B4. Gauge Field Theory

It is a striking fact about Nature that there exist gauge fields which play a key role in mediating interactions. At the "fundamental level" of particle physics one has the electromagnetic field, the various fields involved in the standard model, and the gravitational field. In various areas of condensed matter physics it has also been found useful to introduce gauge field descriptions of certain kinds of collective modes.

It is not possible to survey all of these. In what follows I will (i) discuss some of the underlying general features of all gauge fields, and why we have them, then (ii) discuss how they are described in a path integral formulation, using the ideas first developed by Fadeev and Popov. I will show how this works in both Abelian gauge theories (like QED) and in non-Abelian theories (the Yang-Mills model), along with brief remarks about the electroweak theory. I will then give a brief description of gauge theory for a condensed matter system. It is not possible to go into great detail here - there is no space - but we can at least see how these applications come about.

B.4.1 GAUGE FIELDS - GENERAL FEATURES

The long history of this topic means that before we launch into the modern formulation of the theory, it is necessary to look at some more general aspects of the theory. We begin by a quick resum*é* of some of the history, followed by a comparison of the classical and quantum versions of electromagnetic theory. A key point emerges from this comparison: there is a crucial difference in the role played by the electrodynamic gauge potential $A^\mu(x)$ in the 2 versions, as shown in the Aharonov-Bohm effect, which is discussed thoroughly here.

B.4.1 (a) SOME HISTORICAL REMARKS

the discovery of the electromagnetic field and of the General Relativistic description of spacetime also ushered in the idea of gauge fields, in early work by Weyl - we have known since the 1960's that they are here to stay. The discovery of QM, and the slow development of QFT made it clear that the quantized versions of such fields mediated interactions in physics - the EM and spacetime fields being bosonic, with spin-1 and spin-2 respectively. Once the spin-statistics theorem was first given (by Pauli) it was also clear that only bosonic fields could be associated with classical macroscopic force fields. By the early 1960's (notably in the work of Weinberg) it was known that the only consistent theories of massless bosonic fields had to be either spin-0 (Higgs field - given a mass by the "Higgs" mechanism, actually first discussed by PW Anderson and N. Bogoliubov), spin-1 (EM field), or spin-2 (gravitational field). In condensed matter theory, the key mechanism needed to give these gauge fields a mass (i.e., a finite energy gap) was found - the appearance of an "order parameter", a concept first defined by Landau and Lifshitz in 1935, and developed with great effect by Landau, and later by others, including London(1938), and BCS and Bogoliubov (1956-1962). Anderson's paper, with the hypothesis of the Higgs boson, appeared in 1963. The papers of Higgs, and of Englert and Brout, were all written and published in 1964, and very quickly followed by papers of Kibble (1964) and Guralnilk and Hagen (1964). The appelation "Higgs boson" is a misnomer.

However, there were serious mathematical problems, already noted in the 1950's in the context of QED. These concerned the renormalizability of QED, and the practical task of renormalizing calculations of specific physical quantities. The very small coupling constant in QED made practical calculations tractable, but the same was not true of the weak or strong interactions - by the late 1950's many physicists were in despair over the application of QFT to these interactions, and Landau and others led the way to alternative formulations, such as the S-matrix theory (or "bootstrap" theory), reminiscent of behaviourist "black box" psychology. This detour wasted the time of many physicists (although it produced some useful mathematics), until the tide began to turn in 1966-1967.

This change of attitude, and the "rediscovery of field theory" happened in a curious fashion. There were two key developments. First, the remarkable paper of Yang and Mills on 1954, which generalized the idea of the $U(1)$ gauge transformation used in QED to an $SU(N)$ gauge transformation. This idea was largely ignored at the time since it predicted massless bosons as intermediaries of the strong force (and mesons are not massless!). Second was the slow development of General Relativity (GR). the late 1950's, very few physicists (with the exception of astrophysicists in the UK, Dirac, the Russians surrounding E. M. Lifshitz, and those in the former circle of Einstein in the USA) paid any attention to gravity or to GR. This astonishing neglect showed how narrow-minded some communities in science can be; but it also arose because GR seemed to have little relevance to earth-bound physics, and moreover, seemed to be quite irreconcilable with QM. At that time, only Einstein was talking publicly about his idea of a "unified field theory".

Curiously it was Feynman who first broke away from this mindset, at the Chapel Hill conference in 1957, where he argued, using thought experiments, that gravity had to be quantized. In a period of intense work between 1957-1963 he tried to do this, but ran into a fundamental problem - gravity is non-renormalizable. It is a matter of some mystery why, after the remarkable sucess Feynman had using path integral methods on both superfluid ^4He and the polaron problem, in the period 1952-1956 (diagrams are almost useless for these systems), that in the study of gravity he should abandon path integrals for the "SHut Up And Calculate" (SHUAC) method of diagrams - and this is what he did. His approach led to the discovery of ghosts in gauge field theory, and was pursued to a successful but almost unreadable conclusion by B. de Witt (1964-1967); but it was completely unusable for any calculations.

All this changed in the period 1967-1970 with three key developments. First, in 1966-1967, Fadeev and Popov succeeded in formulating gauge field theory in path integral language - not just for QED, but also for Yang-Mills theory, and even in principle, for gravity. The result was equivalent to deWitt's (published a week later), but in contrast, was simple to understand and use. Second, in 1967-1968, Salam and Weinberg separately published theories unifying the weak and EM interactions into one "electroweak" gauge theory. This theory incorporated the yang-Mills idea of non-Aberlian gauge fields, and the Higgs mechanism to give the gauge bosons a mass. Nobody paid any attention to it until 1970, when 't Hooft, a beginning PhD student in Utrecht, showed - to everyone's astonishment - that the Salam-Weinberg theory was renormalizable; and then, in a tour de force, he showed how to do calculations with it, using a combination of path integral methods and "dimensional regularization", a technique introduced by 't Hooft and his supervisor Veltman (and found by others independently at around the same time).

It is hard to imagine what Feynman's reaction must have been when he saw what he might have achieved, had he stuck with his own path integral methods!

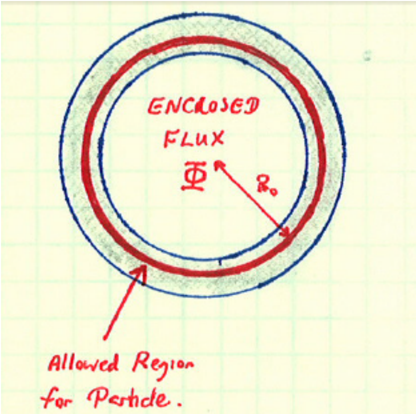
All of this work was subsequently developed into what is now the "standard model". The key further element to be found was the discovery of asymptotic freedom in the strong interactions (mediated by gluons, with a role also played by other bosonic fields). This discovery was actually made by 't Hooft in 1972, but he was discouraged from publishing it by Veltman; it was rediscovered by Gross and Wilczek in 1973, and also by Politzer in 1973 (whose supervisor, Coleman, was in contact with 't Hooft).

At the same time in condensed matter physics, the idea of the "gauge principle" was being applied to various systems. Here the impact was less clear, because we always have a "more microscopic" theory which does not require the gauge formulation; and gauge theories are often hard to work with. Thus, e.g., the gauge theories of high-T_c superconductivity and of the FQH (Fractional Quantum Hall Liquid) have had little practical impact. However they have motivated mainly interesting new developments, and the idea of "spontaneously broken gauge symmetry", with the appearance of an order parameter, is central to all of condensed matter physics. Indeed, in focussing on this, CM physicists are really returning to the basic original idea of gauge theory, which goes back to Weyl in the 1920's, viz., that there had to be an apparent arbitrariness in the way that we parametrize physical variables such as phase, or even length and time in spacetime. In classical physics these variables really are redundant - they are eliminated by fixing a "standard of measurement". But this is not so in a quantum theory.

This difference between classical and quantum physics, in the role played by a gauge field, is crucial. So let us now see how it happens, in the context of one of the simplest gauge theories, viz., QED and the quantized EM field. The remarks we make here arose originally from the analyses of Yang-Mills in 1954, of Aharonov and Bohm in 1957-59, and their subsequent elaborations by many authors.

B.4.1 (b) THE AHARONOV-BOHM EFFECT

When first proposed in the late 1950's, this analysis by Aharonov and Bohm caused huge controversy (NB: the mathematical analysis was actually done by MHL Pryce in Bristol, who later spent the years 1968-2001 at UBC). This was not least because Bohm, a brilliant young protegé of Oppenheimer during the war years, had been expelled from Princeton and



from the USA in the early 1950's, accused of being a communist during the McCarthy years. With the help of Einstein and Pryce, then both at Princeton, Bohm went first to Brazil and then to Israel; Pryce then recruited him to Bristol in the UK in the late 1950's.

There are several key works by Aharonov and Bohm; the one we will be looking at concerns gauge fields. You may also find it interesting to look at Feynman's lectures in Physics, Vol. 3, on this - Feynman was one of the early supporters of Bohm and his work.

(i) Particle on a Ring: Consider a situation where a single non-relativistic particle is forced to move on a 2-dimensional circle of radius R_0 - such a situation is now easy to organize with mesoscopic rings or superconducting SQUIDS, or quantum wells, or even optically with photons. However when Chambers did the first experiment in 1961 it was not so simple.

The quantum mechanics of the problem is quite straightforward, and easily done using the Schrodinger equation. The Hamiltonian is assumed static; then

$$H \;=\; \frac{1}{2m}\big(\mathbf{p} + q\,\mathbf{A}(\mathbf{r})\big)^2 + q\,\phi(\mathbf{r}) \tag{1}$$

and we will assume that a flux Φ is enclosed inside the ring, i.e., that

$$\oint_c d\mathbf{l}\cdot\mathbf{A}(\mathbf{l}) = R_0 \int d\theta A(\theta) = \Phi \tag{2}$$

This is not actually the problem solved by Aharonov and Bohm, and it is easier to solve then their scattering problem. Notice that the full EM Lagrangian, derived from H by canonical transformation, viz.,

$$\mathcal{L} \;=\; \frac{1}{2}m\dot{\mathbf{r}}^2 + q\,\mathbf{A}(\mathbf{r})\cdot\dot{\mathbf{r}} - q\,\phi(\mathbf{r}) \tag{3}$$

where the momentum is

$$\mathbf{p} \;=\; \frac{\partial\mathcal{L}}{\partial\dot{\mathbf{r}}};=\; m\dot{\mathbf{r}} + q\,\mathbf{A}(\mathbf{r}) \tag{4}$$

can be replaced by a truncated version only valid on the 2-dimensional ring of radius R_0. Let us assume that the electrostic potential $\phi(\mathbf{r}) = \phi_0$, a constant (no electric field). Then, on the ring, we have

$$\begin{aligned}\mathcal{L} &\to \frac{1}{2}m\dot{\mathbf{r}}^2 + q\,\dot{\mathbf{r}}\cdot\mathbf{A}(\mathbf{r}) \\ &\to \frac{1}{2}I_0\dot{\theta}^2 + q\,R_0\dot{\theta}A(\theta) \qquad\qquad (\text{if } |\mathbf{r}| = R_0)\end{aligned} \tag{5}$$

where the "moment of inertia"

$$I_0 = mR_0^2 \tag{6}$$

(**i**) Consider first this problem when $\Phi = 0$; we just have a particle circulating on a ring, with

$$H_0 = L_0 = \frac{1}{2}I_0\dot{\theta}^2 \tag{7}$$

Then the normalized solutions to the Schrodinger equation:

$$H\psi_l(\theta) = \epsilon_l\psi_l(\theta) \tag{8}$$

are given by

$$\psi_l(\theta) = \frac{1}{\sqrt{2\pi}}\,e^{il\theta} \tag{9}$$

where $l = 0,\ \pm1,\ \pm2,\dots$ etc., is an integer (the angular momentum quantum number), and θ is, as above the angular coordinate. Here we can treat it as compact, i.e., $0 \le \theta \le 2\pi$.

