under coordinate changes. A Lorentz tensor of type (p, q) transforms with p factors of Λ on upper indices and q factors of Λ^{-T} on lower indices (Section **??**). Not every multi-dimensional array is a physics tensor: the Dirac spinor $\Psi \in \mathbb{C}^4$ is a 4-component array but transforms with $S(\Lambda) \neq \Lambda$, so it is a spinor, not a Lorentz 4-vector.

The connection. The tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$ produces an ML tensor (a multi-index array of coefficients w_{ij}) whose transformation law under the symmetry group is determined by the representations on $\mathcal{H}_A$ and $\mathcal{H}_B$. When both factors carry Lorentz representations, the resulting coefficient array is a physics tensor. When one factor carries a Lorentz representation and another carries a gauge representation, the array transforms with both. Adding a transformation law to an ML tensor makes it a physics tensor. The ML tensor is the container; the physics tensor is the container plus the specification of how it transforms.

16 The Dirac Field: Electrons and Positrons (Spin 1/2)

16.1 From diagram 3 to the electron field

In diagram 3 the Dirac field sits at $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$: the field value is a 4-component complex vector $\Psi(X) = (\psi_L, \psi_R)^T \in \mathbb{C}^4$ whose two halves transform under opposite $SL(2, \mathbb{C})$ factors. The transformation matrix is block-diagonal:

$$M(\Lambda) = S(\Lambda) = \begin{pmatrix} L & 0 \\ 0 & \bar{L}^\dagger \end{pmatrix}, \qquad \Psi(X) \longmapsto S(\Lambda)\, \Psi(\Lambda^{-1} X). \tag{140}$$

The left block L acts on ψ_L (left-handed); the right block $\bar{L}^\dagger$ acts on ψ_R (right-handed). Mass will couple the two halves. The 4-component structure encodes spin ($\pm\frac{1}{2}$) and particle/antiparticle—four degrees of freedom per spacetime point.

16.2 Gamma matrices and Clifford algebra

The Dirac equation requires 4×4 matrices that simultaneously encode Lorentz covariance and the relativistic dispersion relation $p^\mu p_\mu = m^2$. These are the *gamma matrices* γ^μ, and they are built directly from the Pauli matrices.

16.2.1 From Pauli matrices to gamma matrices

Recall the Pauli matrices from Section 8:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{141}$$

satisfying $\sigma_i \sigma_j = \delta_{ij} I + i\epsilon_{ijk}\sigma_k$. They act on $\mathbb{C}^2$ spinors. For the Dirac equation we need objects that act on $\mathbb{C}^4$—the four-component Dirac spinor $\Psi = (\psi_L, \psi_R)^T$ that pairs both chiralities.

The standard construction (Weyl or chiral representation) builds 4×4 gamma matrices as 2×2 block matrices of Pauli matrices:

$$\gamma^0 = \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix}, \qquad \gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}, \quad i = 1, 2, 3, \tag{142}$$

where I_2 is the 2×2 identity matrix. The index μ runs over $0, 1, 2, 3$ in Einstein notation, so γ^μ denotes the collection $\{\gamma^0, \gamma^1, \gamma^2, \gamma^3\}$.

Why four components? The Dirac spinor $\Psi \in \mathbb{C}^4$ carries four degrees of freedom: two spins ($s = \pm\frac{1}{2}$) for the particle, and two spins for the antiparticle. The 4×4 gamma matrices are the minimal matrices that can mix all four components in a Lorentz-covariant way. The Pauli matrices, being 2×2, act within each chirality block; the gamma matrices, by placing Pauli matrices off-block-diagonal, mix left- and right-handed components and produce the mass coupling.

16.2.2 The Clifford algebra: the defining relation

The gamma matrices are defined not by any specific representation (142) but by the algebraic relation they satisfy. The *Clifford algebra* in Minkowski space is:

$$\boxed{\{\gamma^\mu,\, \gamma^\nu\} \equiv \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\,\eta^{\mu\nu}\, I_4.} \tag{143}$$

This single relation is all that is needed. Any representation of (143) in any dimension gives a valid set of gamma matrices. The Weyl representation (142) is one choice; the Dirac representation (with $\gamma^0 = \mathrm{diag}(I, -I)$) is another. Physical predictions are independent of the choice.

Reading the Clifford algebra. For $\mu = \nu$: $\{\gamma^\mu, \gamma^\mu\} = 2(\gamma^\mu)^2 = 2\eta^{\mu\mu} I_4$, so:

$$(\gamma^0)^2 = I_4, \qquad (\gamma^i)^2 = -I_4, \quad i = 1, 2, 3. \tag{144}$$

For $\mu \neq \nu$: $\gamma^\mu\gamma^\nu = -\gamma^\nu\gamma^\mu$—the gamma matrices anticommute when their indices differ.

Connection to Pauli anticommutation. Equation (143) is the four-dimensional generalisation of the Pauli anticommutation relation $\{\sigma_i, \sigma_j\} = 2\delta_{ij}I$. The Pauli matrices generate the Euclidean Clifford algebra $\{\sigma_i, \sigma_j\} = 2\delta_{ij}$; the gamma matrices generate the Minkowski Clifford algebra $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$. The sign difference between δ_{ij} (all positive) and $\eta^{\mu\nu}$ (mixed signs) is why the timelike gamma matrix (144) squares to $+I$ while the spacelike ones square to $-I$.

Verifying the Clifford algebra in the Weyl representation.
Check $\{\gamma^0, \gamma^1\} = 0$ (since $\eta^{01} = 0$):

$$\gamma^0\gamma^1 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}\begin{pmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{pmatrix} = \begin{pmatrix} -\sigma_1 & 0 \\ 0 & \sigma_1 \end{pmatrix},$$
$$\gamma^1\gamma^0 = \begin{pmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{pmatrix}\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & -\sigma_1 \end{pmatrix}.$$

Summing: $\gamma^0\gamma^1 + \gamma^1\gamma^0 = 0 = 2\eta^{01}I_4$. ✓
Check $(\gamma^1)^2 = -I_4$ (since $\eta^{11} = -1$):

$$(\gamma^1)^2 = \begin{pmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{pmatrix}^2 = \begin{pmatrix} -\sigma_1^2 & 0 \\ 0 & -\sigma_1^2 \end{pmatrix} = \begin{pmatrix} -I & 0 \\ 0 & -I \end{pmatrix} = -I_4. \quad ✓$$

16.2.3 Properties and gamma technology

Hermiticity. In the Weyl representation:

$$(\gamma^0)^\dagger = \gamma^0, \qquad (\gamma^i)^\dagger = -\gamma^i. \tag{145}$$

These two cases are unified by the identity:

$$(\gamma^\mu)^\dagger = \gamma^0\,\gamma^\mu\,\gamma^0, \tag{146}$$

which holds in any standard representation. This is used to construct the Dirac conjugate $\bar{\Psi} = \Psi^\dagger\gamma^0$, which transforms correctly under Lorentz transformations.

Covariance. The gamma matrices transform as a Lorentz 4-vector in the sense that conjugation by the Dirac representation matrix $S(\Lambda)$ gives:

$$S^{-1}(\Lambda)\,\gamma^\mu\,S(\Lambda) = \Lambda^\mu{}_\nu\,\gamma^\nu. \tag{147}$$

This is what makes $\bar{\Psi}\gamma^\mu\Psi$ a Lorentz 4-vector and $\bar{\Psi}\gamma^\mu\partial_\mu\Psi$ a Lorentz scalar in the Dirac Lagrangian.

Products of gamma matrices. Using the Clifford algebra (143) repeatedly, any product of gamma matrices can be reduced to one of the 16 independent 4×4 matrices:

$$\Gamma^A \in \{I,\ \gamma^\mu,\ \gamma^{\mu\nu},\ \gamma^{\mu\nu\rho},\ \gamma^5\}, \tag{148}$$

where antisymmetric products are defined as:

$$\gamma^{\mu\nu} = \tfrac{1}{2}[\gamma^\mu, \gamma^\nu] = \gamma^{[\mu}\gamma^{\nu]}, \qquad \gamma^{\mu\nu\rho} = \gamma^{[\mu}\gamma^\nu\gamma^{\rho]}, \tag{149}$$

and γ^5 is defined below. These 16 matrices ($1+4+6+4+1=16$) form a complete basis for all 4×4 complex matrices.

The fifth gamma matrix γ^5. The *chirality matrix* is:

$$\gamma^5 = i\,\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} -I_2 & 0 \\ 0 & I_2 \end{pmatrix} \quad \text{(Weyl representation)}. \tag{150}$$

It satisfies:

$$(\gamma^5)^2 = I_4, \qquad \{\gamma^5, \gamma^\mu\} = 0, \qquad (\gamma^5)^\dagger = \gamma^5. \tag{151}$$

The anticommutation $\{\gamma^5, \gamma^\mu\} = 0$ means γ^5 flips between the two chirality blocks. Its eigenvalues are ± 1; the corresponding projection operators:

$$P_L = \frac{I - \gamma^5}{2} = \begin{pmatrix} I_2 & 0 \\ 0 & 0 \end{pmatrix}, \qquad P_R = \frac{I + \gamma^5}{2} = \begin{pmatrix} 0 & 0 \\ 0 & I_2 \end{pmatrix}, \tag{152}$$

project the Dirac spinor onto its left- and right-handed components: $\psi_L = P_L\Psi$, $\psi_R = P_R\Psi$. The weak force only couples to $P_L\Psi$ – left-handed fermions – which in the Lagrangian appears as $\bar{\Psi}\gamma^\mu P_L \Psi$.

16.2.4 Trace technology

Feynman diagram calculations in QED and QCD reduce cross-sections to traces of products of gamma matrices, via the completeness relation for spinor sums:

$$\sum_s u(\mathbf{p}, s)\bar{u}(\mathbf{p}, s) = \not{p} + m, \qquad \sum_s v(\mathbf{p}, s)\bar{v}(\mathbf{p}, s) = \not{p} - m. \tag{153}$$

Here $\not{p} = \gamma^\mu p_\mu = p_\mu \gamma^\mu$ is the *Feynman slash*—contracting a 4-vector with the gamma matrices via Einstein notation. The slash notation is convenient shorthand; the full expression $\gamma^\mu p_\mu$ is always what is meant.

The fundamental trace identities follow from (143):

$$\mathrm{Tr}[I_4] = 4, \tag{154}$$

$$\mathrm{Tr}[\gamma^\mu] = 0, \tag{155}$$

$$\mathrm{Tr}[\gamma^\mu\gamma^\nu] = 4\,\eta^{\mu\nu}, \tag{156}$$

$$\mathrm{Tr}[\text{odd number of } \gamma\text{'s}] = 0, \tag{157}$$

$$\mathrm{Tr}[\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma] = 4\,(\eta^{\mu\nu}\eta^{\rho\sigma} - \eta^{\mu\rho}\eta^{\nu\sigma} + \eta^{\mu\sigma}\eta^{\nu\rho}), \tag{158}$$

$$\mathrm{Tr}[\gamma^5\gamma^\mu\gamma^\nu] = 0, \tag{159}$$

$$\mathrm{Tr}[\gamma^5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma] = 4i\,\epsilon^{\mu\nu\rho\sigma}, \tag{160}$$

where $\epsilon^{\mu\nu\rho\sigma}$ is the Levi-Civita tensor with $\epsilon^{0123} = +1$.

Deriving (156): $\mathrm{Tr}[\gamma^\mu\gamma^\nu] = 4\eta^{\mu\nu}$.
From the Clifford algebra: $\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu} I_4$. Taking the trace of both sides:

$$\mathrm{Tr}[\gamma^\mu\gamma^\nu] + \mathrm{Tr}[\gamma^\nu\gamma^\mu] = 2\eta^{\mu\nu}\,\mathrm{Tr}[I_4] = 8\eta^{\mu\nu}.$$

Using cyclicity of trace ($\mathrm{Tr}[AB] = \mathrm{Tr}[BA]$): $\mathrm{Tr}[\gamma^\mu\gamma^\nu] = \mathrm{Tr}[\gamma^\nu\gamma^\mu]$. Therefore $2\,\mathrm{Tr}[\gamma^\mu\gamma^\nu] = 8\eta^{\mu\nu}$, giving $\mathrm{Tr}[\gamma^\mu\gamma^\nu] = 4\eta^{\mu\nu}$. ✓

A typical QED trace: spin-averaged $|M|^2$ for $e^-\mu^-$ scattering.
The lowest-order amplitude for electron-muon scattering via a photon is $M \propto \bar{u}(p')\gamma^\mu u(p) \cdot \bar{u}(k')\gamma_\mu u(k)$. Squaring and summing over initial/final spins using (153):

$$\sum_{\text{spins}} |M|^2 \propto \mathrm{Tr}\Big[(\not{p}' + m_e)\gamma^\mu(\not{p} + m_e)\gamma^\nu\Big] \cdot \mathrm{Tr}\Big[(\not{k}' + m_\mu)\gamma_\mu(\not{k} + m_\mu)\gamma_\nu\Big].$$

Expanding and applying (157) (odd traces vanish) and (158):

$$\mathrm{Tr}\Big[\not{p}'\gamma^\mu\not{p}\gamma^\nu\Big] = 4(p'^\mu p^\nu + p'^\nu p^\mu - \eta^{\mu\nu} p'\cdot p),$$
$$\mathrm{Tr}\Big[m_e^2\gamma^\mu\gamma^\nu\Big] = 4m_e^2\eta^{\mu\nu}.$$

The result—the Rosenbluth cross-section—follows by contracting the two traces over μ, ν. Every step is index algebra, powered by the Clifford relation (143) and the trace identities (154)–(160).

16.2.5 Contraction identities

In $d = 4$ dimensions, the Clifford algebra also gives contraction identities for gamma matrices with themselves. These are used to simplify numerators in Feynman diagram calculations:

$$\gamma^\mu\gamma_\mu = 4\,I_4, \tag{161}$$
$$\gamma^\mu\gamma^\nu\gamma_\mu = -2\,\gamma^\nu, \tag{162}$$
$$\gamma^\mu\gamma^\nu\gamma^\rho\gamma_\mu = 4\,\eta^{\nu\rho}\,I_4, \tag{163}$$
$$\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\gamma_\mu = -2\,\gamma^\sigma\gamma^\rho\gamma^\nu. \tag{164}$$

These follow by using $\gamma^\mu\gamma_\mu = \gamma^\mu\eta_{\mu\nu}\gamma^\nu = \frac{1}{2}\{\gamma^\mu,\gamma^\nu\}\eta_{\mu\nu} = \eta_{\mu\nu}\eta^{\mu\nu} I_4 = 4I_4$ for (161), and the Clifford relation to commute the contracted γ^μ past the remaining matrices for the others.

In dimensional regularisation with $d = 4 - 2\varepsilon$ dimensions, the 4 in (161)–(163) is replaced by d, which generates the ε-dependent terms that cancel UV divergences in renormalisation.

Summary: the gamma matrix toolkit.

Object	Definition	Role
γ^μ	4×4 matrices satisfying (143)	Encode Lorentz structure in Dirac eq.
$\{\gamma^\mu,\gamma^\nu\}=2\eta^{\mu\nu}I$	Clifford algebra	Master relation; implies all others
$\gamma^5=i\gamma^0\gamma^1\gamma^2\gamma^3$	Chirality matrix	Separates ψ_L from ψ_R
$P_{L,R}=(I\mp\gamma^5)/2$	Chiral projectors	Project onto chirality eigenstates
$\gamma^\mu p_\mu$ (slash notation)	Contraction with 4-momentum	Dirac operator in momentum space
$\bar\Psi=\Psi^\dagger\gamma^0$	Dirac conjugate	Makes $\bar\Psi\Psi$ Lorentz scalar
$\mathrm{Tr}[\gamma^\mu\gamma^\nu]=4\eta^{\mu\nu}$	Trace identity	Core of spin-sum calculations
$\gamma^\mu\gamma_\mu=4I$	Contraction identity	Simplifies numerators in loops

16.3 The mode expansion: electrons and positrons

Quantising the Dirac field, each Fourier mode splits into a *positive-frequency* part carrying an electron annihilation operator $b(\mathbf{p},s)$ and a *negative-frequency* part carrying a positron creation operator $d^\dagger(\mathbf{p},s)$:

$$\Psi(x)=\sum_{s=\pm\frac12}\int\frac{d^3p}{(2\pi)^3\sqrt{2\omega_{\mathbf{p}}}}\Big[b(\mathbf{p},s)\,u(\mathbf{p},s)\,e^{-ip\cdot x}\;+\;d^\dagger(\mathbf{p},s)\,v(\mathbf{p},s)\,e^{ip\cdot x}\Big]. \tag{165}$$

Here $\omega_{\mathbf{p}}=\sqrt{|\mathbf{p}|^2+m^2}$ and $p\cdot x=\omega_{\mathbf{p}}t-\mathbf{p}\cdot\mathbf{x}$. The spinors $u(\mathbf{p},s)$ and $v(\mathbf{p},s)$ are specific 4-component vectors in $\mathbb{C}^4$—the eigenvectors of the Dirac operator in momentum space that correspond to positive- and negative-energy solutions respectively:

$$(\gamma^\mu p_\mu-m)\,u(\mathbf{p},s)=0,\qquad(\gamma^\mu p_\mu+m)\,v(\mathbf{p},s)=0. \tag{166}$$

The operators obey fermionic anticommutation relations. There are two independent sets—one for electrons, one for positrons:

$$\Big\{b(\mathbf{p},s),\;b^\dagger(\mathbf{p}',s')\Big\}=(2\pi)^3\,\delta^{(3)}(\mathbf{p}-\mathbf{p}')\,\delta_{ss'}, \tag{167}$$

$$\Big\{d(\mathbf{p},s),\;d^\dagger(\mathbf{p}',s')\Big\}=(2\pi)^3\,\delta^{(3)}(\mathbf{p}-\mathbf{p}')\,\delta_{ss'}, \tag{168}$$

with all other anticommutators zero. The Fock space is the tensor product of the electron Fock space and the positron Fock space, each fermionic:

$$\mathcal{F}_{\text{Dirac}}=\mathcal{F}_{e^-}\otimes\mathcal{F}_{e^+}, \tag{169}$$

with basis states $b^\dagger(\mathbf{p}_1,s_1)\cdots d^\dagger(\mathbf{q}_1,r_1)\cdots|0\rangle$.

Why both b and $d^\dagger$ appear in $\Psi(x)$.
In the scalar field, $\hat{\phi} = ae^{-ip\cdot x} + a^\dagger e^{ip\cdot x}$: the field at x is a superposition of destroying a particle (positive frequency) and creating one (negative frequency). The Dirac field does the same, but the positive-frequency mode is a different particle from the negative-frequency mode: destroying an electron is not the same as creating a positron. The field $\Psi(x)$ is therefore the operator that simultaneously "could destroy an electron here" or "could create a positron here." This is Dirac's insight: the negative-energy solutions of his equation are not unphysical; they describe positrons. The field $\Psi(x)$ at a single spacetime point is a coherent superposition of both possibilities.

16.4 The Dirac Lagrangian

The Lagrangian density is:

$$\mathcal{L}_{\text{Dirac}} = \bar{\Psi}\,(i\gamma^\mu\partial_\mu - m)\,\Psi, \qquad \bar{\Psi} \equiv \Psi^\dagger\gamma^0. \tag{170}$$

The gamma matrices γ^μ satisfy the Clifford algebra $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}I$, which is the algebraic packaging of the relativistic dispersion relation $p^\mu p_\mu = m^2$. The combination $\bar{\Psi} = \Psi^\dagger\gamma^0$ is the "Dirac conjugate"; the γ^0 is necessary to make the mass term $\bar{\Psi}\Psi$ Lorentz invariant rather than just Hermitian.

Lorentz invariance. Under diagram 3, $\Psi \to S(\Lambda)\Psi$. A key algebraic identity for the Dirac representation is:

$$S^{-1}(\Lambda)\,\gamma^\mu\,S(\Lambda) = \Lambda^\mu{}_\nu\,\gamma^\nu, \tag{171}$$

which says that conjugation by $S(\Lambda)$ rotates γ^μ as a 4-vector. From this:

$$\bar{\Psi} \to \bar{\Psi}\,S^{-1}(\Lambda), \qquad \bar{\Psi}\Psi \to \bar{\Psi}\,S^{-1}S\,\Psi = \bar{\Psi}\Psi. \quad \checkmark \tag{172}$$

For the kinetic term, using (171) and the fact that $\partial_\mu \to \Lambda^{-1}{}_\mu{}^\nu\partial_\nu$ under $x \to \Lambda x$:

$$\bar{\Psi}\,\gamma^\mu\partial_\mu\,\Psi \;\to\; \bar{\Psi}\,S^{-1}\,\gamma^\mu\,S\,\Lambda^{-1}{}_\mu{}^\nu\,\partial_\nu\,\Psi = \bar{\Psi}\,\Lambda^\mu{}_\rho\gamma^\rho\,\Lambda^{-1}{}_\mu{}^\nu\,\partial_\nu\,\Psi = \bar{\Psi}\,\gamma^\nu\partial_\nu\,\Psi. \quad \checkmark \tag{173}$$

Both the mass term and kinetic term are Lorentz scalars; the action $S = \int d^4x\,\mathcal{L}$ is Lorentz invariant.

16.5 From Lagrangian to Hamiltonian

The conjugate momentum to Ψ is:

$$\pi = \frac{\partial\mathcal{L}}{\partial\dot{\Psi}} = i\Psi^\dagger. \tag{174}$$

The Hamiltonian density follows from the Legendre transform $\mathcal{H} = \pi\dot{\Psi} - \mathcal{L}$:

$$\mathcal{H}_{\text{Dirac}} = \Psi^\dagger\Big(-i\vec{\alpha}\cdot\nabla + \beta\,m\Big)\Psi, \qquad \vec{\alpha} = \gamma^0\vec{\gamma}, \quad \beta = \gamma^0. \tag{175}$$

Substituting the mode expansion (165) and normal-ordering (placing all creation operators to the left, with a sign change per anticommutation), the Hamiltonian simplifies to:

$$H_{\text{Dirac}} = \sum_s \int \frac{d^3p}{(2\pi)^3}\, \omega_{\mathbf{p}} \Big[b^\dagger(\mathbf{p}, s)\, b(\mathbf{p}, s) \;+\; d^\dagger(\mathbf{p}, s)\, d(\mathbf{p}, s) \Big]. \tag{176}$$

The Hamiltonian counts the total energy of all electrons and positrons present, weighted by their individual energies $\omega_{\mathbf{p}}$. It is diagonal in the Fock basis—the free Dirac field is an RNN with a phase-rotating hidden state, one phase per electron mode and one per positron mode.

16.6 What the Schrödinger equation tells us

The Fock hidden state for the Dirac field is a concatenation over all occupation-number sectors:

$$h_t = c_0(t)\; |0\rangle \;\oplus\; \bigoplus_s \int \frac{d^3p}{(2\pi)^3}\, c_e(\mathbf{p}, s;\, t)\; \Big|e^-(\mathbf{p}, s)\Big\rangle \;\oplus\; \bigoplus_s \int \frac{d^3p}{(2\pi)^3}\, c_{\bar{e}}(\mathbf{p}, s;\, t)\; \Big|e^+(\mathbf{p}, s)\Big\rangle \;\oplus\; \cdots \tag{177}$$

Under H_{Dirac}, each coefficient phase-rotates independently:

$$c_e(\mathbf{p}, s;\, t) = e^{-i\omega_{\mathbf{p}} t}\, c_e(\mathbf{p}, s;\, 0), \qquad \omega_{\mathbf{p}} = \sqrt{|\mathbf{p}|^2 + m^2}. \tag{178}$$

The story. The free Dirac RNN tells a quiet story: each electron and each positron drifts through momentum space while its amplitude accumulates phase at its own Compton frequency $\omega_{\mathbf{p}}$. The minimum frequency is $\omega = m$, achieved only for a particle at rest. This is the rest-mass energy: even a stationary electron oscillates at the Compton frequency m (restoring $\hbar$ and c: $mc^2/\hbar$). The mass m is the energy gap between the vacuum and any one-particle state; you cannot excite the field for free.

The spin quantum number $s = \pm\frac{1}{2}$ is carried by the 4-component spinor structure of the mode—the $u(\mathbf{p}, s)$ and $v(\mathbf{p}, s)$ vectors point in different directions in $\mathbb{C}^4$ for different s. The anticommutation relations ensure that no two electrons (or positrons) can share the same $(\mathbf{p}, s)$ mode—the Pauli exclusion principle is a structural consequence of fermionic statistics, not an add-on.

The story becomes rich when we add interactions: the free theory is the baseline that interactions perturb around.

16.7 Path integral for the Dirac field

The path integral for the Dirac field follows the same time-slicing argument, but with one new ingredient: the fermionic anticommutation relations require that the "classical" variables in the path integral are not ordinary numbers but *Grassmann variables*—anticommuting numbers satisfying $\theta_1\theta_2 = -\theta_2\theta_1$.

$$K[\bar{\Psi}_f, T;\, \Psi_i, 0] = \int_{\Psi(0)=\Psi_i}^{\bar{\Psi}(T)=\bar{\Psi}_f} \mathcal{D}\bar{\Psi}\,\mathcal{D}\Psi\;\exp\!\Big(i\int_0^T dt \int d^3x\;\mathcal{L}_{\text{Dirac}}(\bar{\Psi},\Psi)\Big). \tag{179}$$

The Grassmann nature of Ψ and $\bar{\Psi}$ in the path integral automatically enforces Pauli exclusion: the anticommutativity of the integration variables mirrors the anticommutativity of the creation and annihilation operators. The derivation is otherwise identical to the scalar case—time-slice, insert resolutions of identity, evaluate short-time matrix elements via Trotter splitting. The Lagrangian $\mathcal{L}_{\text{Dirac}}$ appears in the exponent, as always.

Remark. The Gaussian Grassmann integral $\int \mathcal{D}\bar{\Psi}\,\mathcal{D}\Psi\, e^{\int \bar{\Psi} M \Psi} = \det M$ is the fermionic analogue of the bosonic Gaussian $\int \mathcal{D}\phi\, e^{-\frac{1}{2}\phi^T M \phi} = (\det M)^{-1/2}$. The determinant appears in the numerator rather than the denominator: fermions contribute the *inverse* of the bosonic weight. This determinant—$\det(i\gamma^\mu\partial_\mu - m)$ for the free field, $\det(i\gamma^\mu D_\mu - m)$ in QED—generates all fermion loops when expanded perturbatively.

17 The Photon Field (Spin 1)

17.1 From diagram 3 to the photon

In diagram 3, the photon sits at $(\frac{1}{2}, \frac{1}{2})$: the field value is a 2×2 complex matrix. Writing the 4-potential as $A = A_\mu \sigma^\mu$, the transformation law (106) with $M_L = L$ and $M_R = L^\dagger$ gives:

$$A(X) \;\longmapsto\; L\,A(X)\,L^\dagger. \tag{180}$$

This is identical to the transformation of the spacetime matrix X itself (diagram 2). The photon field at each point transforms as a copy of spacetime—a 4-vector A_μ that rotates and boosts exactly as x^μ does.

In the conventional index notation, $A_\mu(x) \to \Lambda_\mu{}^\nu A_\nu(\Lambda^{-1}x)$, but the 2×2 matrix form $A = A_\mu\sigma^\mu$ makes the $SL(2,\mathbb{C})$ structure manifest: A transforms by conjugation, just as X does.

17.2 The mode expansion and photon helicity

In Coulomb gauge ($\nabla \cdot \mathbf{A} = 0$, $A_0 = 0$), the two physical transverse polarisations are parametrised by *helicity* $\lambda = \pm 1$ (circular polarisations). The mode expansion is:

$$A_\mu(x) = \sum_{\lambda=\pm1} \int \frac{d^3k}{(2\pi)^3\sqrt{2\omega_{\mathbf{k}}}} \Big[a(\mathbf{k},\lambda)\,\epsilon_\mu(\mathbf{k},\lambda)\,e^{-ik\cdot x} \;+\; a^\dagger(\mathbf{k},\lambda)\,\epsilon_\mu^*(\mathbf{k},\lambda)\,e^{ik\cdot x}\Big], \tag{181}$$

where $\omega_{\mathbf{k}} = |\mathbf{k}|$ (massless) and $\epsilon_\mu(\mathbf{k}, \lambda)$ are polarisation vectors orthogonal to $\mathbf{k}$. For momentum along $\hat{z}$:

$$\epsilon^\mu(\mathbf{k}, +1) = \tfrac{1}{\sqrt{2}}(0,\, 1,\, i,\, 0), \qquad \epsilon^\mu(\mathbf{k}, -1) = \tfrac{1}{\sqrt{2}}(0,\, 1,\, -i,\, 0). \tag{182}$$

The ladder operators obey bosonic commutation relations:

$$\left[a(\mathbf{k}, \lambda),\; a^\dagger(\mathbf{k}', \lambda')\right] = (2\pi)^3\, \delta^{(3)}(\mathbf{k} - \mathbf{k}')\, \delta_{\lambda\lambda'}. \tag{183}$$

The photon Fock space $\mathcal{F}_\gamma$ is the bosonic Fock space with two modes (helicities) per momentum $\mathbf{k}$.

17.3 The Maxwell Lagrangian

Define the field-strength tensor as the antisymmetric part of the potential's gradient:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{184}$$

Its six independent components carry the electric and magnetic fields: $F_{0i} = E_i$ and $\epsilon_{ijk} F_{jk} = B_i$. The Maxwell Lagrangian is:

$$\mathcal{L}_{\text{Maxwell}} = -\frac{1}{4}\, F_{\mu\nu} F^{\mu\nu} = \frac{1}{2}(E^2 - B^2). \tag{185}$$

The Euler–Lagrange equations are Maxwell's equations in vacuum: $\partial_\mu F^{\mu\nu} = 0$.

