QFT: The Scenic Route

Oscar Barrera

It's hard to appreciate a really elaborate joke the first time you hear it. You might catch the punchline, but only after you know where it's going do all the little details in the setup start to make sense. I think learning theoretical physics is a lot like that. Once you've seen the big picture, the rigor and formalism start to feel more natural, and you can enjoy how all the pieces build toward it.

These notes are meant as a tour through some of the most beautiful ideas in modern theoretical physics and the math that ties them all together. They're written for someone who's already met quantum mechanics and a bit of field theory, but I don't assume you remember all the details. Whenever possible, I try to reintroduce key ideas from scratch, with enough context to make them feel familiar again.

We start from the simple idea that symmetries imply something fundamental about the structure of nature, and from that principle watch photons emerge, gauge invariance take shape, and higher-dimensional structures reveal the geometry hidden inside our physical laws. Along the way, I try to point out some of the mathematical scenery that tends to get mentioned in passing but rarely unpacked: things like *topology*, *differential geometry*, and *representation theory*. The goal isn't to make you an expert, but to make the next seminar you sit through just a little less confusing.

This is not a rigorous reference in any sense. It's meant as a friendly stroll through quantum field theory, a way to build the intuition first so that the rigor feels like the natural next step.

Contents

1 Motivation and Definitions

Physics is full of conserved quantities: energy, momentum, angular momentum, electric charge. These are not separate miracles; they all come from one principle. Whenever the laws of physics look the same after some continuous transformation, there is a conserved quantity. Shift time forward by a few seconds and the rules stay the same $\Rightarrow$ energy is conserved. Slide the whole system in space $\Rightarrow$ momentum is conserved. Rotate it $\Rightarrow$ angular momentum.

This connection between symmetry and conservation is the content of **Noether's theorem**, one of the most beautiful results in theoretical physics. It tells us that the quantities we think of as fundamental are simply the shadows cast by deeper symmetries.

But the world is far richer than that. The interactions that dictate our lives are much more complicated than a few conserved numbers. Quantum states come with many "tags": spin, charge, parity... Certain particles come in pairs or families, have antiparticles, or differ by handedness. Some couple to the Higgs field and gain mass; others stay massless. Massless particles have distinct polarizations; some forces act only on left-handed fields. All of these patterns cry out for a systematic way to organize and categorize them.

That organizing principle is *group theory.* It is the mathematical language of symmetry - the tool that lets us understand why particles arrange themselves the way they do, and why the interactions among them take the forms they do.

1.1 What is a Group?

A **group** is a set of elements $G = \{g_1, g_2, \dots\}$ together with a rule, called the *group operation*, that tells you how to combine any two of them:

$$g_1 \circ g_2 \ \in \ G.$$

The operation must satisfy four axioms:

(1) **Closure:** For any $g_1, g_2 \in G$, the product $g_1 g_2$ is also in G.

(2) **Associativity:** $g_1(g_2 g_3) = (g_1 g_2) g_3$ for all $g_i \in G$.

(3) **Identity element:** There exists an element $I \in G$ such that $Ig = gI = g$ for all g.

(4) **Inverse element:** For each $g \in G$ there exists an element $g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = I$.

That's it. Whenever you have a collection of objects with a well-defined way to combine them that never takes you outside the collection itself, you have a group.

Example 1: Rotations in the plane.

All rotations in a *two-dimensional* plane about the origin by some angle θ form a group. The operation of "do one rotation, then do another," is equivalent to a single rotation by the sum of the angles:

$$R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2).$$

The identity is the rotation by 0 (do nothing), and the inverse of $R(\theta)$ is $R(-\theta)$.

Example 2: Complex phases.

The set of all complex numbers with unit magnitude,

$$U(1) = \{\, e^{i\theta} \mid \theta \in [0, 2\pi) \,\},$$

forms a group under ordinary multiplication. This group has *infinitely many* elements - one for each possible value of the continuous angle θ. Each element $e^{i\theta}$ represents a single rotation by an angle θ in the complex plane. The entire set of these elements - i.e. the group $U(1)$ - collectively describes *all* possible rotations about the origin in a flat two-dimensional plane. In other words, while a single $e^{i\theta}$ tells you how to rotate once, the full group $U(1)$ encodes the entire family of rotations that can be performed about the origin.

Example 3: Permutations.

If you have N distinct objects labeled $1, 2, \ldots, N$, all possible ways to rearrange them form the *symmetric group* S_N. Each element of S_N is a *permutation* - a rule that sends each label to another one. The group operation is simply "do one permutation, then another." The identity element is the permutation that leaves every object where it is. Every permutation has an inverse that undoes its action.

Unitary, Orthogonal, and Special Orthogonal

If a matrix M satisfies

$$MM^\dagger = 1,$$

it is said to be *unitary*, and thus M preserves inner products, so lengths and probabilities remain unchanged under its action. Here the dagger ($\dagger$) denotes the conjugate transpose: take the transpose of M and complex-conjugate each entry. If all entries of M are real, this reduces to $MM^T = 1$, and the matrix is called *orthogonal* instead. An orthogonal matrix cannot have linearly dependent rows or columns. Its columns (and rows) form an orthonormal set - each has unit length and they are all mutually perpendicular.

The set of all real orthogonal $N \times N$ matrices forms the **orthogonal group** $O(N)$. The group $O(N)$ includes all rotations and reflections that preserve the lengths of vectors in N-dimensional space. If we restrict to those with determinant $+1$ (pure rotations, no reflections), we obtain the **special orthogonal group** $SO(N)$. The group in *Example 1* is called $SO(2)$, the special orthogonal group in two dimensions.

1.2 Representations

A group becomes physically meaningful when it acts on something. In physics, that "something" is usually a state, a vector, or a field. Intuitively, a **representation** is a way to make the abstract elements of a group concrete by expressing them as matrices or linear transformations that act on a vector space. In other words, instead of thinking of $g \in G$ as an abstract symbol, we represent it by a matrix $\rho(g)$ that actually does something: it moves or mixes components of a vector.

Formally, a representation of a group G on a vector space V is a map

$$\rho : G \to \mathrm{GL}(V)$$

such that

$$\rho(g_1 g_2) = \rho(g_1)\rho(g_2)$$

for all $g_1, g_2 \in G$. This means that doing two group operations in sequence corresponds to applying the two matrices in sequence. $\mathrm{GL}(V)$ is the group of all invertible linear transformations of V (we will return to this and make it more precise in §7).

For example, if G is the group of rotations in space, then a representation might be a set of matrices that rotate vectors or fields in the same way. Different representations describe how different kinds of physical objects "feel" the same symmetry - a scalar stays unchanged, a vector rotates, and more complicated objects transform in their own characteristic ways.

1.3 Lie Groups

Some groups, like the permutations S_N, have elements that are distinct and separate from one another; you can list them one by one. Others, like the phase group $U(1)$, have elements that vary continuously as you change a parameter (the angle θ). Groups of this second kind are called **Lie groups**. Well... sort of.

More formally, a Lie group (pronounced "Lee") is a group that is also a differentiable manifold, meaning that group multiplication and taking inverses are smooth (differentiable) operations. In other words, a Lie group is a continuous family of transformations where you can move smoothly from one element to another, without any jumps or gaps. This smoothness lets us use geometric intuition - the kind we use for smooth surfaces - to study how the group behaves near the identity element.

1.4 From Groups to Lie Algebras

Every Lie group contains an identity[1] element I, and thus we can look at how the group behaves in its immediate neighborhood. Because the group is smooth, we can describe small transformations as tiny steps away from the identity. The directions you can move in - the independent ways to make infinitesimal changes - are called the **generators** of the group.

Mathematically, a small group element can be written as

$$g(\epsilon) = e^{i\epsilon^a T_a},$$

where the real parameters ϵ^a are small and the T_a are the generators. Each generator corresponds to one possible "direction" of continuous change within the group. The generators don't form a group by themselves, but together they describe how the group behaves infinitesimally close to the identity.

We exponentiate the generators because small transformations near the identity can be written as

$$g(\epsilon) \approx I + i\,\epsilon^a T_a,$$

where T_a specify the direction in which we move away from the identity element. They are the "infinitesimal" versions of the full group transformations. Exponentiating them is the natural way to combine an infinite number of such infinitesimal steps into a single, finite transformation while preserving the group composition law. For example, in the case of $U(1)$,

$$g(\epsilon_1)\, g(\epsilon_2) = e^{i\epsilon_1 T}\, e^{i\epsilon_2 T} = e^{i(\epsilon_1+\epsilon_2)T} = g(\epsilon_1 + \epsilon_2),$$

so exponentiation ensures that sequential small transformations add up to the correct finite one.

The set of all generators, along with the rule for how they combine, is called the **Lie algebra** of the group. In general, an **algebra** is a vector space equipped with a product - here, the product

[1] In general I denotes the group identity, but for matrix Lie groups - which are the ones we'll consider here - it is represented by the identity matrix.

is given by the **commutator**, which tells us how two generators combine to make another. In general, doing one infinitesimal transformation and then another is not the same as doing them in reverse. Their difference is captured by the commutator:

$$[T_a, T_b] = if^{abc}T_c.$$

The numbers f^{abc} are the **structure constants**; they tell us exactly how and when the basic transformations fail to commute and therefore encode the local geometry of the group.

For example, in three-dimensional space, the generators of rotations are the angular momentum operators J_x, J_y, and J_z. Each one generates an infinitesimal rotation about its corresponding axis. They satisfy

$$[J_x, J_y] = iJ_z,$$

(with $\hbar = 1$) which expresses the fact that rotating around x and then y is not the same as rotating around y and then x. The result involves J_z because the combined effect of those two small rotations is itself a rotation around the remaining axis, z.

In physics, most practical calculations - such as adding angular momenta or classifying particle spins - are done at the algebra level. The group describes the full, global symmetry, while the algebra tells us how the system responds to infinitesimal transformations - which is usually all we need in quantum theory.

1.5 Notation and Conventions

It's common to use different fonts (and cases) to distinguish between a Lie group and its Lie algebra. The group itself is written with an uppercase letter, such as G, while its algebra is written in lowercase Gothic script, $\mathfrak{g}$. Individual elements of the group are written in lowercase, like $g \in G$. For example, the Lie algebra of $SU(2)$ is denoted $\mathfrak{su}(2)$, and its elements - the generators $T_a \in \mathfrak{su}(2)$ - describe infinitesimal transformations. By exponentiating these generators,

$$g = e^{i\theta^a T_a},$$

we obtain actual group elements g belonging to the group G. The number of independent generators (the dimension of $\mathfrak{g}$) equals the number of continuous parameters needed to specify an element of G.

Physicists often include a factor of i so that the generators T_a are Hermitian (see §3.1 for a refresher). This makes physical sense because Hermitian operators correspond to observables with real eigenvalues, while the resulting group elements $e^{i\theta^a T_a}$ are unitary and therefore preserve probabilities. Mathematicians take the opposite approach: they define the algebra itself to consist of anti-Hermitian elements $X_a = iT_a$, so that e^X is automatically unitary without an extra i. Both conventions guarantee the same unitarity property - they just differ in whether the factor of i lives in the algebra or in the exponential. Either way, when you see $\mathfrak{g}$, think "the infinitesimal version of G" - the space of directions you can move away from the identity.