The propagator for the particle is also easily found; we have

$$\begin{aligned}G(\theta_1\theta_2;t_1t_2) = G(\theta,t) &= \sum_l \psi_l(\theta)\psi_l^*(\theta)e^{\frac{i}{\hbar}\epsilon_l^0 t} \\ &= \sum_l \int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\ \frac{\psi_l(\theta)\psi_l^*(\theta)}{\omega - \epsilon_l^0}\end{aligned} \tag{10}$$

where the eigenvalues are just

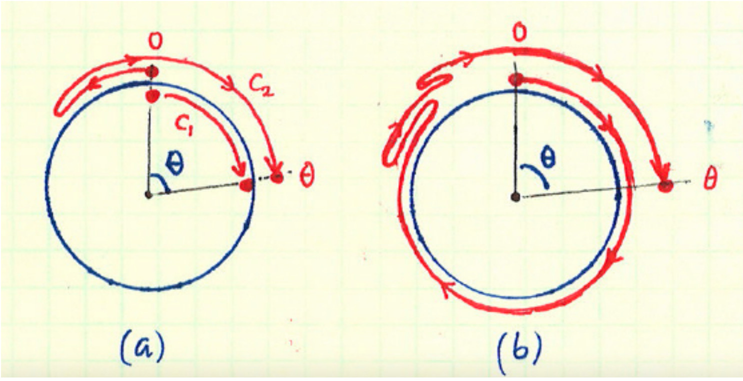
$$\epsilon_l^0 = \frac{\hbar^2}{2I_0}l^2 \tag{11}$$

Using either form in (10), we get the result

$$G(\theta,t) = \left(\frac{I_0}{2\pi i\hbar t}\right)^{\frac{1}{2}}\sum_{l=-\infty}^{\infty}\exp\left\{\frac{i}{\hbar}I_0\frac{(\theta+2\pi l)^2}{2t}\right\} \tag{12}$$

which we can write in more compact form as

$$G(\theta,t) = \left(\frac{I_0}{2\pi i\hbar t}\right)^{\frac{1}{2}}\ e^{\frac{i}{\hbar}I_0\frac{\theta^2}{2t}}\ \mathbf{\Theta}_3\left(\frac{\pi I_0\theta}{\hbar t};\ \frac{2\pi I_0}{\hbar t}\right) \tag{13}$$



where $\mathbf{\Theta}_3(z,t)$ is the biperiodic Jacobi θ-function, defined in the complex plane z, and with the series representation

$$\mathbf{\Theta}_3(z,t) = \sum_{n=-\infty}^{\infty} e^{i(\pi t n^2 + 2zn)} \tag{14}$$

i.e., a discretized version of a Gaussian integral.

Why do we have such a complicated result? The first part of (13) looks just like a free particle of mass I_0; so where does the Jacobi θ-function come from? In this calculation it comes from the compactness of the variable θ; we are dealing with a 1-d "particle in a box", with periodic boundary conditions. But now we can rederive this result in a quite different way, extending the domain of θ so that $-\infty < \theta < \infty$. Let's rederive the result (13) using path integrals. Then we have

$$G(\theta,t) = \int_{\theta(0)=0}^{\theta(t)=\theta} \mathcal{D}\,\mathbf{r}(t)\; e^{\frac{i}{\hbar}S[\mathbf{r},\dot{\mathbf{r}}]} = A(t)\; e^{\frac{i}{\hbar}S_{cl}(\theta,t)} \tag{15}$$

where $A(t)$ is just the fluctuation determinant for a particle of mass I_0 in free particle motion:

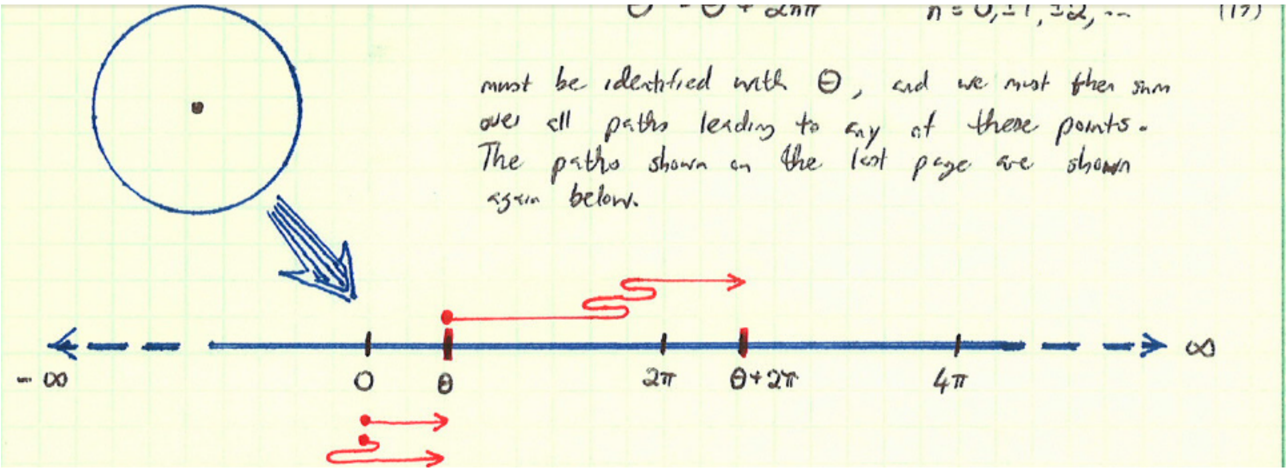
$$A(t) = \left(\frac{I_0}{2\pi i\hbar t}\right)^{\frac{1}{2}} \tag{16}$$

Now, however, we must be careful with the classical action. Recall that we are supposed to sum over all paths. In Fig(a) at the top of the page we see 2 simple paths which begin at $\theta(0) = 0$, and terminate at $\theta(t) = \theta$. Notice that in both cases the *winding number* n is zero.

However, this is not true of the path in Fig(b). This has winding number $n = 1$. And yet this path must also be included in $G(\theta,t)$.

The simplest way to then derive the answer for $S_{cl}(\theta,t)$ is to either (a) note that any path beginning at $\theta(0) = 0$ and ending at $\theta(t) = \theta$ can be decomposed into a path going from $\theta(0) = 0$ to $\theta(\tau) = 0$, where $\tau \leq t$, and then another path going to $\theta(t) = \theta$; but we now sum over all possible winding numbers for the first at these 2 paths. Or else (b), we just "unfold" the ring, as shown in the figure below. Now in this figure we see that any point on the line at

$$\theta' = \theta + 2n\pi, \qquad n = 0, \pm 1, \pm 2, \ldots \tag{17}$$



must be identified with θ, and we must then sum over all paths leading to any of these points. The paths shown on the previous figure are shown again below.

However it is clear from this diagram that we are dealing with a free particle on the line, and so we immediately find that

$$e^{\frac{i}{\hbar}S_{cl}(\theta,t)} = \sum_{n=-\infty}^{\infty} e^{\frac{i}{\hbar}I_0\frac{(\theta+2n\pi)^2}{2t}} \tag{18}$$

have summed over the different winding numbers. Inserting (18) and (16) into (15), we again recover (13).

(ii) Now let's go to the finite flux case. The energy levels are shifted, and the eigenfunctions change; we have

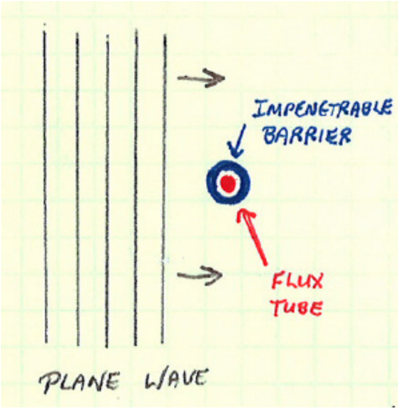
$$\begin{aligned} \psi_l(\theta,\bar{\varphi}) &= \frac{1}{\sqrt{2\pi}}\, e^{i(l+\bar{\varphi}/2\pi)\theta} \\ \epsilon_l(\bar{\varphi}) &= \frac{\hbar^2}{2I_0}\,(l+\bar{\varphi}/2\pi)^2 \end{aligned} \tag{19}$$

and from this we can derive the new answer. Before doing so, notice a crucial point. The answer does not depend on how the flux is distributed inside the rings, or even on whether the flux density (i.e., the magnetic field) is finite at the ring itself - it depends only on the total flux, or rather, the "dimensionless flux" $\bar{\varphi}$, defined as

$$\bar{\varphi} \;=\; \frac{q}{\hbar}\oint d\mathbf{l}\cdot\mathbf{A}(\mathbf{l}) \;=\; \frac{2\pi q}{\hbar}\Phi \;=\; \Phi/\Phi_0, \qquad \text{where } \Phi_0 = h/q \tag{20}$$

and $\Phi_0 = h/q$ is the flux quantum for charge q.

From here on it is clear that for a path with winding number n, we must add a phase $n\bar{\varphi}$



to the action exponent, and so now we get

$$\begin{aligned} G(\theta,t;\bar{\varphi}) &= \left(\frac{I_0}{2\pi i\hbar t}\right)^{\frac{1}{2}} \sum_{n=-\infty}^{\infty} \exp\left\{ in\bar{\varphi} + \frac{i}{\hbar} I_0 \frac{(\theta+2n\pi)^2}{2t} \right\} \\ &= \left(\frac{I_0}{2\pi i\hbar t}\right)^{\frac{1}{2}} e^{\frac{i}{\hbar} I_0 \frac{\theta^2}{2t}} \, \mathbf{\Theta}_3 \left(\frac{\pi I_0 \theta}{\hbar t} - \frac{\bar{\varphi}}{2}; \frac{2\pi I_0}{\hbar t} \right) \end{aligned} \tag{21}$$

and in both forms of this answer, we see what is already obvious from the new eigenfunctions and energies in (19), i.e., that the flux has shifted the answer - it is a "phase shift" operator.

(ii) Charged Particle scattering off a flux tube: Now let's go to the problem that Aharonov and Bohm actually looked at in their famous paper. This problem is shown below - we have an infinitesimal flux tube, still carrying flux Φ, but now confined to a very thin filament, which we will treat as a δ-function in 2-d space. We then have

$$H = \frac{1}{2m}\left(\rho + q\,\mathbf{A}(\mathbf{r})\right)^2 \tag{22}$$

where

$$\mathbf{A}(\mathbf{r}) = \Phi\frac{\hat{z}\times\hat{r}}{2\pi r} = \bar{\varphi}\Phi_0\frac{\hat{z}\times\hat{r}}{2\pi r} = \hat{\theta}\frac{\Phi}{2\pi r} \tag{23}$$

where $\hat{z}$ and $\hat{r}$ are unit vectors, as is $\hat{\theta}$. To make the point even more clearly, let's surround the flux tube at the origin by an infinite potential barrier outside the flux tube, but with a radius r_0 which we will also take to be infinitesimal. Then nothing from outside can penetrate, and the flux tube is isolated from the outside world.