Lorentz invariance. Under $A_\mu \to \Lambda_\mu{}^\nu A_\nu$, the tensor $F_{\mu\nu}$ transforms as $F_{\mu\nu} \to \Lambda_\mu{}^\rho \Lambda_\nu{}^\sigma F_{\rho\sigma}$. The contraction $F_{\mu\nu}F^{\mu\nu} = F_{\mu\nu}\eta^{\mu\rho}\eta^{\nu\sigma}F_{\rho\sigma}$ has all four Lorentz indices contracted with the metric, making it a Lorentz scalar. The action $S = \int d^4x\, \mathcal{L}$ is therefore Lorentz invariant.

17.4 Gauge invariance: why the photon is massless

The Maxwell Lagrangian has a crucial symmetry: it is unchanged under the *gauge transformation*

$$A_\mu(x) \;\longmapsto\; A_\mu(x) + \partial_\mu \chi(x), \tag{186}$$

for any smooth scalar function $\chi(x)$. This holds because $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is unchanged: the $\partial_\mu\partial_\nu\chi$ terms cancel by symmetry of partial derivatives.

Why gauge invariance implies masslessness. A mass term $\frac{1}{2}m^2 A_\mu A^\mu$ would appear natural from a Lorentz-invariance perspective, but under (186):

$$A_\mu A^\mu \;\longmapsto\; (A_\mu + \partial_\mu\chi)(A^\mu + \partial^\mu\chi) \;\neq\; A_\mu A^\mu. \tag{187}$$

A mass term breaks gauge invariance. Gauge invariance therefore forbids the photon from acquiring a mass, and this is not a fine-tuning—it is a symmetry protection. The photon is exactly massless because the theory is exactly gauge invariant.

Gauge invariance as redundancy. The four components of A_μ are not four independent physical degrees of freedom. Configurations related by a gauge transformation (186) describe the same physics. Modding out this redundancy removes two components (the gauge direction and the longitudinal component), leaving exactly two physical degrees of freedom—the two helicities.

17.5 From Lagrangian to Hamiltonian

The conjugate momentum to A_i is $\pi^i = \partial\mathcal{L}/\partial\dot{A}_i = -E^i = F^{0i}$. The Legendre transform gives the Hamiltonian density:

$$\mathcal{H}_{\text{Maxwell}} = \frac{1}{2}(E^2 + B^2). \tag{188}$$

This is the total electromagnetic field energy—the sum of electric and magnetic energy densities. Substituting the mode expansion (181) and normal-ordering:

$$H_{\text{Maxwell}} = \sum_\lambda \int \frac{d^3k}{(2\pi)^3}\, \omega_{\mathbf{k}}\, a^\dagger(\mathbf{k},\lambda)\, a(\mathbf{k},\lambda), \qquad \omega_{\mathbf{k}} = |\mathbf{k}|. \tag{189}$$

Each photon mode contributes energy $\omega_{\mathbf{k}} = |\mathbf{k}|$ when occupied.

17.6 What the Schrödinger equation tells us

The free photon Hamiltonian (189) is diagonal in Fock space. The Schrödinger equation phase-rotates each photon-number sector independently. There are two qualitatively different types of states:

Number (Fock) states. A state $|n_1, n_2, \ldots\rangle$ with definite photon numbers in each mode phase-rotates at a fixed frequency. The expected value of A_μ is identically zero: $\langle A_\mu \rangle = 0$ for all t. A definite number of photons does not produce a classical field.

Coherent states. A coherent state—the eigenstate of the annihilation operator $a(\mathbf{k},\lambda)\,|\alpha\rangle = \alpha_{\mathbf{k}}\,|\alpha\rangle$—is a superposition of Fock states across all photon numbers, exactly as the harmonic oscillator coherent state of Section 6.5. Under Schrödinger evolution, a coherent state remains coherent: $\alpha_{\mathbf{k}}(t) = \alpha_{\mathbf{k}}\, e^{-i\omega_{\mathbf{k}}t}$. The expected value of A_μ oscillates sinusoidally:

$$\big\langle A_\mu(\mathbf{x},t)\big\rangle = \sum_\lambda \int \frac{d^3k}{(2\pi)^3\sqrt{2\omega_{\mathbf{k}}}}\Big[\alpha_{\mathbf{k}}\,\epsilon_\mu\, e^{i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}t)} + \text{c.c.}\Big]. \tag{190}$$

This is a classical electromagnetic wave satisfying Maxwell's equations. *Classical electromagnetism is the coherent-state limit of the photon RNN.* A laser is a coherent state of the photon field; a single photon is a Fock state. The Schrödinger equation of the quantum theory contains both, and interpolates between them continuously.

The story in one sentence. The free photon RNN tells this story: transverse electromagnetic waves propagate at the speed of light, carrying quantised energy $\omega_{\mathbf{k}}$ per photon; coherent superpositions of photon numbers are classical radiation; the vacuum is the ground state of all modes, with zero-point fluctuations that will become relevant when we couple to matter.

17.7 Path integral for the photon

The photon path integral follows the scalar recipe, with one complication: gauge invariance means the measure $\mathcal{D}A_\mu$ over-counts, since configurations related by (186) are physically equivalent. Integrating over all of them would give an infinite answer.

The resolution—the Faddeev–Popov procedure—amounts to choosing one representative configuration per gauge orbit ("fixing a gauge"), inserting a delta function to enforce the gauge condition $G(A) = 0$, and compensating with a Jacobian determinant. In Lorenz gauge $\partial^\mu A_\mu = 0$, the result is:

$$Z_{\text{Maxwell}} = \int \mathcal{D}A_\mu \; \delta(\partial^\mu A_\mu) \; \exp\!\Big(i \int d^4x \; \mathcal{L}_{\text{Maxwell}} \Big). \tag{191}$$

The delta function projects the path integral onto gauge-inequivalent configurations. In RNN terms: the Fock space of the photon already has only two physical modes per $\mathbf{k}$; the path integral over A_μ must be restricted to the same two-dimensional subspace by gauge fixing.

18 Quantum Electrodynamics: Gauge Invariance Forces the Coupling

The Dirac field and the photon field each have their own Lagrangian and their own Fock space. They could in principle be completely independent. The remarkable fact is that a single symmetry principle—*local gauge invariance*—forces them to interact, determines the form of the interaction uniquely, and fixes the coupling constant to be the electric charge e.

18.1 Global symmetry and the conserved current

The free Dirac Lagrangian (170) is invariant under a *global* U(1) phase rotation:

$$\Psi(x) \;\longmapsto\; e^{i\alpha}\Psi(x), \qquad \alpha \in \mathbb{R} \text{ (constant)}, \tag{192}$$

because the phase cancels between $\bar{\Psi}$ and Ψ. By Noether's theorem, this symmetry implies a conserved 4-current:

$$J^\mu = \bar{\Psi}\,\gamma^\mu\,\Psi, \qquad \partial_\mu J^\mu = 0. \tag{193}$$

The conserved charge is $Q = \int d^3x \, J^0 = \int d^3x \, \Psi^\dagger\Psi$, which counts (electrons minus positrons)—the electric charge.

18.2 Local gauge invariance demands a gauge field

The global symmetry (192) with constant α is a rigid rotation of the entire field. Now promote α to a function of spacetime: $\alpha = \alpha(x)$. The Dirac Lagrangian is no longer invariant:

$$\bar{\Psi}\, i\gamma^\mu \partial_\mu \Psi \;\longmapsto\; \bar{\Psi}\, i\gamma^\mu \partial_\mu \Psi \;-\; \bar{\Psi}\, \gamma^\mu (\partial_\mu \alpha)\, \Psi. \tag{194}$$

The extra term $-\bar{\Psi}\gamma^\mu(\partial_\mu\alpha)\Psi$ breaks invariance. To restore it, we introduce a gauge field A_μ and replace the ordinary derivative with the *covariant derivative*:

$$\partial_\mu \;\longmapsto\; D_\mu = \partial_\mu - ieA_\mu, \tag{195}$$

where e is the electric charge. If A_μ simultaneously transforms as:

$$A_\mu(x) \;\longmapsto\; A_\mu(x) + \frac{1}{e}\partial_\mu \alpha(x), \tag{196}$$

then the covariant derivative transforms covariantly:

$$D_\mu \Psi \;\longmapsto\; e^{i\alpha(x)}\, D_\mu \Psi, \tag{197}$$

and the extra term in (194) is exactly cancelled. The field A_μ that compensates for local phase rotations of the electron is precisely the electromagnetic potential. *Demanding local U(1) invariance predicts the existence of the photon.*

Verifying covariant transformation of $D_\mu\Psi$.

$$\begin{aligned} D_\mu\Psi &= (\partial_\mu - ieA_\mu)\Psi \\ &\longmapsto \Big(\partial_\mu - ie\Big(A_\mu + \tfrac{1}{e}\partial_\mu\alpha\Big)\Big) e^{i\alpha}\Psi \\ &= e^{i\alpha}\Big(\partial_\mu + i\partial_\mu\alpha - ieA_\mu - i\partial_\mu\alpha\Big)\Psi = e^{i\alpha}(\partial_\mu - ieA_\mu)\Psi = e^{i\alpha}\, D_\mu\Psi. \quad \checkmark \end{aligned}$$

The $i\partial_\mu\alpha$ terms cancel exactly.

18.3 The QED Lagrangian

Replacing $\partial_\mu \to D_\mu$ in the Dirac Lagrangian and adding the Maxwell kinetic term for A_μ gives the complete QED Lagrangian:

$$\mathcal{L}_{\text{QED}} = \bar{\Psi}\,(i\gamma^\mu D_\mu - m)\,\Psi \;-\; \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{198}$$

Expanding $D_\mu = \partial_\mu - ieA_\mu$:

$$\mathcal{L}_{\text{QED}} = \underbrace{\bar{\Psi}\,(i\gamma^\mu\partial_\mu - m)\,\Psi}_{\text{free electron}} \;-\; \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{free photon}} \;+\; \underbrace{e\,\bar{\Psi}\,\gamma^\mu\,\Psi\,A_\mu}_{\text{interaction}}. \tag{199}$$

The interaction term $e\,J^\mu A_\mu$ is the product of the conserved current (193) with the photon field—a vertex at which one photon and one electron/positron line meet. The coupling constant is e, the electric charge. Everything is determined by gauge invariance; there is no freedom to choose a different form of coupling.

18.4 The QED Hamiltonian

The Legendre transform of (199) gives:

$$H_{\text{QED}} = H_{\text{Dirac}} + H_{\text{Maxwell}} + H_{\text{int}}, \tag{200}$$

where H_{Dirac} (176) and H_{Maxwell} (189) are the free parts, and:

$$H_{\text{int}} = -e \int d^3x \; J^\mu A_\mu = -e \int d^3x \; \bar{\Psi}\, \gamma^\mu \, \Psi \, A_\mu. \tag{201}$$

Expanding Ψ via (165) and A_μ via (181), the interaction Hamiltonian is a sum of terms of the form $b^\dagger b\, a$, $d^\dagger d\, a^\dagger$, $b^\dagger d^\dagger a$, $dba^\dagger$, and so on. Each term shifts the particle content of the Fock hidden state.

Term in H_{int}	**Acts on**	**Physical process**
$b^\dagger(\mathbf{p}')\, b(\mathbf{p})\, a(\mathbf{k})$	$\lvert e^-, \gamma\rangle \to \lvert e^-\rangle$	Photon absorption by electron
$b^\dagger(\mathbf{p}')\, b(\mathbf{p})\, a^\dagger(\mathbf{k})$	$\lvert e^-\rangle \to \lvert e^-, \gamma\rangle$	Photon emission by electron
$b^\dagger(\mathbf{p})\, d^\dagger(\mathbf{q})\, a(\mathbf{k})$	$\lvert \gamma\rangle \to \lvert e^-, e^+\rangle$	Pair creation from photon
$d(\mathbf{q})\, b(\mathbf{p})\, a^\dagger(\mathbf{k})$	$\lvert e^-, e^+\rangle \to \lvert \gamma\rangle$	Pair annihilation to photon

Each of these is a specific off-diagonal block of the Hamiltonian matrix on $\mathcal{F}_{e^-} \otimes \mathcal{F}_{e^+} \otimes \mathcal{F}_\gamma$. The Hamiltonian H_{QED} is the sum of all such blocks. It is still a linear operator on the combined Fock space—still a weight matrix of the RNN.

18.5 What the Schrödinger equation tells us: the QED story

The Schrödinger equation $\dot{h}_t = -iH_{\text{QED}}\, h_t$ now mixes the electron, positron, and photon Fock sectors. Starting from a simple initial state, the hidden state evolves into a superposition of all particle-content configurations that are connected to it by the interaction Hamiltonian.

Consider an initial state with one electron at rest and no photons: $h_0 = \left|e^-(\mathbf{0}, +\tfrac{1}{2})\right\rangle$. The interaction Hamiltonian (201) immediately begins driving transitions:

$$\left|e^-\right\rangle \;\to\; \left|e^-\right\rangle + \epsilon\Big[c_1 \left|e^-, \gamma\right\rangle + c_2 \left|e^-, \gamma, \gamma\right\rangle + \cdots\Big] + O(\epsilon^2).$$

At second order in e, the emitted photon can be reabsorbed, or it can produce a virtual e^+e^- pair that immediately annihilates. These virtual processes—the "quantum fluctuations" around the electron—are what cause the electron's charge and mass to be renormalised.

The *Born rule* gives the probability of each outcome upon measurement. If we measure particle content at time T, we find an electron and n photons with probability $|c_n(T)|^2$. The RNN hidden state carries all these amplitudes simultaneously;

measurement collapses it to one outcome (the Bohr layer) and re-embeds the system at the corresponding Fock basis state.

In one sentence. The QED story is: electrons and positrons emit and absorb photons, photons create and destroy electron-positron pairs, and the Fock hidden state rotates through all these possibilities simultaneously under $e^{-iH_{\rm QED}t}$, with the Born rule assigning probabilities to each outcome.

18.6 Path integral for QED

Combining the scalar, Dirac, and photon path integrals:

$$Z_{\rm QED} = \int \mathcal{D}\bar{\Psi}\,\mathcal{D}\Psi\,\mathcal{D}A_\mu\;\delta(\partial^\mu A_\mu)\;\exp\!\Big(i\int d^4x\;\mathcal{L}_{\rm QED}\Big). \tag{202}$$

The fermion fields are Grassmann variables; the photon field is bosonic; the delta function fixes the gauge. Since $\mathcal{L}_{\rm QED}$ is quadratic in Ψ and $\bar{\Psi}$, the fermion path integral is Gaussian and can be performed exactly:

$$\int \mathcal{D}\bar{\Psi}\,\mathcal{D}\Psi\;e^{i\int \bar{\Psi}(i\gamma^\mu D_\mu - m)\Psi} = \det(i\gamma^\mu D_\mu - m). \tag{203}$$

The result is the *fermion determinant*—a functional of A_μ that encodes all effects of virtual electron-positron pairs:

$$Z_{\rm QED} = \int \mathcal{D}A_\mu\;\det(i\gamma^\mu D_\mu - m)\;e^{-\frac{i}{4}\int F_{\mu\nu}F^{\mu\nu}}. \tag{204}$$

Expanding $\det = \exp\operatorname{Tr}\log$ perturbatively in e:

$$\det(i\gamma^\mu D_\mu - m) = \det(i\gamma^\mu\partial_\mu - m)\cdot\exp\!\Big[\sum_{n=1}^{\infty}\frac{(ie)^n}{n}\operatorname{Tr}\Big[(i\not{\partial}-m)^{-1}A\Big]^n\Big]. \tag{205}$$

Each term in this expansion is a fermion loop with n photon insertions—a Feynman diagram. The $n=2$ term is the photon self-energy (vacuum polarisation); the $n=1$ term vanishes by Furry's theorem. Every Feynman diagram is a term in the perturbative expansion of the path integral, which is itself the time-sliced matrix product of the QED evolution operator on the Fock hidden state.

Object	**Mathematical form**	**RNN meaning**
Free electron propagator	$(i\not{\partial}-m)^{-1}$	Green's function of $W=-iH_{\rm Dirac}$
Free photon propagator	$-i\eta_{\mu\nu}/k^2$	Green's function of $W=-iH_{\rm Maxw}$
QED vertex	$ie\gamma^\mu$	Off-diagonal block of $H_{\rm int}$
Fermion loop	$\operatorname{Tr}\log\big(i\not{D}-m\big)$	Virtual e^+e^- sector in Fock path
Feynman diagram	Product of propagators and vertices	Term in $(U_{\phi'\phi})^N$ matrix product

The path integral is the master tool for analysis: it organises the perturbation theory, makes symmetries (including gauge invariance and Lorentz invariance) manifest at every order, and provides a direct bridge from the quantum Hamiltonian back to the classical Lagrangian. But the physical content is always in the Fock hidden state h_t, rotating under the unitary RNN $e^{-iH_{\text{QED}}t}$.

19 Electroweak Theory and the Higgs Mechanism

Quantum electrodynamics is built on a single gauge group $U(1)_{em}$. The electroweak theory unifies electromagnetism with the weak nuclear force under a larger gauge group $SU(2)_L \times U(1)_Y$, spontaneously broken to $U(1)_{em}$ by the Higgs field. The result: three massive gauge bosons ($W^{\pm}$, Z^0), one massless photon, and massive fermions—all from a single symmetry principle and a single field acquiring a vacuum expectation value. Every step follows the recipe of the previous sections: Lagrangian → Hamiltonian → Schrödinger (what story) → path integral (for analysis).

19.1 Extending diagram 3: internal symmetries

So far, diagram 3 tracked how a field value $\Phi(X)$ transforms when spacetime is Lorentz-transformed: the left index of Φ transforms by $M_L(L)$ and the right index by $M_R(L^\dagger)$. Gauge symmetry adds a new kind of transformation—one that acts on the field value at a single point without moving the point. These *internal* indices are orthogonal to the Lorentz (external) indices.

For the electroweak theory, the field $\Phi(x)$ carries both:

- **Lorentz indices**—spinor indices acted on by $M_L(L)$, $M_R(L^\dagger)$ as in diagram 3.
- **Isospin indices**—$SU(2)$ indices acted on by a local gauge transformation $U(g(x)) \in SU(2)_L$.

The two actions commute: $(U\Phi)M_R = U(\Phi M_R)$ since they act on different indices. The extended transformation law is:

$$\Phi^{\alpha}{}_{s}(X) \;\longmapsto\; U(g)^{\alpha}{}_{\beta}\; M_L(L)^{s}{}_{s'}\; \Phi^{\beta}{}_{s'}(\Lambda^{-1}X), \tag{206}$$

where $\alpha, \beta \in \{1, 2\}$ are isospin indices and $s, s' \in \{1, 2\}$ are Weyl spinor indices. The field is a 2×2 complex matrix in which one $\mathbb{C}^2$ is the Lorentz spinor space (diagram 3) and the other is the isospin space (new).

ML reading. The field $\Phi(x)$ is now an embedding table with three independent index structures: (1) spacetime position x, (2) Lorentz spinor index (acted on by $SL(2,\mathbb{C})$ from diagram 3), (3) isospin index (acted on by the gauge group $SU(2)_L$). Classifying all fields is classifying all representations of the combined group $SL(2,\mathbb{C}) \times SU(2)_L \times U(1)_Y$—the group of "what can happen to an embedding" in the electroweak theory.

19.2 The electroweak gauge group $SU(2)_L \times U(1)_Y$

The gauge group has two factors.

$SU(2)_L$ has generators $\tau^a = \sigma^a/2$ ($a = 1, 2, 3$) and three gauge bosons W_μ^a. The subscript L is the essential asymmetry of the weak force: *this group acts only on left-handed fields*. Right-handed fields are singlets under $SU(2)_L$ (their isospin index is absent). Parity violation is built in at the level of the gauge group.

$U(1)_Y$ has a single gauge boson B_μ and a quantum number called *hypercharge* Y, different from electric charge. After symmetry breaking, electric charge is:

$$Q = T_3 + Y, \tag{207}$$

where T_3 is the third component of isospin (eigenvalue of τ^3).

The field content of one generation is:

Field	**Symbol**	**Lorentz rep.**	$SU(2)_L$	Y	Q
Left lepton doublet	$L = (\nu_L, e_L)^T$	$(\frac{1}{2}, 0)$	doublet	$-\frac{1}{2}$	$(0, -1)$
Right charged lepton	e_R	$(\frac{1}{2}, 0)$	singlet	-1	-1
Left quark doublet	$Q = (u_L, d_L)^T$	$(\frac{1}{2}, 0)$	doublet	$+\frac{1}{6}$	$(+\frac{2}{3}, -\frac{1}{3})$
Right up quark	u_R	$(\frac{1}{2}, 0)$	singlet	$+\frac{2}{3}$	$+\frac{2}{3}$
Right down quark	d_R	$(\frac{1}{2}, 0)$	singlet	$-\frac{1}{3}$	$-\frac{1}{3}$
Higgs doublet	$\phi = (\phi^+, \phi^0)^T$	$(0, 0)$	doublet	$+\frac{1}{2}$	$(+1, 0)$

The covariant derivatives encode the gauge couplings. For a left-handed doublet (e.g., L):

$$D_\mu L = \Big(\partial_\mu - ig\, W_\mu^a \tau^a - ig'\, Y B_\mu\Big) L, \tag{208}$$

where g is the $SU(2)_L$ coupling and g' is the $U(1)_Y$ coupling. For a right-handed singlet (e.g., e_R):

$$D_\mu e_R = \big(\partial_\mu - ig' Y B_\mu\big) e_R. \tag{209}$$

The $SU(2)_L$ term is absent for right-handed fields: they are invisible to W_μ^a.

19.3 The electroweak Lagrangian

The Lagrangian has four parts:

$$\mathcal{L}_{\text{EW}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{fermion}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}. \tag{210}$$

Gauge kinetic terms. For each gauge field we form a field-strength tensor (the non-Abelian generalization of $F_{\mu\nu}$):

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\,\epsilon^{abc} W^b_\mu W^c_\nu, \tag{211}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{212}$$

The extra term $g\epsilon^{abc}W^b_\mu W^c_\nu$ in (211) arises because $SU(2)$ is non-Abelian: the gauge bosons carry isospin charge and interact with each other. The gauge kinetic Lagrangian is:

$$\mathcal{L}_{\text{gauge}} = -\tfrac{1}{4} W^a_{\mu\nu} W^{a\,\mu\nu} - \tfrac{1}{4} B_{\mu\nu} B^{\mu\nu}. \tag{213}$$

Fermion kinetic terms. For each fermion, replace ∂_μ with the appropriate covariant derivative:

$$\mathcal{L}_{\text{fermion}} = \bar{L}\, i\gamma^\mu D_\mu L + \bar{e}_R\, i\gamma^\mu D_\mu e_R + \bar{Q}\, i\gamma^\mu D_\mu Q + \bar{u}_R\, i\gamma^\mu D_\mu u_R + \bar{d}_R\, i\gamma^\mu D_\mu d_R. \tag{214}$$

There are no bare mass terms: $m\bar{\Psi}\Psi = m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$ would mix L (an $SU(2)$ doublet) with e_R (an $SU(2)$ singlet), violating gauge invariance. Fermion masses require the Higgs.

Higgs kinetic and potential terms. The Higgs field $\phi \in \mathbb{C}^2$ is a scalar ($j_L = j_R = 0$) under Lorentz but a doublet under $SU(2)_L$. Its Lagrangian is:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \tag{215}$$

with the *Higgs potential*:

$$V(\phi) = -\mu^2\, \phi^\dagger \phi + \lambda\, (\phi^\dagger \phi)^2, \qquad \mu^2 > 0,\ \lambda > 0. \tag{216}$$

Yukawa couplings. The only gauge-invariant way to give fermions mass is to couple them to the Higgs:

$$\mathcal{L}_{\text{Yukawa}} = -y_e\, \bar{L}\, \phi\, e_R - y_u\, \bar{Q}\, \tilde{\phi}\, u_R - y_d\, \bar{Q}\, \phi\, d_R + \text{h.c.}, \tag{217}$$

where $\tilde{\phi} = i\sigma_2 \phi^*$ is the charge-conjugate Higgs doublet (with $Y = -\frac{1}{2}$, needed to give up-type quarks mass). The couplings y_e, y_u, y_d are dimensionless numbers—the Yukawa coupling constants.

19.4 The Higgs mechanism: spontaneous symmetry breaking

The Lagrangian (210) is invariant under $SU(2)_L \times U(1)_Y$. The Higgs potential (216), however, has a *degenerate minimum*: it vanishes not at $\phi = 0$ but on a sphere of minima at $|\phi|^2 = v^2/2$, where

$$v = \sqrt{\mu^2/\lambda} \approx 246\,\text{GeV}. \tag{218}$$

The quantum ground state (the Fock vacuum $|0\rangle$) must choose one point on this sphere. Any choice breaks the symmetry: the Lagrangian is symmetric but the vacuum is not. This is *spontaneous symmetry breaking* (SSB).

The Mexican hat.
Think of $V(\phi)$ as the height of a Mexican hat, with the rim at radius $|\phi| = v/\sqrt{2}$. Rolling a marble from the centre to the rim costs energy (the direction perpendicular to the rim). Rolling the marble *along* the rim costs nothing (the tangential directions). These zero-cost directions are the Goldstone bosons. But we have a gauge symmetry: different points on the rim are gauge-equivalent. The Goldstone bosons are not physical particles; they are gauge artifacts absorbed into the longitudinal polarisations of the W and Z bosons.

Unitary gauge. Using the $SU(2)_L$ gauge freedom, we can always rotate the Higgs doublet to point in the lower component:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{219}$$

where $h(x)$ is the physical *Higgs field*—the radial fluctuation around the vacuum. The three Goldstone directions have been gauge-transformed away; they will reappear as the longitudinal polarisations of $W^\pm$ and Z^0.

Mass generation for gauge bosons. Substituting (219) into the Higgs kinetic term $(D_\mu\phi)^\dagger(D^\mu\phi)$ and expanding around v:

$$D_\mu\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} -igW^1_\mu \cdot \frac{v+h}{\sqrt{2}} + W^2_\mu \cdot \frac{v+h}{\sqrt{2}} \\ \partial_\mu h - i\frac{g}{2}W^3_\mu(v+h) + i\frac{g'}{2}B_\mu(v+h) \end{pmatrix} + \cdots \tag{220}$$

The terms quadratic in the gauge fields at $h = 0$ give mass terms:

$$\mathcal{L}^{\text{gauge}}_{\text{mass}} = m_W^2\, W^+_\mu W^{-\mu} + \tfrac{1}{2} m_Z^2\, Z_\mu Z^\mu + 0 \cdot A_\mu A^\mu, \tag{221}$$

where we define the physical mass eigenstates by a rotation by the *Weinberg angle* θ_W:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp i W^2_\mu), \tag{222}$$

$$Z_\mu = W^3_\mu \cos\theta_W - B_\mu \sin\theta_W, \tag{223}$$

$$A_\mu = W^3_\mu \sin\theta_W + B_\mu \cos\theta_W, \tag{224}$$

with $\cos\theta_W = g/\sqrt{g^2 + g'^2}$. The masses are:

$$m_W = \frac{gv}{2}, \qquad m_Z = \frac{m_W}{\cos\theta_W}, \qquad m_\gamma = 0. \tag{225}$$

The photon is the unique linear combination of W^3 and B that remains massless after SSB. Its gauge invariance is the residual $U(1)_{em}$ left unbroken by the Higgs vacuum.

Mass generation for fermions. Substituting (219) into the Yukawa coupling (217):

$$-y_e\,\bar{L}\,\phi\,e_R \;\to\; -\frac{y_e(v+h)}{\sqrt{2}}\,\bar{e}_L e_R + \text{h.c.} = -m_e\,\bar{e}e \;-\; \frac{y_e h}{\sqrt{2}}\,\bar{e}e, \tag{226}$$

where $m_e = y_e v/\sqrt{2}$. The first term is the electron mass; the second is the Higgs–electron coupling (proportional to m_e/v). Every fermion mass is generated this way:

$$m_f = \frac{y_f\, v}{\sqrt{2}}. \tag{227}$$

The vacuum expectation value $v \approx 246\,\text{GeV}$ is the single scale that sets all masses in the Standard Model. The Yukawa couplings y_f are the only remaining free parameters for fermion masses.

The Higgs boson mass. The radial fluctuation $h(x)$ has a mass from the curvature of the potential at the minimum:

$$V\!\left(\tfrac{v+h}{\sqrt{2}}\right) = \lambda\, v^2\, h^2 + \lambda\, v\, h^3 + \tfrac{\lambda}{4}\, h^4 + \text{const}, \qquad m_h = v\sqrt{2\lambda}. \tag{228}$$

The Higgs boson, discovered at the LHC in 2012 with $m_h \approx 125\,\text{GeV}$, is the quantum of the radial field $h(x)$—the only particle whose mass sets the scale v directly.

19.5 The Hamiltonian and its Fock space

The Legendre transform of (210) gives:

$$H_{\text{EW}} = H_W + H_Z + H_\gamma + H_{\text{fermion}} + H_h + H_{\text{int}}, \tag{229}$$

where each free part has the standard form $\int \omega_{\mathbf{k}}\, N(\mathbf{k})\, d\mathbf{k}$ with the appropriate mass:

Field	**Mass**	**Statistics**	**d.o.f.**	$\omega_{\mathbf{p}}$
$W^\pm$	$m_W = gv/2$	Boson	3 (massive)	$\sqrt{\lvert\mathbf{p}\rvert^2 + m_W^2}$
Z^0	$m_Z = m_W/\cos\theta_W$	Boson	3 (massive)	$\sqrt{\lvert\mathbf{p}\rvert^2 + m_Z^2}$
γ	0	Boson	2 (massless)	$\lvert\mathbf{p}\rvert$
$e, \nu, q, \ldots$	$m_f = y_f v/\sqrt{2}$	Fermion	4 (Dirac)	$\sqrt{\lvert\mathbf{p}\rvert^2 + m_f^2}$
h	$m_h = v\sqrt{2\lambda}$	Boson	1	$\sqrt{\lvert\mathbf{p}\rvert^2 + m_h^2}$

Massive gauge bosons ($W^\pm$, Z^0) have three polarisations rather than two: the longitudinal polarisation is the Goldstone boson that was "eaten" in unitary gauge. The photon remains massless and transverse (two polarisations).