In physics, a Lie group represents a continuous symmetry of the laws of nature - a smooth transformation that may change the state of a system but leaves the underlying physical laws unchanged. Rotating a particle or shifting a field, for instance, produces a new configuration, yet the equations governing its behavior remain the same. The associated Lie algebra captures the infinitesimal version of these transformations, generated by operators T_a. Through *Noether's theorem*, each continuous symmetry implies a conserved quantity: performing a tiny symmetry transformation changes the system's state but not the action, leading to a conserved current J_a^μ. Each generator T_a therefore corresponds to one conserved quantity, and the number of independent generators equals the number of conserved charges. For example, in ordinary space,

there are three independent rotation generators - one about each axis - and Noether's theorem tells us these correspond to the three components of angular momentum.

This link between symmetry and conservation is one of the cornerstones of modern physics. Lie groups and their algebras provide the natural language for describing it.

2 Photons From $U(1)$

You may have heard that the Standard Model is organized by the symmetry group:

$$SU(3) \times SU(2) \times U(1).$$

Each factor corresponds to a different kind of fundamental interaction. But before diving into the more complicated pieces, it helps to start with the simplest one: $U(1)$.

2.1 Definitions

Formally, $U(1)$ is the group of all complex numbers with unit magnitude:

$$U(1) = \{\, e^{i\alpha} \mid \alpha \in [0, 2\pi) \,\}.$$

It is a one-dimensional Lie group representing rotations in the complex plane. As a manifold, $U(1)$ is topologically a circle, and as a group it is **isomorphic** to the circle group S^1:

$$U(1) \cong S^1.$$

To say that two groups are *isomorphic* means that they have the same structure: there exists a one-to-one correspondence between their elements that preserves how they combine. In other words, if $f : G \to H$ is an isomorphism, then $f(g_1 g_2) = f(g_1) f(g_2)$ for all $g_1, g_2 \in G$.

For $U(1)$ and the circle group $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$, we can define such a map very naturally:

$$f : U(1) \longrightarrow S^1, \qquad f(e^{i\alpha}) = (\cos\alpha,\, \sin\alpha).$$

This map is one-to-one (each element of $U(1)$ maps to a unique point on S^1), onto (every point on S^1 comes from some element of $U(1)$), and satisfies

$$f(e^{i\alpha_1} e^{i\alpha_2}) = (\cos(\alpha_1 + \alpha_2),\, \sin(\alpha_1 + \alpha_2)) = f(e^{i\alpha_1}) \cdot f(e^{i\alpha_2}),$$

where multiplication on S^1 corresponds to angle addition. Hence f is a group isomorphism, and $U(1)$ and S^1 are two equivalent ways of describing the same underlying symmetry: the group of rotations by any angle in a plane.

Each group element corresponds to a rotation by an angle α, and the group operation is just addition of angles modulo 2π:

$$e^{i\alpha_1} \cdot e^{i\alpha_2} = e^{i(\alpha_1 + \alpha_2)}.$$

The corresponding Lie algebra, denoted $\mathfrak{u}(1)$, consists of all purely imaginary numbers $i\theta$ with $\theta \in \mathbb{R}$. It is a one-dimensional real vector space, whose generator we can take to be i itself.

By the way, in general relativity we'll often hear the word **diffeomorphism**. That simply means a smooth, invertible map between manifolds whose inverse is also smooth. Intuitively, a smooth deformation that bends or stretches space without tearing or gluing it.

More broadly, mathematicians use the suffix "$-$morphism" to denote a structure-preserving map, with the precise notion of "structure" depending on context:

- a *homomorphism* preserves the algebraic operations of a structure, e.g. for groups $f(g_1 g_2) = f(g_1) f(g_2)$; it respects how elements combine,
- an *isomorphism* is a homomorphism that is bijective (one-to-one and onto), meaning it preserves and reflects all algebraic relations between elements,
- a *homeomorphism* is a bijective continuous map with a continuous inverse, preserving the topological structure of spaces rather than algebraic operations. It captures when two spaces have the same overall "shape," even if they lack smooth or geometric structure,
- a *diffeomorphism* is a bijective smooth map with a smooth inverse, preserving not just the topological shape but also the differentiable structure. It requires the spaces to have well-defined notions of smoothness.

All of these ideas capture the same spirit: two objects can be considered "the same" if there exists a map between them that preserves the relevant kind of structure; more generally, a *morphism* in **category theory** is any structure-preserving arrow between objects of a given category, abstracting all of the above notions.

Also, the notation S^n refers to the set of points in $(n+1)$-dimensional space that are a unit distance from the origin. So S^1 is the unit circle in $\mathbb{R}^2$, S^2 is the surface of an ordinary sphere in $\mathbb{R}^3$, and S^3 is the three-dimensional "surface" of a four-dimensional ball. Each S^n is called an n-sphere, even though only S^2 looks like a literal sphere in our usual three dimensions.

2.2 What does this have to do with our Lagrangian?

In quantum theory, multiplying a field ψ by an overall phase

$$\psi \longrightarrow e^{i\alpha}\,\psi$$

doesn't change any observable quantities. Think about it: the only things we ever measure are magnitudes squared.

$$|\psi|^2 \to |e^{i\alpha}\psi|^2 = e^{-i\alpha}e^{i\alpha}|\psi|^2 = |\psi|^2.$$

The phase $e^{i\alpha}$ drops out of every measurable quantity. Formally, that means that ordinary quantum mechanics has a $U(1)$ symmetry: it is invariant under global phase rotations.

Now, by Noether's theorem, every continuous symmetry corresponds to a conserved quantity. So this invisible, undetectable $U(1)$ phase freedom must be tracking something real... some conserved quantity.

But now we face a puzzle. If a **global $U(1)$ symmetry** (same phase everywhere) leads to a conserved quantity, what happens if we allow the phase to vary from point to point in spacetime? Shouldn't physics be able to handle that, too? After all, why should a symmetry only hold when every observer uses exactly the same phase convention everywhere? Figuring out how to make a $U(1)$ symmetry local will turn out to be the key that unlocks electromagnetism.

First, a geometric picture

Imagine spacetime as a smooth fabric, where at every point x we've attached a tiny circle-an abstract space representing all possible complex phases of the field. The field's value at that point can be pictured as a little arrow on this circle, specified by an angle $\theta(x)$, the local phase. A $U(1)$ transformation corresponds to rotating each of these circles by some angle. If the same rotation is applied everywhere, we have a global $U(1)$ symmetry; every phase wheel in the universe turns

in unison. If we allow the rotation angle to vary from point to point, those little circles will begin to twist independently. Notationally, if we allow the angle to depend on position,

$$\psi(x) \;\longrightarrow\; e^{i\alpha(x)}\,\psi(x)$$

This says each tiny dial can be turned by a *different* amount at each point of spacetime. But then a new problem appears: how do we compare the phase at x with the phase at a nearby point $x + \mathrm{d}x$? We need a rule that tells us how to "carry" the dial's hand from one point to a neighbor in a consistent way.

2.3 Gauge Fields and Covariant Derivatives

For a complex field ψ, a global $U(1)$ phase rotation is

$$\psi \;\longrightarrow\; e^{i\alpha}\,\psi$$

Because α is the same everywhere, ordinary derivatives behave nicely:

$$\partial_\mu\psi \;\longrightarrow\; e^{i\alpha}\,\partial_\mu\psi,$$

so familiar expressions in our Lagrangian like $|\partial_\mu\psi|^2$ are unchanged.

Now, since a *local* symmetry lets the transformation depend on position,

$$\psi(x) \;\longrightarrow\; e^{i\alpha(x)}\,\psi(x),$$

When we promote the phase rotation α to depend on spacetime, a new term appears when we take derivatives:

$$\partial_\mu(e^{i\alpha(x)}\psi(x)) = e^{i\alpha(x)}(\partial_\mu\psi + i(\partial_\mu\alpha)\,\psi)\,.$$

The extra term $i(\partial_\mu\alpha)\psi$ breaks the simple transformation rule that held under global $U(1)$ symmetry. In other words, the ordinary derivative $\partial_\mu\psi$ no longer "transforms like" ψ itself. This matters because derivatives are what encode dynamics and interactions. If we want our equations of motion and conserved quantities to remain meaningful under local symmetry transformations, then the way we take derivatives must also respect that symmetry.

To restore this symmetry, we introduce a new field $A_\mu(x)$ and define a **covariant derivative**:

$$D_\mu\psi \;\equiv\; (\partial_\mu + i\,e\,A_\mu(x))\psi.$$

If we let the gauge field A_μ transform as

$$A_\mu(x) \;\longrightarrow\; A_\mu(x) - \frac{1}{e}\,\partial_\mu\alpha(x)\;,$$

then the unwanted term cancels perfectly, and the covariant derivative now transforms as neatly as the field itself:

$$D_\mu\psi \;\longrightarrow\; e^{i\alpha(x)}\,D_\mu\psi.$$

You might ask: how are we allowed to just "add" a new field? The guiding principle here is symmetry. Once we insist that a global $U(1)$ phase be promoted to a *local* one, ordinary derivatives spoil invariance. To restore it, we must introduce a **connection**: a mathematical object (here, A_μ) that defines how to compare fields at nearby points in a way consistent with the symmetry. Crucially, this gauge symmetry is a *redundancy* of description rather than a new physical motion: many different pairs (ψ, A_μ) represent the same situation. Intuitively, the gauge potential A_μ is like a choice of reference grid; we can redraw the grid but the underlying landscape is unchanged. This is why we are allowed to *fix a gauge*. Common choices include the

Lorenz gauge $\partial_\mu A^\mu = 0$ (*note: Ludvig Lorenz, not Hendrik Lorentz*), which treats space and time symmetrically, and the **Coulomb gauge** $\nabla \cdot \mathbf{A} = 0$.

It should also be said that Noether's theorem links continuous symmetries to conserved charges - but only for *global* symmetries (those with a single parameter α the same everywhere). When $\alpha \to \alpha(x)$ becomes a *local* symmetry, the "symmetry" acts redundantly on the description of the fields. Thus, local gauge invariance does not yield new conserved charges. The requirement that physical states be invariant under these *local* transformations instead leads to constraints - the most familiar being Gauss's law.

2.4 Quantizing A_μ

Once A_μ is introduced to preserve local $U(1)$ symmetry, it stops being a bookkeeping trick and becomes a genuine field living on spacetime once we give it dynamics. If we move a field first in the x^μ direction and then in the x^ν direction, the total phase change depends on A_μ and A_ν. If the result depends on the order of these infinitesimal moves - that is, going μ then ν gives a different phase than going ν then μ - the connection must possess a kind of "twist" or curvature. That curvature is measured by the field strength

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

When $F_{\mu\nu} = 0$, one can move around any infinitesimal loop and return with exactly the same phase, meaning that there is no physical field, only a choice of gauge (on simply connected domain). When $F_{\mu\nu} \neq 0$, the phase fails to return to its original value, giving a genuine physical effect. In electromagnetism, this corresponds to the presence of electric and magnetic fields.

To let the field actually move, we write

$$\mathcal{L}_{\text{gauge}} = -\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu},$$

which, in the classical limit, ties back to the familiar $\vec{E}$ and $\vec{B}$ fields

$$-\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} = \tfrac{1}{2}(\vec{E}^{\,2} - \vec{B}^{\,2}),$$

as shown in Eq. (8.25) of Schwartz.

Once A_μ exists as part of the theory, it can ripple. Those ripples, small waves in the gauge field moving through spacetime, are quantized in a procedure given by Schwartz Eq. (8.28), where the Fourier-expanded modes of A_μ are promoted to creation and annihilation operators. Those quanta are the **photons**.