And yet, from what we have done above, we know that even though the electric field $\mathbf{E}(\mathbf{r}) = 0$ everywhere, and $\mathbf{B}(\mathbf{r}) = 0$ except inside the flux tube, yet still a quantum particle moving outside will feel the flux! This is utterly different from classical mechanics, where the particle dynamics is governed by the Lorentz equation, which is local:

$$m\ddot{\mathbf{r}}(t) = q\,\mathbf{E}(\mathbf{r}) + q\left(\dot{\mathbf{r}}\times\mathbf{B}(\mathbf{r})\right) \tag{24}$$

Now, we imagine a plane wave state of the particle, incident on the flux tube. If we let the barrier potential radius $r_0 \to 0$, then for some finite wavelength incident wave, with wavelength $\lambda = 2\pi/k$, the scattering cross-section (in 2d) off the potential barrier on its own (ie., with no enclosed flux) is given from elementary scattering theory by

$$\sigma_k \;\sim\; kr_0 \,\ln(kr_0) \tag{25}$$

so that in this limit, the particle does not scatter off the potential barrier. Nevertheless it still scatters off the flux tube, even though it never sees the flux. The Schrodinger equation in 2d, for the Hamiltonian (1) with $\phi(r) = 0$, is

$$\left\{ \partial_r^2 + \frac{1}{r}\partial_r + \left[k^2 - \frac{1}{r^2}(i\partial_\theta + \bar{\varphi})^2 \right] \right\} \; \Psi(r,\theta) = 0 \tag{26}$$

and for $r \neq 0$, the solution can be written in terms of the eigenfunctions of this equation:

$$\Psi(r,\theta) \;=\; \sum_{lk} c_{lk}\psi_{kl}(r,\theta) \;=\; \sum_l c_l e^{il\theta} \, J_{|l+\bar{\varphi}|}(kr), \qquad (\text{if } r = r_0) \tag{27}$$

with the eigenvalues:

$$\begin{aligned} \hat{H}\psi_{lk}(r,\theta) \;&=\; \epsilon_{lk}\psi_{lk}(r,\theta) \\ \epsilon_{lk} \;&=\; \hbar^2 k^2/2m \end{aligned} \tag{28}$$

from which we have

$$G(r_2 r_1;\theta_2\theta_1;t) \;=\; \frac{1}{2\pi}\sum_l \int k\,dk\, J_{|l+\bar{\varphi}|}(kr_1)\, J_{|l+\bar{\varphi}|}(kr_2)\; e^{i[(\theta_2-\theta_1)l-\epsilon_{lk}t]} \tag{29}$$

However this part is easy. More messy is the determination of the scattering amplitude $f_k(\theta)$, defined by the asymmetric form of $\Psi(r,\theta)$ according to

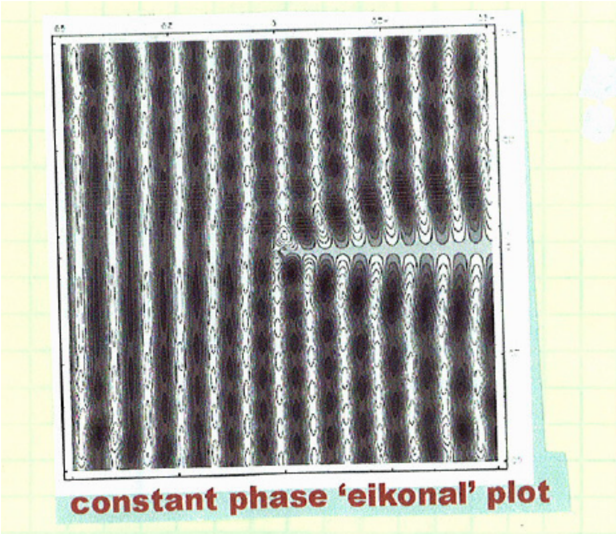
$$\Psi(r,\theta) \;=\; e^{ikx} + \frac{1}{\sqrt{r}} f_k(\theta)\, e^{ikr}, \qquad quad(\text{when } r \to \infty\ ) \tag{30}$$

This was the problem that Pryce solved for Aharonov, to give the result

$$f_k(\theta) \;\simeq\; -\frac{e^{i\pi/4}}{\sqrt{2\pi k}} \sin \pi\bar{\varphi}\; \frac{e^{i\frac{\theta}{2}}}{\cos\frac{\theta}{2}} \tag{31}$$

The form of this result is rather complex, and is illustrated in the figure. From the result in (31) we see that the scattering amplitude is periodic in the flux $\bar{\varphi}$; indeed, when $\bar{\varphi} = n$, an integer, it has no effect at all.

Notice that (31) diverges for $\theta \to \pi$, and in fact it breaks down for both forward and backward scattering. The figure gives an idea of how it actually behaves.



Of course what is so strange about this result is that it shows that the particle dynamics, in the quantum theory of this problem, is controlled not by $\mathbf{E}(\mathbf{r})$ or $\mathbf{B}(\mathbf{r})$, but by $\mathbf{A}(\mathbf{r})$. This means that in the quantum theory, it is $\mathbf{A}(\mathbf{r})$ (and more generally $\mathbf{A}(\mathbf{r}, t)$) that is fundamental object, and not $\mathbf{B}(\mathbf{r}, t)$ or $\mathbf{E}(\mathbf{r}, t)$, which are merely derived from $\mathbf{A}(\mathbf{r}, t)$.

This is the exact opposite of classical EM theory. There, the fundamental physical quantities are $\mathbf{B}(\mathbf{r}, t)$ and $\mathbf{E}(\mathbf{r}, t)$; and from the Lorentz eqtn. (24), we see that these 2 fields entirely control the dynamics of electric charge.

(iii) Classical EM vs. QED: What is the reason for this fundamental difference, between classical EM theory and its quantized version? For many years this was not properly understood, and yet the reason is to be found in equation (21). We notice that the term in the exponential appears without an accompanying $\hbar$, which is actually buried in the definition of $\bar{\varphi}$; let's rewrite (21) as

$$G(\theta, t; \bar{\varphi}) \;=\; A(t) \sum_{n=-\infty}^{\infty} \exp \frac{i}{\hbar} \left\{ n\hbar\bar{\varphi} + I \frac{(\theta + 2\pi n)^2}{2t} \right\} \tag{32}$$

where

$$A(t) \;=\; \left(\frac{I_0}{2\pi i \hbar t} \right)^{\frac{1}{2}} \tag{33}$$

Now suppose we consider the limit $\hbar \to 0$. It is singular; the phase in the exponent diverges. However we notice that the term in $\bar{\varphi}$ is independent of $\hbar$, and it disappears from the exponent. This is because there are actually 2 quantum parameters here, $\hbar$ and q, and we are perfectly entitled to treat them as independent. Thus let's write (32) as

$$
\begin{aligned}
G(\theta,t;\hbar,q) &= A(\hbar t) \sum_{n=-\infty}^{\infty} e^{\frac{i}{2} n q \Phi/\pi} \; e^{\frac{i}{2} I_0 \frac{(\theta+2n\pi)^2}{\hbar t}} \\
&\equiv A(\hbar t) \sum_{n=-\infty}^{\infty} e^{i n \omega_c(q\Phi)} \; e^{i\psi_n(\hbar t,\theta)} \qquad (34)
\end{aligned}
$$

where the prefactor $A(\hbar t)$ and the phase $\psi_n(\hbar t,\theta)$ both diverge as $\hbar t \to 0$:

$$
\begin{aligned}
A(\hbar t) &= \left(\frac{I_0}{2\pi i \hbar t}\right)^{\frac{1}{2}} \xrightarrow[\hbar t \to 0]{} \infty \\
\psi_n(\hbar t,\theta) &= \frac{I_0}{2}\frac{(\theta+2n\pi)^2}{\hbar t} \xrightarrow[\hbar t \to 0]{} \infty \qquad (35)
\end{aligned}
$$

whereas the *topological phase* $\omega_c(q\Phi)$ does not:

$$
\omega_c = \frac{q\Phi}{2\pi}, \qquad \text{(independent of } \hbar,\ t). \qquad (36)
$$

Thus we see already, from looking at a simple non-relativistic problem in the motion of a charged particle coupled to a static EM field, that there is a crucial term in the dynamics that exists in the quantum version of the theory, but not in the classical.

If we think about this a little more, it is not all that obvious why there ought to be any relationship between the quantum and classical versions of electrodynamics. After all, classical EM theory makes no reference to Dirac electrons and holes, which exist in QED, and it is not obvious why the form of the Lagrangian or action for the 2 theories should look the same. Thus we can, if we like, start by asking - why should there be a quantum generalization of a classical gauge theory like QED, and what should it look like?

B.4.2 DERIVATION of GAUGE FIELD THEORIES

In this sub-section we will give some of the well-known arguments that are used to derive the existence and form of gauge theories. A key part of what follows is the comparison between the quantum and classical versions of these theories. I will focus mainly on the general form of these theories, with not too much attention paid to examples. Note, as already emphasized in the previous sub-section, that physics is full of different sorts of gauge theory - why this is so should become clear in the following.

B.4.2 (a) QUANTUM ELECTRODYNAMICS: a $U(1)$ GAUGE THEORY

QED is the simplest sort of gauge theory - we shall see why this is in the course of the derivations to be given. It is important in what follows to compare the classical and quantum versions of electrodynamics, so we begin by reviewing classical EM theory. In what follows it will be assumed that everyone is familiar with special relativity and relativistic notation.

(i) Classical Electrodynamics: Classical Electrodynamics is a theory of sources coupled to fields, and it can be phrased entirely in terms of the fields $\mathbf{E}(x)$ and $\mathbf{B}(x)$, and the charge 4-current $J^\mu(x) = (\rho(x), \mathbf{J}(x))$. However, as is well known, this is not the best way to formulate it, and it is not ideal for the generalization to QED.

Let's see how we can set up classical EM theory starting with an experimental fact, viz., that there exist fields $\mathbf{E}(x)$ and $\mathbf{B}(x)$, and that they act upon electric charges according to the Lorentz force law in (24) above. In 4-vector notation we rewrite the Lorentz equation as

$$f^\mu = qF^{\mu\nu}u_\nu \tag{37}$$

where u_ν is the 4-velocity vector, i.e.,

$$u_\nu = \frac{dx^\nu}{d\tau} \tag{38}$$

where τ is the proper time interval (on a worldline). In what follows will assume the following conventions; the infinitesimal interval $d\mathbf{s}$ is related to the metric by

$$c^2 d\tau^2 \;=\; ds^2 \;=\; g_{\mu\nu} dx^\mu\, dx^\nu \;\xrightarrow[\text{flat space}]{}\; \eta_{\mu\nu} dx^\mu\, dx^\nu \tag{39}$$

where the coordinate 4-vector is $x^\mu = (x^0;\, x^1,\, x^2,\, x^3) = (ct;\, x,\, y,\, z)$, and the flat space Minkowski metric is taken to be

$$\eta^{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{40}$$

with signature -2. If in some general pseudo-Riemannian spacetime we define a set of basis vectors $\mathbf{e}_\mu$ (cotravariant) and $\mathbf{e}_\nu$ (covariant), then

$$\begin{aligned} d\mathbf{s} \;&=\; \mathbf{e}_\mu dx^\mu \;=\; \mathbf{e}^\mu dx_\mu \\ g^{\mu\nu} &= \mathbf{e}^\mu \cdot \mathbf{e}^\nu, \qquad g_{\mu\nu} = \mathbf{e}_\mu \cdot \mathbf{e}_\nu, \qquad g^\mu_\nu = \mathbf{e}^\mu \cdot \mathbf{e}_\mu = \delta^\mu_\nu \end{aligned} \tag{41}$$

The 4-vectors of main interest to us will be:

4-velocity: $\mathbf{u} = u^\nu \mathbf{e}_\nu$, with $u^\nu = \gamma_u(c, \mathbf{u})$, where $\gamma_u = \big(1 - \mathbf{u}^2\big)^{-1/2}$, with $\bar{\mathbf{u}} = \mathbf{u}/c$, and with individual components $\mathbf{u} = \big(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\big)$.