The Fock hidden state is a concatenation over all particle sectors: $h_t \in \mathcal{F}_{W^+} \otimes \mathcal{F}_{W^-} \otimes \mathcal{F}_Z \otimes \mathcal{F}_\gamma \otimes \mathcal{F}_h \otimes \mathcal{F}_e \otimes \cdots$. The free Hamiltonian is block-diagonal; the interaction Hamiltonian H_{int} couples the sectors.

19.6 What the Schrödinger equation tells us

The Weinberg angle is a rotation in field space. The physical fields Z_μ and A_μ in (223)–(224) are linear combinations of the original gauge fields W^3_μ and B_μ. In Fock space, the mass eigenstates $|Z\rangle$ and $|\gamma\rangle$ are rotated from the gauge eigenstates $|W^3\rangle$ and $|B\rangle$ by the same Weinberg angle θ_W. The Schrödinger equation propagates mass eigenstates; the interaction Hamiltonian is written in gauge eigenstates. This is the electroweak analogue of neutrino mixing: the basis in which propagation is diagonal differs from the basis in which coupling is diagonal.

The mass gap as a filter. The minimum energy to excite the $W^\pm$ sector is $m_W \approx 80\,\text{GeV}$. At low energies (say, $E \ll m_W$), the Fock hidden state cannot populate the W sectors: transitions into them are exponentially suppressed by the mass gap. The effective theory at low energies is QED plus a short-range four-fermion interaction (the Fermi theory of weak decays), obtained by integrating out the heavy W and Z from the path integral.

The electroweak story. Start from the vacuum $|0\rangle$. An electron (e^-) and an electron-neutrino (ν_e) are created. The interaction Hamiltonian contains a term $g\,\bar{\nu}_L\,\gamma^\mu\,e_L\,W^+_\mu + \text{h.c.}$: an electron can convert into a neutrino by emitting a W^-. The Fock hidden state evolves as:

$$\left|e^-,\nu_e\right\rangle \;\to\; \left|e^-,\nu_e\right\rangle + \epsilon\,c_1\left|\nu_e,\nu_e,W^-\right\rangle + \epsilon\,c_2\left|e^-,e^-,W^+\right\rangle + \cdots$$

This is beta decay: $e^- \to \nu_e + W^- \to \nu_e + \bar{\nu}_e + e^-$ (the W propagates virtually and decays). The Born rule gives the decay probability. Every weak decay—neutron decay, pion decay, muon decay—is a specific off-diagonal block of H_{int} mixing different Fock sectors.

The Higgs boson in the hidden state. The Higgs field $h(x)$ has its own Fock sector. A state with one Higgs boson is $a^\dagger_h(\mathbf{p})\,|0\rangle$. The Yukawa coupling $-\frac{m_f}{v}\,\bar{f}fh$ in the Hamiltonian allows a Higgs boson to decay into a fermion-antifermion pair: $|h\rangle \to \left|f,\bar{f}\right\rangle$. The Higgs coupling to a fermion is proportional to m_f/v—heavy fermions couple more strongly, which is why $h \to b\bar{b}$ (bottom quark) dominates over $h \to e^+e^-$. This is not an additional assumption; it follows directly from (227): the same Yukawa coupling y_f that generates the mass also generates the Higgs vertex.

19.7 The electroweak path integral

The path integral for the electroweak theory integrates over all field histories:

$$Z_{\text{EW}} = \int \mathcal{D}W^a_\mu\,\mathcal{D}B_\mu\,\mathcal{D}\phi\,\mathcal{D}\bar{\Psi}\,\mathcal{D}\Psi\;\delta(G_{\text{gauge}})\;\exp\!\Big(i\int d^4x\;\mathcal{L}_{\text{EW}}\Big), \tag{230}$$

where $\delta(G_{\text{gauge}})$ fixes the non-Abelian gauge freedom (Faddeev–Popov).

Integrating out the Higgs. In unitary gauge (219), the Higgs path integral splits into an integral over the radial mode $h(x)$ and a trivial integral over the three Goldstone modes (which have been gauge-fixed away). At low energies $E \ll m_h$, we integrate out $h(x)$ as well, generating effective interactions among gauge bosons and fermions suppressed by $1/m_h^2$.

Integrating out W and Z. At energies $E \ll m_W$, integrating out $W^\pm$ and Z^0 from (230) generates the *Fermi effective Lagrangian*:

$$\mathcal{L}_{\text{Fermi}} = -\frac{G_F}{\sqrt{2}}(\bar{\nu}_L\gamma^\mu e_L)(\bar{e}_L\gamma_\mu\nu_L) + \cdots, \qquad \frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2}. \tag{231}$$

The Fermi constant $G_F \approx 1.17 \times 10^{-5}\,\text{GeV}^{-2}$, measured from muon decay, was known decades before the W boson was discovered. The electroweak path integral explains it: G_F is the coefficient of a four-fermion operator generated by integrating out a massive W propagator. In RNN terms: at low energies, the W Fock sector decouples, and its virtual effect is a short-range modification of the fermion-sector weight matrix.

Feynman diagrams of the electroweak theory. The perturbative expansion of (230) in g and g' generates diagrams with W, Z, γ, h, and fermion propagators, connected by the vertices of $\mathcal{L}_{\text{EW}}$. The non-Abelian gauge kinetic term (211) generates WWZ, $WW\gamma$, $WWZZ$, $WW\gamma Z$, and $WWWW$ vertices—self-interactions of the gauge bosons that have no counterpart in QED. Each diagram is, as always, a term in the matrix product $(U_{\phi'\phi})^N$ applied to the electroweak Fock hidden state.

Concept	Physical meaning	RNN / Fock meaning
$SU(2)_L \times U(1)_Y$	Gauge symmetry of SM	Internal index structure of field embedding
$\langle\phi\rangle = v/\sqrt{2}$	Higgs VEV	Preferred direction in Higgs Fock sector
$m_W = gv/2$	W mass from SSB	Energy gap in W Fock sector
Weinberg angle θ_W	W^3, B mixing	Rotation between gauge and mass eigenstates
Yukawa coupling y_f	Fermion-Higgs coupling	Off-diagonal block coupling f and h sectors
Fermion mass $m_f = y_f v/\sqrt{2}$	Mass from SSB	Energy gap set by v and y_f
Fermi constant G_F	Low-energy weak coupling	Effective weight after integrating out W
Higgs boson h	Radial excitation of ϕ	Mode in Higgs Fock sector, mass m_h

The electroweak theory is the same RNN as QED and the scalar field, with a richer field content (two gauge groups, one scalar doublet, three generations of fermions) and a richer vacuum (the Higgs VEV breaks $SU(2)_L \times U(1)_Y$ to $U(1)_{em}$). The Fock hidden state concatenates all sectors; the Schrödinger equation rotates it; the Born rule reads out probabilities; the path integral generates the perturbation theory. *The architecture never changes. Only the weight matrix and the vacuum get richer.*

20 Quantum Chromodynamics: Quarks and Gluons

Quantum chromodynamics (QCD) is the gauge theory of the strong nuclear force. Its gauge group is $SU(3)_c$—the subscript c for *colour*—with three colour charges

(red, green, blue) and eight gauge bosons called *gluons*. Quarks carry colour; leptons do not. Unlike the electroweak theory, QCD does not undergo spontaneous symmetry breaking: the gluons remain massless, and the coupling grows so strong at low energies that quarks are permanently *confined* inside colour-neutral hadrons.

The recipe is the same: Lagrangian → Hamiltonian → Schrödinger (what story) → path integral (for analysis). Diagram 3 extends effortlessly: the quark field simply acquires a third index structure—a colour index acted on by $SU(3)_c$—while the gluon field is, in matrix form, a second copy of the adjoint-representation story we already told for the W bosons.

20.1 Extending diagram 3: the colour index

A quark field $\psi_s^{f\alpha}(X)$ carries three independent index structures:

1. **Lorentz spinor index** $s \in \{1, 2, 3, 4\}$: acted on by the Dirac representation $S(\Lambda)$ from diagram 3, exactly as for the electron.
2. **Colour index** $\alpha \in \{r, g, b\}$: acted on by a local gauge transformation $U(x) \in SU(3)_c$.
3. **Flavour index** $f \in \{u, d, s, c, b, t\}$: a label for the six quark species; not acted on by any exact symmetry.

The combined transformation law is:

$$\psi_s^{f\alpha}(X) \;\longmapsto\; U(x)^{\alpha}{}_{\beta}\, S(\Lambda)^{s}{}_{s'}\, \psi_{s'}^{f\beta}(\Lambda^{-1}X). \tag{232}$$

Colour and Lorentz act on orthogonal index slots and commute.

The crucial difference from electroweak. Both left-handed and right-handed quarks carry colour: the $SU(3)_c$ gauge transformation (232) applies to ψ_L and ψ_R identically. QCD is *parity-symmetric*—the strong force treats left and right equally. This is why the quark mass term $m_q\bar{\psi}\psi$ is $SU(3)_c$-invariant without requiring a Higgs: colour rotations cancel between $\bar{\psi}$ and ψ. Quark masses come from electroweak SSB, not from QCD.

Gluons as adjoint matrices. The gluon field G_μ^a ($a = 1, \ldots, 8$) carries a colour-adjoint index. Writing $\mathbf{G}_\mu = G_\mu^a T^a$ as a traceless Hermitian 3×3 matrix (where $T^a = \lambda^a/2$ are the eight $SU(3)$ generators and λ^a are the Gell-Mann matrices), the gauge transformation is:

$$\mathbf{G}_\mu(x) \;\longmapsto\; U(x)\,\mathbf{G}_\mu(x)\,U^\dagger(x) + \frac{i}{g_s}(\partial_\mu U(x))\,U^\dagger(x). \tag{233}$$

The homogeneous part $U\mathbf{G}_\mu U^\dagger$ is the adjoint-representation action—left multiplication by U and right multiplication by $U^\dagger$, exactly the two-sided structure of diagram 3 with $M_L = U$ and $M_R = U^\dagger$. The inhomogeneous term (proportional to $\partial_\mu U$) is the non-Abelian generalisation of the photon gauge shift $A_\mu \to A_\mu + \partial_\mu\chi$.

In ML terms: the gluon field is an embedding table of 3×3 matrices, and gauge transformations conjugate every matrix simultaneously while adding a derivative correction.

20.2 The QCD Lagrangian

The covariant derivative for a quark in the fundamental representation is:

$$D_\mu \psi = \Big(\partial_\mu - i g_s \, G_\mu^a \, T^a\Big)\psi = \Big(\partial_\mu - i g_s \, \mathbf{G}_\mu\Big)\psi, \tag{234}$$

where g_s is the strong coupling constant. The gluon field-strength tensor is the non-Abelian analogue of $F_{\mu\nu}$:

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s \, f^{abc} \, G^b_\mu \, G^c_\nu, \tag{235}$$

where f^{abc} are the $SU(3)$ structure constants ($[T^a, T^b] = i f^{abc} T^c$). The extra term $g_s f^{abc} G^b G^c$ encodes the fact that gluons carry colour and interact with each other. The QCD Lagrangian is:

$$\mathcal{L}_{\text{QCD}} = \sum_{f=u,d,s,c,b,t} \bar{\psi}^f \Big(i\gamma^\mu D_\mu - m_f\Big)\psi^f \;-\; \frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu}. \tag{236}$$

Expanded, the gluon kinetic term generates three- and four-gluon vertices:

$$-\tfrac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} = -\tfrac{1}{4}(\partial_\mu G^a_\nu - \partial_\nu G^a_\mu)^2 \underbrace{- g_s f^{abc} (\partial_\mu G^a_\nu) G^{b\mu} G^{c\nu}}_{\text{3-gluon vertex}} \underbrace{- \tfrac{g_s^2}{4} f^{abc} f^{ade} G^b_\mu G^c_\nu G^{d\mu} G^{e\nu}}_{\text{4-gluon vertex}}. \tag{237}$$

These self-interactions have no analogue in QED (photons do not carry electric charge and do not interact with each other at tree level).

Lorentz and gauge invariance. The Lorentz invariance of (236) follows by the same argument as for the Dirac and Maxwell Lagrangians: all Lorentz indices are contracted, and the Dirac representation $S(\Lambda)$ satisfies $S^{-1}\gamma^\mu S = \Lambda^\mu{}_\nu \gamma^\nu$. The gauge invariance follows by construction: $D_\mu \psi \to U D_\mu \psi$ under $U \in SU(3)_c$, so $\bar{\psi} i\gamma^\mu D_\mu \psi$ is invariant; the mass term $m_f \bar{\psi}\psi$ is invariant because $U^\dagger U = I$; and $G^a_{\mu\nu} G^{a\mu\nu} = 2\,\mathrm{Tr}[\mathbf{G}^2_{\mu\nu}]$ is invariant under conjugation $\mathbf{G}_{\mu\nu} \to U \mathbf{G}_{\mu\nu} U^\dagger$.

20.3 Creation operators and the QCD Fock space

Quantising (236), the quark field has mode expansion:

$$\psi^{f\alpha}(x) = \sum_s \int \frac{d^3p}{(2\pi)^3 \sqrt{2\omega_{\mathbf{p}}}} \Big[b^{f\alpha}(\mathbf{p}, s)\, u(\mathbf{p}, s)\, e^{-ip\cdot x} + d^{f\alpha\dagger}(\mathbf{p}, s)\, v(\mathbf{p}, s)\, e^{ip\cdot x} \Big], \tag{238}$$

with anticommutation relations:

$$\Big\{ b^{f\alpha}(\mathbf{p}, s),\; b^{g\beta\dagger}(\mathbf{p}', s') \Big\} = (2\pi)^3 \delta^{(3)}(\mathbf{p} - \mathbf{p}')\, \delta^{fg}\, \delta^{\alpha\beta}\, \delta_{ss'}. \tag{239}$$

The creation operator $b^{f\alpha\dagger}(\mathbf{p}, s)$ creates a quark of flavour f, colour α, momentum $\mathbf{p}$, spin s. Similarly $d^{f\alpha\dagger}$ creates an antiquark.

The gluon field has mode expansion:

$$G^a_\mu(x) = \sum_{\lambda=\pm1} \int \frac{d^3k}{(2\pi)^3\sqrt{2\omega_{\mathbf{k}}}} \Big[a^a(\mathbf{k},\lambda)\, \epsilon_\mu(\mathbf{k},\lambda)\, e^{-ik\cdot x} + a^{a\dagger}(\mathbf{k},\lambda)\, \epsilon^*_\mu(\mathbf{k},\lambda)\, e^{ik\cdot x} \Big], \tag{240}$$

with $\omega_{\mathbf{k}} = |\mathbf{k}|$ (massless gluons) and bosonic commutation relations:

$$\Big[a^a(\mathbf{k},\lambda),\ a^{b\dagger}(\mathbf{k}',\lambda') \Big] = (2\pi)^3 \delta^{(3)}(\mathbf{k}-\mathbf{k}')\, \delta^{ab}\, \delta_{\lambda\lambda'}. \tag{241}$$

The creation operator $a^{a\dagger}(\mathbf{k},\lambda)$ creates a gluon with colour-adjoint index $a \in \{1,\dots,8\}$, momentum $\mathbf{k}$, helicity λ.

The QCD Fock space is:

$$\mathcal{F}_{\text{QCD}} = \mathcal{F}_{\text{quark}} \otimes \mathcal{F}_{\text{gluon}}, \tag{242}$$

where $\mathcal{F}_{\text{quark}}$ has $6 \times 3 \times 2$ modes per momentum (6 flavours, 3 colours, 2 spins for quarks and antiquarks) and $\mathcal{F}_{\text{gluon}}$ has 8×2 modes per momentum (8 colour indices, 2 helicities). The Fock hidden state is a concatenation over all particle-number sectors of this combined space.

20.4 The QCD Hamiltonian

The Legendre transform of (236) gives:

$$H_{\text{QCD}} = H_{\text{quark}} + H_{\text{gluon}} + H_{\text{int}}, \tag{243}$$

where the free parts are:

$$H_{\text{quark}} = \sum_{f,\alpha,s} \int \frac{d^3p}{(2\pi)^3}\, \omega_{\mathbf{p}} \Big[b^{f\alpha\dagger} b^{f\alpha} + d^{f\alpha\dagger} d^{f\alpha} \Big], \tag{244}$$

$$H_{\text{gluon}} = \sum_{a,\lambda} \int \frac{d^3k}{(2\pi)^3}\, |\mathbf{k}|\, a^{a\dagger} a^a, \tag{245}$$

and the interaction Hamiltonian is:

$$H_{\text{int}} = -g_s \int d^3x\, \bar{\psi}^f\, \gamma^\mu\, T^a\, \psi^f\, G^a_\mu \ + \ H_{3\text{g}} + H_{4\text{g}}, \tag{246}$$

where $H_{3\text{g}}$ and $H_{4\text{g}}$ are the three- and four-gluon self-interaction terms from the non-Abelian field-strength tensor. The quark-gluon vertex $g_s \bar{\psi} T^a \psi G^a_\mu$ is the QCD analogue of the QED vertex $e\bar{\Psi}\gamma^\mu \Psi A_\mu$, with T^a rotating the colour index. The gluon self-interactions are new: they couple different gluon-colour sectors with no counterpart in electromagnetism.

20.5 What the Schrödinger equation tells us: asymptotic freedom and confinement

The running coupling. The strength of the quark-gluon interaction is not a fixed number; it *runs* with the energy scale μ at which the system is probed:

$$\alpha_s(\mu) = \frac{g_s(\mu)^2}{4\pi} = \frac{12\pi}{(33 - 2N_f)\ln\!\left(\mu^2/\Lambda_{\rm QCD}^2\right)}, \tag{247}$$

where N_f is the number of active quark flavours and $\Lambda_{\rm QCD} \approx 200\,\text{MeV}$ is the intrinsic QCD scale. As $\mu \to \infty$, $\alpha_s \to 0$: the coupling vanishes at high energies. This is *asymptotic freedom*, discovered by Gross, Politzer, and Wilczek (Nobel Prize 2004).

The running has a clean Fock-space interpretation. The weight matrix $H_{\rm QCD}$ can be thought of as defined at a reference scale; the effective weight matrix at a different scale μ is obtained by integrating out the Fock sectors with energies between μ and the reference scale. At high μ, the off-diagonal blocks (interactions) are small and the RNN update is nearly a phase rotation. At low μ, the blocks are large and the dynamics are non-perturbative.

Why gluons cause anti-screening. In QED, virtual electron-positron pairs screen the electric charge: a test charge appears smaller at long distances. In QCD, virtual quark loops also screen the colour charge. But gluons carry colour themselves, and virtual gluon loops *anti-screen*: they spread the colour field more efficiently than quarks concentrate it. For $N_f \leq 16$ quark flavours ($N_f = 6$ in the Standard Model), the anti-screening wins: $33 - 2N_f = 21 > 0$ in (247), and the coupling falls at high energies. Photons carry no electric charge, which is why QED does not have this effect.

Confinement. As $\mu \to \Lambda_{\rm QCD}$, $\alpha_s \to \infty$: perturbation theory breaks down entirely. The non-perturbative dynamics of QCD lead to *colour confinement*—no free coloured particle has ever been observed. The mechanism is heuristically clear: the colour field between two quarks does not spread out (as in QED) but forms a narrow *flux tube*, with energy proportional to its length:

$$E(r) \approx \sigma r + \cdots, \qquad \sigma \approx (440\,\text{MeV})^2, \tag{248}$$

where σ is the *string tension*. Pulling two quarks apart costs energy linearly in their separation. When $E(r)$ exceeds $2m_q$, it is energetically favourable for the vacuum to create a new quark-antiquark pair—the string breaks and two hadrons appear. A free quark can never be isolated.

String breaking in Fock space.
Start with a meson: one quark and one antiquark, $|q^\alpha, \bar{q}_\alpha\rangle$ (summed over colour to form a singlet). The interaction Hamiltonian $H_{\rm int}$ can transfer gluons

between them. At large separation, the Fock hidden state evolves into a superposition:

$$h_t \;\to\; c_1(t)\,|q\bar{q}\rangle + c_2(t)\,|q\bar{q}\,;\,\text{gluon string}\rangle + c_3(t)\,|q\bar{q}\,;\,q'\bar{q}'\rangle + \cdots$$

The third term describes string breaking: the vacuum has produced a new $q'\bar{q}'$ pair from the gluon string sector. The Born rule gives the probability of observing two mesons rather than one. This is the Fock hidden state "deciding" whether to sustain the string or break it.

20.6 Colour confinement and the physical Hilbert space

Gauge invariance in QCD has a profound consequence for the Fock space: *only colour-singlet states are physical*. A colour-singlet state is one invariant under all $SU(3)_c$ gauge transformations—it transforms in the trivial (singlet, $(0,0)$) representation of $SU(3)_c$.

The two simplest colour-singlet combinations of quarks are:

Mesons ($q\bar{q}$):

$$\left|\pi^+\right\rangle = \frac{1}{\sqrt{3}} \sum_{\alpha} \int \frac{d^3p\, d^3q}{(2\pi)^6}\, f(\mathbf{p},\mathbf{q})\; b^{u\alpha\dagger}(\mathbf{p},s)\; d^{d\alpha\dagger}(\mathbf{q},s')\,|0\rangle\,. \tag{249}$$

The sum $\sum_\alpha$ over colour produces the singlet: $|r\bar{r}\rangle + |g\bar{g}\rangle + \left|b\bar{b}\right\rangle$ (normalised).

Baryons (qqq):

$$|p\rangle = \frac{1}{\sqrt{6}}\,\epsilon_{\alpha\beta\gamma} \int [\cdots]\; b^{u\alpha\dagger} b^{u\beta\dagger} b^{d\gamma\dagger}\,|0\rangle\,. \tag{250}$$

The totally antisymmetric $\epsilon_{\alpha\beta\gamma}$ tensor contracts three colours into a singlet: $|rgb\rangle - |rbg\rangle + |gbr\rangle - \cdots$ (the colour wave function of a proton).

The physical Hilbert space of QCD is the subspace of $\mathcal{F}_{\text{QCD}}$ consisting of all colour-singlet states. In RNN terms: the Schrödinger equation rotates h_t within this physical subspace. The interaction Hamiltonian (246) preserves colour neutrality: quark-gluon vertices transfer colour from quark to gluon and back, always maintaining the overall singlet structure.

Hadron	**Quark content**	**Colour singlet**	**Mass (MeV)**
Proton p	uud	$\epsilon_{\alpha\beta\gamma} q^\alpha q^\beta q^\gamma$	938
Neutron n	udd	$\epsilon_{\alpha\beta\gamma} q^\alpha q^\beta q^\gamma$	940
Pion π^+	$u\bar{d}$	$\delta^{\alpha\beta} q_\alpha \bar{q}_\beta$	140
Kaon K^+	$u\bar{s}$	$\delta^{\alpha\beta} q_\alpha \bar{q}_\beta$	494
J/ψ	$c\bar{c}$	$\delta^{\alpha\beta} q_\alpha \bar{q}_\beta$	3097

Chiral symmetry breaking. The QCD vacuum also spontaneously breaks an approximate symmetry. For massless quarks, $\mathcal{L}_{\text{QCD}}$ has a chiral symmetry $SU(N_f)_L \times SU(N_f)_R$ (independent flavour rotations of left- and right-handed quarks). The strong coupling at low energies causes the QCD vacuum to develop a *quark condensate*:

$$\langle \bar{\psi}\psi \rangle = \langle \bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L \rangle \neq 0, \tag{251}$$

mixing left and right and breaking $SU(N_f)_L \times SU(N_f)_R \to SU(N_f)_V$. The Goldstone bosons of this breaking are the *pions* (and kaons, eta): they are light because the symmetry is approximate (light quarks have small but non-zero masses). Pions are pseudo-Goldstone bosons—much lighter than the proton because they are protected by approximate symmetry.

20.7 The QCD path integral and lattice QCD

The QCD path integral integrates over quark and gluon field histories:

$$Z_{\text{QCD}} = \int \mathcal{D}\bar{\psi}\, \mathcal{D}\psi\, \mathcal{D}G^a_\mu\, \delta(G_{\text{gauge}})\, \exp\!\Big(i \int d^4x\, \mathcal{L}_{\text{QCD}}\Big). \tag{252}$$

The quark path integral is Gaussian (as in QED) and can be performed exactly, yielding a fermion determinant:

$$\int \mathcal{D}\bar{\psi}\, \mathcal{D}\psi\; e^{i\int \bar{\psi}(i\not{D}-m)\psi} = \det(i\gamma^\mu D_\mu - m)^{N_f}. \tag{253}$$

The gluon path integral is non-Gaussian (because $G_{\mu\nu}$ is quadratic in G^a_μ, making $\mathcal{L}_{\text{QCD}}$ quartic in the gluon field) and cannot be evaluated exactly.

Lattice QCD: path integral as literal matrix multiplication. At low energies where perturbation theory fails, the path integral is evaluated numerically by discretising spacetime onto a four-dimensional lattice with spacing a:

$$Z_{\text{lattice}} = \int \prod_{x,\mu} dU_\mu(x)\; \det(D_{\text{lat}} + m)^{N_f}\; e^{-S_{\text{Wilson}}[U]}, \tag{254}$$

where $U_\mu(x) = e^{ig_s a G^a_\mu(x) T^a} \in SU(3)$ is the parallel transporter along the link from x to $x + a\hat{\mu}$, and S_{Wilson} is the Wilson action built from products of Us around elementary plaquettes.

In RNN terms, lattice QCD is the most literal realisation of matrix multiplication: the Fock space is truncated to a finite-dimensional Hilbert space (one degree of freedom per lattice site), the path integral is a finite product of integrals over $SU(3)$ matrices, and each lattice step is one application of the discrete evolution operator. The continuum limit $a \to 0$ recovers QCD. Monte Carlo sampling over $SU(3)$ link variables draws samples from the measure $e^{-S_{\text{Wilson}}}$, giving non-perturbative estimates of hadron masses, form factors, and decay constants.

Proton mass from lattice QCD.
The proton mass m_p can be extracted from the two-point function:

$$C(t) = \langle 0 |\, \mathcal{O}_p(t)\, \mathcal{O}_p^\dagger(0)\, | 0 \rangle \xrightarrow{t\to\infty} e^{-m_p t}.$$

Here $\mathcal{O}_p = \epsilon_{\alpha\beta\gamma}(u^{\alpha T} C\gamma^5 d^\beta) u^\gamma$ is the proton interpolating operator (a colour-singlet combination of three quarks). On the lattice, this correlator is a matrix element of (U_{lat}^N) —the RNN evolution operator raised to the Nth power— evaluated between initial and final states both created by $\mathcal{O}_p^\dagger$. The exponential decay in t directly reads off the mass from the Schrödinger eigenvalue. Modern lattice calculations reproduce $m_p \approx 938\,\text{MeV}$ from first principles, with the quark masses ($m_u \approx 2.3\,\text{MeV}$, $m_d \approx 4.8\,\text{MeV}$) as inputs. The remaining $\approx 99\%$ of the proton mass comes from the kinetic and binding energy of gluons and virtual quark-antiquark pairs—from the dynamics of the Fock hidden state, not from the input masses.

20.8 The Complete Standard Model: One Fock Hidden State

We now have all three gauge sectors: $SU(3)_c$ (QCD), $SU(2)_L$ (weak isospin), and $U(1)_Y$ (hypercharge). Together they form the gauge group of the Standard Model:

$$G_{\text{SM}} = SU(3)_c \times SU(2)_L \times U(1)_Y. \tag{255}$$

Every field in the Standard Model is a representation of G_{SM} and of the Lorentz group—it carries a collection of indices, one for each factor, acted on independently. From diagram 3:

- **Lorentz indices**: acted on by $SL(2,\mathbb{C})$ via $M_L(L)$ and $M_R(L^\dagger)$.
- **Colour indices**: acted on by $U_c \in SU(3)_c$.
- **Isospin indices**: acted on by $U_w \in SU(2)_L$ (left-handed fields only).
- **Hypercharge**: a $U(1)_Y$ phase.

All four actions commute. The full transformation law is:

$$\Phi \;\longmapsto\; U_c\, U_w\, e^{ig'Y\alpha}\, M_L(L)\, \Phi\, M_R(L^\dagger). \tag{256}$$

The field content of one generation:

Field	**Lorentz**	$SU(3)_c$	$SU(2)_L$	Y	**Statistics**
$Q_L = (u_L, d_L)^T$	$(\frac{1}{2}, 0)$	**3**	**2**	$+\frac{1}{6}$	Fermion
u_R	$(\frac{1}{2}, 0)$	**3**	**1**	$+\frac{2}{3}$	Fermion
d_R	$(\frac{1}{2}, 0)$	**3**	**1**	$-\frac{1}{3}$	Fermion
$L_L = (\nu_L, e_L)^T$	$(\frac{1}{2}, 0)$	**1**	**2**	$-\frac{1}{2}$	Fermion
e_R	$(\frac{1}{2}, 0)$	**1**	**1**	-1	Fermion
G^a_μ (gluons)	$(\frac{1}{2}, \frac{1}{2})$	**8**	**1**	0	Boson
W^a_μ (W, Z, γ before SSB)	$(\frac{1}{2}, \frac{1}{2})$	**1**	**3**	0	Boson
B_μ	$(\frac{1}{2}, \frac{1}{2})$	**1**	**1**	0	Boson
ϕ (Higgs)	$(0, 0)$	**1**	**2**	$+\frac{1}{2}$	Boson

The Standard Model has three identical generations of quarks and leptons (the table above is for one generation); the only difference between generations is the Yukawa coupling constants—the masses and mixing angles.