Allow each point in spacetime to choose its own phase $\Rightarrow$ *the universe must contain photons.*

Isn't that crazy? Take a moment and appreciate what just happened: we didn't *put* light into the theory by hand. All we did was give every point in space its own phase, and the universe had no choice but to produce electromagnetism. More precisely, localizing the symmetry *requires* introducing the connection A_μ; giving it the kinetic term above *promotes* it to a propagating field whose quanta are photons.

We can tie this back to our group theory discussion. The group $U(1)$ has a one-dimensional Lie algebra with a single generator. Making the symmetry local introduces one gauge connection to preserve invariance, and quantizing its dynamical degrees of freedom produces one gauge boson: the photon.

Polarization

To understand where **polarization** comes from, recall that the gauge potential A_μ has four components. At first glance, one might expect four independent oscillations, but an electromagnetic wave in vacuum only carries two physical degrees of freedom. The others are removed either by constraints or by gauge redundancy.

The time component A_0 does not describe a propagating field. Its equation of motion isn't dynamical (it has no time derivatives) so instead of evolving freely, it enforces Gauss's law $\nabla \cdot \mathbf{E} = 0$ (in vacuum). In Coulomb gauge this appears as an instantaneous constraint; in covariant gauges it still does not add a physical propagating degree of freedom. Even among the spatial components $\mathbf{A}$, not all are independent. For a plane wave

$$A_\mu(x) = \varepsilon_\mu e^{-ik\cdot x},$$

with $k^2 = 0$, Maxwell's equations together with the Lorenz gauge condition $\partial_\mu A^\mu = 0$ (see Schwartz ch. 8) lead to

$$k_\mu \varepsilon^\mu = 0.$$

Physically, this tells us that the electromagnetic wave's oscillations are **transverse**: the electric and magnetic fields wiggle in directions *perpendicular* to the direction of propagation. If the wave moves along $\hat{\mathbf{k}}$, its polarization vector $\boldsymbol{\varepsilon}$ lies in the plane orthogonal to $\hat{\mathbf{k}}$. These perpendicular directions are exactly what we call the two polarization states of light. The wave doesn't "shake" forward and backward along its line of travel; it only vibrates sideways, consistent with both the gauge symmetry and Maxwell's constraints.

So although A_μ begins with four components, the combination of Gauss's law, the gauge condition, and the transversality constraint leaves only *two* independent, physical degrees of freedom. These correspond to the two possible transverse polarizations of a photon. This counting applies specifically to a *massless* spin-1 field like the photon. For massive vector fields the situation changes, but for light in vacuum, gauge invariance ensures that only these two transverse modes can carry energy and information as radiation.

3 The Weak Force: SU(2)

For $SU(2)$, the story is the same but richer. By definition, SU(2) is the set of 2×2 unitary matrices with unit determinant,

$$U \in \mathrm{SU}(2) \quad \Longleftrightarrow \quad U^\dagger U = I, \;\; \det U = 1,$$

where I is the 2×2 identity matrix. The letter "S" stands for **special**, indicating the additional condition $\det U = 1$ that restricts the group to transformations with unit determinant.

A simple way to enforce both conditions is to write any element as an exponential,

$$U = e^{\,iX},$$

and then ask what properties X must have. If we take $U = e^{\,iX}$, then $U^\dagger = (e^{\,iX})^\dagger = e^{-\,iX^\dagger}$. For $U^\dagger U$ to equal the identity (unitarity), the exponents must cancel[2], which happens only if $X^\dagger = X$. In other words, X must be Hermitian.

[2]In general, $e^{A^\dagger} e^{-A} = I$ does not imply $A^\dagger = A$ by exponent addition (the exponential map isn't injective, and the Baker–Campbell–Hausdorff formula matters).

3.1 Quantum Review

Just for completeness, it might be useful to review unitarity, hermiticity, observables, and traces.

A measurable quantity (like energy or spin) has the same possible outcomes no matter how you describe the system - whether you write the wavefunction in position space, momentum space, or any rotated coordinate system. For any linear operator X acting on a vector space, if there exists a vector v such that Xv is proportional to v itself, we call v an **eigenvector** and the proportionality constant λ an **eigenvalue**. This relationship

$$Xv = \lambda v$$

tells us that v is a "special direction" in which the operator doesn't rotate or mix with others; it only stretches or shrinks it by a factor λ.

In quantum mechanics, the state of a system is represented by a vector $|v\rangle$ (in Hilbert space), and a physical quantity that you can observe is represented by an operator X. When you measure that observable, the system collapses into one of the operator's eigenstates - and the value you observe is the corresponding eigenvalue. So the measurement postulate says:

$$X|v\rangle = \lambda|v\rangle \quad \Rightarrow \quad \text{measurement result } = \lambda.$$

Only in such "eigenstates" is the observable well-defined - the system has a definite value for that quantity. If the state is not an eigenstate, measurement outcomes are uncertain and occur with probabilities determined by how much the state overlaps with each eigenvector.

A Hermitian matrix (or operator) $X = X^\dagger$ has the special property that all its eigenvalues are real and its eigenvectors form an orthonormal basis. Mathematically, if $Xv = \lambda v$, then taking the Hermitian conjugate gives

$$v^\dagger X^\dagger = \lambda^* v^\dagger.$$

Using $X = X^\dagger$ then implies $\lambda = \lambda^*$, so the eigenvalue λ is real.

The eigenvalues of any operator are basis independent, but for Hermitian operators they are also guaranteed to be real, which is why such operators represent observables in quantum mechanics. The operator X itself does not depend on the basis we choose - only the numerical values of its matrix elements change. Its eigenvalues, by contrast, are intrinsic properties of the operator: they remain the same under any unitary change of basis. Let's now see explicitly and tie back into the $\det U = 1$ condition for SU(2). A general 2×2 Hermitian matrix has real diagonal entries and complex-conjugate off-diagonal entries:

$$X = \begin{pmatrix} a & c - id \\ c + id & b \end{pmatrix}, \qquad a, b, c, d \in \mathbb{R}.$$

In general, a transformation changes both the length and direction of a vector, but along certain special directions (its eigenvectors) the vector only stretches or shrinks without turning. Rewriting $Xv = \lambda v$ as $(X - \lambda I)v = 0$, we see that such a solution exists when $\det(X - \lambda I) = 0$, so the determinant must vanish at each eigenvalue, implying $\det(X - \lambda I) = (\lambda - \lambda_1)(\lambda - \lambda_2)$.

We find the characteristic polynomial

$$\det(X - \lambda I) = \lambda^2 - (a + b)\lambda + (ab - c^2 - d^2).$$

If the eigenvalues are λ_1 and λ_2, then

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2.$$

The individual diagonal entries a and b depend on the choice of basis, so they are not themselves invariant quantities. However, matching the coefficients of $-\lambda$ in these two expressions shows $\lambda_1 + \lambda_2 = a + b$, and likewise $\lambda_1\lambda_2 = \det X$. Thus, the sum of the diagonal entries of a square matrix is equal to $(\lambda_1 + \lambda_2)$ and remains the same under any change of basis. In other words, while the diagonal elements may change, their total (and hence the sum of the eigenvalues) does not. We define the **trace**, $\mathrm{tr}X$, as the sum of the eigenvalues of X.

In linear algebra, the trace measures the total "weight" of a transformation - it tells you how much the transformation, on average, stretches or shrinks space. Physically, traces matter because they pick out the parts of a transformation that act equally on all components: those global, indistinguishable shifts that don't change relative orientations.

In quantum mechanics, traces appear constantly when describing statistical mixtures and expectation values. The key object here is the **density matrix** (or density operator) ρ, which encodes the full statistical state of a system. For a pure state $|\psi\rangle$, it is defined as $\rho = |\psi\rangle\langle\psi|$; for a mixed state (an ensemble of states $|\psi_i\rangle$ occurring with probabilities p_i) we write $\rho = \sum_i p_i \, |\psi_i\rangle\langle\psi_i|$. The expectation value of any observable O in such a state is then

$$\langle O\rangle = \mathrm{tr}(\rho\, O),$$

which generalizes the familiar expression $\langle\psi|O|\psi\rangle$ and, crucially, does not depend on the basis used to represent the system. The trace here plays the role of an average over all possible outcomes, much like a sum over probabilities in classical statistics. When dealing with composite systems $A\otimes B$, the reduced state of subsystem A is obtained by taking a **partial trace** over B:

$$\rho_A = \mathrm{tr}_B(\rho_{AB}),$$

which corresponds to "forgetting" the degrees of freedom of B. This is precisely how we describe entanglement and decoherence. Normalization of probabilities appears through the simple condition $\mathrm{tr}(\rho) = 1$.

3.2 Pauli Matrices and the $\mathfrak{su}(2)$ Algebra

For any square matrix M, the determinant of its exponential is related to its trace:

$$\det\Big(e^{M}\Big) = e^{\,\mathrm{tr}\,M}.$$

Remember that the determinant is the product of a matrix's eigenvalues, while the trace is their sum. If M has eigenvalues $\{\lambda_i\}$, then e^M has eigenvalues $\{e^{\lambda_i}\}$. So

$$\det\Big(e^{M}\Big) = \prod_i e^{\lambda_i} = e^{\sum_i \lambda_i} = e^{\,\mathrm{tr}\,M}.$$

Applying this to our case, with $U = e^{\,iX}$,

$$\det U = e^{\,i\,\mathrm{tr}X}.$$

Requiring $\det U = 1$ means

$$e^{\,i\,\mathrm{tr}X} = 1 \quad \Rightarrow \quad \mathrm{tr}X \in 2\pi\mathbb{Z}.$$

For elements smoothly connected to the identity, we take the branch with $\mathrm{tr}X = 0$. (In SU(2), this subtlety doesn't cause trouble because the entire group is connected. Every element can be reached smoothly from the identity. But in some other groups, like $\mathrm{O}(n)$, which includes both rotations ($\det = 1$) and reflections ($\det = -1$), only part of the group (the rotation subgroup $\mathrm{SO}(n)$)is connected to the identity.)

Requiring the trace of X to vanish removes any uniform rescaling or global phase factor, ensuring that $\det U = 1$. In this sense, a *traceless* generator produces a transformation that is purely rotational in the internal space; it changes orientation without introducing any overall rescaling. Looking back at our general Hermitian matrix, the traceless condition, $\mathrm{tr}X = a + b = 0$, fixes $b = -a$, leaving three real parameters a, c, d. Substituting $b = -a$ gives

$$X = \begin{pmatrix} a & c - id \\ c + id & -a \end{pmatrix}.$$

We can then separate this into real coefficients multiplying three fixed matrices:

$$X = a \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + c \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + d \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

These three independent matrices form a natural basis for all Hermitian traceless 2×2 matrices. Up to an overall normalization factor, the three fixed matrices we obtained are exactly the Pauli matrices!