Note we then have $d\tau = dt/\gamma_u$, and also

$$u^2 \;=\; u^\nu u_\nu \;=\; \left(\frac{ds}{d\tau}\right)^2 \;=\; c^2 \tag{42}$$

4-acceleration:

$$\mathbf{a} = a^\nu \mathbf{e}_\nu, \quad \text{where} \quad a^\nu \;=\; \frac{d^2x^\nu}{d\tau^2} \;=\; \frac{du^\nu}{d\tau} \tag{43}$$

so that

$$u_\nu a^\nu \;=\; 0 \tag{44}$$

4-current density:

$$\mathbf{J}(x) \;=\; \rho_0(x)\mathbf{u}(x) \;=\; J^\nu(x)\mathbf{e}_\nu(x) \tag{45}$$

where $\rho_0(x)$ is the proper charge density; then

$$J^\nu \;=\; \rho_0\gamma_u(c, \mathbf{u}) \;=\; (\rho c,\, \jmath) \tag{46}$$

where $\rho(x) = \gamma_u\rho_0(x)$, and $\mathbf{j}(x) \;=\; \rho(x)\mathbf{u}(x)$ is the 3-current density; from eq. (42), we have

$$\mathbf{J}^2 \;=\; J^\mu J_\nu \;=\; \rho^2c^2 - \jmath^2 \tag{47}$$

Returning now to the Lorentz force equation, we notice that a contraction of (37) with a 4-vector u_μ gives $f^\mu u_\mu = qF^{\mu\nu}u_\mu u_\nu = 0$, because a 4-force, like a 4-acceleration, must be perpendicular to its associated velocity (cf. eq. (44)); thus $F_{\mu\nu}$ must be antisymmetric, i.e.,

$$F_{\mu\nu} = -F_{\nu\mu} \tag{48}$$

We have introduced the tensor $F_{\mu\nu}(x)$ as a way of describing the Lorentz force law in a relativistically invariant way - the presence of the velocity vector $\dot{\mathbf{r}} = \mathbf{u}$ in (24) naturally leads to a formulation in terms of the velocity 4-vector $\mathbf{u}(x)$, and then to an equation of the form in (37). If we now want to recover (24), we write $F_{\mu\nu}(x)$ in the form

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & -B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & B_y & B_x & 0 \end{pmatrix} \tag{49}$$

Now, we also wish to find a source equation for the EM field - we may treat this as a general theoretical requirement, or again base it on experiment (which show that EM fields act on charges, and are acted on by them as well). Rather than appeal to experiment, we can instead simply ask what is the natural relativistically invariant form required for electric charge to generate the field. Since the charge $\rho(x)$ is simply one component of the 4-current $\mathbf{J}(x)$, we look for a linear equation relating $F^{\mu\nu}(x)$ to $J^\mu(x)$. The obvious way to do this is to write

$$\partial_\mu F^{\mu\nu}(x) \;=\; \kappa J^\nu(x), \qquad \kappa = \begin{Bmatrix} 4\pi/c & \text{(cgs)} \\ \mu_0 & \text{(MKS)} \end{Bmatrix} \tag{50}$$

(any other 4-vector A_μ contracted with $F_{\mu\nu}(x)$ would do, but experiment reveals no other such quantities); the constant κ is given by experiment.

At first it might seem that eqs. (37) and (50) characterize the theory properly. But actually the components of $F_{\mu\nu}$ are not independent of each other - there are 6 apparently independent components in (49), but only 4 independent quantities in (50). One can either argue that $F_{\mu\nu}(x)$ has to be therefore constructed from a 4-vector, or appeal to experiment for the relationship between the components of $F_{\mu\nu}(x)$. Both lines of argument lead to the conclusion that we can write $F_{\mu\nu}(x)$ in the form

$$F_{\mu\nu}(x) \;=\; \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) \tag{51}$$

where

$$A^\mu(x) \;=\; \left\{ \begin{array}{lr} \big(\phi(x)/c,\ \mathbf{A}(x)\big) & \text{(MKS)} \\ \big(\phi(x),\ \mathbf{A}(x)\big) & \text{(cgs)} \end{array} \right\} \tag{52}$$

and $\phi(x)$, $\mathbf{A}(x)$ are the electric and magnetic potentials. Another way to enforce the restriction to 4 independent components is via the Bianchi identity, written as

$$\partial_{[\lambda} F_{\mu\nu]} \;\equiv\; \partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} + \partial_\mu F_{\nu\lambda} \;=\; 0 \tag{53}$$

which actually follows directly from (51). There are thus 2 different ways we can write the field equations of classical electromagnetism; we assume:

$$\begin{array}{c} \partial_\mu F^{\mu\nu}(x) \;=\; \mu_0 J^\nu(x) \\ \textbf{and} \\ F^{\mu\nu} \;=\; (\partial^\mu A^\nu - \partial^\nu A^\mu) \qquad \textbf{OR} \qquad \partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} + \partial_\mu F_{\nu\lambda} \;=\; 0 \end{array} \tag{54}$$

So far so good. Now observe that we have a rather peculiar theory, for the key observable variables are $\mathbf{E}(x)$ and $\mathbf{B}(x)$, and of course $\mathbf{J}(x)$, but the underlying field variable for the EM field, which is not observable, is $A^\mu(x)$.

It then follows that if we make any changes of variables, coordinate transformations, etc., the quantities $F_{\mu\nu}(x)$ and $\mathbf{J}(x)$ ought to be invariant, but $A^\mu(x)$ does not have to be. This immediately leads to the possibility of Gauge Transformations. Suppose we make the transformation

$$A^\mu(x) \;\longrightarrow\; \tilde{A}^\mu(x) \;=\; A^\mu(x) + \alpha^\mu(x) \tag{55}$$

Then we have

$$\tilde{F}_{\mu\nu} \;=\; (\partial_\mu A_\nu - \partial_\nu A_\mu) \;+\; (\partial_\mu \alpha_\nu - \partial_\nu \alpha_\mu) \tag{56}$$

and for $\tilde{F}_{\mu\nu} = F_{\mu\nu}$, we require $\alpha^\mu = \partial^\mu \chi$, so:

$$\tilde{A}^\mu(x) \;=\; A^\mu + \partial^\mu \chi(x) \tag{57}$$

Thus any one of an infinite set of functions $\tilde{A}^\nu(x)$ is just as valid as $A^\mu(x)$ for the treatment of the classical EM fields.

We now wish to write an action functional for the classical EM theory, in preparation for the transition to QED. There will be 2 parts to this action, the matter part and the field part.

Particle Action in EM theory: For a set of particles of mass m_j, at spacetime coordinate x_j, it is well known that the action has the form

$$S_p \;=\; -\sum_j m_j c \int ds_j \;=\; -c^2 \sum_j \int dt_j\,(1-u_j^2/c^2)^{1/2} \tag{58}$$

Suppose we ignore this term, concentrating only on the electromagnetic part that comes from the interaction of the current $J^\mu(x)$ with the gauge field $A_\mu(x)$, which from now on we take to be the fundamental EM field. The simplest scalar term for a Lagrangian density combining these two is just $J_\mu A^\mu$, and actually experiment tells us that we should have

$$S_{int} \;=\; -\int d^4x\; J^\mu(x) A_\mu(x) \tag{59}$$

At first glance this coupling seem problematic, because it is apparently not invariant under the gauge transformation of (57). However, one can actually show that it is gauge invariant, as follows. Suppose we gauge transform (59):

$$\begin{aligned} A^\mu(x) \;\to\; A^\mu + \partial^\mu \psi(x) \qquad\qquad &\Longrightarrow \qquad\qquad S_{int} \to \tilde{S}_{int} \\ \tilde{S}_{int} \;=\; -\int d^4x\, J^\mu(A_\mu + \partial_\mu \chi) \;&=\; -\int d^4x\; [J^\mu A_\mu + \partial_\mu(J^\mu \chi) - (\partial_\mu J^\mu)\chi] \end{aligned} \tag{60}$$

Now the 2nd term, $\partial_\mu(J^\mu\chi)$, can be rewritten as a surface integral of $J^\mu\psi$ at infinity, and we will assume no current sources at infinity. This leaves the 3rd term, and we see the coupling term (59) will be gauge invariant if the current conservation equation

$$\partial_\mu J^\mu(x) = 0 \tag{61}$$

is satisfied. But the truth of this is easily demonstrated, starting from the 1st equation of motion in (54); for if we take the derivative ∂_ν of (54), the left side of (54) must vanish because of the antisymmetric property (48) of the field $F_{\mu\nu}(x)$.

Field action in EM theory: We want to find a scalar quantity which is gauge invariant for the EM field action. This time we cannot get away with using combinations of $A_\nu(x)$, because any combination like $A^\nu A_\nu$ will not be gauge invariant. Thus we must resort to $F_{\mu\nu}$, and the correct form of the field action is in fact

$$S_{EM} \;=\; -\frac{1}{4\mu_0}\int d^4x\, F^{\mu\nu}(x) F_{\mu\nu}(x) \qquad \text{(MKS units)} \tag{62}$$

where the prefactor is determined by experiment.

Thus we find that the total EM action is

$$S_{cl}[J^\mu,\, A^\mu] \;=\; -\int d^4x \left\{ \frac{1}{4\mu_0} F^{\mu\nu}(x) F_{\mu\nu}(x) + J_\mu(x) A^\mu(x) \right\} \tag{63}$$

and, by varying this action, we can recover the equations of motion.

(ii) Quantum Electrodynamics: : Now let us start again, this time with the classical electrons of EM theory replaced by the electrons described by the Dirac equation for a spin-1/2 fermionic field having the bare action

$$S_0 \; = \; \int d^4x \, \bar{\psi}(x) \, [i\gamma^\mu \partial_\mu - m] \, \psi(x) \tag{64}$$

What we now wish to show is that a quantum action corresponding to the classical action above can be derived for the Dirac electron coupled to an EM field - but the arguments used are a little different (at least at first glance). They have the great advantage of being easily generalizable to a variety of matter fields (although one can quibble rather strongly about this in the case of quantum gravity).

We begin by noting that one can realize a simple "gauge transformation" on (64), because the phase of $\psi(x)$ is arbitrary. Thus we make the global gauge transformation

$$\psi(x) \; \longrightarrow \; e^{-i\theta} \psi(x) \tag{65}$$

Under this transformation, S_0 is invariant; the system possesses a global $U(1)$ symmetry. However, suppose we allow θ to depend on the spacetime coordinates; this move is of course highly non-trivial, and is suggested by the following considerations:

(i) the original motivation of Weyl, in the 1920's; since one can envisage, in general relativity, a charge of "coordinate measure" as one moves around in space, why not the same for phase measure (here the name "gauge", meaning a measurement scale).

(ii) Quantum Mechanics naturally suggests, particularly in its path integral form, the role of phase as a fundamental variable in the theory. This idea reappears in the Aharonov-Bohm effect.