The complete Lagrangian.

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{Yukawa}}, \tag{257}$$

where $\mathcal{L}_{\text{QCD}}$ (236), $\mathcal{L}_{\text{EW}}$ (210), and $\mathcal{L}_{\text{Yukawa}}$ (217) are as derived above. The complete Lagrangian fits on one line; it is the most compressed description of all known particle physics.

One Fock hidden state. The Hilbert space of the Standard Model is the tensor product of all Fock spaces:

$$\mathcal{F}_{\text{SM}} = \mathcal{F}_{\text{quark}} \otimes \mathcal{F}_{\text{lepton}} \otimes \mathcal{F}_{\text{gluon}} \otimes \mathcal{F}_W \otimes \mathcal{F}_Z \otimes \mathcal{F}_\gamma \otimes \mathcal{F}_h. \tag{258}$$

A state of the universe is a unit vector $h_t \in \mathcal{F}_{\text{SM}}$—a concatenation of blocks for every possible particle content: zero quarks, one quark, one quark and one gluon, a proton and a photon, and so on to arbitrary complexity. The Schrödinger equation:

$$\frac{d\,h_t}{dt} = -i\,H_{\text{SM}}\,h_t, \qquad H_{\text{SM}} = H_{\text{QCD}} + H_{\text{EW}} + H_{\text{Yukawa}}, \tag{259}$$

rotates h_t unitarily. The Born rule reads out probabilities. The path integral $\int \mathcal{D}\Phi\, e^{iS_{\text{SM}}[\Phi]}$ is the time-sliced matrix multiplication of this evolution operator, with $\mathcal{L}_{\text{SM}}$ emerging from the short-time Trotter expansion.

Sector	Gauge group	Key feature
QCD	$SU(3)_c$	Asymptotic freedom; confinement
Electroweak	$SU(2)_L \times U(1)_Y$	SSB via Higgs; W, Z massive
QED (residual)	$U(1)_{em}$	Unbroken; photon massless
Architecture	Fock space + Schrödinger equation + Born rule	
Changes across sectors	Gauge group, field representations, vacuum	
Unchanged	The RNN. Every sector is the same linear unitary update.	

The Standard Model describes every experiment ever performed in particle physics, from the Lamb shift to the Higgs discovery at the LHC, using exactly this framework. It has free parameters—19 numbers: 3 gauge couplings, 6 quark masses, 3 lepton masses, 4 CKM mixing angles, the Higgs mass and VEV, and the QCD vacuum angle—but no free architecture. The architecture is always the same.

Quantum mechanics is an RNN. Quantum field theory extends the hidden state to Fock space. The Standard Model specifies which representations of $SU(3)_c \times SU(2)_L \times U(1)_Y \times SL(2,\mathbb{C})$ *the fields carry, how the Higgs breaks the gauge group, and what the coupling constants are. The Schrödinger equation, the Born rule, and the path integral are not features of any particular theory: they are the architecture shared by all of them. Fock space is the only departure from single-particle quantum mechanics. Everything else is the same RNN.*

21 Gauge Orbits: Faddeev–Popov and BRST

21.1 Recall: The path integral is a matrix product

The single identity behind everything

Every path integral in this book—scalar field, Dirac field, photon, QED, QCD, the full Standard Model—follows from one algebraic identity:

$$\boxed{e^{-iHT} = \lim_{N\to\infty} \underbrace{\left(e^{-iH\epsilon}\right)\cdots\left(e^{-iH\epsilon}\right)}_{N \text{ factors}}, \qquad \epsilon = T/N.} \tag{260}$$

This is a tautology. It requires no new physics and no new assumptions. It holds for any Hamiltonian H acting on any Hilbert space—finite or infinite dimensional, quantum mechanics or quantum field theory, particles or strings.

The path integral is what you get when you evaluate the matrix element $\langle q_f|\, e^{-iHT}\, |q_i\rangle$ by inserting a resolution of identity $I = \int dq\ |q\rangle\langle q|$ between every consecutive pair

of factors in (260):

$$\langle q_f|\, e^{-iHT}\, |q_i\rangle = \int dq_1 \cdots dq_{N-1} \prod_{k=0}^{N-1} \underbrace{\langle q_{k+1}|\, e^{-iH\epsilon}\, |q_k\rangle}_{\text{one matrix entry}}. \tag{261}$$

This is matrix multiplication. Each factor $U_{q_{k+1},q_k} = \langle q_{k+1}|\, e^{-iH\epsilon}\, |q_k\rangle$ is one entry of the short-time evolution matrix U; integrating over $q_1, \ldots, q_{N-1}$ is the (q_f, q_i) entry of the matrix product U^N.

What the "classical field" is

The labels $q_0, q_1, \ldots, q_N$ in (261) are eigenvalues of the observable used to define the resolution of identity. In quantum mechanics one uses position eigenstates: $q_k = x_k \in \mathbb{R}$. In quantum field theory one uses field-configuration eigenstates: $q_k = \phi_k(\cdot)$, a function from $\mathbb{R}^3$ to $\mathbb{R}$.

The "classical field configurations" integrated over in $\int \mathcal{D}\phi$ are *labels for basis states of Fock space.* The state $|\phi(\cdot)\rangle$ is defined by $\hat{\phi}(\mathbf{x})\, |\phi(\cdot)\rangle = \phi(\mathbf{x})\, |\phi(\cdot)\rangle$ for all $\mathbf{x}$; it is a coherent superposition across all particle-number sectors. The integral $\int \mathcal{D}\phi$ is a sum over a complete basis of $\mathcal{F}$—nothing more exotic than a matrix trace.

Path integral language	**Matrix language**	**RNN language**
Field configuration $\phi(\cdot)$	Row/column label of U	Basis state index of h_t
Sum over paths $\int \mathcal{D}\phi$	Matrix multiplication U^N	Unrolled RNN for N steps
Short-time amplitude $e^{i\epsilon\mathcal{L}}$	One matrix entry $U_{q'q}$	One-step transition amplitude
Action $S[\phi]$	Log-amplitude of the path	Cumulative log-weight
Stationary-phase path	Dominant matrix entry	Greedy / Viterbi path
Classical field equation	Saddle-point of U^N	Most probable trajectory
Quantum corrections	Off-diagonal contributions	All non-greedy paths

Why the Lagrangian appears

The short-time matrix entry $U_{q'q} = \langle q'|\, e^{-iH\epsilon}\, |q\rangle$ is evaluated for small ϵ by the Trotter splitting:

$$e^{-iH\epsilon} = e^{-i(T+V)\epsilon} \approx e^{-iT\epsilon}\, e^{-iV\epsilon} + O(\epsilon^2). \tag{262}$$

The potential V is diagonal in the q-basis; the kinetic term T is diagonal in the conjugate momentum basis. Inserting a momentum-space resolution of identity to evaluate the kinetic factor gives a Gaussian integral that produces $e^{i\epsilon(T_{\text{kin}}-V)} = e^{i\epsilon\mathcal{L}}$. The Lagrangian in the exponent of the path integral is a consequence of Trotter

splitting—the short-time approximation to one matrix entry, not an independent postulate.

There is one object, the unitary operator e^{-iHT}, and two ways to compute it: as the exponential of the Hamiltonian (Schrödinger), or as a matrix product with the Lagrangian in the exponent (Feynman). They are the same RNN written in different coordinates.

21.2 The over-counting problem

A gauge theory has a local symmetry: the physical configuration is not a single field $A_\mu(x)$ but an *equivalence class* of fields related by gauge transformations. For $U(1)$:

$$[A_\mu] = \Big\{ A_\mu + \partial_\mu \chi : \chi(x) \in \mathbb{R} \Big\}. \tag{263}$$

For $SU(N)$:

$$[A_\mu] = \Big\{ U A_\mu U^\dagger + \tfrac{i}{g}(\partial_\mu U) U^\dagger : U(x) \in SU(N) \Big\}. \tag{264}$$

Each equivalence class is a *gauge orbit*—the set of all field configurations physically indistinguishable from A_μ.

The path integral $\int \mathcal{D}A_\mu \, e^{iS[A]}$ integrates over all field configurations, including all members of each gauge orbit. Since $S[A]$ is gauge-invariant, every element of an orbit contributes the same factor $e^{iS[A]}$. The integral therefore counts each physical state infinitely many times—once per gauge transformation, of which there are infinitely many.

ML framing. The model has a symmetry (gauge invariance), and naively summing over all configurations counts each equivalence class many times. The solution is to integrate over *orbits* rather than individual configurations—to choose one canonical representative per equivalence class with the correct Jacobian.

21.3 Gauge fixing: choosing a section

A *gauge condition* $G(A) = 0$ selects exactly one representative from each orbit—a *section* of orbit space. Common choices:

$$\text{Lorenz gauge:} \quad \partial^\mu A_\mu = 0, \tag{265}$$

$$\text{Coulomb gauge:} \quad \nabla \cdot \mathbf{A} = 0, \tag{266}$$

$$\text{Axial gauge:} \quad A_3 = 0. \tag{267}$$

The Faddeev–Popov procedure converts the orbit-space integral into an ordinary field path integral. The key identity is:

$$1 = \int \mathcal{D}\chi \; \delta\Big(G(A^\chi)\Big) \; \left| \det \frac{\delta G(A^\chi)}{\delta \chi} \right|, \tag{268}$$

where A^χ is the gauge transform of A by χ. Inserting (268) into $\int \mathcal{D}A\, e^{iS[A]}$ and using gauge invariance:

$$\int \mathcal{D}A\, e^{iS[A]} = \underbrace{\int \mathcal{D}\chi}_{\text{orbit volume (cancels)}} \times \int \mathcal{D}A\; \delta\Big(G(A)\Big)\; \Delta_{\text{FP}}[A]\; e^{iS[A]}. \tag{269}$$

The orbit volume factors out and cancels in all physical observables. The remaining integral is over one representative per orbit, weighted by the *Faddeev–Popov determinant*:

$$\Delta_{\text{FP}}[A] = \left|\det M_{\text{FP}}[A]\right|, \qquad M_{\text{FP}}[A] = \left.\frac{\delta G(A^\chi)}{\delta \chi}\right|_{G(A)=0}. \tag{270}$$

For Lorenz gauge in $U(1)$: $M_{\text{FP}} = \Box$ (the d'Alembertian, independent of A—no ghost coupling in abelian theories). For non-Abelian gauge theories: $M_{\text{FP}} = \partial^\mu D_\mu$ depends on A, generating ghost-gluon vertices.

21.4 Ghost fields: the determinant as a path integral

The Faddeev–Popov determinant $\det M_{\text{FP}}[A]$ depends on the gauge field and must be kept inside the path integral. It is written as a path integral over Grassmann *ghost fields* $c(x)$ and $\bar{c}(x)$:

$$\det M_{\text{FP}}[A] = \int \mathcal{D}\bar{c}\,\mathcal{D}c\; \exp\!\Big(i \int d^4x\; \bar{c}(x)\, M_{\text{FP}}[A]\, c(x)\Big). \tag{271}$$

This uses the standard Grassmann identity $\det M = \int \mathcal{D}\bar{\theta}\mathcal{D}\theta\, e^{\int \bar{\theta} M \theta}$, which is the fermionic analogue of the bosonic Gaussian $(\det M)^{-1/2} = \int \mathcal{D}\phi\, e^{-\frac{1}{2}\phi^T M \phi}$. The full gauge-fixed path integral is:

$$Z = \int \mathcal{D}A\,\mathcal{D}\bar{c}\,\mathcal{D}c\; \exp\!\Big(i \int d^4x\; \Big[\mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{gf}} + \bar{c}\, M_{\text{FP}}[A]\, c\Big]\Big), \tag{272}$$

where $\mathcal{L}_{\text{gf}} = -\frac{1}{2\xi}(\partial^\mu A_\mu)^2$ is the gauge-fixing term (a soft delta function parametrised by ξ).

Ghost fields are Grassmann scalars—they violate the spin-statistics theorem and are not physical particles. They appear in internal Feynman diagram lines (loops) but never as external states. They are a bookkeeping device: their path integral is the Jacobian of the change of variables from "all gauge fields" to "one representative per orbit $\times$ orbit direction."

21.5 BRST symmetry: the residual invariance after gauge fixing

After Faddeev–Popov gauge fixing, the action in (272) is no longer invariant under the original gauge symmetry (the gauge condition breaks it). However, it is

invariant under a new global fermionic symmetry called *BRST symmetry* (Becchi, Rouet, Stora, Tyutin), parametrised by a constant Grassmann number ε:

$$\delta_{\mathrm{BRST}}\, A^a_\mu = \varepsilon\, (D_\mu c)^a, \tag{273}$$

$$\delta_{\mathrm{BRST}}\, c^a = -\tfrac{\varepsilon}{2}\, g\, f^{abc}\, c^b c^c, \tag{274}$$

$$\delta_{\mathrm{BRST}}\, \bar{c}^a = \varepsilon\, B^a, \tag{275}$$

$$\delta_{\mathrm{BRST}}\, B^a = 0. \tag{276}$$

The gauge field transformation (273) is a gauge transformation with the ghost c as the parameter—it is the residue of the original gauge symmetry, now controlled by an anticommuting field.

The crucial property is *nilpotency*: applying δ_{BRST} twice gives zero,

$$\delta^2_{\mathrm{BRST}} = 0, \tag{277}$$

because $c^a c^a = 0$ for Grassmann c^a. This makes BRST a differential on the space of fields.

21.6 Physical states as BRST cohomology

The conserved BRST charge Q satisfies $Q^2 = 0$. The physical Hilbert space is the cohomology of Q:

$$\mathcal{H}_{\mathrm{phys}} = \frac{\ker Q}{\operatorname{im} Q} = \frac{\{|\psi\rangle : Q\,|\psi\rangle = 0\}}{\{Q\,|\chi\rangle : |\chi\rangle \in \mathcal{F}\}}. \tag{278}$$

States annihilated by Q are BRST-closed; states of the form $Q\,|\chi\rangle$ are BRST-exact. Physical states are closed but not exact.

The content of BRST cohomology is: a state is physical if and only if it is gauge-invariant and contains no ghost excitations. The longitudinal polarisation of the photon (unphysical) is exact; the two transverse polarisations are closed and not exact. The cohomology (278) precisely selects the two physical helicities and discards the gauge and ghost directions.

ML framing, summarised.

Gauge theory concept	ML / linear algebra concept
Gauge orbit $[A_\mu]$	Equivalence class under a symmetry
Gauge fixing $G(A) = 0$	Choosing a canonical representative (normal form)
FP determinant $\det M_{\text{FP}}$	Jacobian of the change of variables
Ghost fields c, $\bar{c}$	Auxiliary variables implementing the Jacobian as a Gaussian
BRST charge Q, $Q^2 = 0$	Nilpotent differential (exterior derivative)
Physical states $= \ker Q / \operatorname{im} Q$	Cohomology: gauge-invariant quotient space
BRST Ward identity	Gauge invariance of physical observables despite gauge fixing

22 Renormalisation as Parameter Learning

22.1 The problem: sensitivity to unknown high-energy physics

When we compute loop diagrams (quantum corrections), we encounter integrals over all momenta:

$$I = \int_0^\infty \frac{d^4k}{(2\pi)^4}\, f(k, p, m). \tag{279}$$

For renormalisable theories these diverge as $k \to \infty$—ultraviolet (UV) divergences. The physical interpretation: the theory is being asked to sum over all energy scales, including scales far above where it has been tested. It is sensitive to *unknown high-energy physics*.

A QFT is an effective description valid below some scale Λ. Above Λ there is new physics we do not know. Asking loop integrals to run to ∞ is asking the model to extrapolate arbitrarily far beyond its training domain.

22.2 Regularisation

Make the integral finite by introducing a regulator.

Hard cutoff. Replace the upper limit: $\int_0^\infty \to \int_0^\Lambda$. The integral $I(\Lambda)$ is finite and carries explicit Λ-dependence.

Dimensional regularisation. Evaluate in $d = 4 - 2\varepsilon$ dimensions:

$$I(\varepsilon) = \mu^{2\varepsilon} \int \frac{d^dk}{(2\pi)^d}\, f(k, p, m), \tag{280}$$

where μ is an arbitrary mass scale. UV divergences appear as poles $1/\varepsilon$ as $\varepsilon \to 0$. This preserves gauge and Lorentz symmetry and is the standard choice for the Standard Model.

22.3 Renormalisation: absorbing divergences into parameters

Write each bare parameter as a renormalised parameter plus a counterterm:

$$m_0^2 = m_R^2 + \delta m^2, \qquad \lambda_0 = \lambda_R + \delta\lambda, \qquad \phi_0 = Z_\phi^{1/2}\,\phi_R. \tag{281}$$

The counterterms δm^2, $\delta\lambda$, $(Z_\phi - 1)$ are chosen to cancel the $1/\varepsilon$ poles in loop integrals. The renormalised parameters m_R, λ_R are finite and fixed by matching to experiment (renormalisation conditions). The physical prediction is the loop result minus its divergent part—finite, regulator-independent, and calculable.

ML framing.

Renormalisation concept	**ML concept**
Bare parameters m_0, λ_0	Initial parameters before training
Loop corrections	Gradient signal from quantum fluctuations
Counterterms δm, $\delta\lambda$	Fine-tuning updates to match data
Renormalised parameters m_R, λ_R	Trained parameters matching experiment
Renormalisation condition	Loss function and training objective
UV divergence	Overfitting to out-of-domain (UV) physics
Regularisation cutoff Λ	Maximum training resolution

Renormalisation is parameter fitting: bare parameters are initialised, quantum corrections shift them, counterterms adjust them back so that the renormalised parameters reproduce experimental measurements. This is fine-tuning.

22.4 Renormalisability: finite-dimensional parameter space

A theory is *renormalisable* if all UV divergences at any loop order can be absorbed into a *finite* set of counterterms—precisely the parameters already present in $\mathcal{L}$.

The criterion is power counting in $d = 4$ spacetime dimensions. An operator $\mathcal{O}$ has mass dimension $[\mathcal{O}]$; since $[\mathcal{L}] = 4$, its coupling has dimension $[\lambda] = 4 - [\mathcal{O}]$:

$[\mathcal{O}]$	$[\lambda]$	Type	Divergences	Predictivity
< 4	> 0	Super-renormalisable	Finitely many diagrams	Maximally predictive
$= 4$	$= 0$	Renormalisable	Logarithmic at all loops	Predictive with finite parameters
> 4	< 0	Non-renormalisable	New type at every loop	Requires infinitely many parameters

A renormalisable theory contains only operators of dimension ≤ 4. Every UV divergence has the form of one of these operators; the corresponding counterterms

absorb all divergences; and the theory makes predictions using finitely many parameters.

ML framing. Renormalisability = the model is *finite-dimensional*: parametrised by a fixed, finite number of weights. A non-renormalisable theory requires new counterterms at each loop order—infinitely many parameters—and loses predictive power beyond a fixed approximation order. It is the QFT analogue of a model that must grow unboundedly in size as it is asked for higher precision.

The Standard Model has exactly 19 free parameters. Given these 19 numbers measured from experiment, the theory predicts every other observable in particle physics. It is a 19-parameter model of all known non-gravitational physics.

22.5 Wilsonian renormalisation: coarse-graining across scales

The modern (Wilsonian) view reveals why renormalisation works. Define the QFT with a hard UV cutoff Λ—field modes with $|\mathbf{k}| > \Lambda$ are excluded. Now *integrate out* the high-momentum modes $\Lambda/b < |\mathbf{k}| < \Lambda$:

$$e^{-S_{\text{eff}}[\phi_<]} = \int \mathcal{D}\phi_> \; e^{-S[\phi_< + \phi_>]}, \tag{282}$$

producing an *effective action* S_{eff} at scale Λ/b:

$$S_{\text{eff}}[\phi_<] = \int d^4x \left[\tfrac{Z}{2}(\partial\phi_<)^2 + \tfrac{m_{\text{eff}}^2}{2}\phi_<^2 + \lambda_{\text{eff}}\phi_<^4 + \frac{c_6}{\Lambda^2}\phi_<^6 + \cdots \right]. \tag{283}$$

Higher-dimensional operators (ϕ^6, ϕ^8, ...) are generated by integrating out $\phi_>$, but suppressed by powers of $1/\Lambda$. At energies $E \ll \Lambda$ they are negligible. The effective theory at scale E needs only the renormalisable operators of dimension ≤ 4—a finite-dimensional model.

Repeating (282) iteratively defines the *renormalisation group (RG) flow*: how effective parameters change as the coarse-graining scale μ is lowered. The flow is governed by the *beta function*:

$$\mu \frac{d\lambda}{d\mu} = \beta(\lambda). \tag{284}$$

Wilsonian concept	ML concept	Example
UV cutoff Λ	Training data resolution	Image pixels at Λ dpi
Integrating out $\phi_>$	Marginalising over fine features	Average-pooling
Effective action S_{eff}	Compressed model at scale μ	Distilled / pruned network
Running coupling $\lambda(\mu)$	Scale-dependent parameter	Hyperparameter vs. layer depth
RG fixed point ($\beta = 0$)	Scale-invariant representation	Self-similar / fractal feature
Relevant operator ($[\mathcal{O}] < 4$)	Feature that grows with scale	High-level semantic concept
Irrelevant operator ($[\mathcal{O}] > 4$)	Feature suppressed at low energy	Fine-grained texture detail
Marginal operator ($[\mathcal{O}] = 4$)	Logarithmically drifting feature	Running coupling constant

Three qualitatively different RG behaviours:

Asymptotic freedom ($\beta < 0$): $\lambda \to 0$ as $\mu \to \infty$. The theory becomes free at high energies. QCD: $\alpha_s \to 0$ at short distances. Quarks barely interact inside a proton when probed hard enough.

Infrared freedom ($\beta > 0$): $\lambda \to 0$ as $\mu \to 0$. The theory becomes free at low energies. QED: $\alpha_{\text{em}} \to 0$ at long distances. This is why perturbative QED is so accurate for atomic physics.

Fixed point ($\beta = 0$): $\lambda = \lambda^*$ is scale-independent. The theory is conformal—invariant under rescaling. Fixed points are the attractors of RG flow; they describe second-order phase transitions and play a central role in string theory and the AdS/CFT correspondence.

22.6 Effective field theory: organised non-renormalisability

A non-renormalisable theory is not useless. It is an *effective field theory* (EFT): a description valid below some scale Λ, organised as a series in E/Λ:

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\leq 4} + \frac{1}{\Lambda}\sum_i c_i^{(5)}\mathcal{O}_i^{(5)} + \frac{1}{\Lambda^2}\sum_j c_j^{(6)}\mathcal{O}_j^{(6)} + \cdots \tag{285}$$

At energies $E \ll \Lambda$, terms of order $(E/\Lambda)^n$ are suppressed. The EFT is predictive to any desired precision in E/Λ using a finite number of operators at each order. Every renormalisable theory is an EFT valid to all orders in E/Λ; a non-renormalisable EFT is valid up to corrections of order $(E/\Lambda)^n$ for some finite n.

The Standard Model is itself almost certainly an EFT with a UV completion at some scale Λ_{new}. The leading correction is the dimension-5 Weinberg operator:

$$\mathcal{O}^{(5)} = \frac{c}{\Lambda_{\text{new}}}(\bar{L}\tilde{\phi})(\tilde{\phi}^T L^c), \tag{286}$$

which generates neutrino Majorana masses $m_\nu \sim v^2/\Lambda_{\text{new}}$ after electroweak SSB. The observed neutrino mass scale $m_\nu \sim 0.1\,\text{eV}$ implies $\Lambda_{\text{new}} \sim 10^{14}$–$10^{16}\,\text{GeV}$. The Standard Model Lagrangian is precise and complete at currently accessible energies; the EFT expansion (285) tells us what new physics looks like at the next order.

Final summary. The three technical tools of this chapter are answers to three questions.

1. *How do you compute e^{-iHT} for a field theory?* Time-slice it into N short steps, evaluate each step's matrix entry via Trotter splitting, and multiply. The result is the Feynman path integral with $e^{iS[\phi]}$ in the integrand.
2. *How do you avoid over-counting gauge-equivalent configurations?* Gauge-fix to a section of orbit space using the Faddeev–Popov identity. Represent the resulting Jacobian as a ghost path integral. The residual BRST symmetry ($Q^2 = 0$) identifies the physical states as its cohomology.
3. *How do you fit the parameters of a QFT to experimental data despite UV divergences?* Regularise to make divergences explicit; absorb them into bare parameters via counterterms; fix the renormalised parameters from experiment. A renormalisable theory needs only finitely many counterterms—it is a finite-parameter model, predictive at all energies below its domain of validity.

None of these is a modification of the fundamental framework. The Fock hidden state, the Schrödinger equation, the Born rule, and the path integral remain unchanged. These tools are computational and organisational procedures for extracting predictions from a theory whose underlying architecture—a linear unitary RNN on a Fock space—was fixed from the beginning.

23 Feynman Diagrams as Cumulants: The ϕ^4 Model

Feynman diagrams are often introduced as a collection of pictorial rules handed down from on high. Their true origin is simpler: they are the terms in a Taylor expansion of a log-partition function, exactly as cumulants are in statistics and energy-based models in machine learning. We develop this for the scalar ϕ^4 theory, the simplest interacting QFT, using the path integral with a source.

23.1 The generating functional: QFT meets the energy-based model

Add a *source* $J(x)$ to the scalar field Lagrangian—a position-dependent external field that couples linearly to ϕ:

$$\mathcal{L}_J = \mathcal{L}(\phi) + J(x)\,\phi(x). \tag{287}$$

The *generating functional* is the path integral with this source:

$$Z[J] = \int \mathcal{D}\phi\ \exp\!\Big(-S_E[\phi] + \int d^4x\ J(x)\,\phi(x)\Big), \tag{288}$$

where we work in *Euclidean* signature ($t \to -i\tau$, $p^0 \to ip^0_E$) so that e^{-S_E} is a genuine probability weight. The Euclidean action for ϕ^4 theory is:

$$S_E[\phi] = \int d^4x \left[\tfrac{1}{2}(\partial_\mu \phi)^2 + \tfrac{1}{2}m^2\phi^2 + \frac{\lambda}{4!}\phi^4\right]. \tag{289}$$

This is an energy-based model. An energy-based model (EBM) in ML defines a probability distribution over configurations ϕ via:

$$p[\phi;\, J] = \frac{1}{Z[J]}\, \exp\Big(-S_E[\phi] + \int J\phi\Big). \tag{290}$$

The energy function is $E[\phi] = S_E[\phi]$—the Euclidean action. The source $J(x)$ is an external field (a position-dependent bias). The generating functional $Z[J]$ is exactly the partition function of this EBM. The *free energy* is:

$$W[J] = \log Z[J]. \tag{291}$$

Everything that follows from $W[J]$ in QFT has a direct statistical counterpart.

QFT object	**EBM / statistics object**	**Formula**
Euclidean action $S_E[\phi]$	Energy function $E(\phi)$	$p \propto e^{-E}$
Generating functional $Z[J]$	Partition function with field J	$\int e^{-E+J\cdot\phi}\mathcal{D}\phi$
Free energy $W[J] = \log Z[J]$	Cumulant generating functional	$\log Z$
$\delta W/\delta J(x)$	Expected value	$\langle\phi(x)\rangle_J$
$\delta^2 W/\delta J(x)\delta J(y)$	Covariance / variance	$\langle\phi(x)\phi(y)\rangle_c$
$\delta^n W/\delta J^n$	nth cumulant	Connected n-point function
Feynman diagrams	Cumulant expansion terms	Diagrammatic cumulants

23.2 First and second derivatives: expectation and propagator

Taking the first functional derivative of $W[J]$ with respect to $J(x)$:

$$\frac{\delta W[J]}{\delta J(x)} = \frac{1}{Z[J]}\frac{\delta Z[J]}{\delta J(x)} = \frac{\int \mathcal{D}\phi\ \phi(x)\, e^{-S_E + \int J\phi}}{Z[J]} = \langle\phi(x)\rangle_J. \tag{292}$$

The first derivative of the free energy is the expected field value—the one-point function. This is the exact analogue of $\partial \log Z/\partial J = \langle x\rangle_J$ in a discrete EBM.

The second derivative gives the connected two-point function:

$$\frac{\delta^2 W[J]}{\delta J(x)\, \delta J(y)} = \langle\phi(x)\phi(y)\rangle_J - \langle\phi(x)\rangle_J\langle\phi(y)\rangle_J = \langle\phi(x)\phi(y)\rangle_{J,c}. \tag{293}$$

This is the *covariance* of the field at x and y—exactly the variance formula $\partial^2 \log Z/\partial J^2 = \langle\phi^2\rangle - \langle\phi\rangle^2$ from basic statistics. Setting $J = 0$:

$$\Delta(x,y) \;\equiv\; \frac{\delta^2 W}{\delta J(x)\,\delta J(y)}\bigg|_{J=0} = \langle\phi(x)\phi(y)\rangle_c. \tag{294}$$

This is the *propagator*: the covariance of the quantum field at two spacetime points. In QFT it is drawn as a line connecting x to y. In ML it is the Gram matrix of the field distribution.

The nth derivative of $W[J]$ at $J = 0$ is the nth *connected* correlation function—the nth cumulant of the field distribution. Cumulants, not moments, are the natural objects: they are additive over independent subsystems and vanish for uncorrelated points. Feynman diagrams are the diagrammatic representation of these cumulants.