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

However, when we talk about the Lie algebra $\mathfrak{su}(2)$ - the algebra of infinitesimal generators of the group SU(2) - it is customary to introduce a convenient normalization. We define the generators as

$$T^a = \frac{\sigma^a}{2}, \qquad a = 1, 2, 3.$$

This factor of $\frac{1}{2}$ is chosen so that the algebra takes its simplest, most symmetric form. To see this, recall that the Pauli matrices satisfy

$$[\sigma^a, \sigma^b] = 2i\, \epsilon^{abc}\, \sigma^c,$$

where ϵ^{abc} is the Levi-Civita symbol, defined by

$$\epsilon^{123} = +1, \qquad \epsilon^{abc} = -\epsilon^{bac}, \qquad \epsilon^{abc} = 0 \text{ if any two indices are equal.}$$

It is a completely antisymmetric tensor that encodes the structure constants of $\mathfrak{su}(2)$ and expresses how the basis elements "rotate" into one another under commutation. Defining the generators as $T^a = \frac{1}{2}\sigma^a$ therefore rescales by a factor of $\frac{1}{2}$

$$[T^a, T^b] = i\, \epsilon^{abc}\, T^c,$$

which is the standard commutation relation of the $\mathfrak{su}(2)$ Lie algebra.

3.3 Geometric Picture and the Bloch Sphere

For $U(1)$, each point in spacetime carried a little circle, a single phase angle $\theta(x)$ that could spin independently. To get a picture for SU(2), recall that an element is a 2×2 complex matrix that preserves lengths and has determinant one. Such a matrix can be labeled by two complex numbers (α, β) that satisfy

$$|\alpha|^2 + |\beta|^2 = 1.$$

Because each complex number has two real components, (α, β) lives in four real dimensions, and the equation above carves out a *three-dimensional surface* inside that space. This surface is called a **3-sphere**, written S^3, in the same way that ordinary points (x, y, z) with $x^2 + y^2 + z^2 = 1$ form the familiar 2-sphere S^2. Another way to see this is that the three Pauli matrices define

the independent directions in which you can move while remaining on the S^3 surface[3]. Each generator represents a tiny rotation in one of these directions, and by smoothly combining them, you can reach any point on the three-sphere.

When we define the three-sphere,

$$S^3 = \{(x, y, z, w) \in \mathbb{R}^4 \mid x^2 + y^2 + z^2 + w^2 = 1\},$$

the coordinate w plays the role of an additional spatial dimension beyond the three we can visualize. (By the way, this space can also be described in terms of **quaternions**, an extension of complex numbers with three independent imaginary parts: $q = a + b\,\mathbf{i} + c\,\mathbf{j} + d\,\mathbf{k}$)

Rearranging the defining equation for S^3 gives

$$w^2 = 1 - (x^2 + y^2 + z^2).$$

For w to be a real number, the right-hand side must be non-negative:

$$1 - (x^2 + y^2 + z^2) \geq 0 \quad \Longrightarrow \quad x^2 + y^2 + z^2 \leq 1.$$

This inequality means that for any given value of w, the allowed points (x, y, z) must lie inside a sphere of radius $\sqrt{1 - w^2}$. But this radius is only real if $-1 \leq w \leq 1$. If w went outside this range, $1 - w^2$ would become negative and there would be no real solutions. Thus the three-sphere "lives" entirely within the interval $-1 \leq w \leq 1$.

For each fixed value of w, the remaining coordinates satisfy

$$x^2 + y^2 + z^2 = 1 - w^2,$$

which is the equation of an ordinary two-sphere (the surface of a 3D ball) in three dimensions, with radius $\sqrt{1 - w^2}$.

- When $w = 0$, the radius is 1: we see a full unit sphere in 3D space.
- As w increases toward $+1$ (or decreases toward -1), the radius $\sqrt{1 - w^2}$ shrinks.
- At the endpoints $w = \pm 1$, the radius becomes zero: the sphere collapses to a single point.

As w smoothly varies from -1 to $+1$, these 3D spherical slices expand and contract continuously. A good way to visualize this is to imagine a three-dimensional ball whose size varies smoothly in "time," where we treat time as a stand-in for the hidden fourth spatial direction. Each instant of this motion represents one particular value of w, and the entire evolution together forms the full four-dimensional surface of the three-sphere S^3.

Group Manifold vs. State Vectors

Okay, so we have that the equation

$$x^2 + y^2 + z^2 + w^2 = 1$$

describes the *topology* of the space of SU(2) elements: as a manifold, $\mathrm{SU}(2) \cong S^3$. This is the geometry of the *group* U itself (the set of 2×2 unitary matrices with $\det U = 1$). In many applications, however, we are not only interested in the group elements U, but in how they act on *state vectors* (spinors)

$$|\psi\rangle = \begin{pmatrix} a \\ b \end{pmatrix}, \qquad |a|^2 + |b|^2 = 1.$$

[3]Technically they span the tangent space at the identity; the whole S^3 is obtained by exponentiation

We will go over why spinors take this particular form later. For now, just notice that the set of all normalized spinors also forms a three-sphere S^3 inside $\mathbb{C}^2 \cong \mathbb{R}^4$. But there is an important physical identification: two spinors that differ by an *overall complex phase* represent the same physical state,

$$|\psi\rangle \;\sim\; e^{i\phi}\,|\psi\rangle, \qquad \phi \in [0, 2\pi).$$

Geometrically, this means that through each point of the spinor's S^3 there exists a circle S^1 (the "phase direction") for which moving along this circle changes only the overall phase and leaves all observables unchanged.

Therefore, while the *space of normalized representatives* is S^3, the *space of physically distinct states* is obtained by identifying all points related by that phase circle:

$$\text{(physical state space)} \;=\; S^3/S^1 \;=\; S^2.$$

In words: there is a circular S^1 freedom along S^3 that does not affect measurements; removing (quotienting by) this redundant degree of freedom leaves a two-sphere S^2. This is the usual **Bloch sphere**: each point on S^2 labels one physically distinct spin orientation. There is, however, a subtlety here: while S^2 represents the space of states, it is *not* itself a group. The group SU(2) acts on this sphere, but the sphere is only the geometric stage on which those rotations occur. We will discuss this distinction in §5.1.

Either way, for visualization purposes, imagine that at every spacetime point x you attach a small, ordinary three-dimensional sphere. A local SU(2) transformation then acts as an ordinary spatial rotation of this attached sphere: each point in spacetime can independently rotate its own internal sphere by some angle and about some chosen axis. When we demand local SU(2) invariance (allowing these rotations to vary smoothly from point to point) we must introduce three compensating gauge fields, one for each generator. Just as local phase freedom in $U(1)$ forces the existence of a single gauge field A_μ (and thus photons), local orientation freedom in SU(2) forces three gauge fields $W^1_\mu, W^2_\mu, W^3_\mu$, which turn out to be the $W^\pm$ and Z^0 bosons, where I'm ignoring some subtleties[4] in the spontaneous symmetry breaking of the Higgs.

3.4 Non-Abelian Consequences

The gauge field itself takes values in the Lie algebra:

$$A_\mu = A^a_\mu\, T^a,$$

where the T^a are the generators introduced above. The covariant derivative acting on a field then becomes

$$D_\mu = \partial_\mu + i\, g\, A^a_\mu\, T^a.$$

Here I'm using g for a generic coupling (instead of e for QED). In the Abelian case (like electromagnetism), the generator acts trivially. But, in a non-Abelian theory, the generators T^a do *not* commute (that is the very definition of a **non-Abelian** group):

$$[T^a, T^b] = i\, f^{abc}\, T^c,$$

where f^{abc} are the *structure constants* of the Lie algebra ($f^{abc} = \epsilon^{abc}$ for SU(2)). Because of this noncommutativity, when we compute the commutator of two covariant derivatives,

$$[D_\mu, D_\nu] = i\, g\, (\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)\, T^a - g^2 [A^a_\mu T^a,\, A^b_\nu T^b],$$

[4] In reality, $SU(2)_L$ gives three gauge fields $W^{1,2,3}_\mu$. After electroweak symmetry breaking in $SU(2)_L \times U(1)_Y$, $W^{1,2}$ combine into $W^\pm$, and W^3 mixes with the hypercharge field B_μ to produce the Z^0 and the photon γ.

the last term no longer vanishes. Using the commutation relation of the generators, this becomes

$$[D_\mu, D_\nu] = i\, g\, (\partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g\, f^{abc}\, A^b_\mu A^c_\nu) T^a.$$

By definition, the field strength tensor $F_{\mu\nu}$ is the object that measures the failure of covariant derivatives to commute:

$$[D_\mu, D_\nu] = i\, g\, F^a_{\mu\nu}\, T^a.$$

So we identify:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g\, f^{abc} A^b_\mu A^c_\nu.$$

The extra, nonlinear term proportional to f^{abc} arises precisely because the generators T^a fail to commute-this is what distinguishes a non-Abelian gauge field from an Abelian one.

The non-Abelian nature of the gauge field has profound consequences. Because the gauge potentials A^a_μ themselves carry the group index a and interact through the term $g\, f^{abc} A^b_\mu A^c_\nu$, the gauge bosons are not independent. They can couple to one another, leading to self-interactions that make the theory highly nonlinear. This is the key feature distinguishing **Yang-Mills** theories (QFT with e.g. SU(2) or SU(3)) from the Abelian U(1) case: the force carriers themselves participate in the dynamics, enabling phenomena like confinement.

We have already associated the Abelian U(1) symmetry with electromagnetism, whose gauge boson is the photon. Naturally, one may ask: what physical interaction corresponds to the non-Abelian SU(2) symmetry? Historically, SU(2) was introduced to describe the *weak nuclear force* - the interaction responsible for processes such as **beta decay** (e.g., $n \to p + e^- + \bar{\nu}_e$), where weak interactions change flavor.

4 The Strong Force: $SU(3)$

Here we turn the dial to three components. The symmetry group SU(3) consists of all 3×3 unitary matrices with determinant one:

$$U \in \mathrm{SU}(3) \quad \Longleftrightarrow \quad U^\dagger U = I, \qquad \det U = 1.$$

It therefore describes rotations in an abstract three-dimensional complex space that preserve lengths and orientations.

A general 3×3 complex matrix has

$$3 \times 3 = 9 \text{ complex entries} = 18 \text{ parameters}.$$

Imposing unitarity,

$$U^\dagger U = I,$$

requires each row (and column) of U to be orthonormal. Each row must have unit length,

$$|u_{i1}|^2 + |u_{i2}|^2 + |u_{i3}|^2 = 1, \quad i = 1, 2, 3,$$

giving 3 real conditions. Distinct rows must be orthogonal,

$$u^*_{i1} u_{j1} + u^*_{i2} u_{j2} + u^*_{i3} u_{j3} = 0, \quad i \neq j.$$

There are three independent pairs $(1,2), (1,3), (2,3)$. Each of these complex equations gives two real conditions (its real and imaginary parts must vanish), contributing $3 \times 2 = 6$ real constraints.

Thus unitarity enforces

$$3 + 6 = 9 \text{ real constraints},$$

reducing the count from 18 to $18 - 9 = 9$ independent real parameters.

These 9 parameters describe the group $U(3)$, the set of all 3×3 unitary matrices. But for $SU(3)$, the condition

$$\det U = 1$$

removes one additional degree of freedom corresponding to an overall global phase. These eight independent directions correspond to the eight generators of SU(3), represented by the Gell-Mann matrices:

$$T^a = \frac{\lambda^a}{2}, \qquad a = 1, \ldots, 8,$$

where the λ^a are

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

$$\lambda^4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \qquad \lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \qquad \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Geometrically, this extends the familiar SU(2) picture. In the SU(2) case, a single two-component complex state can be visualized as a point on a Bloch sphere S^2. For SU(3), we introduce the labels red, green, and blue to distinguish the internal components. We can imagine three such "color" directions, but the group does not act on them independently - it mixes them in a single higher-dimensional space. A rotation in this space can tilt one color direction while counter-tilting another, keeping the total internal vector normalized and its overall phase irrelevant.