We shall see that the consequences of introducing such a local gauge transformation give further reason to look at it. So we now consider the transformation:

$$\psi(x) \; \longrightarrow \; e^{i\theta(x)} \psi(x) \tag{66}$$

We immediately see that S_0 is not invariant under this transformation; in fact we get

$$\begin{aligned} S_0 \;\; &\rightarrow \;\; \int d^4x \, \bar{\psi}(x) \, \left[i\gamma^\mu \big(\partial_\mu - i\partial_\mu \theta(x) \big) - m \right] \psi(x) \\ &= \;\; S_0 + \int d^4x \, \bar{\psi}(x) \gamma^\mu \partial_\mu \theta(x) \psi(x) \end{aligned} \tag{67}$$

However, we can have a gauge-invariant theory if we do the following:

- We replace ∂_μ by a "gauge-covariant" derivative D_μ so that $D_\mu \psi(x) \rightarrow \; e^{-i\theta(x)} D_\mu \psi(x)$ is invariant. This will work if

- we write $D_\mu = \partial_\mu + iqA_\mu$, and at the same time suppose that the field $A_\mu(x)$ transforms according to

$$A_\mu(x) \; \longrightarrow \; A_\mu(x) + \frac{1}{q} \partial_\mu \theta(x) \tag{68}$$

thereby cancelling the extra term in (67).

Thus we are led to replace (64) by

$$\begin{aligned} S_0 &= \int d^4x\, \bar{\psi}(x)\, [i\gamma^\mu D_\mu - m]\, \psi(x) \\ &= \int d^4x\, \bar{\psi}(x)\, \left[i\gamma^\mu \big(\partial_\mu + iqA_\mu(x)\big) - m\right] \psi(x) \end{aligned} \quad (69)$$

and then, recognizing that the field $A_\mu(x)$ must have its own individual term in the Lagrangian (it is a now field), which must be gauge invariant, we are finally led to the total QED action:

$$S_{QED} = \int d^4x \left\{ \bar{\psi}(x)\, [i\gamma^\mu \partial_\mu - M]\, \psi(x) - J_\mu(x) A^\mu(x) - \frac{1}{4\mu_0} F^{\mu\nu}(x) F_{\mu\nu}(x) \right\} \quad (70)$$

where the Dirac current $J_\mu(x)$ is

$$J_\mu(x) = q\bar{\psi}(x)\gamma_\mu \psi(x) \quad (71)$$

Thus we have shown that the gauge invariance of the Dirac electron terms in the action, under a local phase transformation, leads naturally to the existence of a gauge field $A^2(x)$ of the same kind as appears in the original classical action for the EM field! We now see another reason to take all this seriously; the gauge invariance naturally prevents the existence of a term $\sim A^\nu(x)$, which would give the photon a mass, something excluded by experiment. It is also interesting to note that the field $F_{\mu\nu}(x)$ acquires a new significance in the quantum mechanical theory (which in a path integral formulation can be related directly back to the classical theory, as we shall see). Consider the action of the *curvature* operator

$$\hat{R}_{\mu\nu} = [D_\mu, D_\nu] \equiv D_\mu D_\nu - D_\nu D_\mu \quad (72)$$

on the wave-function; we immediately find that

$$\hat{R}_{\mu\nu}\, \psi(x) = iqF_{\mu\nu}(x)\, \psi(x) \quad (73)$$

so that the different components of the field intensity $F_{\mu\nu}(x)$ (i.e., the "physical" fields $\mathbf{E}(x)$, $\mathbf{B}(x)$) are just components of the curvature. Thus we can argue that in reality we have

- $A^\mu(x)$: The underlying "EM field", or "EM vacuum", which is not directly accessible to us in classical EM theory; it is accessible as a phase variable in QM, modulo gauge transformation;

- $F_{\mu\nu}(x)$: The "curvature" - a kind of generalized "polarization" - of $\mathbf{A}(x)$, with components $\mathbf{E}(x)$, $\mathbf{B}(x)$, which is directly accessible to us via its effects on charge, in both classical EM and QED.

We note of course that $F_{\mu\nu}(x)$ is gauge invariant, as we have already seen. Notice two key features of this $U(1)$ theory:

(i) The form of the coupling to the gauge field (indeed, the existence of the gauge field, in the argument) arises solely from the transformation of the matter field under local phase transformation; one ends up with a covariant or "minimal" coupling. The requirement of gauge invariance then uniquely determines the coupling (as well as the masslessness of the gauge field).

(ii) The photon does not couple to itself. There is no reason coming from gauge invariance why this "photon self-coupling" does not happen (it could, e.g., come from higher powers of $F_{\mu\nu}(x)$ in the action); nevertheless there are no such terms in the bare action.

There is also a 3rd key feature which we will come to below, concerning the connection between gauge transformations and conservation laws (Noether's Theorem).

B.4.2 (b) NON-ABELIAN GAUGE FIELDS - YANG-MILLS THEORY

In their pioneering paper published in 1954, Yang and Mills vastly extended the ideas of gauge fields beyond the $U(1)$ discussion given above. This paper, inspired by ideas from GR as well as by questions arising in particle physics, was almost entirely ignored for a decade - it was well ahead of its time. There was nothing in classical physics at that time which suggested such a quantum theory, apart from GR - although now we have very nice examples from condensed matter physics (e.g., superfluid ^3He, where the quantum order parameter obeys a set of classical equations of motion which look like a non-Abelian gauge theory on a background dynamics curved spacetime). The main reasons that the Yang-Mills theory received so little attention until the early-mid 1960's were

(i) Almost nobody was interested in or familiar with classical GR; and quantum gravity hardly existed even as an idea.

(ii) The Yang-Mills (YM) theory predicted massless particles. Adding a mass term broke the SU(N) gauge invariance. No mechanism at that time was known, at least in particle physics, that would break this invariance in a physically satisfactory way.

(iii) Nobody knew how to calculate with such theories.

As we will see, objections (i) and (ii) were slowly overcome; the key to solving (ii) was the Anderson-Higgs mechanism, implemented for YM theories by Salam and Weinberg. The key to (iii) was the use of path integral methods, without which even QED remained hard to understand. The main initiative came from 't Hooft, and from 't Hooft and Veltman.

In what follows we will begin by going through the important special case of $SU(2)$ gauge symmetry, which is a nice pedagogical example. I will then sketch how this is generalized to higher non-Abelian groups, and discuss the physical significance of all this.

(i) $SU(2)$ Gauge Theory: Recall that the simple $SU(2)$ group can be represented by the Pauli matrices $\hat{\tau}_j^{\alpha\beta}$, with

$$\begin{aligned} \hat{\tau}_i \hat{\tau}_j &= \delta_{ij} + i\epsilon_{ijk}\hat{\tau}_k \\ [\hat{\tau}_i, \, \hat{\tau}_j] &= 2i\epsilon_{ijk}\hat{\tau}_k \end{aligned} \qquad (81)$$

and we may define the unitary operator $\hat{U}(\mathbf{\Omega}) = \hat{U}(\hat{\mathbf{n}}\Omega)$, where $\hat{\mathbf{n}}$ is a unit vector on the Bloch sphere, as

$$\hat{U}(\mathbf{\Omega}) \;=\; e^{-\frac{i}{2}\mathbf{\Omega}\cdot\boldsymbol{\tau}} \;=\; e^{-\frac{i}{2}\Omega_j\tau_j^{\alpha\beta}} \;\;\equiv\;\; \lim_{N\to\infty}\big(1-\frac{i}{2N}\mathbf{\Omega}\cdot\boldsymbol{\tau}\big)^N \tag{82}$$

where the last form gives us the infinitesimal operator. Note that $\hat{U}(\mathbf{\Omega}) = \hat{U}_{\alpha\beta}(\mathbf{\Omega})$ is a matrix operator in "spin space", and the unit vector $\hat{\mathbf{n}}$ tells us the direction around which $\mathbf{\Omega}$ is effecting a rotation.

We may now go through much the same manoeuvres as we did for the Abelian gauge field. We introduce in this case a 2-component spinor Dirac field, $\psi_\alpha(x)$, and note that the Lagrangian

$$\mathcal{L}_0 \;=\; \bar{\psi}_\alpha(x)\left[i\gamma^\mu\partial_\mu\delta^\alpha_\beta - m\delta^\alpha_\beta\right]\psi^\beta(x)$$

is invariant under a global $SU(2)$ rotation. However, let us now apply the local operator, where $\mathbf{\Omega}\to\mathbf{\Omega}(x)$; then

$$\begin{aligned}\hat{U}\big(\mathbf{\Omega}(x)\big)\;\psi(x) \;&=\; U^{\alpha\beta}\big(\mathbf{\Omega}(x)\big)\;\psi_\beta(x)\\ &= e^{-\frac{i}{2}\mathbf{\Omega}(x)\cdot\boldsymbol{\tau}^{\alpha\beta}}\psi_\beta(x) \;=\; \tilde{\psi}_\alpha(x)\end{aligned} \tag{83}$$

and so we find that $\mathcal{L}_0$ is not invariant, for the same reason as before, i.e., we get an extra gradient term:

$$S_0 \quad\longrightarrow\quad S_0 + \int d^4x\;\bar{\psi}^\alpha(x)\left[U^{-1}_{\alpha\gamma}\big(\mathbf{\Omega}(x)\big)\partial_\mu U^{\gamma\beta}\big(\mathbf{\Omega}(x)\big)\right]\psi_\beta(x) \tag{84}$$

We therefore introduce the gauge covariant derivative, with a charge g_0:

$$\begin{aligned}D_\mu \;&=\; \big(\partial_\mu + i\frac{g_0}{2}\boldsymbol{\tau}\cdot\mathbf{A}_\mu(x)\big)\\ D^{\alpha\beta}_\mu \;&=\; \big(\partial_\mu\delta^{\alpha\beta} + i\frac{g_0}{2}\tau_j^{\alpha\beta}\cdot A^j_\mu(x)\big)\end{aligned} \tag{85}$$

where we use a vector notation $\boldsymbol{\tau}^{\alpha\beta} \;=\; (\tau_j^{\alpha\beta},\tau_x^{\alpha\beta},\tau_y^{\alpha\beta}) \;\equiv\; \tau_j^{\alpha\beta} \;\to\; \boldsymbol{\tau}$, so as to suppress the clutter of indices, with the boldface indicating a vector in real spacetime. Since $\bar{\psi}^\alpha(x)$ transforms to $\tilde{\bar{\psi}}_\alpha = u^{-1}_{\alpha\beta}\bar{\psi}_\beta$, we clearly want a transformation such that

$$D_\mu\psi \;\longrightarrow\; U\big(\mathbf{\Omega}(x)\big)D_\mu\psi \tag{86}$$

and, going through the algebra, analogous to that for $U(1)$ gauge fields, we find that

$$\frac{1}{2}\boldsymbol{\tau}\cdot\mathbf{A}_\mu \;\;\longrightarrow\;\; U(\mathbf{\Omega})\frac{1}{2}\boldsymbol{\tau}\cdot\mathbf{A}_\mu U^{-1}(\mathbf{\Omega}) \;+\; ig_0^{-1}\big(\partial_\mu U(\mathbf{\Omega})\big)U^{-1}(\mathbf{\Omega}) \tag{87}$$

or, for an infinitesimal transformation $\delta\mathbf{\Omega}(x)$ (cf. eq. (82)), we get