23.3 The free theory: exact Gaussian integral

For $\lambda = 0$, S_E is quadratic and the path integral is a Gaussian. Writing the action in momentum space ($\phi(x) = \int \frac{d^4p}{(2\pi)^4}\tilde{\phi}(p)\,e^{ip\cdot x}$):

$$S_E^{(0)}[\phi] = \frac{1}{2}\int \frac{d^4p}{(2\pi)^4}\,(p^2+m^2)\,|\tilde{\phi}(p)|^2, \tag{295}$$

the Gaussian integral over each mode $\tilde{\phi}(p)$ gives:

$$W_0[J] = \log Z_0[J] = \frac{1}{2}\int \frac{d^4p}{(2\pi)^4}\,\frac{|\tilde{J}(p)|^2}{p^2+m^2} = \frac{1}{2}\int d^4x\,d^4y\;J(x)\,\Delta_0(x-y)\,J(y), \tag{296}$$

where the *free propagator* is:

$$\Delta_0(x-y) = \int \frac{d^4p}{(2\pi)^4}\,\frac{e^{ip\cdot(x-y)}}{p^2+m^2}, \qquad \tilde{\Delta}_0(p) = \frac{1}{p^2+m^2}. \tag{297}$$

Confirming (293):

$$\frac{\delta^2 W_0}{\delta J(x)\,\delta J(y)}\bigg|_{J=0} = \Delta_0(x-y). \quad \checkmark \tag{298}$$

The free propagator is the Green's function of the Klein–Gordon operator: $(-\partial^2 + m^2)\Delta_0(x-y) = \delta^4(x-y)$. In Feynman diagram language it is a single line from x to y. In ML language it is the kernel of the Gaussian process defined by $S_E^{(0)}$.

The propagator as a Gaussian process kernel.
The free scalar field with action $S_E^{(0)}$ defines a Gaussian process over $\mathbb{R}^4$: the field $\phi(x)$ is a random function with mean zero and covariance $\langle\phi(x)\phi(y)\rangle = \Delta_0(x-y)$. The kernel $\Delta_0(x-y)$ depends only on $|x-y|$ (translational invariance) and decays as $e^{-m|x-y|}/|x-y|$ at large separation. The mass m is the inverse correlation length: heavy fields are short-range correlated, light fields are

long-range correlated, and massless fields ($m = 0$) have power-law correlations $\sim 1/|x-y|^2$. The Feynman propagator is the kernel function of the quantum field's Gaussian process.

23.4 Wick's theorem: computing free-theory cumulants

Because the free theory is Gaussian, all connected n-point functions with n odd vanish, and for n even:

$$\langle \phi(x_1)\cdots\phi(x_{2k})\rangle_0 = \sum_{\text{pairings}} \prod_{\text{pairs}(i,j)} \Delta_0(x_i - x_j). \tag{299}$$

This is *Wick's theorem*: the nth moment of a Gaussian is the sum over all $(n-1)!!$ perfect pairings of the n points, each pairing contributing one propagator per pair. The number of pairings grows as: $(2k-1)!! = 1\times 3\times 5\times\cdots\times(2k-1)$.

Wick's theorem for four fields.

$$\begin{aligned}\langle \phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\rangle_0 &= \Delta_0(x_1-x_2)\Delta_0(x_3-x_4)\\ &+ \Delta_0(x_1-x_3)\Delta_0(x_2-x_4)\\ &+ \Delta_0(x_1-x_4)\Delta_0(x_2-x_3).\end{aligned}$$

Three pairings of four points; each pairing is one diagram. Since none of these is connected (each is a product of two independent propagators), the connected four-point function $\langle\phi_1\phi_2\phi_3\phi_4\rangle_c$ vanishes in the free theory—as expected, since a Gaussian has no connected cumulants beyond order 2.

23.5 Perturbative expansion: turning on the interaction

For $\lambda > 0$, split $S_E = S_E^{(0)} + S_E^{\text{int}}$ where $S_E^{\text{int}} = \frac{\lambda}{4!}\int d^4x\,\phi(x)^4$. The key observation is that $\phi(x)$ inserted into the free-theory path integral is equivalent to a functional derivative with respect to $J(x)$:

$$\phi(x)\, e^{-S_E^{(0)} + \int J\phi} \;\equiv\; \frac{\delta}{\delta J(x)}\, e^{-S_E^{(0)} + \int J\phi}. \tag{300}$$

Therefore the full generating functional can be written:

$$Z[J] = \exp\!\left(-\frac{\lambda}{4!}\int d^4x\,\left(\frac{\delta}{\delta J(x)}\right)^4\right) Z_0[J], \tag{301}$$

and expanding the exponential in powers of λ:

$$Z[J] = \left[1 - \frac{\lambda}{4!}\int d^4z\left(\frac{\delta}{\delta J(z)}\right)^4 + \frac{1}{2}\left(\frac{\lambda}{4!}\right)^2\int d^4z_1 d^4z_2\,\cdots + \cdots\right] Z_0[J]. \tag{302}$$

Each term is evaluated by acting with functional derivatives on $Z_0[J] = e^{\frac{1}{2}\int J\Delta_0 J}$, which brings down factors of $\Delta_0(x-y)J(y)$. Setting $J=0$ at the end selects the vacuum correlators.

23.6 Feynman rules for ϕ^4

The expansion (302) generates a systematic set of diagrams. Each term in the expansion corresponds to a diagram built from two elementary pieces:

Propagator (line): one factor of $\Delta_0(x-y)$, drawn as a line connecting point x to point y. In momentum space:

$$\tilde{\Delta}_0(p) = \frac{1}{p^2+m^2}. \tag{303}$$

Vertex (dot with four legs): one factor of $(-\lambda)\int d^4z$, drawn as a point z with four lines meeting. The factor $-\lambda$ comes from differentiating $e^{-\lambda\phi^4/4!}$ once; the $\int d^4z$ integrates over the vertex position.

The *Feynman rules* for ϕ^4 theory in position space:

1. Draw all topologically distinct diagrams with the specified external points and the required number of vertices.
2. Assign a propagator $\Delta_0(x_i - x_j)$ to each line.
3. Assign a factor $(-\lambda)\int d^4z$ to each vertex.
4. Multiply by the symmetry factor $1/S$ (the number of ways to permute internal lines and vertices leaving the diagram unchanged).
5. Sum over all diagrams at each order in λ.

23.7 Explicit diagrams

Order λ^0: free propagator.

The two-point function at zeroth order is simply:

$$\langle\phi(x)\phi(y)\rangle^{(0)} = \Delta_0(x-y). \tag{304}$$

Diagram: a single line from x to y.

Order λ^1: tadpole correction to the propagator.

Acting with $-\frac{\lambda}{4!}(\delta/\delta J)^4$ at a point z on $Z_0[J]$ and extracting the connected two-point function gives:

$$\langle\phi(x)\phi(y)\rangle_c^{(1)} = -\frac{\lambda}{2}\int d^4z\;\Delta_0(x-z)\,\Delta_0(z-z)\,\Delta_0(z-y). \tag{305}$$

The combinatorial factor $1/2$ is the symmetry factor (swapping the two ends of the loop leaves the diagram unchanged). The factor $\Delta_0(z-z) = \int \frac{d^4k}{(2\pi)^4} \frac{1}{k^2+m^2}$ is the loop integral—a single propagator with both ends at the same point, integrated over all loop momenta k. This integral diverges at large k: it is the first UV divergence we encounter, absorbed by the mass counterterm δm^2 in renormalisation.

Diagram: a line from x to a vertex z, a closed loop at z, and a line from z to y.

Order λ^1: tree-level four-point vertex.

The connected four-point function at first order in λ has no loops:

$$\langle\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\rangle_c^{(1)} = (-\lambda)\int d^4z\; \Delta_0(x_1-z)\,\Delta_0(x_2-z)\,\Delta_0(x_3-z)\,\Delta_0(x_4-z). \tag{306}$$

Four external points each connected by a propagator to a single vertex z. In momentum space (Fourier transforming $x_i \to p_i$, with momentum conservation $p_1+p_2+p_3+p_4=0$ at the vertex):

$$\tilde{G}_c^{(4)}(p_1,p_2,p_3,p_4) = (-\lambda)\,\frac{1}{p_1^2+m^2}\cdot\frac{1}{p_2^2+m^2}\cdot\frac{1}{p_3^2+m^2}\cdot\frac{1}{p_4^2+m^2}. \tag{307}$$

This is the *tree-level scattering amplitude*: four particles meeting at a point and interacting with strength λ.

Diagram: an $\times$ shape—four lines meeting at a central vertex.

Order λ^2: one-loop correction to the four-point function.

At order λ^2, two vertices z_1 and z_2 are connected by two internal propagators (a loop), while the four external lines attach two to z_1 and two to z_2:

$$\begin{aligned}\langle\phi_1\phi_2\phi_3\phi_4\rangle_c^{(2)} \supset \frac{(-\lambda)^2}{2}\int d^4z_1\,d^4z_2\;&\Delta_0(x_1-z_1)\Delta_0(x_2-z_1)\\ &\times[\Delta_0(z_1-z_2)]^2\,\Delta_0(z_2-x_3)\Delta_0(z_2-x_4).\end{aligned} \tag{308}$$

The factor $[\Delta_0(z_1-z_2)]^2$ is a double propagator between the two vertices—a loop. In momentum space, the loop is an integral $\int \frac{d^4k}{(2\pi)^4}\tilde{\Delta}_0(k)\tilde{\Delta}_0(q-k)$ where $q=p_1+p_2$ is the external momentum flowing through the loop. This integral is logarithmically divergent—the first UV divergence in the four-point function, absorbed by the coupling counterterm $\delta\lambda$.

Diagram: two vertices connected by a double line (a bubble), with two external legs on each vertex.

23.8 Connected diagrams from $W[J]$: why $\log Z$?

The full generating functional $Z[J]$ includes both connected and disconnected diagrams. For example, at order λ^0 the four-point function includes the three

disconnected pairings from Wick's theorem (299) in addition to the connected tree diagram.

The free energy $W[J] = \log Z[J]$ *selects only connected diagrams.* This is the field-theory statement of the cumulant theorem: log of the moment generating function gives the cumulant generating function, which encodes only irreducible (connected) correlations. Explicitly:

$$W[J] = \sum_{\text{connected diagrams}} \frac{1}{S} \times (\text{diagram value}), \tag{309}$$

where the sum is over all connected vacuum diagrams with J insertions, and S is the symmetry factor of each diagram.

Physical scattering amplitudes depend only on connected diagrams: disconnected contributions factor into independent sub-processes with no mutual interaction. $W[J]$ is therefore the physically relevant object, and its derivatives—the connected correlation functions—are the quantities computed in experiments.

23.9 The effective action: 1PI diagrams and the Legendre transform

One further step gives the most compact representation. Define the *classical field* $\phi_c(x) = \langle \phi(x) \rangle_J = \delta W / \delta J(x)$ and perform the Legendre transform:

$$\Gamma[\phi_c] = W[J] - \int d^4x \; J(x)\, \phi_c(x). \tag{310}$$

The functional $\Gamma[\phi_c]$ is the *effective action* (or quantum effective action). It generates only *one-particle irreducible* (1PI) diagrams—diagrams that remain connected after cutting any single internal line. Its equation of motion $\delta\Gamma/\delta\phi_c = -J$ is the full quantum equation of motion (including all loop corrections), and its second derivative is the inverse propagator:

$$\frac{\delta^2 \Gamma}{\delta\phi_c(x)\, \delta\phi_c(y)} = \Delta^{-1}(x, y), \tag{311}$$

the full (dressed) propagator Δ being the inverse of this.

ML framing. The effective action (310) is the *Legendre transform of the log-partition function*—a central object in statistical mechanics and variational inference. In variational Bayes, $\Gamma[\phi_c]$ is the negative ELBO (evidence lower bound) with the mean field ϕ_c as the variational parameter. The stationarity condition $\delta\Gamma/\delta\phi_c = 0$ (for $J = 0$) is the self-consistency equation of the mean field theory. Computing Γ perturbatively in λ is equivalent to improving the mean field approximation by adding loop corrections—exactly as variational inference is improved by incorporating higher-order correlations.

Generating functional	Generates	ML analogue
$Z[J]$	All diagrams (connected + disconnected)	Moment generating function
$W[J] = \log Z[J]$	Connected diagrams only	Cumulant generating function
$\Gamma[\phi_c]$	1PI diagrams only	Negative ELBO / free energy

Summary. The entire Feynman diagram machinery for ϕ^4 theory follows from one object—the generating functional $Z[J]$—and three operations:

1. **Take** log: $W[J] = \log Z[J]$ selects connected diagrams.
2. **Differentiate with respect to** J: each derivative inserts a field, producing the n-point cumulants. The propagator is the second cumulant; the ϕ^4 vertex is the fourth.
3. **Legendre transform**: $\Gamma[\phi_c] = W[J] - \int J\phi_c$ selects 1PI diagrams and gives the quantum equation of motion.

Feynman diagrams are not a new framework. They are the terms in the Taylor expansion of a log-partition function, drawn pictorially. The lines are propagators (covariances); the vertices are interaction terms (higher cumulants of the non-Gaussian theory); the loops are the integrals that generate UV divergences and require renormalisation. An ML practitioner who has worked with EBMs, Gaussian processes, or variational inference has already handled every mathematical object that appears here.

24 Recap: The Universe Runs a Schrödinger RNN

A reader who knows GPT and Mamba has, without realising it, already understood the mathematical skeleton of all of fundamental physics. This section makes that correspondence explicit and closes the loop.

24.1 The one equation

Everything in this book follows from one equation:

$$\boxed{\frac{d\,h_t}{dt} = -i\,H\,h_t.} \tag{312}$$

This is the Schrödinger equation. In ML terms it is a continuous-time linear RNN with weight matrix $W = -iH$. The constraint that H is Hermitian ($H = H^\dagger$) makes W skew-Hermitian ($W^\dagger = -W$), which is exactly the condition that the hidden state norm is preserved:

$$\frac{d}{dt}\|h_t\|^2 = \langle \dot{h}_t, h_t\rangle + \langle h_t, \dot{h}_t\rangle = \langle -iHh_t, h_t\rangle + \langle h_t, -iHh_t\rangle = -i\langle H\rangle + i\langle H\rangle = 0. \tag{313}$$

A Hermitian weight matrix is the unique choice that makes a linear RNN norm-preserving. Norm preservation is the requirement that probabilities sum to one. The Schrödinger equation is the unique linear, norm-preserving, continuous-time RNN.

The connection to Mamba is direct. Mamba's core is a linear state-space model:

$$\dot{h}(t) = A\,h(t) + B\,x(t), \qquad A \in \mathbb{C}^{n\times n}. \tag{314}$$

Quantum mechanics is this model with $A = -iH$ (skew-Hermitian), $B = 0$ (no external input for an isolated system), and $h \in \mathbb{C}^n$ a unit vector. Mamba's structured state-space matrices are, in the quantum case, constrained to be minus-i times a Hermitian operator. The physics constraint (unitarity) is the ML constraint (norm preservation), stated in two languages.

24.2 The hidden state: from token embedding to Fock space

In GPT, the hidden state $h \in \mathbb{R}^d$ is a dense vector that encodes rich semantic content: syntactic role, factual associations, discourse context. We do not find it surprising that a single high-dimensional vector can carry so much structure.

The quantum hidden state $h_t \in \mathbb{C}^n$ is the same idea. For a single qubit, $n = 2$: the hidden state is a unit vector in $\mathbb{C}^2$, encoding the probability amplitude for each of two outcomes. For a harmonic oscillator, $n = \infty$: the hidden state encodes the amplitude for each energy level.

Quantum field theory makes one change: the hidden state lives in *Fock space*

$$\mathcal{F} = \mathcal{H}^{(0)} \oplus \mathcal{H}^{(1)} \oplus \mathcal{H}^{(2)} \oplus \cdots, \tag{315}$$

a direct sum (concatenation) of spaces for zero particles, one particle, two particles, and so on. This is a richer embedding: not only quantum amplitudes for each physical configuration, but quantum amplitudes over *all possible particle contents* simultaneously. A state with uncertain particle number is a superposition across blocks of the concatenated vector.

When the weight matrix H has off-diagonal blocks connecting different particle-number sectors, the hidden state rotates across them: particles are created and annihilated. The interaction terms in the Standard Model Hamiltonian are precisely these off-diagonal blocks. Particle physics is what happens when a linear RNN has a weight matrix that mixes particle-number blocks.

24.3 The weight matrix: diagram 3 and the Standard Model

The weight matrix H must respect the symmetries of the theory. In relativistic quantum mechanics, the relevant symmetries are:

- The Lorentz group $SL(2,\mathbb{C})$: spacetime symmetry.

- The gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$: internal symmetry.

Diagram 3 makes the constraint explicit. A field $\Phi(X)$ is an embedding table—a map from spacetime points to vectors—and every Lorentz transformation Λ induces a two-sided action on the field value:

$$\Phi(X) \;\longmapsto\; M_L(L)\,\Phi(X)\,M_R(L^\dagger), \tag{316}$$

where $X = x^\mu \sigma_\mu$ is the spacetime point written as a 2×2 Hermitian matrix and $L \in SL(2,\mathbb{C})$ represents Λ. Different choices of (M_L, M_R) give different particle species: $(1,1)$ is a scalar, $(L,1)$ is a left Weyl spinor, $(1,L^\dagger)$ is a right Weyl spinor, $(L,L^\dagger)$ is a 4-vector. The gauge group adds a third action on an internal index, orthogonal to the Lorentz indices.

The Standard Model is the unique renormalisable theory built from these fields, invariant under $SU(3)_c \times SU(2)_L \times U(1)_Y \times SL(2,\mathbb{C})$, with the gauge symmetry broken by a Higgs doublet. Its weight matrix H_{SM} is completely determined by 19 dimensionless numbers: 3 gauge couplings, 9 fermion masses, 4 mixing angles, the Higgs mass and vacuum expectation value, and the QCD vacuum angle.

24.4 The path integral: computing with matrix products

The Schrödinger equation tells us what the universe computes. The path integral tells us how we compute with it.

We cannot run (312) forward directly: $\mathcal{F}_{\text{SM}}$ is too large and H_{SM} too complex. Instead, we use the identity:

$$e^{-iHT} = \lim_{N\to\infty}\left(e^{-iH\epsilon}\right)^N, \qquad \epsilon = T/N. \tag{317}$$

Time-slicing the evolution operator and inserting a complete basis between every step yields:

$$\langle q_f | e^{-iHT} | q_i \rangle = \int \mathcal{D}q\; e^{iS[q]}, \tag{318}$$

where $S[q] = \int_0^T \mathcal{L}(q,\dot{q})\,dt$ is the classical action and the integral sums over all paths $q(t)$ connecting q_i to q_f. The Lagrangian $\mathcal{L}$ emerges from the short-time Trotter approximation to one matrix entry—it is not a postulate, it is a consequence of matrix multiplication.

The saddle point of $e^{iS[q]}$ is the classical trajectory (greedy decoding). Quantum corrections are all other paths (the full sum). Feynman diagrams are the perturbative expansion of the sum around the saddle, organised by powers of the coupling constants.

24.5 The complete correspondence

ML / Mamba / GPT	**Quantum mechanics / QFT**
Linear RNN $\dot{h} = Wh$	Schrödinger equation
Skew-Hermitian weight $W = -iH$	Hermitian Hamiltonian
Norm preservation $\|h_t\| = 1$	Unitarity / probability sum to 1
Token embedding $h \in \mathbb{R}^d$	Quantum state $h \in \mathbb{C}^n$
Softmax readout	Born rule $\|\langle q_o, h\rangle\|^2$
Hard embedding lookup	Wavefunction collapse
Evolve activations / evolve weights	Schrödinger / Heisenberg picture
Rich semantic embedding	Fock hidden state (all particle contents)
Concatenation of vector blocks	Direct sum $\mathcal{H}^{(0)} \oplus \mathcal{H}^{(1)} \oplus \cdots$
Off-diagonal weight blocks	Creation / annihilation operators
Equivariant embedding table	Quantum field $\Phi(x)$
Two-sided matrix action $M_L \Phi M_R$	Lorentz representation (diagram 3)
Internal index structure	Gauge group representation
Iterated matrix multiplication	Feynman path integral
Most probable trajectory	Classical field equation
Perturbation around saddle	Feynman diagram expansion
Symmetry-equivariant model	Gauge-invariant Lagrangian
Equivalence class of configs	Gauge orbit
Canonical representative per class	Gauge fixing (Faddeev–Popov)
Finite-parameter model	Renormalisable theory
Fine-tuning / parameter fitting	Renormalisation
Coarse-graining / distillation	Wilsonian RG / effective field theory
19 hyperparameters	Standard Model free parameters
One weight matrix	H_{SM}
One hidden state	$h_t \in \mathcal{F}_{\mathrm{SM}}$

24.6 What changes and what does not

Across the entire journey from a single qubit to the full Standard Model, exactly one thing changes and everything else stays fixed.

What changes: the Hilbert space and the weight matrix.

- Qubit: $h_t \in \mathbb{C}^2$, $H = \frac{\omega}{2}\sigma_z$.
- Harmonic oscillator: $h_t \in \ell^2(\mathbb{N})$, $H = \omega(a^\dagger a + \frac{1}{2})$.
- Scalar field: $h_t \in \mathcal{F}_{\text{boson}}$, $H = \int \omega_k a_k^\dagger a_k \, dk$.
- QED: $h_t \in \mathcal{F}_{e^-} \otimes \mathcal{F}_{e^+} \otimes \mathcal{F}_\gamma$, $H = H_{\text{Dirac}} + H_{\text{Maxwell}} + eJ^\mu A_\mu$.
- Standard Model: $h_t \in \mathcal{F}_{\text{SM}}$, $H = H_{\text{SM}}$.

What does not change:

- The equation: $\dot{h}_t = -iHh_t$.
- The readout: $p(o) = |\langle q_o, h_t\rangle|^2$.
- The reset: $h_t \leftarrow q_o$ upon measurement.
- The path integral: $\langle q_f|e^{-iHT}|q_i\rangle = \int \mathcal{D}q \, e^{iS[q]}$.

The architecture is fixed at the start and never revised. The Standard Model is not a more complicated theory than the qubit; it is the same theory with a richer hidden state and a more structured weight matrix.

24.7 The universe runs on physical variables, not gauge potentials

There is a conceptual distinction that the path integral formalism can obscure: the universe does not run in gauge variables. It runs in *physical* variables.

The photon field A_μ is not physical: two field configurations related by $A_\mu \to A_\mu + \partial_\mu \chi$ are identical states. The physical content resides in the gauge-invariant quantities—in electromagnetism, the electric and magnetic fields:

$$\mathbf{E} = -\nabla\phi - \partial_t \mathbf{A}, \qquad \mathbf{B} = \nabla \times \mathbf{A}. \tag{319}$$

$\mathbf{E}$ and $\mathbf{B}$ are invariant under $A_\mu \to A_\mu + \partial_\mu \chi$; they are what a physical detector measures. The Hamiltonian written in terms of $\mathbf{E}$ and $\mathbf{B}$:

$$H_{\text{Maxwell}} = \frac{1}{2}\int d^3x \, \Big(|\mathbf{E}|^2 + |\mathbf{B}|^2\Big), \tag{320}$$

is a perfectly well-defined operator on the physical (gauge-invariant) Fock space. The Schrödinger equation

$$\frac{d\, h_t}{dt} = -i\, H_{\text{Maxwell}}\, h_t \tag{321}$$

runs on this physical Hilbert space directly. No gauge fixing is needed. No Faddeev–Popov determinant appears. No ghost fields. The universe evolves $\mathbf{E}$ and $\mathbf{B}$ — not A_μ.

The same principle holds in QCD. The physical variables are the colour electric and magnetic fields $\mathbf{E}^a$ and $\mathbf{B}^a$ (and colour-singlet combinations of quark fields). The Hamiltonian written in these variables acts on the physical colour-singlet Hilbert space without any gauge redundancy.

Then why do we use A_μ at all?

Because A_μ makes Lorentz invariance and the gauge symmetry *manifest* in the Lagrangian. Writing the Maxwell Lagrangian as $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ in terms of A_μ packages the dynamics in a covariant, compact form. But covariance of notation and covariance of physics are different things. The universe's dynamics are Lorentz-invariant and gauge-invariant whether or not we choose to describe them using a gauge potential.

The path integral is an analytical tool, not the operating system.

The Feynman path integral integrates over field histories $\phi(\mathbf{x}, t)$ or $A_\mu(\mathbf{x}, t)$—in the case of gauge theories, over gauge-equivalent configurations as well, corrected by Faddeev–Popov. This is not how the universe runs. It is how *we compute*: the path integral is a mathematical identity that reorganises the matrix product e^{-iHT} into a sum over trajectories in field configuration space, evaluated perturbatively around a saddle point. It is an analytical technique for extracting predictions from the Schrödinger RNN.

The distinction matters conceptually:

	Schrödinger RNN	**Feynman path integral**
Role	How the universe runs	How we compute
Variables	Physical: $\mathbf{E}$, $\mathbf{B}$, ψ	Gauge potentials A_μ, ϕ
Gauge fixing	Not needed	Required (Faddeev–Popov)
Ghosts	Absent	Present (bookkeeping)
Output	h_t at time t	Correlation functions, amplitudes
Domain	Exact, nonperturbative	Usually perturbative

The Schrödinger RNN is the ontology. The path integral is the epistemology—our best tool for interrogating the RNN's outputs without being able to run it directly. Gauge fixing, ghosts, and BRST are features of the path integral formalism, not features of the universe.

24.8 Source code of the universe

Suppose the universe's operating system can handle infinite-dimensional Fock vectors natively—as tensors indexed by particle-content tuples, stored in an infinite-dimensional complex Hilbert space—and can apply infinite-dimensional Hermitian operators to them exactly. The universe's source code is then remarkably short.

universe.py

```
import fock                             # infinite-dim Fock space over C
import operators as op                  # infinite-dim Hermitian operators
from scipy.linalg import expm           # matrix exponential (exact)
from constants import SM_PARAMS         # 19 Standard Model parameters

# ---------------------------------------------------------------------
# Build the weight matrix
# ---------------------------------------------------------------------

def build_hamiltonian(params):
    """
    Construct H_SM from the 19 Standard Model parameters.
    H acts on F_SM = F_quark x F_lepton x F_gluon x F_W x F_Z x F_gamma x F_h.
    """
    g3, g2, g1          = params.gauge_couplings          # SU(3), SU(2), U(1)
    m_u, m_d, m_s, \
    m_c, m_b, m_t       = params.quark_masses
    m_e, m_mu, m_tau    = params.lepton_masses
    theta12, theta13, \
    theta23, delta_CP   = params.CKM_angles
    m_h, v              = params.higgs_sector
    theta_QCD           = params.qcd_vacuum_angle

    H_QCD = op.build_QCD(g3, m_u, m_d, m_s, m_c, m_b, m_t, theta_QCD)
    H_EW  = op.build_EW(g2, g1, m_e, m_mu, m_tau, m_h, v,
                        theta12, theta13, theta23, delta_CP)

    return H_QCD + H_EW                 # acts on F_SM; Hermitian by construction


# ---------------------------------------------------------------------
# Initialise the hidden state
# ---------------------------------------------------------------------

H = build_hamiltonian(SM_PARAMS)
h = fock.vacuum()                       # |0>: unit vector, all sectors empty
                                        # ||h||^2 == 1.0  (guaranteed)
```

```
# ------------------------------------------------------------------
# The universe loop
# ------------------------------------------------------------------

dt = 5.391e-44                          # Planck time [seconds]

while True:

    # --- Schrodinger update (exact unitary rotation) ---
    h = expm(-1j * H * dt) @ h          # h_{t+dt} = e^{-iH dt} h_t
                                        # ||h||^2 preserved to machine precision

    # --- Born-rule measurement (if an observer is present) ---
    if observer.is_present():
        outcome = observer.measure(h)   # sample o ~ |<q_o, h>|^2
        h = fock.embed(outcome)         # Bohr layer: h <- q_o  (hard reset)
        observer.record(outcome)
```

A few observations about this code.

The state is a Fock tensor. `fock.vacuum()` returns a unit vector in $\mathcal{F}_{\text{SM}}$, represented as a tensor whose indices are occupation numbers $(n_{u,r,\uparrow}, n_{u,r,\downarrow}, n_{u,g,\uparrow}, \ldots, n_\gamma^{+1}, n_\gamma^{-1}, n_h)$ for every mode of every field. The tensor is infinite-dimensional but sparse at finite energy: all but finitely many occupation numbers are zero for any physically realisable state. In practice the universe's Fock tensor is an extremely sparse object—most of the Hilbert space is empty.

The weight matrix is a sparse operator. `H_SM` is a Hermitian operator on $\mathcal{F}_{\text{SM}}$. Its free part H_{free} is diagonal in the Fock basis—it phase-rotates each occupation-number sector independently. Its interaction part H_{int} is off-diagonal, coupling sectors that differ by the creation or annihilation of one or a few particles. Each interaction vertex changes at most a handful of occupation numbers at once. The operator is sparse in exactly the sense that neural network weight matrices are sparse after pruning: most entries are zero; the nonzero entries are structured by the gauge symmetry.

The update is exact. `expm(-1j * H * dt)` is the exact matrix exponential—no approximation. The universe does not time-slice itself; it applies $e^{-iH\,dt}$ exactly at each Planck step. We time-slice when we compute (the path integral), because we cannot apply e^{-iHT} exactly. The universe has no such limitation.

Measurement is stochastic. `observer.measure(h)` samples an outcome o from the distribution $p(o) = |\langle q_o, h\rangle|^2$ and collapses $h \leftarrow q_o$. The randomness lives

entirely in this line. The Schrödinger update is deterministic; the Born rule is stochastic. This is the only source of randomness in the universe.

The loop has no convergence criterion. It runs forever. The universe does not stop when it reaches a fixed point. The hidden state h_t rotates continuously; its norm is exactly preserved at every step; the rotation never ends. There is no loss function being minimised. The universe is not training—it is running inference on a fixed weight matrix H_{SM}.

The 19 parameters are compile-time constants. `SM_PARAMS` is read once when the Hamiltonian is built and never updated. The universe does not learn its own parameters online. Renormalisation—from our perspective—is the procedure of inferring those 19 constants from experimental measurements. From the universe's perspective there is nothing to infer: the constants are fixed, the Hamiltonian is fixed, and the loop runs.