The eight generators of SU(3) describe the possible motions on this combined surface. The collection of such directions collapses to a **complex projective space**, where each point represents all complex vectors that differ only by an overall phase. Complex projective spaces are the natural arenas for quantum states: they keep track of relative orientations between components while ignoring overall normalization and phase. For SU(3), this space is called $\mathbb{CP}^2$, a four-dimensional curved manifold that generalizes the Bloch sphere of SU(2).

The group SU(3) turns out to be the gauge symmetry underlying the *strong nuclear force*, also known as quantum chromodynamics (QCD). Its gauge bosons are the *gluons*, which mediate interactions between quarks by exchanging color charge. Gluons themselves carry color and thus interact with one another. These self-interactions are so strong at low energies that quarks and gluons cannot exist freely (*confinement*), so we observe them only inside composite particles such as protons, neutrons, and mesons.

5 SU(2), SO(3), and Double Covering

Let's go back to the subtlety of the Bloch sphere. At this point it is important to notice that while both $SU(2)$ and the sphere of normalized spinors have the same underlying shape (S^3),

once we quotient out by the phase S^1 to get the Bloch sphere S^2, *we no longer have a group.* The reason is simple: once we quotient out by the phase, we have lost the algebraic structure that made $SU(2)$ a group. There is no meaningful way to "multiply" two points on the Bloch sphere and stay within the sphere such that the order of multiplication is well defined or preserved. The Bloch sphere represents *states*, not *symmetries*. It is the space *on which* the symmetry group acts, not a symmetry group itself.

5.1 The Bloch Sphere

Even though S^2 is not a group, it inherits a beautiful property from $SU(2)$: the group acts **transitively** on it. That is, for any two points on the Bloch sphere, there exists some element of $SU(2)$ that rotates one state into the other. This is exactly what we mean when we say that $SU(2)$ is the group of spin rotations: it acts as the symmetry group of the sphere of spin directions. We can connect $SU(2)$ to something more familiar: the group $SO(3)$, which consists of all real 3×3 rotation matrices acting on ordinary spatial vectors. Both groups describe rotations, but they act on different kinds of objects. $SO(3)$ acts on *"arrows"* in three-dimensional space, while $SU(2)$ acts on *spinors*, the two-component complex objects that describe quantum spin.

To make the connection precise, we introduce the idea of a **group homomorphism**. A group homomorphism is simply a rule that associates to each element of one group a corresponding element of another group in a way that respects composition:

$$f(U_1 U_2) = f(U_1)\, f(U_2) \quad \text{for all } U_1, U_2.$$

In other words, doing two $SU(2)$ transformations in sequence and then mapping to $SO(3)$ is the same as mapping each one individually and then composing the resulting spatial rotations.

There exists such a homomorphism between $SU(2)$ and $SO(3)$,

$$f : SU(2) \longrightarrow SO(3),$$

and we can define it concretely by looking at how $SU(2)$ acts on the Pauli matrices. Given any unit vector $\vec{n}$ in $\mathbb{R}^3$, we can associate to it the operator

$$\vec{\sigma} \cdot \vec{n} = \sigma_x n_x + \sigma_y n_y + \sigma_z n_z.$$

Now take any element $U \in SU(2)$. We can see how it acts on the operator $\vec{\sigma}\cdot\vec{n}$ by conjugating it:

$$U\,(\vec{\sigma}\cdot\vec{n})\,U^\dagger.$$

Because U is unitary, this operation does not change the overall form of the operator-it simply mixes the Pauli matrices among themselves. The result can therefore be written as another linear combination of the same three matrices:

$$U\,(\vec{\sigma}\cdot\vec{n})\,U^\dagger = \vec{\sigma}\cdot\vec{n}\,'.$$

This defines a new vector $\vec{n}\,'$ in $\mathbb{R}^3$, obtained by rotating the original vector $\vec{n}$ to some new direction:

$$\vec{n}\,' = R\,\vec{n}, \qquad R \in SO(3).$$

In other words, every $SU(2)$ matrix U determines a unique rotation R in real three-dimensional space, according to how it moves the Pauli matrices around under conjugation. We can therefore define a map

$$f : SU(2) \longrightarrow SO(3), \qquad f(U) = R,$$

which associates each U with the corresponding spatial rotation R that it produces.

The key point is that this mapping is *onto* - every possible spatial rotation R comes from some U - but it is not *one-to-one*. Two distinct matrices in $SU(2)$, differing only by an overall sign,

$$U \text{ and } -U,$$

produce exactly the same rotation R in $SO(3)$. Therefore, $SU(2)$ is said to be a **double cover** of $SO(3)$:

$$SO(3) \cong SU(2)/\{\pm I\}.$$

Geometrically, this means that the three-sphere S^3 of $SU(2)$ "wraps around" the space of ordinary rotations twice. A continuous path in $SU(2)$ corresponding to a 2π rotation does not bring you back to the same group element, but rather to its negative $-U$. Only after a full 4π rotation do you return to your starting point.

This is the origin of the familiar spinor behavior: a spin-$\frac{1}{2}$ state changes sign under a 2π rotation, even though all observable quantities remain the same. The double covering by $SU(2)$ is what allows half-integer spin representations to exist in quantum mechanics.

5.2 Topological View: π_1 and $S^3/\{\pm 1\}$

There is also a deeper, topological way to understand the relation between $SU(2)$ and $SO(3)$. The group manifold of $SU(2)$ is the three-sphere S^3. Topologically, this space is **simply connected**, which means that every closed loop drawn on it can be continuously shrunk down to a single point without leaving the surface. In other words, there are no "holes" or obstructions that prevent a loop from collapsing.

By contrast, $SO(3)$ has a slightly trickier shape. It can be obtained from S^3 by identifying opposite points:

$$SO(3) \simeq S^3/\{\pm 1\}.$$

Geometrically, you can imagine that each point on the three-sphere and its **antipode** (the point directly opposite it) represent the same rotation in $SO(3)$. When we make this identification, we end up with a space that is *not* simply connected. Consider the path in $SO(3)$ that represents a full 2π rotation in ordinary three-dimensional space. Physically, such a rotation brings every object back to its original orientation, so one might expect the corresponding loop in the space of all rotations to be trivial. However, from a topological point of view, this loop cannot be continuously deformed-"shrunk"-to a single point while staying within $SO(3)$. In topology, to shrink a loop means to slide and deform it smoothly until it collapses to the identity element, without ever leaving the space itself.

The obstruction comes from how $SO(3)$ is related to $SU(2)$. The group $SU(2)$ has the topology of a 3-sphere, S^3, while $SO(3)$ is obtained from it by identifying opposite points on that sphere. A 2π rotation in $SO(3)$ corresponds to a path on S^3 that begins at some point and ends at its antipode. Because those two points are identified in $SO(3)$, the path appears to close there-but underneath, on S^3, its endpoints are distinct. That hidden separation prevents the loop in $SO(3)$ from being contracted to a point: doing so would require "pulling" the path through the identified antipodal region, which is not possible within the space.

If you instead follow the same path twice-a total 4π rotation-the lifted path on S^3 starts and ends at the same point. This time the loop can be smoothly shrunk to the identity, showing that in $SO(3)$ a 2π rotation is topologically distinct from the identity, but a 4π rotation is not. This is precisely what it means for $SU(2)$ to be a *double cover* of $SO(3)$.

Formally, this idea is captured by the concept of **homotopy**. Two loops are said to be *homotopic* if one can be smoothly deformed into the other without cutting or leaving the space. This relation divides all possible loops into *equivalence classes*, where each class represents a single "type" of loop. An **equivalence class** is simply a collection of objects that are considered the same according to some rule of equivalence. In this case, all loops that can be smoothly deformed into one another belong to the same class.

These classes can be combined in a natural way: if two loops both start and end at the same point, we can first travel around one loop and then immediately around the other. This operation - called **concatenation** of loops - defines how the classes combine, giving the set of classes a genuine group structure. The resulting object is called the **fundamental group** of the space, denoted π_1. It captures, in a very precise way, how the space is "connected" at a global level: whether it has holes, twists, or identifications that prevent certain loops from shrinking to a point. In this sense, π_1 acts as a kind of topological fingerprint, telling us how paths can wrap around the space.

For $SU(2)$ and $SO(3)$, these groups are:

$$\pi_1(SU(2)) = 0, \qquad \pi_1(SO(3)) = \mathbb{Z}_2.$$

where $\mathbb{Z}_2$ is the two-element group describing a system with only two distinct states, such that doing the operation twice brings you back to the start. The first equation says that every loop on S^3 can be shrunk away. The second says that in $SO(3)$ there are two distinct classes of loops: one that can be contracted (the 4π rotation), and one that cannot (the 2π rotation). This $\mathbb{Z}_2$ structure is another way of expressing the same "double cover" relationship: each rotation in $SO(3)$ corresponds to two opposite points in $SU(2)$-one for each element of this two-element group.

6 Products of Groups and the Standard Model

When we write a product such as $\mathrm{U}(1) \times \mathrm{U}(1)$, we mean that each element of the combined group is specified by a pair (g_1, g_2), one from each factor. Geometrically, this means we are combining two independent spaces: the resulting manifold is the Cartesian product of the two shapes that represent each group.

6.1 $U(1) \times U(1)$ as a Torus

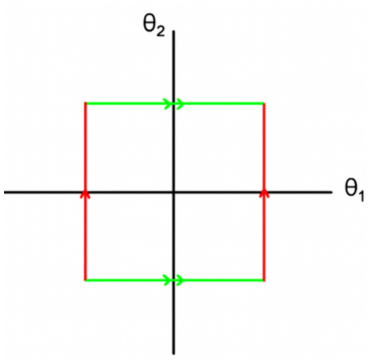
For example, $\mathrm{U}(1)$ is the set of all complex numbers of unit magnitude, forming a one-dimensional circle, S^1. The product $\mathrm{U}(1) \times \mathrm{U}(1)$ is therefore the product of two circles, which geometrically is a two-dimensional torus. To see this, think of S^1 as all points on a circle - you can describe each point by an angle θ, where $\theta \in [0, 2\pi)$. Now take two circles. A point in $S^1 \times S^1$ is described by two angles:

$$(\theta_1, \theta_2), \quad \text{with } \theta_1, \theta_2 \in [0, 2\pi).$$

So geometrically, you can think of all pairs (θ_1, θ_2) as forming a square parameter space:

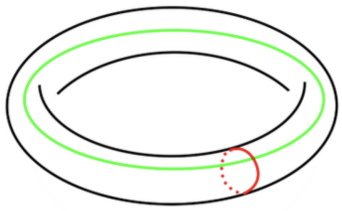
$$0 \leq \theta_1 < 2\pi, \quad 0 \leq \theta_2 < 2\pi.$$

Each point in this square corresponds to one point in $S^1 \times S^1$.



Because $\theta_1 = 0$ and $\theta_1 = 2\pi$ represent the same point on the first circle, the *left and right* edges of the square are glued together. Similarly, $\theta_2 = 0$ and $\theta_2 = 2\pi$ represent the same point on the second circle, so the *top and bottom* edges of the square are glued together.

If you glue the left/right edges together, you get a cylinder. Then, if you glue the cylinder's two circular ends together, you get a torus. That's exactly the topology of $S^1 \times S^1$.