$$\begin{aligned}\frac{1}{2}\boldsymbol{\tau}\cdot\mathbf{A}_\mu \;&\longrightarrow\; \frac{1}{2}\boldsymbol{\tau}\cdot\mathbf{A}_\mu \;+\; \frac{1}{2}\boldsymbol{\tau}\cdot(\delta\mathbf{\Omega}\times\mathbf{A}_\mu) \;+\; \frac{1}{2g_0}\boldsymbol{\tau}\cdot\partial_\mu\delta\mathbf{\Omega}\\ \frac{1}{2}\tau_i A^i_\mu \;&\longrightarrow\; \frac{1}{2}\tau_i A^i_\mu \;+\; \frac{1}{2}\epsilon_{ijk}\tau^i\delta\Omega^j A^k_\mu \;+\; \frac{1}{2g_0}\tau_i\partial_\mu\delta\Omega^i\end{aligned} \tag{88}$$

where we write out all the components in the 2nd form. Thus we can now write everything in terms of a new non-Abelian gauge field, given by (85), with the transformation property

$$\begin{aligned} \mathbf{A}_\mu(x) &\rightarrow \mathbf{A}_\mu(x) + \delta\mathbf{A}_\mu(x) \\ \delta\mathbf{A}_\mu(x) &= \big(\delta\mathbf{\Omega}(x) \times \mathbf{A}_\mu(x)\big) \; + \; \frac{1}{g_0}\partial_\mu \delta\mathbf{\Omega}(x) \end{aligned} \quad (89)$$

and we can immediately generalize this from the infinitesimal transformation to the general transformation

$$\begin{aligned} \mathbf{A}_\mu(x) &\longrightarrow \mathbf{A}_\mu(x) + \frac{1}{g_0}\partial_\mu \delta\mathbf{\Omega}(x) \; + \; \big(\mathbf{\Omega}(x) \times \mathbf{A}_\mu(x)\big) \\ A^i_\mu(x) &\longrightarrow A^i_\mu(x) + \frac{1}{g_0}\partial_\mu \delta\Omega^i(x) \; + \; \big(\epsilon^i_{jk}\Omega^j(x) A^k_\mu(x)\big) \end{aligned} \quad (90)$$

which we should compare with the Abelian transformation in (68). We can also relate this to a "curvature tensor", or field intensity tensor; as with (72) above, we have an operator

$$R^{\alpha\beta}_{\mu\nu} \;=\; D^{\alpha\gamma}_\mu D^{\gamma\beta}_\nu - D^{\alpha\gamma}_\nu D^{\gamma\beta}_\mu \quad (91)$$

and applying this to $\psi_\alpha(x)$, we find that

$$R^{\alpha\beta}_{\mu\nu}\; \psi_\beta(x) \;=\; \frac{ig_0}{2}(\boldsymbol{\tau}^{\alpha\beta} \cdot \mathbf{F}_{\mu\nu})\; \psi_\beta(x) \quad (92)$$

where we define

$$\begin{aligned} \mathbf{F}_{\mu\nu} &= \big(\partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu\big) \; + \; g(\mathbf{A}_\mu \times \mathbf{A}_\nu) \\ F^i_{\mu\nu} &= \big(\partial_\mu A^i_\nu - \partial_\nu A^i_\mu\big) \; + \; g\epsilon^i_{jk} A^j_\mu A^k_\nu \end{aligned} \quad (93)$$

Unlike the Abelian $F_{\mu\nu}(x)$, this tensor is not gauge-invariant, as we see by making the transformation; we have

$$\begin{aligned} \boldsymbol{\tau} \cdot \mathbf{F}_{\mu\nu}(x) &\longrightarrow U\big(\mathbf{\Omega}(x)\big)\; (\boldsymbol{\tau} \cdot \mathbf{F}_{\mu\nu}(x))\; U^{-1}\big(\mathbf{\Omega}(x)\big) \\ \mathbf{F}_{\mu\nu}(x) &\longrightarrow \mathbf{F}_{\mu\nu}(x) \; + \; \Omega(x) \times \mathbf{F}_{\mu\nu}(x) \end{aligned} \quad (94)$$

However, the analogue of the Abelian field action term in (63) and (70) is gauge-invariant; i.e., the term

$$\mathrm{Tr}\,\{(\boldsymbol{\tau} \cdot \mathbf{F}_{\mu\nu})\; (\boldsymbol{\tau} \cdot \mathbf{F}^{\mu\nu})\} \;=\; \frac{1}{2} F^i_{\mu\nu} F^{\mu\nu}_i \quad (95)$$

is gauge-invariant. Thus we are led to the form of our spinor generalization of the Abelian gauge theory, taking the form of an action in which a vector gauge field $\mathbf{A}_\mu(x)$ is coupled to a spinor fermion field $\psi(x)$, with the action

$$S\,[\bar{\psi}, \psi;\, \mathbf{A}_\mu] \;=\; \int d^4x\; \psi(x)\, [i\gamma^\mu D_\mu - m]\, \psi(x) \;-\; \frac{1}{4} F^i_{\mu\nu}(x) F^{\mu\nu}_i(x) \quad (96)$$

with D_μ given by (85). Now we would like to write this in a way analogous to (63), using a current operator. We can do this if we define

$$\begin{aligned} \mathbf{J}^\mu(x) \;\equiv\; J_i^\mu(x) &= \frac{g}{2}\bar{\psi}(x)\gamma^\mu\,\boldsymbol{\tau}\,\psi(x) \\ &= \frac{g}{2}\bar{\psi}_\alpha(x)\gamma^\mu\tau_i^{\alpha\beta}\psi_\beta(x) \end{aligned} \tag{97}$$

A little later we will see how such a choice can be justified (and in the same way justify the choice (71) for Abelian QED). In any case, with this choice we can write the final form for the action:

$$S \;=\; \int d^4x \left\{ \bar{\psi}_\alpha(x)\left[i\gamma^\mu\partial_\mu - m\right]\psi^\alpha(x) \;-\; \mathbf{J}^\mu(x)\cdot\mathbf{A}_\mu(x) \;-\; \frac{1}{4}\mathbf{F}_{\mu\nu}(x)\mathbf{F}^{\mu\nu}(x) \right\} \tag{98}$$

with the quantities defined as we have already seen.

This concludes the discussion of a simple $U(1)$ non-Abelian gauge theory, which has the advantage of being easily understandable in temrs of spinors. Let us now move to a more general discussion.

(ii) General Non-Abelian Gauge Theories: We can reformulate all of this for a general non-Abelian gauge group. Thus, one can imagine some simple Lie group G with generators $\{g_a\}$ satisfying the algebra

$$[g_a,\, g_b] = i f_{abc} g^c \tag{99}$$

and we will represent this Lie algebra with matrices $\mathbf{T}$ which operate on a n-dimensional fermion field $\boldsymbol{\Psi}(x) = \Psi_a(x)$. The matrices $\mathbf{T}$ then have the commutation relation

$$[T_a,\, T_b] = i f_{abc} T^c \tag{100}$$

and the general local gauge transformation will act on $\boldsymbol{\Psi}(x)$ according to

$$\hat{U}\big(\boldsymbol{\Lambda}(x)\big)\boldsymbol{\Psi}(x) = e^{-ig_0\mathbf{T}\cdot\boldsymbol{\Lambda}(x)} \tag{101}$$

where we can think of $\boldsymbol{\Lambda}(x)$ as a "hyperangle" in the n-dimensional space. We introduce a generalized gauge field $\mathbf{A}^\mu$ with components $A_a^\mu(x)$, which transform according to

$$A_a^\mu \;\longrightarrow\; A_a^\mu + \partial^\mu\Lambda_a \;+\; g_0 f_{abc}\Lambda^b A_c^\mu \tag{102}$$

and we define the gauge covariant derivative

$$D^\mu \;=\; \big(\partial^\mu + ig_0\mathbf{T}\cdot\mathbf{A}^\mu\big) \tag{103}$$

Then the action of the commutator $R^{\mu\nu} = [D^\mu,\, D^\nu]$ on the fermion field is given by

$$R^{\mu\nu} \;=\; [D^\mu,\, D^\nu] \;=\; ig_0\mathbf{T}\cdot\mathbf{F}^{\mu\nu} \tag{104}$$

with

$$\mathbf{F}^{\mu\nu}(x) \;=\; \left(\partial^\mu \mathbf{A}^\nu - \partial^\nu \mathbf{A}^\mu\right) \;+\; ig_0\,[\mathbf{A}^\mu,\, \mathbf{A}^\nu] \tag{105}$$

which transforms according to

$$\mathbf{F}^{\mu\nu}(x) \;\longrightarrow\; \hat{U}(\mathbf{\Lambda})\mathbf{F}^{\mu\nu}(x)\hat{U}^{-1}(\mathbf{\Lambda}) \tag{106}$$

Finally, we write the total action for this theory as

$$\begin{aligned} S_{YM} \;&=\; \int d^4x \;\left\{ \bar{\mathbf{\Psi}}(x)\,[i\gamma^\mu D_\mu - m]\,\mathbf{\Psi}(x) \;-\; \frac{1}{4}\mathbf{F}_{\mu\nu}(x)\cdot\mathbf{F}^{\mu\nu}(x) \right\} \\ &=\; \int d^4x \;\left\{ \bar{\mathbf{\Psi}}(x)\,[i\gamma^\mu \partial_\mu - m]\,\mathbf{\Psi}(x) \;-\; \mathbf{J}^\mu(x)\cdot\mathbf{A}^\mu(x) \;-\; \frac{1}{4}\mathbf{F}_{\mu\nu}(x)\cdot\mathbf{F}^{\mu\nu}(x) \right\} \end{aligned} \tag{107}$$

of which the $SU(2)$ theory is obviously a special case.

B.4.2 (c) PHYSICAL PROPERTIES of GAUGE FIELDS

There are many interesting things that one can say about gauge fields, particularly about non-Abelian gauge fields. In the following I will confine the remarks to some fairly basic points.

(i) Equations of Motion and Current Conservation: Up until now we have just looked at the action and the quantities in it. But an obvious physical question is - how do the fields affect each other's motion?

To answer this we find the equations of motion of the 2 fields, by functionally differentiating the action. We then have, as usual, that

$$\begin{aligned} \delta S_{YM} \;&=\; \int d^4x \;\left[\partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial^\mu \mathbf{A}^\nu)} - \frac{\partial\mathcal{L}}{\partial\mathbf{A}^\nu}\right]\delta\mathbf{A}^\mu \\ &\qquad + \;\left[\partial^\mu \frac{\partial\mathcal{L}}{\partial(\partial^\mu \bar{\mathbf{\Psi}})} - \frac{\partial\mathcal{L}}{\partial\bar{\mathbf{\Psi}}}\right]\delta\bar{\mathbf{\Psi}} \;+\; \left[\partial^\mu \frac{\partial\mathcal{L}}{\partial(\partial^\mu \mathbf{\Psi})} - \frac{\partial\mathcal{L}}{\partial\mathbf{\Psi}}\right]\delta\mathbf{\Psi} \end{aligned} \tag{108}$$

giving us equations of motion for $\mathbf{A}^\nu$, $\bar{\mathbf{\Psi}}$, $\mathbf{\Psi}$. The simplest of these equations is for the fermion field $\mathbf{\Psi}(x)$; we have

$$\left(i\gamma^\mu D_\mu - m\right)\mathbf{\Psi}(x) \;=\; \left[\left(i\gamma^\mu \partial_\mu - m\right) \;+\; ig_0\bar{\mathbf{\Psi}}(x)\gamma^\mu\,\mathbf{T}\cdot\mathbf{A}^\nu(x)\right]\mathbf{\Psi}(x) \;=\; 0 \tag{109}$$

so that the fermion field is acted upon by a "source" which combines the anti-field $\bar{\mathbf{\Psi}}(x)$ and the gauge field term $\mathbf{T}\cdot\mathbf{A}^\nu(x)$. A similar equation is obeyed by the anti-field $\bar{\mathbf{\Psi}}(x)$.