Everything we call physics is a readout. A particle detector, a bubble chamber, a LHC calorimeter—all of these are implementations of `observer.measure(h)`. The detector samples from the Born distribution; the outcome is a classical record. The Feynman diagram calculation predicts the distribution from which that sample is drawn. We compute the distribution; the universe draws the sample.

24.9 Why an ML reader can learn this in a day

The standard physics curriculum presents quantum mechanics, quantum field theory, gauge theory, and renormalisation as a multi-year sequence of increasingly difficult subjects, each requiring new conceptual frameworks. What the RNN lens reveals is that the new frameworks were never necessary.

The conceptual content of the entire curriculum is:

1. A linear RNN with a skew-Hermitian weight matrix is called the Schrödinger equation. Its hidden state is called the quantum state. Its norm-squared readout is called the Born rule. Its hard reset is called wavefunction collapse. *This is all of quantum mechanics.*
2. Promoting the hidden state to Fock space—a concatenation of fixed-particle-number subspaces—allows particle number to vary. Off-diagonal blocks in the weight matrix create and annihilate particles. *This is quantum field theory.*
3. The field (the embedding table) must transform equivariantly under the Lorentz group and the gauge group. The two-sided action $M_L \Phi M_R$ classifies all particle species. *This is the representation theory of the Standard Model.*
4. Computing e^{-iHT} by time-sliced matrix multiplication yields the path integral, with the classical Lagrangian in the exponent as a consequence of the Trotter

approximation. *This is the Feynman formalism.*

5. The path integral over a gauge theory over-counts gauge orbits; Faddeev–Popov corrects the Jacobian; BRST identifies the physical subspace. *This is gauge quantisation.*
6. UV divergences reflect sensitivity to unknown high-energy physics; counterterms absorb them into finitely many parameters; the renormalisation group tracks how those parameters flow across scales. *This is renormalisation.*

An ML practitioner who is comfortable with linear RNNs, embedding tables, equivariant networks, and matrix exponentials has the complete mathematical toolkit. The physics terminology is a second language for structures already familiar. The goal of this book has been to provide the dictionary.

The universe solves (312) continuously, for the Hilbert space $\mathcal{F}_{\text{SM}}$ and the weight matrix H_{SM}. We cannot run it forward. We time-slice it, expand around saddle points, and read off predictions. Those predictions, derived from a single linear equation with 19 free parameters, agree with every experiment ever performed in particle physics.

The universe is a linear recurrent network. Its hidden state is a unit vector in Fock space. Its weight matrix is Hermitian. Its outputs are Born-rule probabilities. Everything else is a consequence.

25 Entanglement and the Bell Inequality

When two systems coexist, the joint hidden state lives in the *tensor product* $\mathbb{C}^{n_A} \otimes \mathbb{C}^{n_B}$. For two qubits this is $\mathbb{C}^4$, with basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ where $|ab\rangle \equiv |a\rangle_A \otimes |b\rangle_B$.

25.1 Entangled states cannot be factored

A *product state* is one that factorises: $h = h_A \otimes h_B$. In such a state the two qubits are independent: measuring A provides no information about B beyond what the marginal distribution already encoded. In ML terms, a product state is a fully factored probability model.

The *Bell states* are the four maximally entangled two-qubit states. The simplest is:

$$\left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}}\Big(|00\rangle + |11\rangle\Big) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}. \tag{322}$$

Why $|\Phi^+\rangle$ cannot be factored.
Suppose $|\Phi^+\rangle = (\alpha_0 |0\rangle + \beta_0 |1\rangle) \otimes (\alpha_1 |0\rangle + \beta_1 |1\rangle)$. Expanding:

$$= \alpha_0\alpha_1 |00\rangle + \alpha_0\beta_1 |01\rangle + \beta_0\alpha_1 |10\rangle + \beta_0\beta_1 |11\rangle .$$

Matching coefficients with $\frac{1}{\sqrt{2}}(1,0,0,1)^T$ requires $\alpha_0\beta_1 = 0$ and $\beta_0\alpha_1 = 0$, while $\alpha_0\alpha_1 \neq 0$ and $\beta_0\beta_1 \neq 0$. The first two imply that one factor in each pair is zero, but then the last two cannot be satisfied. Contradiction. ✓

25.2 Perfect correlations from entanglement

Measure both qubits in $|\Phi^+\rangle$ using $\sigma_z \otimes \sigma_z$, whose 4×4 matrix is $\mathrm{diag}(+1,-1,-1,+1)$. The Born rule gives:

$$\langle \sigma_z \otimes \sigma_z \rangle = \left\langle \Phi^+ \right| (\sigma_z \otimes \sigma_z) \left| \Phi^+ \right\rangle = \tfrac{1}{2}(+1+0+0+1) = +1. \tag{323}$$

The outcomes are perfectly correlated: whenever Alice's qubit is $+1$, Bob's is also $+1$, and vice versa. Yet the marginal distribution for each qubit alone is $p(+1) = p(-1) = \frac{1}{2}$—maximally uncertain.

The correlation is not because the qubits carry pre-agreed hidden values. The Bell inequality rules that out.

25.3 The CHSH inequality

Let Alice choose between measurements A (along $\hat{a}$) and A' (along $\hat{a}'$), and Bob between B and B', each returning ±1. Define the *CHSH quantity*:

$$S = E(A,B) + E(A,B') + E(A',B) - E(A',B'), \tag{324}$$

where $E(A,B) = \langle AB \rangle$ is the correlation.

Classical bound. If outcomes are determined by a shared classical hidden variable λ, then for any fixed λ:

$$A(\lambda)\big[B(\lambda) + B'(\lambda)\big] + A'(\lambda)\big[B(\lambda) - B'(\lambda)\big] = \pm 2,$$

because $B(\lambda), B'(\lambda) \in \{+1,-1\}$ forces one bracket to be ±2 and the other to be 0. Averaging over λ gives $|S| \leq 2$.

Quantum value. For the Bell state $|\Phi^+\rangle$, the correlation between measurements along directions making angle θ is

$$E(\hat{a},\hat{b}) = \cos\theta_{ab}. \tag{325}$$

Violating the Bell inequality.
Choose $\hat{a} = 0\check{\text{r}}$, $\hat{a}' = 90\check{\text{r}}$, $\hat{b} = 45\check{\text{r}}$, $\hat{b}' = -45\check{\text{r}}$.

$$\begin{aligned} E(A, B) &= \cos 45\check{\text{r}} = \tfrac{1}{\sqrt{2}}, \\ E(A, B') &= \cos(-45\check{\text{r}}) = \tfrac{1}{\sqrt{2}}, \\ E(A', B) &= \cos 45\check{\text{r}} = \tfrac{1}{\sqrt{2}}, \\ E(A', B') &= \cos 135\check{\text{r}} = -\tfrac{1}{\sqrt{2}}. \end{aligned}$$

$$S = \tfrac{1}{\sqrt{2}} + \tfrac{1}{\sqrt{2}} + \tfrac{1}{\sqrt{2}} - \left(-\tfrac{1}{\sqrt{2}}\right) = \frac{4}{\sqrt{2}} = 2\sqrt{2} \approx 2.83.$$

This exceeds the classical bound of 2. No model with pre-stored, local hidden variables can reproduce this. Quantum mechanics is provably not a factored model with classical hidden state.

The ML framing makes the surprise vivid: entanglement is the difference between a *joint* hidden state (one vector in $\mathbb{C}^4$) and an *independent* hidden state (two vectors in $\mathbb{C}^2$ each). Bell's theorem says these two model classes are empirically distinguishable with any finite number of measurements.

26 The Density Matrix and Mixed States

So far h_t has always been a single definite unit vector—a *pure state*. In practice we may not know the initial state exactly, or we may hold only part of a larger entangled system. The right object to describe such a situation is the *density matrix*.

26.1 From pure states to distributions

If the system is in state $|h_i\rangle$ with probability p_i, the density matrix is:

$$\rho = \sum_i p_i \, |h_i\rangle \, \langle h_i| \, . \tag{326}$$

For a pure state $|h\rangle$ this collapses to $\rho = |h\rangle \, \langle h|$.

Key properties of every density matrix:

- **Hermitian**: $\rho = \rho^\dagger$.
- **Positive semi-definite**: $\langle v| \, \rho \, |v\rangle \geq 0$ for all $|v\rangle$.
- **Unit trace**: $\text{tr}(\rho) = 1$.
- **Purity**: $\text{tr}(\rho^2) \leq 1$, with equality iff ρ is a pure state.

In ML terms, the density matrix is the *second-moment matrix* of the hidden-state distribution. A pure state is a point mass; a mixed state is any other distribution.

26.2 The Born rule for mixed states

For a measurement with eigenvectors $\{q_o\}$, the Born rule becomes a trace:

$$p(o) = \mathrm{tr}\Big(\rho \; |q_o\rangle \, \langle q_o|\Big) = \langle q_o|\rho|q_o\rangle. \tag{327}$$

This recovers the pure-state rule when $\rho = |h\rangle \, \langle h|$, since then $\langle q_o|\rho|q_o\rangle = |\langle q_o|h\rangle|^2$.

26.3 Evolution and measurement update

Schrödinger evolution acts on both sides of ρ:

$$\rho_t = e^{-iHt} \, \rho_0 \, e^{+iHt}. \tag{328}$$

After measuring outcome o, the Bohr layer updates the density matrix:

$$\rho \;\leftarrow\; |q_o\rangle \, \langle q_o| \, . \tag{329}$$

Pure state vs. mixed state on a qubit.
Pure superposition $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$:

$$\rho_{\text{pure}} = |+\rangle \, \langle +| = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad \mathrm{tr}(\rho^2) = 1.$$

Maximally mixed state (equal uncertainty over $|0\rangle$ and $|1\rangle$):

$$\rho_{\text{mixed}} = \frac{1}{2} \, |0\rangle \, \langle 0| + \frac{1}{2} \, |1\rangle \, \langle 1| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathrm{tr}(\rho^2) = \frac{1}{2}.$$

Both give $p(0) = p(1) = \frac{1}{2}$ when measured in the σ_z basis. They differ in the *off-diagonal elements* (the *coherences*): ρ_{pure} has non-zero off-diagonals; ρ_{mixed} does not. Measuring in the σ_x basis distinguishes them: ρ_{pure} gives $p(+1) = 1$, while ρ_{mixed} gives $p(+1) = p(-1) = \frac{1}{2}$.

26.4 Entanglement and partial traces

Density matrices also arise when we hold only one part of a bipartite entangled state. For the Bell state $|\Phi^+\rangle$, the joint density matrix is $\rho_{AB} = |\Phi^+\rangle \, \langle \Phi^+|$. Alice's local state is obtained by *tracing out* Bob's subsystem:

$$\rho_A = \mathrm{tr}_B(\rho_{AB}) = \sum_{b \in \{0,1\}} \langle b|_B \; \rho_{AB} \; |b\rangle_B = \frac{1}{2} I. \tag{330}$$

Alice's reduced state is the maximally mixed state—she sees complete randomness. Yet the joint state ρ_{AB} is pure. All the structure lives in the *correlations* between Alice and Bob, not in either subsystem alone.

27 Decoherence: Why the Classical World Appears

The density matrix framework resolves a long-standing puzzle: if the universe is governed by the unitary Schrödinger equation, why do we never observe macroscopic objects in superposition? The answer is *decoherence*.

27.1 The system-environment model

Suppose a qubit S (the *system*) interacts with a large environment E. The joint initial state is:

$$|\Psi_0\rangle = \Big(\alpha\,|0\rangle_S + \beta\,|1\rangle_S\Big) \otimes |e_0\rangle_E\,. \tag{331}$$

Under a joint unitary evolution that entangles S with E:

$$|\Psi_t\rangle = \alpha\,|0\rangle_S \otimes |e_0(t)\rangle_E + \beta\,|1\rangle_S \otimes |e_1(t)\rangle_E\,, \tag{332}$$

where $|e_0(t)\rangle\,, |e_1(t)\rangle \in \mathbb{C}^{N_E}$ are environment states that branch apart as S and E interact.

27.2 Coherence decays as the environment branches

We do not observe the environment, so we trace it out. The reduced density matrix of S is:

$$\rho_S(t) = \mathrm{tr}_E\Big(|\Psi_t\rangle\,\langle\Psi_t|\Big) = \begin{pmatrix} |\alpha|^2 & \alpha\bar{\beta}\,\gamma(t) \\ \bar{\alpha}\beta\,\overline{\gamma(t)} & |\beta|^2 \end{pmatrix}, \qquad \gamma(t) = \langle e_1(t)|e_0(t)\rangle\,. \tag{333}$$

The *coherence* $\gamma(t) = \langle e_1(t)|e_0(t)\rangle$ is the overlap between the two environment branches. As S and E interact, the environment states become more and more orthogonal—they encode different information about the state of S.

Exponential decay of coherence.
For many physically relevant couplings the overlap decays as:

$$|\gamma(t)| = e^{-\Gamma t/2},$$

where Γ is the *decoherence rate* (proportional to the coupling strength and environment size). Then:

$$\rho_S(t) \xrightarrow{t\to\infty} \begin{pmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{pmatrix}.$$

The off-diagonal coherences vanish. The qubit transitions from the pure superposition ρ_{pure} to the mixed state $|\alpha|^2\,|0\rangle\,\langle 0| + |\beta|^2\,|1\rangle\,\langle 1|$—a classical probability distribution over outcomes.

27.3 The pointer basis

Not all observables decohere at the same rate. The states that are most *robust* to environmental entanglement—the ones for which $\gamma(t)$ decays slowest—are called the *pointer states* or *pointer basis* of the system. For a particle coupled to its spatial environment through position, the pointer basis is approximately the position eigenbasis: position becomes classical fastest, which is why macroscopic objects have definite locations.

27.4 Decoherence in the RNN picture

In RNN terms, decoherence is *environment-induced noise* on the hidden state. The full system+environment state is a pure hidden state in a vast space $\mathbb{C}^{n_S} \otimes \mathbb{C}^{N_E}$; decoherence is what happens when we marginalise over the environment dimensions. The resulting dynamics on ρ_S are no longer unitary—they are a *quantum channel*, the quantum analogue of a Markov kernel. Crucially, decoherence does not require any modification to the Schrödinger equation. It is simply entanglement, viewed from one subsystem's perspective.

28 Quantum Computing

Quantum computing exploits the RNN hidden state directly as a computational resource. A register of n qubits has a hidden state $h \in \mathbb{C}^{2^n}$—exponentially large in the number of physical bits. A quantum algorithm is a designed sequence of unitary gates (recurrent weight matrices) whose interference pattern concentrates the final Born-rule probability on the desired answer.

28.1 The quantum circuit model

A quantum circuit applies a sequence of unitary matrices $U_1, U_2, \ldots, U_d$ to an initial state (typically $|0 \cdots 0\rangle$):

$$h_{\text{out}} = U_d \cdots U_2 \, U_1 \; |0 \cdots 0\rangle \, . \tag{334}$$

Each U_k typically acts on one or two qubits at a time; the rest of the register is left unchanged (tensor-producted with I). The Born-rule measurement at the end reads out the answer. The art of quantum algorithm design is arranging the gates so that wrong answers interfere destructively and the right answer interferes constructively.

28.2 The no-cloning theorem

A fundamental constraint separates quantum computing from classical: *you cannot copy an unknown quantum hidden state.*

Proof (three lines).
Suppose a unitary U copies any state: $U|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle$. Apply U to the two computational basis states:

$$U|0\rangle|0\rangle = |0\rangle|0\rangle,$$
$$U|1\rangle|0\rangle = |1\rangle|1\rangle.$$

By linearity of U, the superposition $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ maps to:

$$U|+\rangle|0\rangle = \tfrac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big) = \big|\Phi^+\big\rangle.$$

But cloning would require $|+\rangle|+\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$. These are different states ($|\Phi^+\rangle$ is entangled; $|+\rangle|+\rangle$ is a product state). Contradiction. U cannot exist. ✓

No-cloning has deep consequences: eavesdropping on a quantum channel is detectable (reading a state disturbs it), and quantum error correction must protect information without copying it. In ML terms, you can save and restore classical model weights, but you cannot "checkpoint" an arbitrary quantum hidden state.

28.3 Grover's algorithm: quadratic search speedup

Grover's algorithm searches an unstructured database of N items in $O(\sqrt{N})$ quantum queries versus $O(N)$ classical queries. We work out the $N = 4$ case, which is fully solvable with explicit 4×4 matrices and requires exactly *one* query to find the target with certainty.

Setup. The four-item database is indexed by $\{00, 01, 10, 11\}$. The target is $|11\rangle$. The algorithm uses two 4×4 unitary matrices:

1. **Oracle** O: marks the target by flipping its amplitude sign.

$$O = I - 2|11\rangle\langle 11| = \mathrm{diag}(1, 1, 1, -1). \tag{335}$$

2. **Diffusion** D: reflects about the uniform superposition $|s\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$.

$$D = 2|s\rangle\langle s| - I = \frac{1}{2}\begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}. \tag{336}$$

The Grover operator is $G = D\,O$.

One-step Grover search on $N = 4$ items.
Step 1. Prepare the uniform superposition with the Hadamard gate:

$$|s\rangle = \tfrac{1}{2}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}.$$

Step 2. Apply the oracle O:

$$O\,|s\rangle = \tfrac{1}{2}\begin{pmatrix}1\\1\\1\\-1\end{pmatrix} \equiv |s'\rangle\,.$$

The target's amplitude is negated; the others are unchanged.
Step 3. Apply the diffusion D. Compute $\langle s|s'\rangle = \frac{1}{4}(1+1+1-1) = \frac{1}{2}$. Then:

$$\begin{aligned} D\,|s'\rangle &= 2\langle s|s'\rangle\,|s\rangle - |s'\rangle \\ &= |s\rangle - |s'\rangle = \tfrac{1}{2}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \tfrac{1}{2}\begin{pmatrix}1\\1\\1\\-1\end{pmatrix} = \begin{pmatrix}0\\0\\0\\1\end{pmatrix} = |11\rangle\,. \end{aligned}$$

Readout. The Born layer gives $p(11) = |\langle 11|\,G\,|s\rangle\rangle|^2 = 1$. The target is found with certainty after a single oracle query.

Why it works. The oracle inverts the target amplitude (negative phase), making it the "odd one out". The diffusion operator is a reflection about the mean: it amplifies whichever state is below the mean and suppresses everything above. Together they act like a pair of mirrors that rotate the hidden state from the uniform superposition $|s\rangle$ toward the target $|11\rangle$, one angular step per Grover iteration.

For general N, the optimal number of iterations is $\lfloor \frac{\pi}{4}\sqrt{N} \rfloor$, giving success probability ≈ 1 and total query complexity $O(\sqrt{N})$. A classical algorithm must query $O(N)$ items on average. The quadratic speedup is not a trick of parallelism—it is constructive interference in the amplitude space, the same mechanism that produces the oscillating Born probabilities we saw in Section 2.

RNN interpretation. The Grover circuit is a short-horizon RNN: the hidden state $h \in \mathbb{C}^N$ is updated by repeating the same unitary $G = DO$ for $O(\sqrt{N})$ steps. The readout is a single Born-layer measurement at the end. The algorithm illustrates that even a linear, weight-tied unitary RNN can solve problems exponentially faster than any classical sequential computation, provided the hidden state is initialised

and measured correctly.

29 What About Gravity?

All of the foregoing—the Schrödinger RNN, Fock space, gauge fields, the Standard Model—lives on a fixed flat Minkowski spacetime. Gravity is conspicuously absent. This is not an accident. Our contention is that gravity is *not a force to be added to the RNN at the same level as electromagnetism or the strong force*. It is something qualitatively different: a collective, emergent, statistical-mechanical phenomenon, analogous to pressure or temperature. The spacetime metric $g_{\mu\nu}$ is a macroscopic average, not a fundamental field.

29.1 Four analogies for emergence

Before any equations, four pictures to build intuition.

(1) Rolling circle from straight-line motions. Place a collection of balls, each constrained to move along its own straight line. Arrange the lines so that the balls' phases are offset by equal increments. The individual motions are all linear, but their collective envelope traces a perfect rolling circle. No ball moves in a circle; the circle is a property of the ensemble, not of any individual.

(2) Wave from local rotations. Place dots on a two-dimensional surface, each rotating in a small local circle. Arrange the phases so that neighbouring dots are slightly offset. The individual motions are purely local and circular, but their collective pattern propagates as a travelling wave across the surface. The wave is not a property of any dot; it is a property of the phase field.

(3) Flocking birds (Gestalt). Each bird responds only to its nearest neighbours: match velocity, avoid collision, maintain cohesion. The individual rules are local and simple. The collective exhibits a global shape, a global direction, and a global velocity—a Gestalt that belongs to the flock, not to any bird. The shape is an emergent macroscopic variable.

(4) Navier–Stokes from water molecules. A glass of water consists of $\sim 10^{25}$ molecules undergoing quantum mechanical collisions. The Navier–Stokes equations describe the collective flow of this enormous ensemble using macroscopic variables: velocity field $\mathbf{v}(\mathbf{x}, t)$, pressure $p(\mathbf{x}, t)$, density $\rho(\mathbf{x}, t)$. These variables do not exist at the molecular level. They are statistical averages over many molecules in a small volume element. Turbulence, vortices, and waves are all collective phenomena.

The claim. The spacetime metric $g_{\mu\nu}(x)$ is to the pre-geometric quantum fields as Navier–Stokes variables are to water molecules. It is a macroscopic statistical average of more fundamental degrees of freedom. The Einstein field equations

are the effective equations of motion for this average, in the same sense that Navier–Stokes is the effective equation of motion for the coarse-grained molecular flow. Geometry is collective behaviour; it cannot exist at the level of individual quantum events.

29.2 Pre-geometric QFT: the Lagrangian before spacetime

We start with a QFT defined on a flat internal Minkowski space, carrying a Lorentz symmetry. The fields are the same matter fields of the Standard Model plus one new ingredient: a *pre-geometric tensor field* $T^{\mu\nu}(x)$, a symmetric rank-2 tensor built from the matter fields themselves. The natural candidate is a *Noether current squared*: the stress-energy tensor of the matter fields.

Recall that for any matter field with Lagrangian $\mathcal{L}_{\text{matter}}$, the Noether current associated with translation symmetry is the stress-energy tensor:

$$T^{\mu\nu} = \frac{\partial \mathcal{L}_{\text{matter}}}{\partial(\partial_\mu \phi_a)} \partial^\nu \phi_a \; - \; \eta^{\mu\nu} \mathcal{L}_{\text{matter}}. \tag{337}$$

Adding a *tensor-square* (current-current) interaction to the Lagrangian:

$$\mathcal{L}_{\text{pre-geo}} = \mathcal{L}_{\text{SM}} \; + \; \frac{\kappa}{2} T^{\mu\nu} T_{\mu\nu}, \tag{338}$$

where κ is a coupling with dimensions of length^2. This term is a *current-current interaction*: it couples the stress-energy of matter at the same spacetime point. It is the pre-geometric seed from which spacetime curvature will emerge.

Lorentz invariance. The term $T^{\mu\nu}T_{\mu\nu}$ contracts all Lorentz indices, making it a scalar. It is invariant under the Lorentz transformations of diagram 3. No curved spacetime is assumed; the pre-geometric Lagrangian is entirely defined on flat Minkowski space.

ML framing. In an energy-based model, the energy function $E[\phi]$ is typically written in terms of local filters or operators applied to the field ϕ—convolutions, derivatives, local response functions. The stress-energy tensor $T^{\mu\nu}$ is exactly such an operator: it is a local function of ϕ and $\partial_\mu \phi$, encoding the local energy and momentum flux. Adding $T^{\mu\nu}T_{\mu\nu}$ to the energy is adding a *current-current interaction*—a product of two local filters at the same point. This is the pre-geometric analogue of a squared-gradient penalty or a structured energy function in an EBM. The term is a *generative seed*: it sources the emergent metric.

29.3 Hubbard–Stratonovich: introducing the metric as a latent variable

The Hubbard–Stratonovich (HS) transformation is the field-theory version of completing the square. It introduces an *auxiliary field* $g_{\mu\nu}(x)$ to linearise the quadratic term $T^{\mu\nu}T_{\mu\nu}$:

$$\exp\Big(-\frac{\kappa}{2}\int d^4x \; T^{\mu\nu}T_{\mu\nu}\Big) = \int \mathcal{D}g_{\mu\nu} \; \exp\Big(-\int d^4x \left[\frac{1}{2\kappa} g_{\mu\nu} g^{\mu\nu} \; - \; g_{\mu\nu} \, T^{\mu\nu}\right]\Big). \tag{339}$$

The auxiliary field $g_{\mu\nu}(x)$ is a symmetric rank-2 tensor—the *metric perturbation*. Its coupling to matter is $g_{\mu\nu}T^{\mu\nu}$, which is exactly the minimal coupling of a metric to the stress-energy tensor in linearised gravity.

After the HS transformation, the partition function is:

$$Z = \int \mathcal{D}\phi_{\rm SM}\, \mathcal{D}g_{\mu\nu}\ \exp\Big(-S_{\rm SM}[\phi; g] - S_{\rm HS}[g]\Big), \tag{340}$$

where $S_{\rm SM}[\phi; g]$ is the Standard Model action with the metric $g_{\mu\nu}$ coupled to matter, and $S_{\rm HS}[g] = \int \frac{1}{2\kappa} g_{\mu\nu} g^{\mu\nu}$. Now integrate out the matter fields $\phi_{\rm SM}$:

$$Z = \int \mathcal{D}g_{\mu\nu}\, e^{-S_{\rm HS}[g]} \underbrace{\int \mathcal{D}\phi_{\rm SM}\, e^{-S_{\rm SM}[\phi;g]}}_{\det(\cdots)^{\pm 1}}. \tag{341}$$

The matter path integral is a functional determinant that depends on $g_{\mu\nu}$. Its logarithm generates an *effective action for the metric*:

$$S_{\rm eff}[g] = S_{\rm HS}[g] - \log \int \mathcal{D}\phi_{\rm SM}\, e^{-S_{\rm SM}[\phi;g]}. \tag{342}$$

ML framing: latent variable model. The HS transformation is precisely the introduction of a latent variable. In a latent variable model:

$$p(x) = \int dz\, p(x \mid z)\, p(z), \tag{343}$$

the data x (here: the matter fields $\phi_{\rm SM}$) and the latent variable z (here: the metric $g_{\mu\nu}$) have a joint distribution. Marginalising over x gives the marginal $p(z)$: the model for the emergent collective pattern. The metric is the latent variable that encodes the collective gravitational behaviour of matter.

ML / statistics	**Emergent gravity**	**Formula**
Data x	Matter fields $\phi_{\rm SM}$	Integrated out
Latent variable z	Metric $g_{\mu\nu}$	Emergent
Energy $E(x, z)$	$S_{\rm SM}[\phi; g] + S_{\rm HS}[g]$	Pre-geo action
Marginal $p(z) = \int p(x, z)dx$	$e^{-S_{\rm eff}[g]}$	Eq. (342)
HS transformation	Restricted Boltzmann machine	Linearise quadratic
Current-current $T^{\mu\nu}T_{\mu\nu}$	Visible-hidden coupling in RBM	Generative seed

The restricted Boltzmann machine (RBM) is a direct ML analogue of the HS transformation. In an RBM, visible units v and hidden units h have an energy $E(v, h) = v^T W h$; marginalising over h gives $p(v) \propto e^{v^T W W^T v/2}$—a Gaussian with covariance WW^T. The HS transformation does the same in reverse: it introduces hidden units (the metric) to linearise a quadratic coupling of the visible units (the stress-energy tensor).

29.4 Heat kernel expansion: the Einstein-Hilbert action emerges

The functional determinant in (341) is evaluated by the *heat kernel expansion*—the QFT analogue of the saddle-point (asymptotic) expansion in statistics. For a field operator $\mathcal{M}[g]$ (the kinetic operator of the matter fields in the background metric $g_{\mu\nu}$):

$$\log\det\mathcal{M}[g] = \operatorname{Tr}\log\mathcal{M}[g] = -\int_0^\infty \frac{ds}{s}\ \operatorname{Tr}\ e^{-s\mathcal{M}[g]}. \tag{344}$$

The heat kernel $K(x,y;s) = \langle x|\, e^{-s\mathcal{M}}\,|y\rangle$ has an asymptotic expansion for small s (short proper time, high momentum):

$$K(x,x;s) = \frac{1}{(4\pi s)^2}\Big[a_0(x) + a_2(x)\,s + a_4(x)\,s^2 + \cdots\Big], \tag{345}$$

where the *Seeley–DeWitt coefficients* $a_{2k}(x)$ are local geometric invariants built from the Riemann curvature:

$$a_0(x) = 1, \tag{346}$$

$$a_2(x) = \tfrac{1}{6}R, \tag{347}$$

$$a_4(x) = \tfrac{1}{180}(R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - R_{\mu\nu}R^{\mu\nu} + \tfrac{5}{2}R^2 + \cdots). \tag{348}$$

Here R is the Ricci scalar, $R_{\mu\nu}$ the Ricci tensor, and $R_{\mu\nu\rho\sigma}$ the Riemann tensor.

Substituting into (342) and integrating over s with a UV cutoff Λ:

$$S_{\text{eff}}[g] = \int d^4x\,\sqrt{g}\,\Big[\underbrace{c_0\Lambda^4}_{\text{cosmological const.}} - \underbrace{\frac{c_2\Lambda^2}{16\pi G}\,R}_{\text{Einstein–Hilbert}} + \underbrace{c_4\left(R_{\mu\nu\rho\sigma}^2 + \cdots\right)}_{\text{higher curvature}} + \cdots\Big], \tag{349}$$

where G is Newton's gravitational constant (related to κ and Λ), and c_0, c_2, c_4 are numerical coefficients. The *Einstein–Hilbert action* $\frac{1}{16\pi G}\int\sqrt{g}\,R$ emerges at second order in the heat kernel expansion. It is the leading curvature term in the effective action for the emergent metric.