Alternatively, we could have constructed $S^1 \times S^1$ in a more geometric way. Start with a single circle S^1. Imagine moving along this circle - this will represent the first coordinate, θ_1. Now, at every point on this circle, draw another circle in a direction perpendicular to it. This second circle represents the second coordinate, θ_2.

In this picture, the first S^1 forms the "large" circular path (the ring around the donut), while the second S^1 forms the "small" circular cross-section (the tube of the donut). As you travel once around the large circle, the smaller circle comes along for the ride - remaining attached and perpendicular at every point. In a similar way, we can construct a plane from two lines. Start with one line $\mathbb{R}$, and at every point on it attach another line in a perpendicular direction. The resulting space is the Cartesian product $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$, which is the familiar Euclidean plane.

6.2 $SU(3) \times SU(2) \times U(1)$

The same idea of taking products extends naturally from simple spaces to more complicated *group manifolds*. Just as $S^1 \times S^1$ combines two circles into a torus, we can form new symmetry spaces by taking the product of several continuous groups. Each factor contributes its own independent "direction" of symmetry, and together they describe systems whose transformations can act separately or in combination on different internal properties.

Now that we have built up an intuition for all the components, let's go back to the Standard Model, where all fundamental forces are manifestations of local gauge symmetries described by

$$SU(3) \times SU(2) \times U(1),$$

and perhaps unified in larger groups like $SU(5)$. This idea, known as a **grand unified theory** (GUT), suggests that the strong, weak, and electromagnetic forces are not truly separate, but

rather different expressions of a single, hidden symmetry woven into the structure of reality itself. The group $SU(5)$ provides a natural extension that can *embed* the entire Standard Model gauge group:

$$SU(3) \times SU(2) \times U(1) \ \subset \ SU(5).$$

In such grand unified theories, all gauge interactions arise from a single, larger symmetry that is spontaneously broken at very high energies into the familiar three factors. Geometrically, one can think of $SU(5)$ as a higher-dimensional analogue of $SU(3)$ and $SU(2)$, whose internal manifold contains the structure of all three forces at once. At everyday energy scales, this symmetry appears "fractured," revealing only the distinct $SU(3)$, $SU(2)$, and $U(1)$ components that govern the strong, weak, and electromagnetic interactions.

In either case, I do think it is important to make a clear distinction here, though. When we say that the Standard Model lives in $SU(3) \times SU(2) \times U(1)$, we do *not* mean that physical space has extra coordinates corresponding to these group manifolds. Spacetime itself remains the familiar four-dimensional manifold.

However, at each point in spacetime there exists an *internal fiber* - a tiny copy of the gauge group manifold. That is why these are called *internal* or *gauge* degrees of freedom. Each point in spacetime carries its own local copy of the gauge group, and the gauge fields (gluons, W/Z bosons, photons) describe how these internal fibers are "twisted" and connected across space and time. The way the fibers connect - the *connection* on the bundle - is what we perceive as the fundamental interactions.

In this sense, while we do not "live" inside a higher-dimensional manifold, the fields of the Standard Model take values on the internal product space $SU(3) \times SU(2) \times U(1)$, which acts as a higher-dimensional scaffolding for their symmetry behavior.

7 Representations: Fields as Multiplets

Up to now we've only talked about the generators of a symmetry and how they act *infinitesimally*- that is, how they give tiny nudges to a field. But a symmetry group contains not only its generators, it also contains full, finite transformations obtained by exponentiating them. To understand how these finite transformations act, we need to describe more precisely what it is they act *on*.

We've already been talking about fields $\Phi(x)$, so let's make their structure explicit. At each spacetime point x, the field takes values in some internal vector space V. In other words, a field is a function

$$\Phi : \text{spacetime} \longrightarrow V,$$

assigning to each point x a vector $\Phi(x) \in V$ (e.g., in $U(1)$ each point in spacetime carries a copy of $\mathbb{C}$). Different kinds of fields (scalars, spinors, vectors) correspond to different choices of this internal space and different ways the symmetry acts on it.

A **representation** is the rule that tells us how a full group element $g \in G$ acts on that internal space. Formally, it is a map

$$\rho : G \longrightarrow \mathrm{GL}(V),$$

which assigns to each group element g an invertible linear transformation $\rho(g)$ acting on V. In other words, $\rho(g)$ is a rule that moves vectors around within V without changing their overall size or independence - it reshuffles the components of a field in a reversible way. Here $\mathrm{GL}(V)$, the **general linear group** of V, means the set of all invertible linear maps from V to itself; equivalently, all invertible $n \times n$ matrices if $V \simeq \mathbb{C}^n$. In compact form,

$$\mathrm{GL}(V) = \{\, T : V \to V \mid T \text{ is linear and invertible} \,\}.$$

The key property of a representation is that it preserves the group structure:

$$\rho(g_1 g_2) = \rho(g_1)\,\rho(g_2).$$

Under such an action, the field transforms as

$$\Phi(x) \longrightarrow \Phi'(x) = \rho(g)\,\Phi(x).$$

7.1 The Fundamental Representation

For $U(1)$, the group element is just a complex phase $e^{i\alpha}$, and the field is a single complex number $\psi(x) \in \mathbb{C}$. The representation is one-dimensional:

$$\rho(e^{i\alpha})\,\psi = e^{i\alpha}\psi.$$

In this case the internal vector space $V \simeq \mathbb{C}$ is just a single complex line, and the group acts by rotating that line around the origin. This is the smallest nontrivial way for the group to act.

In $SU(2)$, instead of a single complex phase, we now have three independent directions of rotation, each generated by one of the Pauli matrices σ^a. These generators act on a two-dimensional complex vector space known as the **fundamental representation** of $SU(2)$ - the smallest space on which the group acts nontrivially.

Every element of $SU(2)$ is a 2×2 unitary matrix with determinant one. The most direct way for such matrices to act is by ordinary matrix multiplication on two-component column vectors in $\mathbb{C}^2$. That is, they naturally act on:

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

where ψ_1 and ψ_2 are complex numbers. Under an $SU(2)$ transformation, the spinor transforms as

$$\Psi \;\longrightarrow\; U(\boldsymbol{\alpha})\,\Psi = \exp\!\left(i\,\frac{\alpha_a \sigma^a}{2}\right)\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ are real parameters specifying the rotation in the internal $SU(2)$ space. The two complex components, ψ_1 and ψ_2, can be viewed as the amplitudes for a system to be in the "spin up" or "spin down" state along a chosen axis.

For $SU(2)$, the minimal nontrivial representation acts on two components, forming a **doublet**. Larger groups require correspondingly larger internal spaces because there are more independent directions in which the system can rotate. In $SU(3)$, for example, the fundamental representation is three-dimensional: each group element is a 3×3 unitary matrix with determinant one, acting on a three-component complex vector. This three-component structure provides the natural home for quarks, while the complex-conjugate space serves as the representation for antiquarks.

7.2 Adjoint Representation

So far we have seen how a group can act on some external vector space. But every Lie algebra also carries a natural representation acting on itself: the **adjoint representation**.

In the fundamental representation, the generators T^a act as matrices that rotate an external vector, say a spinor Ψ. In the adjoint representation, we instead let those same generators act on the *set of generators themselves*. In other words, the algebra acts by taking commutators:

$$(\mathrm{ad}_{T^a})\,T^b \;=\; [T^a,\, T^b] \;=\; i\,f^{abc}\,T^c.$$

Here the symbol ad_{T^a} (read "the adjoint action of T^a") does *not* refer to the Hermitian adjoint[5]. In group theory, "adjoint" means something else entirely: it describes how the algebra acts on itself. For the Lie algebra, this action is defined by taking commutators, so for any element X of the algebra,

$$\mathrm{ad}_{T^a}(X) \;=\; [T^a,\, X].$$

This map ad_{T^a} is a linear operator that tells us how X changes when we "rotate" it using the generator T^a. If we now use the generators $\{T^b\}$ themselves as a basis for the algebra, we can express the operator ad_{T^a} in matrix form. This tells us exactly how the generator T^a mixes the basis directions labeled by b and c. In this basis, the linear operator ad_{T^a} is represented by a matrix whose components are

$$(\mathsf{T}^a_{\mathrm{adj}})_{bc} = -\,i\,f^{abc}.$$

The values in f^{abc} encode precisely how the generators fail to commute, and therefore determine how the adjoint action mixes them with one another. The dimension of this adjoint representation is therefore equal to the number of independent generators of the group. This representation has a clear physical meaning. For $U(1)$, there is only one generator, which commutes with itself, so the adjoint representation is one-dimensional and trivial. **That is why the photon carries no electric charge:** it does not "rotate" into anything else.

For $SU(2)$, the situation is richer. The algebra has three generators $T^a = \frac{1}{2}\sigma^a$ satisfying

$$[T^a,\, T^b] = i\,\epsilon^{abc}\,T^c,$$

where ϵ^{abc} are the Levi-Civita symbols. The structure constants are therefore $f^{abc} = \epsilon^{abc}$, and so the adjoint representation takes the form

$$(\mathsf{T}^a_{\mathrm{adj}})_{bc} = -\,i\,\epsilon^{abc}.$$

Each $\mathsf{T}^a_{\mathrm{adj}}$ is now a 3×3 purely imaginary antisymmetric matrix acting on a three-dimensional vector space spanned by the generators themselves:

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

These are precisely the matrices that generate ordinary spatial rotations in three dimensions (up to isomorphism). Thus, in the adjoint representation of $SU(2)$, the three basis directions (or equivalently, the three gauge bosons) can rotate into one another just like the three spatial axes.

This is very different from the fundamental representation built from the Pauli matrices. In the fundamental case, the generators $T^a = \frac{1}{2}\sigma^a$ act on two-component spinors Ψ; the algebra acts on an *external* two-dimensional space of matter fields. In the adjoint case, the same algebra now acts on a *three-dimensional internal space* formed by its own generators. The fundamental representation describes how matter fields transform under the symmetry, while the adjoint representation describes how the gauge fields themselves transform.

For $SU(3)$, there are eight generators, giving rise to eight gluons that transform as an octet. This is what makes non-Abelian gauge fields unique: their force carriers not only mediate interactions, they also carry the very charges to which they couple.

[5] Historically, both notions trace back to the Latin *adjungere* ("to join to"), but they evolved in parallel fields: in quantum mechanics, "adjoint" meant the operator that pairs nicely under an inner product (so $\langle Ax, y\rangle = \langle x, A^\dagger y\rangle$), while in group theory, "adjoint" meant the group acting "on itself" (the group is joined to itself by conjugation).

Do we need the Pauli matrices?

A natural question to ask is: "Wait, but the generators are the Pauli matrices... which is what we used for the fundamental representation... yet the adjoint representation acts on the generators (and thus on the Pauli matrices) themselves. Couldn't we have represented the generators in some different way?"

The Pauli matrices are merely *one convenient realization* of the abstract $\mathfrak{su}(2)$ generators, corresponding to the fundamental representation (just like the Gell-Mann matrices λ^a for $\mathfrak{su}(3)$). What actually defines the algebra are the commutation relations

$$[T^a,\, T^b] = i\, f^{abc}\, T^c,$$

not the specific matrices chosen to represent them. Because the adjoint representation is defined entirely in terms of these structure constants,

$$(\mathsf{T}^a_{\text{adj}})_{bc} = -\,i\, f^{abc},$$

it is *independent of the particular matrix realization* of the generators.