The gauge field equation of motion is taken from the 1st variation in (108), and it gives

$$D_\mu \mathbf{F}^{\mu\nu}(x) \;=\; g_0\bar{\mathbf{\Psi}}(x)\,\gamma^\mu\mathbf{T}\,\mathbf{\Psi}(x) \;=\; \mathbf{J}^\mu(x) \tag{110}$$

which is just the generalization of the usual sourced equation of motion for the EM field.

It is very interesting and useful to look more closely at the connection between currents like $\mathbf{J}_\nu(x)$ in (110) and the symmetries that exist in the field theory of interest. Let us recall where an expression like (108) comes from; our theory has a Lagrangian $\mathcal{L}\big(\mathbf{X}_p,\, \partial_\mu \mathbf{X}_p\big)$, where $\mathbf{X}_p = \big(\phi(x),\, \bar{\psi}_\alpha(x),\, \psi(x),\, \mathbf{A}^\nu(x),\, \ldots\big)$ is the set of all fields that $\mathcal{L}$ depends on, collected into one "superfield" $\mathbf{X}_p$. We then have

$$\begin{aligned} \delta S \;&=\; \int d^D x\, \left[\frac{\partial \mathcal{L}}{\delta \mathbf{X}_p} \delta \mathbf{X}_p + \frac{\partial \mathcal{L}}{\delta(\partial_\mu \mathbf{X}_p)} \delta(\partial_\mu \mathbf{X}_p) \right] \\ &=\; \int d^D x\, \left[\frac{\partial \mathcal{L}}{\delta \mathbf{X}_p} + \frac{\partial \mathcal{L}}{\delta(\partial_\mu \mathbf{X}_p)} \partial_\mu \right] \delta \mathbf{X}_p \end{aligned} \tag{111}$$

Now from this we derive a form like (108) by throwing away boundary terms, arguing that at the boundary of our spacetime they vanish - this is done using integration by parts, to show that

$$\partial_\mu \big(\frac{\delta \mathcal{L}}{\delta(\partial_\mu \mathbf{X}_p)} \big) - \frac{\delta \mathcal{L}}{\delta \mathbf{X}_p} \;=\; 0 \tag{112}$$

once we set $\delta S = 0$. However, substituting this into (111), we immediately find that δS in (111) can be written as a total derivative, i.e.,

$$\delta S \;=\; \int d^D x\, \partial_\mu \left[\frac{\delta \mathcal{L}}{\delta(\partial_\mu \mathbf{X}_p)} \delta \mathbf{X}_p \right] \;=\; \int d^D x\, \delta \mathcal{L}(\mathbf{X}_p, \partial_\mu \mathbf{X}_p) \tag{113}$$

where we now identify the quantity in bracelets as a "current". Before continuing with argument, let's just consider what form the current $\mathbf{J}^\mu(x)$ might take. This clearly depends on what field $\mathbf{X}_p$ we are dealing with. In what follows we will assume that all transformations of the Lagrangian we are interested in can be effected by unitary operators acting on the field $\mathbf{X}_p$, i.e., that we can write

$$\tilde{\mathbf{X}}_p \;=\; e^{i \mathbf{G}^a_{pq} \omega_a} \mathbf{X}_q \tag{114}$$

for the transformed field, where $\mathbf{G}^a_{pq}$ is the generator of the transformation and ω_a is the "angle" by which it is effected. Then, e.g., we have

$$\begin{aligned} \mathbf{G}^a_{pq}\omega_a \;&\longrightarrow\; g_0 \mathbf{T}^a_{\alpha\beta} \cdot \boldsymbol{\Lambda}_a && \text{(general non-Abelian transformation)} \\ \mathbf{G}^a_{pq}\omega_a \;&\longrightarrow\; \frac{g}{2} \boldsymbol{\tau}^a_{\alpha\beta} \cdot \boldsymbol{\Omega}_a && \text{(SU(2) transformation)} \\ \mathbf{G}^a_{pq}\omega_a \;&\longrightarrow\; q\theta && \text{(U(1) transformation)} \end{aligned} \tag{115}$$

where in the case of a global transformation, ω_a is independent of x, whereas for a local transformation, $\omega_a \to\; \omega_a(x)$.

From (114) we have

$$\delta \mathbf{X}_p \;=\; i \delta\omega_a(x) \mathbf{G}^a_{pq} \mathbf{X}^q \;=\; \left(\frac{\partial \mathbf{X}_p}{\partial \omega_a} \right) \delta\omega_a(x) \tag{117}$$

and so now we can write $\delta\mathcal{L}$ in (113) as

$$\delta\mathcal{L} \;=\; \partial_\mu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \mathbf{X}_p)} \frac{\partial \mathbf{X}_p}{\partial \omega_a} \right] \delta\omega_a \tag{118}$$

Now, the key point. Suppose we make a transformation of the fields, which will be assumed infinitesimal, parametrized by the infinitesimal $\delta\omega_a$, and we find that $\mathcal{L}$ is unchanged. It then immediately follows that

$$\begin{aligned} \partial_\mu \mathbf{J}^\mu_a(x) \;&=\; 0 \\ \mathbf{J}^\mu_a(x) \;&=\; \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \mathbf{X}_p)} \frac{\partial \mathbf{X}_p}{\partial \omega_a} \right] \end{aligned} \tag{119}$$

This is usually called "Noether's theorem", and evaluation of it for any of the fields we have looked at so far immediately gives us a conservation law for the currents we have defined. As an example of (119), consider the bare Lagrangian

$$\mathcal{L}_0 \;=\; \bar{\mathbf{\Psi}}^\alpha(x) \left[i\gamma^\mu \partial_\mu - m \right] \mathbf{\Psi}^\alpha(x) \tag{120}$$

The infinitesimal field transformation is

$$\mathbf{\Psi}^\alpha(x) \;=\; \left[-ig_0 \mathbf{T}^a_{\alpha\beta} \mathbf{\Psi}^\beta(x) \right] \delta\mathbf{\Lambda}_a \tag{121}$$

so that the current is

$$\begin{aligned} \mathbf{J}^\mu_a(x) \;&=\; \big(i\bar{\mathbf{\Psi}}^\alpha(x)\gamma^\mu \big) \big(-g_0 \mathbf{T}^a_{\alpha\beta} \mathbf{\Psi}^\beta(x) \big) \\ &=\; g_0 \mathbf{\Psi}^\alpha(x) \gamma^\mu \mathbf{T}^{\alpha\beta}_a \mathbf{\Psi}_\beta(x) \end{aligned} \tag{1}$$

as previously derived in (110). From (119) we then see that for the Lagrangian in (120), $\mathbf{J}^\mu_a(x)$ is conserved.

One can say a lot more about such conserved currents, but the basic message here is clear - symmetries lead to conservation laws, just as in classical physics and in ordinary QM.

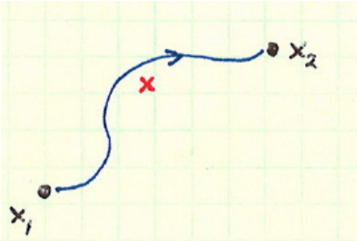
(ii) Physical Interpretation of $\mathbf{F}_{\mu\nu}(x)$: It has already been stated that $\mathbf{F}_{\mu\nu}(x)$ can be thought of as a kind of curvature, and here we amplify on this statement.

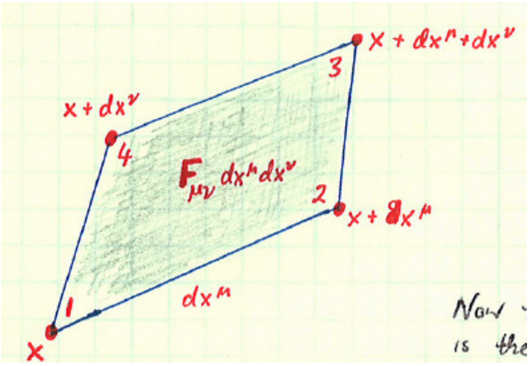
Consider the case where the transformation of the field we have been talking about now consists in looking at the change of the field as we move from one point to another. Thus we are interested in the transformation

$$\psi(x_2) \;=\; \hat{U}(x_2, x_1)\psi(x_1) \tag{123}$$

and more generally in the correlator

$$g_2(x_2, x_1) \;=\; \langle 0 \mid \hat{T}\left\{ \psi(x_2)\psi(x_1) \right\} \mid 0 \rangle \tag{124}$$





Now let's consider the effect of adding a gauge field into the unitary transformation $\hat{U}(x_2, x_1)\, \psi(x_1)$. In path integral language, we can write an expression of the form

$$\hat{U}(x_2, x_1)\, \psi(x_1) \;=\; e^{\frac{i}{\hbar}\int_{x_1}^{x_2} dx\, (\mathcal{L}_0 + \mathcal{L}_A)} \;=\; G^0_{21}[x]\, e^{\frac{i}{\hbar}\int_{x_1}^{x_2} dx\, \mathcal{L}_A} \tag{125}$$

where the "amplitude" $G^0_{21}[x]$ is the result of making the transformation in the absence of the gauge field, along a special path x.

Let's now focus on the extra contribution here, which we call

$$\mathcal{P}_A(x_2, x_1 \mid x) \;=\; e^{\frac{i}{\hbar}\int dx\, \mathcal{L}_A(\bar{\psi}, \psi; \mathbf{A}^\mu)} \tag{126}$$

for a general gauge field; thus, e.g., for the non-Abelian Yang-Mills theory of (107), we have

$$\mathcal{P}_A(x_2, x_1 \mid x) \;=\; e^{ig_0 \int_{x_1}^{x_2} dx_\mu\, \mathbf{T}\cdot\mathbf{A}^\mu(\mathbf{x})} \tag{127}$$

Now the interesting question here is - what happens if we take the system through a circuit, and bring it back to the same place? This question leads us to a specific example of a "Berry phase" argument, which is more usually discussed for a simple wave-function, as in QM.

To extract the curvature it is sufficient to look at an infinitesimal circuit, which we assume to be oriented arbitrarily in spacetime. Let us imagine following this circuit along the counterclockwise path $1 \to 2 \to 3 \to 4$; i.e., we wish to calculate the contribution

$$\mathcal{P}(x, x' \mid 4\,3\,2\,1) \;=\; e^{ig_0 \oint dx^\mu \mathbf{T}\cdot\mathbf{A}_\mu(x)} \;=\; \mathcal{P}(4,3)\mathcal{P}(3,2)\mathcal{P}(2,1)\mathcal{P}(1,4) \tag{128}$$

Now there are 2 obvious ways to do this. One is the simple and quick method of using Stokes's theorem, i.e., we write

$$e^{ig_0 \oint dx^\mu \mathbf{T}\cdot\mathbf{A}_\mu(x)} \;=\; e^{ig_0 \oint dx^\mu \, dx^\nu \left\{ \left(\partial_\mu \mathcal{A}_\nu(x) - \partial_\nu \mathcal{A}_\mu(x)\right) + g_0 [\mathcal{A}_\mu(x), \, \mathcal{A}_\nu(x)] \right\}} \tag{129}$$

where $\mathcal{A}_\mu(x) = \mathbf{T} \cdot \mathbf{A}_\mu(x)$. This result is obtained by noting the identity

$$e^{(\hat{A}+\hat{B})x} \;=\; e^{\hat{A}x}\, e^{\hat{B}x}\, e^{-\left[\hat{A},\,\hat{B}\right]x^2/2} + \mathcal{O}(x^3) \tag{130}$$

for the non-commuting operators $\mathcal{A}_\mu(x)$ and $\mathcal{A}_\nu(x)$, and then using two of the four integrations in the commutator term to remove the derivatives from the integral $[\partial_\mu \mathcal{A}_\nu, \, \partial_\nu \mathcal{A}_\mu,]$ in the exponent.