The variation of (349) with respect to $g_{\mu\nu}$ gives:

$$G_{\mu\nu} + \Lambda_{\text{cc}}\,g_{\mu\nu} = 8\pi G\,\langle T_{\mu\nu}\rangle, \tag{350}$$

the *Einstein field equations*, where $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ is the Einstein tensor and $\langle T_{\mu\nu}\rangle$ is the expectation value of the matter stress-energy. This equation is not postulated; it is derived as the saddle-point equation of the effective action for the emergent metric.

Heat kernel as asymptotic expansion. The expansion (345) in powers of s is the QFT analogue of a saddle-point or steepest-descent expansion in statistics. The coefficients a_{2k} play the role of cumulants: a_0 is the zeroth-order (volume) term, a_2 is the first curvature correction (analogous to the mean), a_4 is the second curvature correction (analogous to the variance), and so on. The Einstein–Hilbert term is the *first non-trivial cumulant* of the matter fluctuation spectrum in a curved background.

29.5 The tensor-square term from Wilson RG flow

Where does the pre-geometric term $\kappa\, T^{\mu\nu}T_{\mu\nu}$ come from? One natural answer: it is generated by the Wilson renormalisation group flow.

Recall from Section 22 that integrating out high-momentum modes $\phi_>$ with $|\mathbf{k}| > \Lambda/b$ generates an effective action for the low-momentum modes $\phi_<$, with new operators appearing at each step. Among the operators generated by integrating out matter at a UV scale M:

$$S_{\text{eff}}[\phi_<] = S_{\text{SM}}[\phi_<] + \frac{\kappa(M)}{2}\int d^4x\; T^{\mu\nu}[\phi_<]\, T_{\mu\nu}[\phi_<] + O(M^{-4}), \tag{351}$$

with $\kappa(M) \sim 1/M^2$. This is a *dimension-8 operator* (since $T^{\mu\nu}$ has dimension 4 in $d = 4$, and $T^{\mu\nu}T_{\mu\nu}$ has dimension 8). It is irrelevant in the Wilsonian sense—suppressed by $1/M^2$ at energies $E \ll M$. But it is inevitably generated: any interacting QFT with a UV completion at scale M produces this term via loop corrections.

The coupling $\kappa \sim 1/M^2$ sets the scale at which the emergent metric becomes strongly coupled: this is Newton's constant $G \sim \kappa\Lambda^2 \sim \Lambda^2/M^2$. For $M = M_{\text{Planck}} \sim 10^{19}\,\text{GeV}$, gravity is suppressed by $1/M^2_{\text{Planck}}$ at accessible energies—which is why it is so weak compared to the other forces.

The tensor-square term is not added by hand to the Lagrangian; it is the inevitable output of the RG flow, a dimension-8 operator generated when any matter theory is coarse-grained to energies below its UV completion scale. In this sense, the pre-geometric Lagrangian (338) is not an assumption but a consequence of the Wilsonian framework applied to matter in the presence of a UV cutoff.

29.6 Strings as flux tubes

The HS transformation introduces the metric $g_{\mu\nu}$ as a latent variable that captures the collective behaviour of stress-energy correlations. But the current-current interaction $T^{\mu\nu}T_{\mu\nu}$ also sources other collective degrees of freedom.

In QCD, the colour electric field between two quarks forms a *flux tube*: a narrow tube of confined chromoelectric flux, with energy proportional to its length (the string tension σ). An analogous mechanism can operate for the stress-energy current. The stress-energy flux between two massive objects can form a *gravitational flux tube*—a string-like collective excitation of the pre-geometric tensor field.

In the HS-transformed theory, integrating out matter fluctuations generates a kinetic term for $g_{\mu\nu}$ (the Einstein–Hilbert term) but also higher-derivative and non-local terms. These non-local terms, re-summed to all orders, can produce *string-like effective degrees of freedom*: the one-dimensional worldsheet of a string is a gradient flow line of the collective stress-energy field, exactly as a QCD flux tube is a gradient flow line of the colour electric field. Strings, in this picture, are

not fundamental: they are flux tubes of the emergent gravitational field, which is itself an emergent description of collective matter behaviour.

This is consistent with the observation that the string tension is set by $\alpha' \sim 1/M^2_{\text{Planck}} \sim \kappa$—the same coupling that appears in the pre-geometric Lagrangian.

29.7 Why quantising gravity may be misguided

The standard program of quantum gravity attempts to quantise the metric $g_{\mu\nu}$ directly—to treat it as a fundamental field and apply the canonical quantisation or path integral procedure to the Einstein–Hilbert action. This program runs into severe difficulties: the theory is non-renormalisable (the coupling G has dimension length^2, so new counterterms appear at every loop order), and the physical interpretation of quantum fluctuations of spacetime geometry is deeply obscure.

From the emergent-gravity perspective, these difficulties have a clear diagnosis. Consider the Navier–Stokes analogy.

The Navier–Stokes equations describe the collective flow of water molecules. They are derived by coarse-graining the molecular dynamics— integrating out the microscopic degrees of freedom and retaining only the macroscopic variables $\mathbf{v}$, p, ρ. The resulting equations are deterministic and classical at the macroscopic scale, even though the underlying molecular dynamics are quantum mechanical.

Quantising the Navier–Stokes equations—treating $\mathbf{v}(x,t)$ and $p(x,t)$ as quantum operators and writing a path integral over them— would be a category error. The quantum mechanics lives at the molecular level; the Navier–Stokes variables are emergent averages; quantising the averages double-counts the quantum effects already present in the molecules and produces meaningless results.

The Einstein field equations are in exactly the same position. The metric $g_{\mu\nu}$ is an emergent macroscopic average of the pre-geometric stress-energy correlations. The quantum mechanics lives at the level of the matter fields and their Fock-space RNN dynamics. Quantising the metric directly is quantising the average—a category error of the same kind as quantising Navier–Stokes.

Fluid mechanics	**Emergent gravity**	**Status**
Water molecules	SM matter fields ϕ	Fundamental (quantum)
Molecular collisions	Stress-energy correlations	Fundamental (quantum)
Velocity field $\mathbf{v}$	Metric $g_{\mu\nu}$	Emergent (classical avg.)
Pressure p	Ricci scalar R	Emergent (classical avg.)
Navier–Stokes eqn	Einstein field eqn	Emergent (effective)
Quantise N-S?	Quantise gravity?	Category error
Temperature	Cosmological constant?	Statistical in nature

Gravity, on this view, is a statistical mechanical phenomenon. Newton's constant G is not a fundamental coupling but an effective parameter encoding the density of pre-geometric degrees of freedom, exactly as the viscosity η in Navier–Stokes encodes the mean free path of molecules. The cosmological constant is an equilibrium thermodynamic quantity—analogous to temperature or pressure—not a fundamental parameter of the quantum theory. Its smallness is a naturalness problem in the same sense as asking why room temperature is small compared to the Planck temperature: it is a macroscopic collective property, not a fundamental coupling.

What is the right question? The right question is not "how do we quantise $g_{\mu\nu}$?" but "what is the pre-geometric QFT whose stress-energy correlations produce the observed metric as a collective average?" The answer is the pre-geometric Lagrangian (338): matter fields with a current-current interaction, defined on flat internal Minkowski space, with the metric emerging via the Hubbard–Stratonovich transformation and the heat kernel expansion.

In the RNN picture: the universe is a Schrödinger RNN with hidden state $h_t \in \mathcal{F}_{\mathrm{SM}}$. Spacetime geometry is not an input to this RNN but an output—a macroscopic pattern read off from the collective behaviour of the Fock hidden state. The geometry we perceive is the same kind of emergent pattern as the shape of a flock of birds or the pressure of a gas: real, reproducible, and useful as a description, but not fundamental.

The universe does not run on curved spacetime. It runs on Fock space. Curved spacetime is what the Fock hidden state looks like from the outside, averaged over many quantum events.

30 The Observer and the Rendering Layer

We have specified the source code of the universe. But source code alone does not produce output. A game engine without a rendering layer runs silently: hidden states update, matrix products accumulate, Fock vectors rotate—and nothing is ever seen. The measurement problem in quantum mechanics and the observer problem in general relativity are, at their core, the same question: *what is the rendering layer, and what is its relationship to the physics engine?*

30.1 Copenhagen is not an interpretation: it is an axiom system

The Copenhagen interpretation is almost universally described as one "interpretation" of quantum mechanics among others—Many Worlds, Pilot Wave, Relational, QBism, and so on. This framing is misleading. Copenhagen is not an interpretation of quantum mechanics. It is *the axiom system of quantum mechanics*, stated precisely, with the undefined primitives left undefined.

Consider the parallel with Euclidean geometry. Euclid does not define "point" or "line." They are primitive terms, introduced by the axioms themselves—"two points determine a line," "a line can be extended indefinitely." Asking "but what *is* a point, really?" is a category error. A point is whatever satisfies the axioms. The power of the system comes precisely from leaving the primitive undefined: Euclidean geometry applies equally to physical space, to the abstract plane $\mathbb{R}^2$, and to any other structure satisfying the axioms.

The Copenhagen axioms are:

1. A physical system is described by a state h_t in a Hilbert space $\mathcal{H}$, evolving by the Schrödinger equation $\dot{h}_t = -iHh_t$ between measurements.
2. An *observable* is a Hermitian operator O on $\mathcal{H}$, with eigenstates $\{q_o\}$ and eigenvalues $\{o\}$.
3. Upon *measurement* of O, the outcome o occurs with probability $|\langle q_o, h_t\rangle|^2$ (Born rule), and the state resets: $h_t \leftarrow q_o$ (Bohr layer).
4. **An *observer* is whatever performs a measurement. Observer is undefined.**

Axiom 4 is not an oversight. It is a deliberate structural choice, exactly analogous to leaving "point" undefined in Euclid. The observer is a primitive of the axiomatic system. Demanding a definition of "observer" in terms of the other axioms is the same category error as demanding a definition of "point" in terms of lines.

What Copenhagen actually asserts. The substantive claim of Copenhagen is not about observers; it is about reality. Classical reality—a definite value of position, momentum, spin—does *not* exist, even counterfactually, in the absence of a measurement. There is no fact of the matter about where the electron is

before you measure it. There is no hidden variable that the wavefunction conceals. The wavefunction h_t is the complete description of the system, and it is *observer-relative*: different observers with different measurement records assign different wavefunctions to the same system.

This is not a statement of ignorance—it is an ontological claim. The electron does not have a position; the question "where is it?" is malformed until a measurement context is specified. The wavefunction is not a probability distribution over hidden definite values; it is the irreducibly complete description of what *could be* rendered, conditional on an observer choosing to render it.

30.2 The observer as a subroutine of the game engine

We have already written the source code:

```
if observer.is_present():    outcome = observer.measure(h)
```

The observer is not external to the universe. It is a subroutine of the game engine itself—a specialised module with a specific interface. Three properties define this interface.

(1) The observer has a classical memory. The output of `observer.measure(h)` is a classical record—a definite outcome o stored in memory. The observer's memory is not a Fock vector; it is a classical data structure. This is not a limitation of the theory but a specification: the output layer is classical by definition, because classical records are what "observation" means. A detector that is itself quantum-mechanical does not observe; it becomes entangled. An observer, by definition, is a system that writes a classical record.

(2) The Bohr layer is a handshake protocol. The state update $h_t \leftarrow q_o$ is not a physical process happening inside the Schrödinger RNN. It is a *handshake*: the observer sends a query ("measure observable O"), the engine samples an outcome from the Born distribution, returns it to the observer's memory, and resets the hidden state to the corresponding eigenstate. The handshake synchronises the engine's hidden state with the observer's classical record. After the handshake, the engine continues running from q_o; the observer holds outcome o.

(3) The wavefunction is observer-relative. Different observers at different times, with different measurement histories, hold different classical records. Because each measurement resets the hidden state, the engine's state after a measurement is conditioned on the observer's record. Two observers who have made different measurements assign different wavefunctions to the same system —not because they are ignorant of the "true" state, but because the hidden state *is* observer-relative. This is not relativism; it is the consistent consequence of the handshake protocol applied to multiple output interfaces simultaneously.

observer.py

```
class Observer:
    """
    A subroutine of the game engine with a classical memory.
    The observer does not live outside the universe;
    it is a specialised output interface.
    """

    def __init__(self, observable: HermitianOperator):
        self.O       = observable          # what this observer measures
        self.record  = []                  # classical memory (list of outcomes)

    def is_present(self) -> bool:
        """
        True when this observer is coupled to the system.
        Whether an observer 'is present' is itself a physical question
        about the state of the environment -- undefined at the axiom level.
        """
        return self._coupled

    def measure(self, h: FockVector) -> Eigenvalue:
        """
        Handshake protocol between observer and engine.

        1. Compute Born probabilities from current hidden state.
        2. Sample one outcome (irreducibly stochastic).
        3. Write outcome to classical memory.
        4. Return eigenstate q_o so engine can reset h <- q_o.
        """
        probs   = {o: abs(inner(q_o, h))**2
                   for o, q_o in self.O.eigen_decomposition()}
        outcome = sample(probs)            # Born rule: only randomness in universe
        self.record.append(outcome)        # classical write -- irreversible
        return self.O.eigenstate(outcome)  # engine resets h <- q_o
```

30.3 Entanglement: two addresses to the same engine

Consider two electrons created in the singlet state and separated by a billion light-years:

$$h = \frac{1}{\sqrt{2}}\Big(|\uparrow\rangle_A\,|\downarrow\rangle_B - |\downarrow\rangle_A\,|\uparrow\rangle_B\Big). \tag{352}$$

This is a single vector in the tensor-product Fock space $\mathcal{F}_A \otimes \mathcal{F}_B$. There is one hidden state, one engine, one Fock vector.

When observer Alice measures electron A and obtains spin-up, the engine executes:

$$h \leftarrow |\uparrow\rangle_A |\downarrow\rangle_B . \tag{353}$$

The full hidden state is reset. Both addresses—A and B—are updated simultaneously, because they are not two separate systems. They are two subsystem labels within a single Fock vector.

When Bob subsequently measures electron B, the Born rule gives $p(\downarrow_B) = 1$. This is not because a signal travelled from Alice to Bob. No signal travelled. The correlation was already encoded in the entangled Fock vector (352). Alice's measurement updated the single shared hidden state; Bob is reading from the same shared memory.

Spooky action at a distance is a misdiagnosis. Einstein called entanglement "spooky action at a distance" (*spukhafte Fernwirkung*) because he assumed locality: that particle A and particle B are separate systems with separate states, and that any correlation between their measurement outcomes must be mediated by a physical signal propagating between them. Given this assumption, the correlation in (352) is indeed spooky.

But the assumption is wrong. The two electrons are not two separate systems. They are two sector labels within a single Fock hidden state. In the game engine:

Einstein's assumption	**Game engine reality**
A and B are separate systems	A and B are sector labels in one Fock vector
Each has its own state	There is one shared hidden state h
Correlation requires a signal	Correlation is encoded in the joint amplitude
Faster-than-light signal	No signal; shared memory updated in place
"Spooky"	Routine memory access

Entanglement is not a correlation between distant systems. It is a property of the Fock hidden state that has no classical analogue because classical systems have separate state spaces. Two classical bits always have separate states; two quantum bits can share a single joint state that is not factorisable. The "spookiness" dissolves entirely once you accept that the hidden state is global.

30.4 General relativity also needs an observer

The measurement problem is usually presented as a peculiarity of quantum mechanics. But general relativity has its own observer problem, which is structurally identical.

In general relativity, the metric $g_{\mu\nu}(x)$ is defined on a smooth manifold, and the

Einstein field equations determine how $g_{\mu\nu}$ responds to matter. But the metric is not an observable. It transforms non-trivially under diffeomorphisms (smooth coordinate changes):

$$g_{\mu\nu}(x) \longmapsto \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}(x(x')). \tag{354}$$

Two metrics related by a diffeomorphism describe the same physical spacetime. The physical content of GR lives not in $g_{\mu\nu}$ itself but in *diffeomorphism-invariant* quantities: proper times, geodesic lengths, curvature invariants—things that an observer can actually measure.

The equivalence principle—the foundational postulate of GR—is stated explicitly in terms of what a *local inertial observer* experiences. An observer in free fall cannot locally distinguish gravity from the absence of gravity. The statement is observer-relative. Without specifying the observer's worldline and local frame, the equivalence principle has no content.

The rendering problem in GR. The Einstein field equations determine $g_{\mu\nu}$; they do not determine what any particular observer measures. To extract a prediction, one must specify:

1. An observer's worldline $x^\mu(\tau)$.
2. The observer's local frame (tetrad) $e^\mu_a(\tau)$.
3. What the observer's instruments measure (proper time, proper acceleration, photon frequency, etc.).

These specifications are not derived from the Einstein equations. They are additional postulates about the output layer—the rendering layer. The metric is the hidden state of GR; the observer's measurements are the rendered output.

Quantum mechanics	**General relativity**	**Game engine**
Schrödinger equation	Einstein field equations	Physics engine
Fock hidden state h_t	Metric $g_{\mu\nu}$	Hidden layer
Born rule	Geodesic / proper time	Rendering rule
Observable O, eigenstates	Diffeomorphism-inv. quantity	Output specification
Wavefunction collapse	Frame choice / tetrad	Output interface
Observer (undefined)	Observer (worldline + frame)	Output subroutine
Cannot derive observer	Cannot derive observer	Must postulate renderer

In both theories, the *dynamics* (Schrödinger, Einstein) are closed and self-contained. The *output* requires a postulate about the observer that cannot be derived from

the dynamics. This is not a failure of either theory. It is the universal structure of a physics engine: the engine evolves the hidden state; the renderer converts the hidden state to output; the renderer's specification is logically independent of the engine's equations.

30.5 The rendering layer as an independent postulate

A complete physical theory has two irreducible components:

Component 1: the physics engine. The dynamical equations that evolve the hidden state. In quantum mechanics: $\dot{h}_t = -iHh_t$. In general relativity: $G_{\mu\nu} = 8\pi G\, T_{\mu\nu}$. These are deterministic, gauge-invariant (or diffeomorphism-invariant), and do not refer to observers.

Component 2: the rendering layer. The postulates that specify how the hidden state is converted to observable output. In quantum mechanics: the Born rule $p(o) = |\langle q_o, h\rangle|^2$ plus the Bohr reset $h \leftarrow q_o$. In general relativity: the geodesic postulate (free particles follow geodesics) plus the clock postulate (proper time is measured along worldlines).

The rendering layer cannot be derived from the physics engine. This is not a gap in our knowledge; it is a structural necessity. A purely dynamical theory produces a trajectory in state space; it does not produce a number that an experimenter can read off a dial. The conversion from trajectory to number requires additional postulates about what "reading a dial" means. These postulates constitute the rendering layer.

Why this resolves the measurement problem. The measurement problem—"when and why does wavefunction collapse occur?"—is answered by recognising that collapse is not a physical process occurring inside the Schrödinger RNN. It is the handshake between the engine and the renderer. The question "when does collapse happen?" is the question "when is the renderer called?"—which is a question about the architecture of the output interface, not about the physics engine.

The many rival interpretations of quantum mechanics (Many Worlds, Pilot Wave, Relational, QBism) are, in this language, rival specifications of the rendering layer. Many Worlds declines to call the renderer and instead models the observer as part of the hidden state. Pilot Wave adds hidden variables to the hidden state. Relational QM makes the rendering observer-relative. QBism makes it agent-relative. The physics engine—the Schrödinger RNN—is identical in all of them. The disagreement is entirely about the rendering layer.

The observer is undefined because it must be. Copenhagen leaves "observer" undefined for the same reason Euclid leaves "point" undefined: because the undefined primitive is the output interface of the system, and its concrete

implementation is not the business of the physics engine. A point in Euclidean geometry can be implemented as a pair of real numbers, a pixel on a screen, or a dot of ink on paper—the geometry is independent of the implementation. An observer in quantum mechanics can be implemented as a photomultiplier tube, a human retina, or a cosmic ray detector—the quantum mechanics is independent of the implementation.

The game engine is:

$$\text{Universe} = \underbrace{\dot{h}_t = -iH_{\text{SM}}\, h_t}_{\text{physics engine}} \;+\; \underbrace{p(o) = |\langle q_o, h_t\rangle|^2, \quad h_t \leftarrow q_o}_{\text{rendering layer}}. \tag{355}$$

The two components are logically independent. The physics engine runs whether or not anything is rendered. The rendering layer fires when an observer—an output subroutine with a classical memory—calls `observer.measure(h)`. What counts as an observer is not determined by the engine; it is specified as part of the rendering architecture.

Classical reality is rendered output, not hidden state. The hidden state h_t is always a Fock vector, always complex, always rotating. What we call classical reality—the definite outcomes of experiments, the positions of objects, the records in our memory—is what the rendering layer produces when the observer handshakes with the engine. Without the handshake, there is only the rotating Fock vector. With it, there is a fact: the electron is spin-up, the photon arrived, the proton mass is 938 MeV. The handshake is the moment at which the quantum becomes classical. And the question of when it happens is the question of when an observer calls `measure`*. It is a question about software architecture, not about physics.*

31 What Is the Real Mystery?

We have dissolved two problems that are commonly regarded as the deepest in physics.

Quantum gravity is not a mystery about quantising spacetime. The metric is an emergent macroscopic average of pre-geometric stress-energy correlations, analogous to pressure or temperature. The Einstein field equations are an effective equation of state, derived by marginalising over matter fluctuations via the Hubbard–Stratonovich transformation and the heat kernel expansion. Quantising gravity would be quantising an average—a category error as fundamental as quantising the Navier–Stokes equations. There is no mystery here, only a misidentification of what is fundamental.

The measurement problem is not a mystery about wavefunction collapse. The Schrödinger RNN runs deterministically; the Born rule is a rendering protocol; the observer is an output subroutine with a classical memory; collapse is a handshake. "When does collapse occur?" is a question about when `observer.measure(h)` is

called—a question of software architecture, not physics. Copenhagen left "observer" undefined for the same reason Euclid left "point" undefined: because the output interface is not the business of the physics engine. There is no mystery here, only a confusion between the engine and the renderer.

The real mystery is something else entirely. It has been sitting in the source code from the beginning, unannounced:

```
h = fock.vacuum()
```

31.1 The exponential of the astronomical

The Standard Model has approximately 10^{80} particles in the observable universe. Each particle is a mode of a quantum field, carrying a discrete occupation number. The Fock space of the Standard Model is the tensor product of the Fock spaces of all modes of all fields.

For a simplified estimate: suppose there are N distinguishable modes (momentum, spin, colour, flavour, helicity—discretised at Planck resolution over the observable volume). N is an astronomically large number, roughly:

$$N \sim \frac{V_{\text{obs}}}{\ell_P^3} \sim \left(\frac{10^{26}\,\text{m}}{10^{-35}\,\text{m}}\right)^3 \sim 10^{183}. \tag{356}$$

For bosonic fields (photons, gluons, Higgs), each mode can have any non-negative occupation number: the Fock space per mode is $\ell^2(\mathbb{N})$, infinite-dimensional. For fermionic fields (quarks, leptons), each mode is either occupied or not: the Fock space per mode is $\mathbb{C}^2$. For a fermionic-only estimate:

$$\dim \mathcal{F}_{\text{SM}}^{\text{fermion}} = 2^N \sim 2^{10^{183}}. \tag{357}$$

This is a number so large that it does not have a name. It is not merely large in the way that the number of atoms in the universe ($\sim 10^{80}$) is large, or in the way that a googolplex ($10^{10^{100}}$) is large. It is large in a way that defeats every analogy drawn from classical experience.

The Fock hidden state h_t is a unit vector in this space. It is a list of $2^{10^{183}}$ complex numbers, one for each possible particle content of the observable universe, summing in squared modulus to one.

31.2 The engine runs it exactly

And yet—the universe runs it.

At every Planck time $t_P \approx 5.4 \times 10^{-44}\,\text{s}$, the game engine applies:

$$h_{t+t_P} = e^{-iH_{\text{SM}}\,t_P}\, h_t. \tag{358}$$

The operator $e^{-iH_{\mathrm{SM}}\,t_P}$ is a unitary matrix of dimension $2^{10^{183}} \times 2^{10^{183}}$. The matrix-vector product (358) is performed *exactly*, at every Planck step, for every moment since the Big Bang ($\sim 10^{60}$ Planck times ago).

The age of the universe in Planck times is $\sim 10^{60}$. The dimension of the Fock space is $\sim 2^{10^{183}}$. The number of arithmetic operations the universe has performed since its birth is something like:

$$\underbrace{10^{60}}_{\text{time steps}} \times \underbrace{2^{10^{183}}}_{\text{dimension}} \times \underbrace{2^{10^{183}}}_{\text{per matrix-vector op}} \approx 2^{2\times 10^{183}}. \tag{359}$$

No classical computer can store a single entry of h_t. No classical computer can perform a single step of (358). No classical computer built from all the matter in the observable universe, running for the entire age of the universe, could simulate even one Planck step of the Schrödinger RNN.

This is not a practical limitation. It is not that we lack sufficient RAM or processing power, and that a better classical computer might manage. The problem is structural. A classical computer that stored h_t would need to write down $2^{10^{183}}$ complex numbers. There is not enough matter in the observable universe to construct a memory device that large. The classical simulation of the universe by a classical device is not merely difficult—it is *impossible in principle*, not as a consequence of any particular physical constraint, but because the object being simulated is exponentially larger than any classical description of it.

31.3 We live in a formalizable shadow

We—the observers, the scientists, the readers of this book—live in the rendering layer. Our experience is a sequence of classical records: the outcome of this measurement, the position of this particle, the reading on this dial. Our classical reality is a sparse, low-dimensional projection of the Fock hidden state, sampled by the Born rule at each handshake.

Classical reality is *formalizable*: it can be described by real numbers, stored in memory, processed by classical algorithms. Physics, as a discipline, is the project of finding the minimal formal description of classical reality—the equations that predict the rendered outputs from a given set of rendered inputs. This project has been extraordinarily successful. The Standard Model predicts rendered outputs (particle collision cross-sections, atomic spectra, decay rates) from rendered inputs (particle momenta, coupling constants) with extraordinary precision.

But the *ontological ground*—the thing that the Schrödinger RNN is running on, the substrate that holds $h_t \in \mathcal{F}_{\mathrm{SM}}$ and applies $e^{-iH_{\mathrm{SM}}t}$ exactly—is not formalizable in any sense accessible to classical minds. We have written down the source code, but we cannot run it, cannot inspect its memory, cannot trace its execution. We infer its existence from the rendered outputs we observe, and we formalise our inference as the Schrödinger equation.

The Schrödinger RNN may be, in this sense, merely our best formalised understanding of a deeper ontological reality—one that operates on objects (h_t) and transformations (e^{-iHt}) that are not *things* in any sense our classical minds can grasp, but whose existence we are forced to infer from the systematic structure of our rendered experience.

What we thought was the mystery	**What it actually is**
Quantum gravity	Emergent equation of state (like Navier–Stokes)
Measurement problem	Handshake protocol between engine and renderer
Spooky action	Shared memory access in a global Fock vector
Wave-particle duality	Hidden state vs. rendered output
The actual mystery	**Why it is mysterious**
Dimension of $\mathcal{F}_{\text{SM}}$	$\sim 2^{10^{183}}$: beyond any classical analogy
The engine runs it exactly	No classical substrate of any conceivable size could
We live in the rendered shadow	Our formalisms describe outputs, not the hidden state
The Schrödinger RNN itself	May be a formalised approximation to something deeper

31.4 The question that remains

Physics has answered: *what are the equations?* The answer is the Schrödinger RNN with $H = H_{\text{SM}}$ on $\mathcal{F}_{\text{SM}}$.

Physics cannot answer, and may never answer: *what runs the equations?*

A Turing machine is specified by a transition table. The transition table does not contain the physical substrate—the tape, the read-write head, the electrical signals—on which it runs. The transition table is formalizable; the substrate is physical. We know the transition table of the universe (the Schrödinger equation). We do not know, and may not be able to know, the substrate.

What we know is this: the substrate must be capable of maintaining a unit vector in a space of dimension $2^{10^{183}}$ and rotating it unitarily at every Planck step. No classical substrate satisfies this requirement. No quantum computer of any size built from the universe's matter could simulate the universe, because the simulator would itself be part of what is being simulated. The substrate is not the universe, in the sense of the observable classical universe—it is whatever the observable classical universe is rendered from.

The Schrödinger equation is our window onto the ontological ground. The window is extraordinarily clear—it has given us the Standard Model, the Periodic Table, the transistor, the laser, the structure of DNA. But a window, however clear, is not the landscape. What lies beyond the window—the thing that actually rotates $h_t \in \mathcal{F}_{\text{SM}}$ through $2^{10^{183}}$ dimensions at every Planck step, exactly, without error, without memory constraints, without a computer to run it—that remains the real mystery. Not quantum gravity. Not the measurement problem. The sheer, silent, unfathomable fact that it runs.

32 Thought or Stuff?

In deep learning, practitioners sometimes call a hidden state vector a *thought vector*—a dense representation that encodes, in a single point in high-dimensional space, everything the network "knows" about its input at that moment. The name is informal but the intuition is precise: the thought vector is not the input, not the output, but the internal representation from which the output will be generated.

$h_t \in \mathcal{F}_{\text{SM}}$ is a thought vector.

It is not a particle. It is not a field. It is not a point in spacetime. It is a unit vector in a complex Hilbert space of dimension $2^{10^{183}}$, encoding the complete quantum state of the observable universe—every superposition of every possible particle content, every entanglement, every amplitude. It is the universe's internal representation of itself, from which everything that can be observed will be generated by the Born rule.