7.3 Irreducible Representations

When a group acts on a vector space, it can mix some directions while leaving others untouched. For example, a rotation in the $x-y$ plane leaves the z-axis fixed. If we represent the group action by matrices $\rho(g)$, each matrix tells us how the components of a vector are reshuffled when the symmetry is applied. In general, these matrices may have a block structure: some sets of components only mix among themselves, never communicating with the rest. Mathematically, this means that there exists a nonzero subspace $W \subset V$ such that $\rho(g)W \subseteq W$ for all group elements g. In plain terms, the action of the group never carries a vector out of W: any vector that begins inside this subspace stays there under every transformation. In other words, the symmetry acts entirely within W, leaving it closed under its rotations. When this occurs, we say the representation is **reducible**.

If no such smaller subspace exists - that is, if every direction eventually talks to every other under the symmetry - then the representation is **irreducible**. Irreducible representations (often called "**irreps**") are the fundamental building blocks of symmetry: the smallest possible "mixing spaces" that cannot be decomposed any further without destroying the group structure. In matrix language, an irrep corresponds to a set of matrices that cannot be made block-diagonal by any change of basis. Once a representation is expressed in block-diagonal form, each block is an independent irrep.

7.4 Building Larger Representations

There is a simple and very physical way to *build* larger multiplets from smaller ones: we can combine two representations to form a new one. Geometrically, this means attaching two internal arrows and asking how the group rotates the pair together. Mathematically, this is done by taking a **tensor product** of the two representation spaces.

Consider $SU(2)$. Each doublet can be pictured as a two-component spinor. If we have two such doublets, say Ψ_i and Φ_i, each transforms under an $SU(2)$ rotation as

$$\Psi_i \;\longrightarrow\; U_{ij}\Psi_j, \qquad \Phi_i \;\longrightarrow\; U_{ij}\Phi_j.$$

When we consider them together, the pair $\Psi_i\Phi_j$ lives in the product space of both doublets. Under a group transformation, every index is rotated by its own copy of U:

$$(\Psi_i\Phi_j) \;\longrightarrow\; U_{ik}U_{jl}(\Psi_k\Phi_l).$$

This defines a new representation of $SU(2)$ on the four-dimensional space of pairs (i, j). In other words, we have built a new four-dimensional vector space whose elements are all possible bilinear combinations of two doublets, and the group now acts linearly on this larger space.

Now, not all of these four directions are mixed in the same way. We can separate them into two simpler pieces: one that is symmetric in the indices i, j, and one that is antisymmetric. These two subspaces never mix under any $SU(2)$ rotation. Hence the tensor product representation **decomposes** into two independent parts:

$$2 \otimes 2 = 3 \oplus 1.$$

The symmetric part has three independent components - it transforms like a vector under rotations, a **triplet**. The antisymmetric part has only one independent component - it behaves as a scalar, a *singlet*. In this way, the tensor product of two doublets naturally falls apart into two irreducible representations.

To see this, it's useful to recall what kind of operation this tensor product really is. When we take a tensor product of two representations, we are not adding them together but forming all possible *pairs* of their components.

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \qquad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \qquad u \otimes v = \begin{pmatrix} u_1 v_1 \\ u_1 v_2 \\ u_2 v_1 \\ u_2 v_2 \end{pmatrix}.$$

Now, among the four components labeled by index pairs $(i, j) = (11), (12), (21), (22)$, not all are mixed in the same way by $SU(2)$. Because the same U acts on both indices, the transformation commutes with the exchange of i and j. This means that the space can be divided into two sectors: one consisting of symmetric combinations, which are unchanged under $i \leftrightarrow j$, and one consisting of antisymmetric combinations, which pick up a minus sign under exchange. The group action preserves this distinction: it can never turn a symmetric tensor into an antisymmetric one or vice versa. Thus the product space splits into two invariant subspaces under all $SU(2)$ transformations.

The components (11) and (22) are already symmetric by themselves, since exchanging the two indices leaves them unchanged. (12) and (21) are exchanged under $i \leftrightarrow j$, so to construct quantities that are definite under this exchange we must take their symmetric and antisymmetric combinations: $(12 + 21)$ and $(12 - 21)$.

The symmetric combinations

$$(11), \quad (22), \quad (12 + 21),$$

form a three-dimensional subspace, while the single antisymmetric combination

$$(12 - 21)$$

forms a one-dimensional subspace. Hence the tensor product of two doublets decomposes as

$$2 \otimes 2 = 3 \oplus 1,$$

If this isn't clear, we can visualize it explicitly. Although $u \otimes v$ is really an element of the four-dimensional vector space $V \otimes V$, we can arrange its four components into a 2×2 grid because of the double index structure:

$$u \otimes v = \begin{pmatrix} (u \otimes v)_{11} & (u \otimes v)_{12} \\ (u \otimes v)_{21} & (u \otimes v)_{22} \end{pmatrix} = \begin{pmatrix} u_1 v_1 & u_1 v_2 \\ u_2 v_1 & u_2 v_2 \end{pmatrix}.$$

This is not a matrix in the usual sense, but simply a convenient way to *arrange* the four basis components of the tensor product space into a square shape. In this form, it becomes straightforward to see how to separate the tensor into its symmetric and antisymmetric parts. For any rank-2 tensor T_{ij}, we can always write

$$T_{ij}^{(\text{sym})} = \tfrac{1}{2}(T_{ij} + T_{ji}), \qquad T_{ij}^{(\text{antisym})} = \tfrac{1}{2}(T_{ij} - T_{ji}),$$

which project onto the subspaces that are even or odd under the exchange of indices $i \leftrightarrow j$. Applying this to $T_{ij} = u_i v_j$ gives

$$(u \otimes v)_{\text{sym}} = \frac{1}{2}\begin{pmatrix} 2u_1v_1 & u_1v_2 + u_2v_1 \\ u_1v_2 + u_2v_1 & 2u_2v_2 \end{pmatrix}, \qquad (u \otimes v)_{\text{antisym}} = \frac{1}{2}\begin{pmatrix} 0 & u_1v_2 - u_2v_1 \\ -(u_1v_2 - u_2v_1) & 0 \end{pmatrix}.$$

The symmetric part contains three independent components (corresponding to the triplet) and the antisymmetric part has one (corresponding to the singlet). Because $SU(2)$ acts identically on both indices, these symmetric and antisymmetric subspaces remain invariant under all group rotations. The decomposition is exact:

$$(u \otimes v)_{\text{sym}} + (u \otimes v)_{\text{antisym}} = u \otimes v,$$

since by construction $\frac{1}{2}(T_{ij} + T_{ji}) + \frac{1}{2}(T_{ij} - T_{ji}) = T_{ij}$.

From the matrix point of view, this means that the 4×4 matrices representing the tensor product action can be rearranged into two independent blocks: one 3×3 block for the symmetric subspace and one 1×1 block for the antisymmetric subspace. Each block is an irrep.

If you want one picture to keep in mind, think of every field as living not only in spacetime but also in an *internal space* whose geometry is determined by a symmetry group. A *representation* specifies exactly how a field's internal components transform when that symmetry is applied. The *fundamental* representation describes the simplest type of object the group can rotate nontrivially. The *adjoint* representation describes how the symmetry acts on its own directions. And an *irreducible* representation is one whose internal motion cannot be decomposed into smaller, disconnected parts.

The Pauli matrices σ^a and the Gell-Mann matrices λ^a are not the symmetries themselves; they are coordinate axes chosen within the internal spaces of the fundamental representations. Changing the basis of a multiplet simply corresponds to rotating these axes, but it does not alter the underlying geometry. These matrices are the rulers we lay down on the same curved surface, not the surface itself.

8 Spacetime Symmetries

The next great symmetry in physics is not an internal one at all; it is the symmetry of spacetime itself. In special relativity, all inertial observers agree on the spacetime interval

$$ds^2 \;=\; dt^2 - dx^2 - dy^2 - dz^2\,.$$

The transformations that leave this quantity unchanged form the **Lorentz group**. They include rotations and boosts that mix space and time, changing how one observer's axes are oriented relative to another's. These transformations, however, are all performed about a single point in spacetime. Rotations pivot around an origin, and boosts tilt the axes through that same fixed point. However, since the laws of physics should not depend on where or when we choose that point, it is natural to extend the symmetry to include spacetime **translations**. Together, the Lorentz transformations and translations form the full **Poincaré group**, which encompasses both the orientation and the placement of an observer's coordinate system in spacetime.

Definition of $\mathbb{R}^{1,3}$ and $O(1,3)$

To be concrete, the translation group $\mathbb{R}^{1,3}$ consists of all four-vectors

$$a^\mu = (a^0, a^1, a^2, a^3),$$

with the group operation given by simple addition,

$$a^\mu + b^\mu = (a^0 + b^0,\, a^1 + b^1,\, a^2 + b^2,\, a^3 + b^3).$$

This is an *Abelian* group: performing two translations in any order gives the same result. The Lorentz group $O(1,3)$ is the set of all real 4×4 matrices Λ that preserve the Minkowski metric,

$$\Lambda^T \eta \Lambda = \eta, \qquad \eta = \mathrm{diag}(+,-,-,-).$$

These are precisely the linear transformations that leave the spacetime interval $ds^2 = dt^2 - dx^2 - dy^2 - dz^2$ unchanged. The subgroup with $\det \Lambda = +1$ and ${\Lambda^0}_0 \geq 1$ is called the *proper, orthochronous Lorentz group* $SO^+(1,3)$, which preserves both orientation and the direction of time.

In practice, not every Lorentz transformation can be reached by exponentiating the generators of the algebra $\mathfrak{so}(1,3)$. The algebra only generates the part of the group that is smoothly connected to the identity: the **proper, orthochronous Lorentz group**, denoted $SO^+(1,3)$. The full Lorentz group $O(1,3)$ also includes discrete operations that flip spatial orientation (P) or reverse the direction of time (T). These additional transformations are not connected to the identity and therefore lie in separate components of the group manifold. For example, the time-reversing component is centered on the matrix $T = \mathrm{diag}(-1,\, 1,\, 1,\, 1)$, which flips the sign of the time coordinate and cannot be reached continuously from the identity. That is, we cannot reach T from $I = \mathrm{diag}(1,\, 1,\, 1,\, 1)$, because flipping the sign of the time coordinate cannot be achieved through any continuous rotation or boost.

In other words, while the Lorentz algebra encodes the *infinitesimal*, continuously connected symmetries of spacetime, the discrete P and T operations represent distinct, disconnected symmetries. To keep things concrete, we will follow the same convention used in Schwartz: unless stated otherwise, "the Lorentz group" will refer to the connected component $SO^+(1,3)$.

8.1 Semidirect Products

Mathematically, the *full* Poincaré group is a **semidirect product**,

$$\mathbb{R}^{1,3} \rtimes O(1,3),$$

meaning it is built by combining two types of symmetries: translations, forming the additive group $\mathbb{R}^{1,3}$, and Lorentz transformations, forming $O(1,3)$. The crucial point is that these two parts do not commute. Performing a Lorentz transformation and then a translation is not the same as doing them in the opposite order. A Lorentz transformation "rotates" or "boosts" the translation vector itself before the shift is applied.

Formally, a *semidirect product*[6] $N \rtimes G$ is built from two subgroups, N and G, that interact in a specific way. The subgroup N (in our case, the spacetime translations) is required to be **normal**, meaning that if you conjugate an element of N by any element of the full group, you end up back in N again. Intuitively, this means N sits inside the group in a way that remains stable under the action of G.

[6] Here we used the *outer* semidirect product

The price of this stability is that moving elements of N past elements of G changes them slightly. It "twists" them by a map called an **automorphism**. An automorphism of N is an *isomorphism from N to itself*: a one-to-one map that preserves the group operation,

$$\varphi(n_1 n_2) = \varphi(n_1)\,\varphi(n_2), \qquad \text{for all } n_1, n_2 \in N.$$

It represents a symmetry of the group N seen from within. In a semidirect product, each element of G induces such an automorphism of N; this is how G "acts" on N. In other words: when you do a Lorentz transformation, you change what "direction" a translation means.

Symbolically, the group multiplication in a semidirect product takes the form

$$(g_1, a_1)\,(g_2, a_2) = \big(g_1 g_2,\ a_1 + \varphi_{g_1}(a_2)\big),$$

where φ_{g_1} denotes the automorphism of N induced by $g_1 \in G$. In words: when composing two elements, the second translation a_2 is first "rotated" by the group element g_1 before being added to a_1. If the two parts were totally independent, you'd have a direct product:

$$(g_1, a_1)(g_2, a_2) = (g_1 g_2,\ a_1 + a_2) \qquad \text{(if independent)}$$

In the Poincaré group, the automorphism φ_Λ is just the usual Lorentz action on four-vectors:

$$a^\mu \ \longmapsto\ \Lambda^\mu{}_\nu\, a^\nu.$$

Thus, when we multiply two Poincaré elements (Λ_1, a_1) and (Λ_2, a_2), we obtain

$$(\Lambda_1, a_1)\,(\Lambda_2, a_2) = \big(\Lambda_1\Lambda_2,\ a_1 + \Lambda_1 a_2\big),$$

showing explicitly that the Lorentz part "acts on" the translation part before the two are combined. This intertwining between translations and Lorentz transformations is exactly what makes the product *semi*-direct rather than direct.

8.2 Isometries

Formally, a **metric space** is an ordered pair $(\mathcal{M}, d)$, where $\mathcal{M}$ is a set and d is a *metric* on $\mathcal{M}$; that is, a function

$$d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$$

which assigns to each pair of points $x, y \in \mathcal{M}$ a non-negative real number $d(x, y)$, the *"distance"* between x and y. The function d satisfies the following properties for all $x, y, z \in \mathcal{M}$:

$$\begin{aligned}
&d(x,x) = 0, && \text{(identity)}\\
&x \neq y \implies d(x,y) > 0, && \text{(positivity)}\\
&d(x,y) = d(y,x), && \text{(symmetry)}\\
&d(x,z) \leq d(x,y) + d(y,z). && \text{(triangle inequality)}
\end{aligned}$$

An **isometry** of a metric space $(\mathcal{M}, d)$ is a map

$$\phi : \mathcal{M} \longrightarrow \mathcal{M}$$

that preserves the metric, meaning that

$$d(\phi(x), \phi(y)) = d(x, y) \quad \text{for all } x, y \in \mathcal{M}.$$

When $\mathcal{M}$ is a smooth manifold equipped with a Riemannian metric g, we call $(\mathcal{M}, g)$ a **Riemannian manifold**. Geometrically, this means that at every point of $\mathcal{M}$, the metric g allows us to measure lengths of tangent vectors and angles between them, so that concepts like distance and curvature make sense on the manifold.

An *isometry* of a Riemannian manifold $(\mathcal{M}, g)$ is a smooth map that preserves these geometric measurements. This is expressed by

$$\phi^* g = g,$$

where the **pullback** $\phi^* g$ means we compare the metric before and after applying ϕ. The comparison agrees everywhere on the manifold since $\phi^* g = g$. That is, If $F : \mathcal{M} \to \mathcal{N}$ is a smooth map and $\mathcal{N}$ carries a metric g, then F *pulls back* the metric g to $\mathcal{M}$, meaning it allows us to measure lengths and angles on $\mathcal{M}$ using the geometry of $\mathcal{N}$.

For Minkowski space[7], the metric is constant everywhere, so the most general transformation that preserves it must be linear up to a constant shift:

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu}\, x^{\nu} + a^{\mu},$$

where the matrix Λ satisfies $\Lambda^T \eta\, \Lambda = \eta$. The set of all such transformations (Λ, a) forms the *Poincaré group*,

$$\mathrm{ISO}(1,3) \;=\; \mathbb{R}^{1,3} \rtimes O(1,3),$$

which we saw earlier as the semidirect product of translations and Lorentz transformations.

8.3 The Mirror Hidden in the Lorentz Algebra

We can study the the translation and Lorentz parts of $\mathrm{ISO}(1,3)$ independently. The translation generators P_μ commute among themselves,

$$[P_\mu,\, P_\nu] = 0,$$

so they form an Abelian **subalgebra** corresponding to the additive group $\mathbb{R}^{1,3}$. This part of the algebra carries no intrinsic structure beyond linear addition; it does not affect the internal geometry of spacetime. A *subalgebra* is simply a smaller algebra sitting inside a larger one: a subset of generators that is closed under the same commutation relations. In other words, if you take any two elements of the subalgebra and compute their commutator, the result stays within that same set. All of the interesting structure therefore lives in the Lorentz sector, which tells us how directions are related and how boosts and rotations intertwine.

Now, recall that the Lorentz algebra $\mathfrak{so}(1,3)$ is generated by three spatial rotations and three boosts:

$$J_i = \tfrac{1}{2}\epsilon_{ijk} M_{jk}, \qquad K_i = M_{0i}, \qquad i = 1,2,3.$$

The J_i generate ordinary rotations among spatial axes, while the K_i generate Lorentz boosts that mix space and time. Using the commutation relations above, we can work out how these operators interact:

$$[J_i,\, J_j] = i\,\epsilon_{ijk}\, J_k, \qquad [J_i,\, K_j] = i\,\epsilon_{ijk}\, K_k, \qquad [K_i,\, K_j] = -\,i\,\epsilon_{ijk}\, J_k.$$

The first line shows that the J_i alone form a closed $\mathfrak{so}(3)$ subalgebra: the familiar rotation algebra. The second and third lines show that boosts transform as vectors under rotations, and that the boosts themselves do not commute: performing two boosts in different directions does not simply add up to a single, larger boost; it also introduces a small rotation of the spatial axes.

[7] Here "Riemannian" is used loosely; in spacetime the metric has Lorentzian (pseudo-Riemannian) signature, so the same definition of isometry applies with g replaced by the Minkowski metric.

The minus sign in

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

means that the boosts cannot by themselves form a compact rotation algebra; instead, they "mix" with the J_i. A group is called **compact** if its parameters are bounded, so that the group manifold has finite volume. Ordinary spatial rotations, described by $SO(3)$, are compact because their angles are periodic: rotating by 2π brings you back to where you started, so it involves ordinary trigonometric functions. For Lorentz boosts, the flipped sign of the metric means that exponentiating now produces hyperbolic functions instead (the usual $\cosh\beta$ that appears in boosts).

Geometrically, a boost is a "rotation" in a plane that includes time, and because hyperbolic functions never repeat, the rapidity can grow without bound. This is why boosts form a **non-compact** group. Compact groups such as $SO(3)$ or $SU(2)$ have finite parameter ranges and admit finite-dimensional *unitary* representations, giving rise to discrete spectra like the quantized spins of particles. Non-compact groups like $SO(1,3)$ instead have unbounded parameters and infinite-volume manifolds. Their only non-trivial unitary representations are infinite-dimensional. Note that "non-compact" does *not* mean the algebra fails to close: the full set of six generators $\{J_i, K_i\}$ is perfectly closed under commutation. The boosts alone do not form a subalgebra since their commutator produces a rotation, but the Lorentz algebra as a whole remains fully consistent; its non-compactness refers instead to the unbounded, hyperbolic geometry of the group manifold.

The pattern of the signs in the Lorentz algebra hints that there might be a hidden simplicity if we choose a different basis of generators. The commutator of two boosts produces a rotation with an opposite sign, suggesting that we could cancel this mismatch by mixing the boosts with an imaginary factor. Define the new combinations

$$J_i^+ \equiv \tfrac{1}{2}(J_i + i\,K_i), \qquad J_i^- \equiv \tfrac{1}{2}(J_i - i\,K_i)$$

These particular linear combinations are chosen because they *diagonalize* the Lorentz commutation relations - each set closes on itself:

$$[J_i^+,\, J_j^+] = i\,\epsilon_{ijk}\,J_k^+, \qquad [J_i^-,\, J_j^-] = i\,\epsilon_{ijk}\,J_k^-, \qquad [J_i^+,\, J_j^-] = 0.$$

Thus, we discover that the complexified Lorentz algebra splits neatly into two independent pieces:

$$\mathfrak{so}(1,3)_{\mathbb{C}} \;\cong\; \mathfrak{su}(2)_L \oplus \mathfrak{su}(2)_R.$$

Each copy behaves exactly like the familiar rotation algebra $\mathfrak{su}(2)$, with the same structure constants and representation theory. The two halves commute, acting independently.

This decomposition, however, applies to the *complexified* Lorentz algebra. The real algebra $\mathfrak{so}(1,3)$ is not literally a sum of two compact $\mathfrak{su}(2)$'s: one cannot exponentiate both J_i^+ and J_i^- to obtain compact subgroups at the same time. In the real form, boosts still remain non-compact. The complexification simply reveals the hidden symmetry structure - how the Lorentz algebra behaves *as if* it contained two copies sewn together by complex combinations.

Only afterwards do we recognize the physical meaning: fields that transform under $\mathfrak{su}(2)_L$ alone are **left-handed spinors**, while those transforming under $\mathfrak{su}(2)_R$ are **right-handed spinors**. Together they form the full, four-component **Dirac spinor**, transforming under both chiral halves simultaneously.

8.4 Particles as irreps

In ordinary quantum mechanics, we describe systems by states $|\psi\rangle$ that live in a Hilbert space and transform under the symmetries of the theory. In quantum field theory, the relevant symmetry is the Poincaré transformation, which we denote by $\mathcal{P}$, and acts on a quantum state as

$$|\psi\rangle \longrightarrow \mathcal{P}\,|\psi\rangle.$$

Recall that a collection of states that mix only among themselves under all such transformations forms a *representation* of the Poincaré group. If no smaller subset of states remains closed under the action of the group, the representation is said to be *irreducible*. These are the smallest possible sets of states that can be connected to one another by spacetime symmetries, and hence correspond to the most elementary physical entities.

Because probabilities must be preserved under any symmetry operation, we require these representations to be *unitary*:

$$\mathcal{P}^\dagger \mathcal{P} = 1.$$

This ensures that inner products such as matrix elements

$$\mathcal{M} = \langle \psi_1 | \psi_2 \rangle$$

remain invariant under Poincaré transformations, guaranteeing that measurable quantities do not depend on the observer's inertial frame.

Putting these ingredients together gives a precise statement:

Particles transform under irreducible unitary representations of the Poincaré group.

This is not just a useful description; as Schwartz says, it is effectively the *definition* of a **particle** in quantum field theory. Each type of particle corresponds to one irrep, and every state within that irrep can be reached from any other by a suitable combination of translations, rotations, and boosts. Different irreps describe fundamentally different species of particles: those with different masses or spins cannot be connected by any Poincaré transformation.