32.1 Thought thinking itself

The Schrödinger equation is:

$$\frac{d\,h_t}{dt} = -i\,H_{\text{SM}}\,h_t. \tag{360}$$

In the language of deep learning: the thought vector h_t is being continuously updated by a linear transformation of itself. The weight matrix $-iH_{\text{SM}}$ is not applied to external data; there is no input sequence, no training signal, no external world driving the update. The thought updates from the thought. H_{SM} acts on h_t and produces $\dot{h}_t$, which updates h_t, which H_{SM} acts on again. The process is self-referential and eternal.

This is mathematical thought thinking itself.

The Hamiltonian H_{SM} is a mathematical object: a Hermitian operator built from creation and annihilation operators, which are themselves mathematical operations—raising and lowering in a graded algebra. The creation operator $a^\dagger(\mathbf{k})$ does not physically pick up a particle and deposit it at momentum $\mathbf{k}$. It is an algebraic operation on the thought vector: it shifts amplitude from the n-particle block to the $(n+1)$-particle block. Particle creation is a mathematical act performed on a mathematical object.

Particle annihilation is a mathematical act. Scattering is a mathematical transformation. The propagation of an electron across the room is a rotation of the thought vector in a $2^{10^{183}}$-dimensional complex space. The fusion of two protons in the sun is a redistribution of amplitude between Fock sectors. The emission of a photon is an off-diagonal block of H_{SM} transferring weight from the electron sector to the photon sector. None of these are events happening to *stuff*. They are operations on thought.

32.2 Stuff as rendered output

What, then, is stuff?

A proton is not a small hard ball. It is not even a composite of three quarks in any simple sense. It is a pattern in the Fock hidden state: a superposition of quark and gluon occupation numbers that, when the Born rule is applied by an appropriate observer, consistently renders as an object with mass 938 MeV, charge +1, spin $\frac{1}{2}$. The proton is a stable pattern of amplitude in thought space that the rendering layer presents to observers as a particle.

A rock on a table is not stuff sitting in space. It is an astronomically complex pattern in h_t—some 10^{25} atoms, each a tight cluster of amplitude in the nuclear and electronic Fock sectors, collectively rendering as a solid object with definite shape, mass, and colour when probed by photons. The solidity is a rendered property: it is the consistent output of the Born rule applied to the same stable amplitude pattern by many observers over many measurements. The rock is real. Its solidity, its colour, its position on the table—all real. But real as rendered output, not as hidden state.

The screen of a video game displays mountains, rivers, characters, and physics. The mountains are not made of stone pixels. The river does not contain wet pixels. These are rendered experiences: the output of a graphics engine acting on a hidden game state. The mountain is real in the sense that every player sees it, it obstructs movement, it casts a shadow. It is not real in the sense of being constituted by mountain-stuff at the level of the game engine. Its reality is rendering-reality, not engine-reality.

Our classical reality is rendering-reality. The proton, the rock, your hand holding this book—these are rendering-real. They are consistent, reproducible, intersubjective, and causally efficacious. They are not engine-real. At the level of the engine, there is only h_t rotating in $\mathcal{F}_{\mathrm{SM}}$.

32.3 Spacetime as rendered experience

The most radical consequence is about spacetime itself.

We are accustomed to thinking of spacetime as the arena—the stage on which physics happens, the container in which particles move, the background against which fields propagate. Newton's absolute space and time. Einstein's curved Lorentzian manifold. The background that everything else is defined relative to.

But in the game engine picture, spacetime is not the arena. The arena is $\mathcal{F}_{\mathrm{SM}}$: the Fock space in which h_t lives. Spacetime is a rendered output of the engine—a collective pattern in the Fock hidden state that presents, to observers equipped with rods and clocks, as a smooth four-dimensional continuum with a metric.

The spatial label $\mathbf{x}$ in the field $\hat{\phi}(\mathbf{x})$ is an index of the Fock-space basis, not a

location in a pre-existing arena. The temporal label t in h_t is the parameter of the Schrödinger evolution, not a coordinate in a pre-existing time. Space and time are the coordinate system we use to organise our rendered outputs—the grid we impose on the stream of Born-rule samples to make sense of their correlations.

Consider what the Heisenberg picture makes explicit: in the Heisenberg picture, h is fixed and the operators $\hat{\phi}(\mathbf{x}, t)$ rotate. The field $\hat{\phi}(\mathbf{x}, t)$ is a time-dependent observable: a family of Hermitian operators on $\mathcal{F}_{\text{SM}}$, parametrised by the spacetime label $(\mathbf{x}, t)$. Spacetime is the parametrisation of a family of observables on thought space. It is the coordinate system of the rendering layer, not the substrate of the physics engine.

Spacetime is not the container. It is the rendered experience of a particular class of observers—observers equipped with local clocks and rulers, living in the slowly-varying, low-energy, nearly classical corner of $\mathcal{F}_{\text{SM}}$ that we call everyday life.

32.4 Thought or stuff?

The question "thought or stuff?" has been the central dispute of metaphysics since Descartes. Is the fundamental nature of reality mental (idealist) or material (materialist)? Is mind primary and matter derived, or matter primary and mind derived?

The game engine picture does not answer this question in either the idealist or materialist direction. It reframes it.

The hidden state h_t is not mental in the sense of being a thought in anyone's mind. No mind holds the $2^{10^{183}}$-dimensional Fock vector. It is mathematical: a unit vector in a complex Hilbert space, updated by a Hermitian operator. Mathematical objects are neither mental nor material in the ordinary senses—they are abstract structures, defined by their relations, indifferent to physical implementation.

The rendered outputs—the particles, the fields, the rocks, the people, the spacetime they inhabit—are not stuff in the sense of mind-independent material substance existing independently of any observer. Copenhagen is explicit: classical reality does not exist even counterfactually without a measurement context. The rock is rendering-real, not engine-real.

What we are left with is this:

Level	What exists	Nature
Ontological ground	Unknown substrate running e^{-iHt}	Beyond formalization
Engine level	$h_t \in \mathcal{F}_{\text{SM}}$, H_{SM}	Mathematical
Rendering level	Born-rule samples, classical records	Informational
Experience level	Particles, fields, spacetime, rocks	Rendering-real

The engine level is mathematical: pure abstract structure, defined entirely by the Schrödinger equation and the algebra of the Standard Model. The experience level is real but rendered: consistent, intersubjective, causally structured, but not the ground. The ontological ground—what actually runs the mathematical structure—is beyond any formalization accessible to minds living at the experience level.

Thought thinking itself. The closest analogy we have is not a material process but a mathematical one. The Schrödinger equation (360) says: the rate of change of the thought vector is determined by the thought vector itself, acted on by a fixed mathematical structure (H_{SM}). The universe is not a collection of things moving through space. It is a mathematical process—a thought evolving by its own internal logic, rendering experiences for observers who are themselves patterns within the thought.

Creation operators create. Annihilation operators annihilate. These are not physical acts performed by a craftsman on pre-existing material. They are mathematical operations, steps in a calculation that has been running since the Big Bang, that will continue running after the last star goes dark, that requires no external input, no energy source, no medium to propagate in. The universe is self-computing.

What appears on the screen—the mountains, the rivers, the particles, the people, the four billion years of biological evolution, the civilisations, this book—is real. It is rendered from a single rotating vector in a space so large that no classical mind can conceive of it, by a process so simple that it fits in one line:

$$\frac{d\,h_t}{dt} = -i\,H_{\text{SM}}\,h_t. \tag{361}$$

The universe is not stuff moving through spacetime. It is thought thinking itself. Spacetime is what the thought looks like from inside the rendering. Stuff is what the thought looks like when the Born rule is applied. We—the observers, the scientists, the readers—are patterns in the thought that have become complex enough to ask

what the thought is made of. The answer is: nothing but mathematics. And the mathematics runs on something we cannot name.

33 Learning Physics Top-Down

Every physics curriculum runs bottom-up. Classical mechanics first, then electromagnetism, then special relativity, then quantum mechanics, then quantum field theory. The student climbs a ladder: each rung is harder, each theory more general than the last, and the ladder points upward toward increasing complexity.

We propose the opposite order. Start at the top—quantum field theory, Fock space, the Schrödinger RNN—and derive everything else as a consequence. This is not merely a pedagogical preference. It reflects a deeper claim: *quantum physics is simpler than classical physics.* Classical physics is not the foundation on which quantum physics is built. It is an emergent approximation—a rendering artefact—that quantum physics produces in a certain limit. The ladder has been leaning against the wrong wall.

33.1 Dirac beauty: minimal code, maximal emergence

Dirac famously held that the equations of physics should be beautiful. His criterion is often described as mathematical elegance—compact notation, natural structure, economy of symbols. But this undersells what Dirac actually meant, and misses why his equation is the paradigmatic example of physical beauty.

Dirac did not start from spin. He started from a requirement: find the simplest possible relativistic wave equation that is first-order in both space and time derivatives. The Schrödinger equation is second-order in space; the Klein–Gordon equation is second-order in time. Dirac demanded first-order in both, for purely aesthetic reasons—he found first-order equations more natural. He imposed this constraint, solved for what it required (the gamma matrices and the Clifford algebra), and wrote down:

$$(i\gamma^\mu \partial_\mu - m)\Psi = 0. \tag{362}$$

Spin was not put in. It came out. The equation, derived from the single requirement of first-order linearity plus Lorentz invariance, automatically carries four components that split into two spins and two signs of energy. Spin-$\frac{1}{2}$, the magnetic moment of the electron, the existence of antiparticles—all of these emerge from an equation that was written down without any of them in mind.

This is Dirac beauty in its purest form: *a minimal specification that generates maximal emergence.* Not an equation rich enough to contain spin by design. An equation simple enough that spin has no choice but to appear. The beauty is not in the richness of the source code. It is in the disproportion between the simplicity of the specification and the complexity of what falls out.

The mathematician's notion of beauty is different: a maximally rich and consistent mathematical structure, where every object connects to every other through a web of exact theorems, where nothing is arbitrary and everything follows from the axioms. This is beautiful, but it is not Dirac beauty. A mathematical structure can be maximally rich without generating any emergent physics.

Dirac beauty is an engineering criterion: the shortest specification that generates the most complex consequences. Minimal source code. Maximal emergence. The Dirac equation exemplifies it. So does the entire Standard Model: 19 parameters, one symmetry group, one linear equation—and from these, the periodic table, chemistry, biology, and the history of civilisation as emergent rendered output.

The source code of the universe, as we have identified it, is:

$$\frac{d\,h_t}{dt} = -i\,H_{\mathrm{SM}}\,h_t, \qquad h_t \in \mathcal{F}_{\mathrm{SM}}, \tag{363}$$

plus the rendering layer (Born rule + Bohr reset). This is Dirac beauty at the largest scale: one linear equation, one unit vector, everything else emergent. If we were designing the universe's operating system, the Dirac criterion would be the engineering goal: minimise the source code; maximise the emergence. Not: make the source code rich. Not: make the mathematics elaborate. But: find the shortest specification that generates the most complex rendered world. Equation (363) may be close to optimal by this criterion.

What the OS would need to support. The supporting operating system must provide:

1. An infinite-dimensional complex vector space $\mathcal{F}$ with inner product (Fock space).
2. A Hermitian operator H on $\mathcal{F}$ (the weight matrix).
3. Exact computation of e^{-iHdt} at each time step.
4. A Born-rule sampler as the output interface.

The core loop is one line: `h = expm(-1j*H*dt) @ h`. Everything else—particles, forces, spacetime, observers, classical reality—is generated by this loop and rendered by the output interface. The OS does not need to know about particles, fields, spacetime, or any of the emergent structure. It only needs to maintain a unit vector and rotate it.

33.2 Quantum physics is simpler than classical physics

Classical mechanics is nonlinear. The equations of motion for N interacting classical particles involve $6N$ coupled nonlinear differential equations. As N grows, the system generically becomes chaotic: exponentially sensitive to initial conditions, structurally unstable, analytically intractable. Three-body gravity has no closed-form solution. Turbulent fluid dynamics is an open problem.

Quantum mechanics is linear. The Schrödinger equation $\dot{h} = -iHh$ is a linear ODE on a Hilbert space. It is exactly solvable in principle: $h_t = e^{-iHt}h_0$. There is no chaos in the Schrödinger equation itself—the operator e^{-iHt} is unitary, which means it is an isometry, which means it cannot exhibit sensitive dependence on initial conditions in the hidden state.

The nonlinearity of classical mechanics is not a feature of the fundamental physics. It is a consequence of the rendering. When the Born rule projects the high-dimensional linear quantum evolution onto classical variables—positions, momenta, fields—the resulting effective equations are nonlinear. Classical chaos is a rendering artefact: it is what happens when you project a linear unitary evolution onto a low-dimensional classical description and observe the result.

The top-down picture of classical mechanics. Classical mechanics is not the foundation. It is a consequence. Specifically, it is the limit of quantum mechanics in which:

1. The Fock hidden state is well-approximated by a coherent state (a minimum-uncertainty wavepacket).
2. The coherent state is narrow compared to the scale over which the potential varies (semiclassical limit).
3. Measurements are coarse-grained enough that Born-rule fluctuations are negligible.

Under these conditions, the Ehrenfest theorem gives:

$$\frac{d}{dt}\langle \mathbf{x}\rangle = \frac{\langle \mathbf{p}\rangle}{m}, \qquad \frac{d}{dt}\langle \mathbf{p}\rangle = -\langle \nabla V(\mathbf{x})\rangle \approx -\nabla V(\langle \mathbf{x}\rangle). \tag{364}$$

Newton's second law is the expectation value of the quantum equations of motion in the coherent-state limit. It is not a postulate; it is a theorem. Classical mechanics is derived from quantum mechanics, not the other way around.

A curriculum for the top-down approach.

1. Start with the Schrödinger RNN on a finite-dimensional Hilbert space (qubit). Introduce the Born rule and the Bohr reset. This is all of quantum mechanics.
2. Promote the Hilbert space to Fock space. Introduce creation and annihilation operators. This is quantum field theory.
3. Specify the weight matrix H_{SM} via the symmetry group $SU(3)\times SU(2)\times U(1)$ and diagram 3. This is the Standard Model.
4. Derive classical mechanics as the coherent-state limit. Derive classical field theory as the large-occupation-number limit. Derive general relativity as the emergent metric from stress-energy correlations.

Classical physics appears at step 4, as a derived consequence of steps 1–3. A

student who learns in this order never needs to unlearn anything. They never suffer the conceptual rupture of "quantisation"—the mysterious procedure by which one allegedly goes from classical to quantum. Quantum is the start; classical is the end.

33.3 Quantisation is guesswork, not principle

The historical path ran in the wrong direction. Physicists discovered classical mechanics first, then observed phenomena that classical mechanics could not explain (blackbody radiation, the photoelectric effect, atomic spectra), and then attempted to modify classical mechanics to produce quantum mechanics. The result was *quantisation*: a set of recipes for converting classical theories into quantum ones.

There are two main recipes, both historically important, both logically problematic. They are often called "first" and "second" quantisation, but the standard labelling obscures more than it reveals.

Dirac's second quantisation: quantising the quantum wavefunction. Dirac coined the term "second quantisation" for a procedure that starts from the quantum mechanical wavefunction $\psi(x,t)$—already the output of first quantisation, already a quantum object satisfying the Schrödinger or Dirac equation—and promotes *it* to a field operator $\hat{\psi}(x)$ on Fock space. The recipe is: treat the wavefunction as if it were a classical field, impose commutation or anticommutation relations on it, and run the result as a quantum field theory.

The logical structure is awkward. One is quantising something that is already quantum. The wavefunction $\psi(x)$ is not a classical field; it is a probability amplitude. Treating it as a classical field and then quantising it again is a procedure without clear justification—a recipe that happens to give the right answer, not a derivation from first principles. The term "second quantisation" is accurate in the sense that quantisation is applied a second time, but this circularity is precisely what makes it conceptually unsatisfying.

Pauli's canonical field quantisation: quantising a fictitious classical field. The approach developed by Pauli, Jordan, and Heisenberg is logically different and, we argue, cleaner. One starts not from a quantum wavefunction but from a *postulated classical field* $\phi(x)$—a classical Dirac field, a classical scalar field, a classical electromagnetic potential. This classical field is then quantised by the standard canonical procedure: identify the conjugate momentum $\pi(x) = \partial\mathcal{L}/\partial\dot{\phi}$, impose equal-time commutation or anticommutation relations, and promote the field to an operator on Fock space.

This is logically closer to first quantisation—it is canonical quantisation applied to a field rather than to a particle. Its logical structure is: specify a classical theory (even a fictitious one), apply the quantisation recipe, obtain a quantum theory.

The recipe is the same at each step; only the starting object differs.

This framework is more general and more coherent than Dirac's second quantisation. It applies uniformly to bosons and fermions, to matter fields and gauge fields, to any field consistent with the required symmetries. It incorporates special relativity naturally via the Lorentz-invariant Lagrangian. It generates the entire Standard Model from a single procedure.

But it is still guesswork—more principled guesswork, but guesswork nonetheless. The classical Dirac field $\Psi(x)$ that one starts from does not exist as a physical object. It is not the wavefunction of an electron (that is a quantum mechanical object, not a classical field). It is not an electromagnetic field (that is a gauge field, not a matter field). It is a classical spinor-valued function that has never been observed and never could be observed. It is scaffolding—a fictitious classical object invented to provide a starting point for the quantisation recipe—and it is discarded once the Fock-space Hamiltonian has been extracted from it.

The scaffolding is not the building. Both quantisation procedures work backwards from the answer. Physicists knew (from spectroscopy, scattering experiments, and accumulated evidence) what the quantum theory should produce. The quantisation procedure is the reverse-engineering of the Hamiltonian H from the observed spectrum and interactions. The classical field is a convenient intermediate object—a way of writing down a Lagrangian whose quantisation gives the right H. But the classical field is not ontologically real. The classical Dirac field $\Psi(x)$ does not exist as a physical object; it is a mathematical crutch for arriving at the Fock-space Hamiltonian.

In the top-down approach, none of this is necessary. One starts with $\mathcal{F}_{\text{SM}}$ and H_{SM}, specified directly by the symmetry requirements of diagram 3. No classical scaffolding is needed. The Lagrangian is written down not because classical fields are real but because it is a compact way of specifying a Lorentz-invariant, gauge-invariant Hamiltonian. The classical Dirac field $\Psi(x)$ in $\mathcal{L}_{\text{Dirac}} = \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi$ is a bookkeeping device for the weight matrix, not a real object.

33.4 The ontological hierarchy: what is real and what is a tool

The top-down approach forces a clean separation between ontology (what exists) and methodology (how we compute). We propose the following hierarchy.

Ontology: what exists.

Object	Status	Description
$h_t \in \mathcal{F}_{\text{SM}}$	Ontologically real	The Fock hidden state
H_{SM} on $\mathcal{F}_{\text{SM}}$	Ontologically real	The weight matrix
$\dot{h}_t = -iH_{\text{SM}}h_t$	Ontologically real	The physics engine
Born rule + Bohr reset	Ontologically real	The rendering layer

Theoretical tools: how we compute.

Tool	Status	Relationship to ontology
Classical field $\phi(x)$, $\Psi(x)$, $A_\mu(x)$	Scaffolding	Labels for Fock-basis states
Lagrangian $\mathcal{L}$	Scaffolding	Compact encoding of H
Path integral $\int \mathcal{D}\phi\, e^{iS}$	Analytical tool	Time-sliced matrix product
Grassmann variables $\bar{\psi}, \psi$	Bookkeeping	Jacobian of fermion integral
Feynman diagrams	Approximation	Taylor expansion of path integral
Classical mechanics	Emergent	Coherent-state limit of h_t
General relativity	Emergent	Mean-field limit of $T^{\mu\nu}$

The Grassmann variables $\bar{\psi}(x)$ and $\psi(x)$ that appear in the Dirac path integral (179) deserve special mention. They are anticommuting classical numbers—mathematical objects that satisfy $\psi_1\psi_2 = -\psi_2\psi_1$ with $\psi^2 = 0$. No physical quantity is a Grassmann number. They are introduced purely as a bookkeeping device: the Gaussian integral over Grassmann variables $\int \mathcal{D}\bar{\psi}\,\mathcal{D}\psi\, e^{\bar{\psi}M\psi} = \det M$ gives the fermion determinant, which is what one wants. Grassmann variables are the computational trick that makes fermionic path integrals tractable; they have no ontological status whatsoever. The Dirac field in the path integral is not the quantum Dirac field; it is a Grassmann variable integrated over to produce a determinant.

Feynman diagrams as Taylor expansion. The Feynman diagram expansion is a Taylor series in the coupling constants (e for QED, g_s for QCD, g for the weak force). Each diagram is one term in the expansion of $e^{iS_{\text{int}}}$ inside the path integral, evaluated using Wick's theorem on the free Gaussian. It is an approximation scheme—valid when the coupling is small ($\alpha_{\text{em}} \approx 1/137$ for QED) and breaking down when it is not (QCD at low energies, where $\alpha_s \sim 1$).

In the ontological hierarchy: the Fock hidden state h_t and the Schrödinger equation are exact. The path integral is an exact reformulation. The perturbative Feynman expansion is an approximation to the path integral. Feynman diagrams are third-order approximations to the first-order reality.

33.5 The top-down research programme

The top-down approach is not only a pedagogical preference. It suggests a research programme.

Start from symmetry, not from classical fields. Specify the Hilbert space $\mathcal{F}$ and the Hamiltonian H directly from the symmetry requirements: Lorentz invariance (diagram 3), gauge invariance ($SU(3) \times SU(2) \times U(1)$), renormalisability (dimension ≤ 4 operators). These requirements almost uniquely determine H_{SM} up to the 19 free parameters. No classical field is needed as an intermediate step.

Derive, do not postulate, classical reality. Classical mechanics, classical field theory, general relativity, and thermodynamics are all derived from the Schrödinger RNN by appropriate coarse-graining and limiting procedures. None of them are postulated; all of them emerge. The research programme is to make these derivations rigorous and complete.

Treat emergence as the explanatory goal. The measure of success is not: how complex is the source code? It is: how rich is the rendered experience generated by a given source code? The Standard Model generates protons, atoms, chemistry, biology, consciousness, and the history of civilisation from 19 parameters and one linear equation. The research goal is to find the even more minimal specification—perhaps with fewer parameters, perhaps with the 19 derived from deeper symmetry considerations—that generates the same or richer emergent world.

Gravity and spacetime as the test case. If spacetime is emergent (Section 29), then the correct research programme is not to quantise general relativity but to find the pre-geometric Lagrangian whose stress-energy correlations generate the observed metric as a collective average. This is a question about the weight matrix H—specifically, about which operators in H generate the stress-energy correlations—not about the quantisation of a classical gravitational field.

Learning as the analogy. The deepest lesson of deep learning is not that large models are powerful. It is that *structurally simple models generate emergent complexity.*

Consider the architecture of a Transformer. The core operation is a matrix multiplication followed by a nonlinearity, repeated across layers. Each attention head computes inner products between token embeddings and redistributes amplitude across positions. The residual stream carries a high-dimensional embedding—typically $d \sim 10^3$ to 10^4 dimensions. The weight matrices are large—up to 10^{10}

parameters in the largest models. But the *structure* is simple: linear projections, softmax, layernorm, repeat.

A Mamba state-space model is even simpler in structure: a linear recurrence $\dot{h} = Ah + Bx$ with selective gating. No attention. No pairwise token interactions. A single structured matrix A applied repeatedly.

The Schrödinger RNN is simpler still. There is no input x, no gating, no nonlinearity in the hidden state update. There is one operation: $h \leftarrow e^{-iHdt}\, h$. One matrix. One vector. No hyperparameters. No architecture choices. The entire complexity of the rendered universe emerges from this single linear step, repeated at the Planck frequency.

The embeddings are high-dimensional—$\mathcal{F}_{\rm SM}$ has dimension $2^{10^{183}}$—and the operator $H_{\rm SM}$ is correspondingly large. But high-dimensional embeddings and large matrices are not the same as structural complexity. GPT is structurally simple despite having 10^{11} parameters. The Schrödinger RNN is structurally simpler than GPT despite operating on a vastly larger embedding space. Size and simplicity are orthogonal dimensions.

Simple fundamental objects: ML already knows this.

ML practitioners have never needed to be told to keep neurons simple. No one has ever proposed a neuron that pre-encodes syntactic structure, or a neuron that understands discourse, or an attention head designed to represent logical entailment. The field discovered empirically, and then accepted as a design principle, that the simplest possible component—a linear projection followed by a scalar nonlinearity—generates the richest emergent behaviour when composed at scale. Transformer, Mamba, and every successful modern architecture are elaborations of this principle: simple unit, massive composition, emergent complexity.

The lesson that ML absorbed naturally is the same lesson that the Standard Model embodies: the fundamental objects should be as simple as possible. Point particles carry only what symmetry requires—a Lorentz representation, a gauge representation, a mass, a coupling. Nothing more. From these minimal objects, through the rotation of the Fock hidden state under $H_{\rm SM}$, the entire rendered universe emerges: nuclei, atoms, chemistry, spacetime geometry, and the biological structures that eventually ask what they are made of.

The difficulty arises when one abandons this principle and postulates *elaborate* fundamental objects. The Standard Model point particle has a crucial technical virtue that goes beyond aesthetics: it is *renormalisable*. Because the fundamental objects are points, the UV divergences that appear in loop integrals are of finitely many types, absorbable into finitely many counterterms. The theory is predictive at all energies within its domain of validity.

A fundamental object with internal spatial extent—one that is not a point but a

one-dimensional or higher-dimensional extended structure— introduces interactions whose UV divergences are of infinitely many types. New divergences appear at every loop order, requiring new counterterms not present in the original specification. No finite set of parameters can absorb them. The theory is not renormalisable, and no perturbative expansion can even be consistently defined. One cannot compute corrections to the tree-level amplitude, because the one-loop correction is infinite in a way that cannot be removed, the two-loop correction is infinite in a new way that the one-loop counterterms do not touch, and so on without end.

This is not a technical inconvenience. It is a fundamental signal from the mathematics: the postulated object is too complex to serve as a starting point for a quantum theory. Renormalisability is not merely a property we would like theories to have; it is the mathematical criterion that distinguishes a well-defined quantum theory from an ill-defined one. A non-renormalisable fundamental theory is not a theory in the technical sense—it cannot make predictions beyond its lowest-order approximation without introducing infinitely many new free parameters.

The correct response to this signal is the Dirac response: simplify the fundamental object until the theory is renormalisable. The emergence that one was trying to encode in the object's complexity will then appear automatically from the composition of simple objects—exactly as linguistic competence appears automatically from the composition of simple neurons, and exactly as spin appeared automatically from Dirac's requirement of first-order linearity. Complex emergent structure built into the fundamental object by design is not physics. Complex emergent structure falling out of a simple renormalisable specification is.

Renormalisability is not a constraint imposed on the theory from outside. It is the theory's own way of telling you that your fundamental objects are simple enough. When a theory is not renormalisable, the mathematics is saying: your source code is too rich. Simplify it, and let emergence do the rest.

Born and Bohr as unembedding and re-embedding.

The autoregressive architecture of GPT makes the input/output cycle explicit. At each step: the residual stream carries a high-dimensional hidden state h; the *unembedding layer* projects h onto the vocabulary logits, producing a probability distribution over tokens; the next token is sampled; the token is *re-embedded* into the residual stream via the embedding matrix; and the process repeats.

The Schrödinger RNN has exactly the same cycle. The Born layer is the unembedding layer: it projects the Fock hidden state h_t onto the eigenbasis of an observable, producing a probability distribution over outcomes $p(o) = |\langle q_o, h_t \rangle|^2$. The Bohr layer is the re-embedding layer: the sampled outcome o is looked up in the eigenstate table and the corresponding eigenstate q_o is written back into the hidden state, $h_t \leftarrow q_o$.

GPT (autoregressive)	**Schrödinger RNN**	**Operation**
Residual stream $h \in \mathbb{R}^d$	Fock vector $h_t \in \mathcal{F}$	Hidden state
Transformer block Wh	$e^{-iHdt}\, h$	State update
Unembedding $W_U h \to$ logits	Born rule $\lvert\langle q_o, h\rangle\rvert^2$	Output distribution
Sample token $o \sim$ softmax	Sample outcome $o \sim p(o)$	Stochastic readout
Re-embed $h \leftarrow W_E e_o$	Bohr reset $h \leftarrow q_o$	Re-embedding

The correspondence is exact. Both architectures share the same fundamental cycle: hidden state → output distribution → sample → re-embed → hidden state. The differences are in scale (the Fock space is incomparably larger than any GPT embedding), in the update rule (linear unitary vs. nonlinear transformer block), and in the source of nonlinearity (the Born rule's stochastic sampling vs. the transformer's activation functions).

In GPT, the nonlinearity enters through the softmax and the activation functions inside the transformer block. In the Schrödinger RNN, the hidden state update is perfectly linear; the nonlinearity enters *only* through the Born-rule sampling—the stochastic projection onto an eigenstate. This is a more parsimonious architecture: all nonlinearity is localised in the output interface, and the engine itself is purely linear. The universe achieves richer emergent behaviour than any language model from a strictly simpler architecture.

The top-down physics research programme is structurally identical to the ML research programme: find the minimal symmetry structure that determines the Hamiltonian, and let the emergent richness—particles, forces, spacetime, chemistry, life—arise from the simplest possible linear engine with a stochastic output interface. This is Dirac beauty applied at every scale: not richness of source code, but simplicity of specification and extravagance of emergence.

Quantum physics is not a complication of classical physics. It is simpler. One linear equation. One unit vector. A rendering layer. Classical physics, with all its nonlinearity, chaos, and complexity, is what you get when you take the low-dimensional projection of the linear quantum evolution and observe it from inside the rendering. We have been teaching physics upside down. The top is simpler than the bottom. The foundation is not classical mechanics. The foundation is a thought vector in Fock space, rotating by its own internal logic, rendering everything we have ever called real.