If this manoeuvre seems too much of a trick (and it looks much better if we phrase it in terms of differential forms), then we can do the path integral long-hand. We have

$$\begin{aligned}
\mathcal{P}(3,2)\mathcal{P}(2,1) &= e^{ig_0 \mathcal{A}_\mu(x) + dx^\nu)\, dx^\nu} \; e^{ig_0 \mathcal{A}_\mu(x)\, dx^\mu} \\
&= \exp\left\{ ig_0 \left[\mathcal{A}_\mu dx^\mu + \left(\mathcal{A}_\nu dx^\nu + \partial_\mu \mathcal{A}_\nu dx^\mu \, dx^\nu \right) \right] - \frac{g_0^2}{2} \left[\mathcal{A}_\mu, \, \mathcal{A}_\nu\right] dx^\mu \, dx^\nu \right\}
\end{aligned}$$

and so we get

$$\mathcal{P}(x, x' \mid 4\,3\,2\,1) \;=\; \exp\left\{ ig_0 \left[\left(\partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu \right) - ig_0 \left[\mathcal{A}_\mu, \, \mathcal{A}_\nu\right] \right] dx^\mu \, dx^\nu \right\} \tag{131}$$

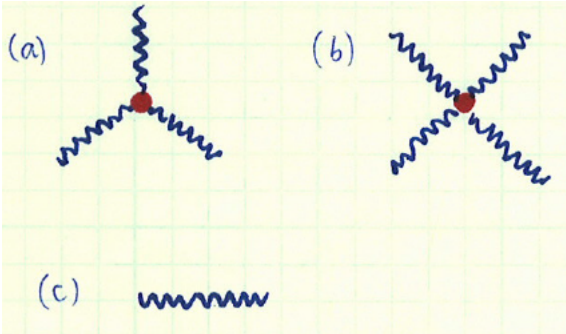
From this result, which agrees with (129), we see that the net effect of moving around this path is to change the amplitude. Now we have already seen this effect in our discussion of the Aharonov-Bohm effect, for which the curvature at a point is just the magnetic field $\mathbf{B}(x)$ (assuming the field $\mathcal{A}^\mu(x)$ is static). What we see in (131) is just the generalization of this to the non-Abelian case (cf. eq. (105)).

Why do we call this a curvature? Actually this is because of the analogy with GR, where we measure the curvature of spacetime by parallel transporting some 4-vector around a loop. Here we are actually transporting a field around the space of field configurations, in the Hilbert space of these configurations, by moving along a circuit in spacetime. Thus we can think of this as a curvature of the gauge field itself.

The effect of the commutator, and of the non-commutating property of the fields, has a profound effect on the field dynamics. Let us go back to the result we have for the total action, eq. (107), then has terms of 3rd and 4th-order in the gauge field:

$$\begin{aligned}
-\frac{1}{4}\mathbf{F}^a_{\mu\nu} \cdot \mathbf{F}^{\mu\nu}_a &= -\frac{1}{2}\mathrm{Tr}\left\{ \left(\mathbf{T}\cdot\mathbf{F}_{\mu\nu}\right)\left(\mathbf{T}\cdot\mathbf{F}^{\mu\nu}\right) \right\} \\
&= -g_0 f_a^{bc} \partial_\mu \mathcal{A}^a_\nu \mathcal{A}^\mu_b \mathcal{A}^\nu_c \;-\; \frac{g_0^2}{4} f^{abc} f_{ade} \mathcal{A}^\mu_b \mathcal{A}^\nu_c \mathcal{A}^d_\mu \mathcal{A}^e_\nu
\end{aligned} \tag{132}$$

where the f_{abc} are, as before, the elements characterizing the Lie group algebra (so for simple angular momentum-style operators, we would have $f_{abc} = \epsilon_{abc}$).



Thus we immediately expect to see terms corresponding to diagram like those in (a) and (b) above, which correspond to the 1st and 2nd self-interaction terms in (132) above. This is in addition to the free gauge line shown in (c).

From this we see a really crucial difference between $U(1)$ gauge fields like those in QED, and the non-Abelian Yang-Mills theory. If we refer back to the discussion just after equation (80), a key feature of ordinary electrodynamics was highlighted, i.e., that photons do not interact with themselves. However, Yang-Mills gauge fields have their own "self-charge", and act as a source for themselves (in addition to having matter fields as their source). The Yang-Mills gauge field is thus fundamentally non-linear. This has profound effects on the dynamics of YM fields; many of these effects have yet to be explored, and we still do not have anything like a full understanding of them.

And of course all this takes no account of the coupling of YM fields to matter itself, which makes it all the more complicated! To properly deal with all of this would take us deep into the standard model.

Finally, notice that we have not yet quantized this theory! To do this we have to adapt path integrals to gauge theory, which we now do.

B.4.3: PATH INTEGRALS for GAUGE FIELDS

It is perfectly possible to deal with QED using a conventional canonical approach, and the results of doing this are strewn across dozens of textbooks and thousands of papers. What is less often emphasized are certain difficulties in such an approach, which proved insurmountable in the case of non-Abelian gauge theories (at least at until the end of the 1960's). For this reason the success of the path integral approach proved decisive for the subsequent development of particle physics.

The technical key which opened the door was the development by Fadeev and Popov in 1966-67 of their method of integrating over gauge-equivalent fields, following the discovery of "ghost" contributions by Feynman in the early 1960's. In the following I will explain the simple picture behind the Fadeev-Popov idea, and then give its formal elaboration.

B.4.3 (a) FUNCTIONALLY INTEGRATING over REDUNDANT VARIABLES

We wish first of all to understand exactly what is the problem that arises when we have to deal with gauge fields, in the paht integral formalism and elsewhere. To do this we will begin with a simpler analogous problem, and see how its solution can be generalized to deal with gauge fields.

(i) Redundant Variables and Jacobians: Let's start by considering the following mathematical problem. Suppose we are given some functional $I[f]$ of a function $f(Q)$ of a variable $Q = \Big(\{x_i\};\ \{q_j\}\Big)$. To focus things, imagine the functional I tells us the total energy on the earth coming from the sun at some time t (we suppress the dependence on t, and $f(Q)$ is the intensity of sunlight above some point on the earth's surface, having angular coordinates $\Omega = (\theta, \phi)$). Then $x_i = \Omega = (\theta, \phi)$, and the $\{q_j\}$ are a set of atmospheric variables like pressure P, temperature T, humidity, etc.

Now as physicists we understand clearly that the intensity depends only on the angle Ω (measured with respect to the sun's direction); all other variables are irrelevant (and we also know that $I[f] = C \int d\Omega f(\Omega)$, where C is a constant). Suppose, however, we did not know this. Then we might have some trouble finding the relevant variables amongst all the others - a common problem in, e.g., medical trials. The question then is - if we have the functional $I[f]$ written in some arbitrary way as a functional of $f(\Omega, \{q_j\})$, how can we extract the meaningful information, ie., how can we get rid of the *redundant variables* $\{q_j\}$?

This problem suggests a rather simpler mathematical question, which I present by way of an example. Suppose we are given the simple integral

$$I = \int d^2\mathbf{R}\, f(\mathbf{R}) \tag{133}$$

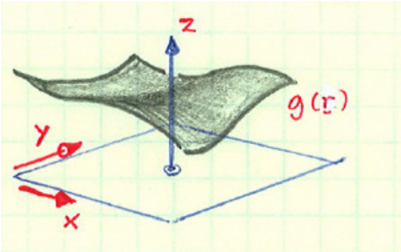
where $\mathbf{R} = (x, y, z)$ is a vector in 3-d space; BUT we have the restriction that we are only allowed to integrate on the surface shown in the figure, defined by $z = g(\mathbf{r})$, where $g(\mathbf{r})$ is just the "height" of the surface, as a function of the 2d planar variable $\mathbf{r} = (x, y,)$. Now we might think we could then write

$$I = \int d^3\mathbf{R}\, f(\mathbf{R})\delta\big(z - g(\mathbf{r})\qquad (??) \tag{134}$$

Now this would work if the surface was a flat horizontal one, ie., if

$$\begin{aligned} f(\mathbf{R}) &= f(\mathbf{r}) \quad \forall z \\ \frac{\partial^n f(\mathbf{R})}{\partial z^n} &= 0 \end{aligned} \tag{135}$$

so that $f(\mathbf{R})$ is independent of z. Obviously in this case the problem is trivial - the variable z is quite irrelevant to our considerations.



Suppose however that it turns out that we don't know the explicit form of the equation for the surface in the form $Z = g(z)$, but we only know that the surface obeys an implicit equation of form

$$G(\mathbf{R}) = 0 \tag{136}$$

so that the surface might have some arbitrary shape, ie., not just be a plane parallel to the xy plane (in the above example, we have $G(\mathbf{R}) = z - g(\mathbf{r})$). Then we would write [1]

$$I = \int d^3\mathbf{R}\; f(\mathbf{R}) \Big| \frac{\partial G(\mathbf{R})}{\partial z} \Big| \delta\big(G(\mathbf{R})\big) \tag{137}$$

to take care of the change of variables - we have introduced a Jacobian which takes care of the variable rate at which we pass through the surface, at different points in the space, when we vary z. All of this is a low-dimensional example of a more general problem; we have a function

$$\begin{aligned} I &= \int d\mathbf{x}\; f(\mathbf{x}) \qquad\qquad \mathbf{x} = (x_1, \ldots, x_n) \\ &= \int d\mathbf{Q}\; f(\mathbf{Q})\delta(\mathbf{q}) \qquad \mathbf{q} = (q_1, \ldots, q_m) \end{aligned} \tag{138}$$

where $\mathbf{Q} = (\mathbf{x}, \mathbf{q}) = (x_1, \ldots, x_n; q_1, \ldots, q_m)$, and $f(\mathbf{Q}) = f(\mathbf{x})$, $\forall \mathbf{q}$; this is the analogue of the above problem where the surface is flat and horizontal. Then if we define a surface in $\mathbf{Q}$-space upon which $f(\mathbf{Q})$ varies, but where it is independent of the other coordinates in $\mathbf{Q}$-space that are orthogonal to the surface variables, we can make the same arguments as above. We define the surface as

$$G(\mathbf{Q}) = 0 \tag{139}$$

and then we have

$$I = \int d\mathbf{Q}\; f(\mathbf{Q}) \det \Big| \frac{\partial G(\mathbf{Q})}{\partial \mathbf{q}} \Big| \delta\big(G(\mathbf{Q})\big) \tag{140}$$

as the solution to our problem.

[1] Note that for some function $f(x)$, with zeros at points $\{x_j^o\}$, with $j = 1, 2, \ldots$, we have $\delta\big(f(x)\big) = |f'(x)|^{-1}\delta\big(x - x_j^o\big)\big)$. For I in (137) we have, in the case where $G(\mathbf{R}) = z - g(\mathbf{r})$ (cf. (124), that $\delta\big(G(\mathbf{R})\big) = \Big|\frac{\partial G}{\partial z}\Big|^{-1} \delta\big(z - g(\mathbf{r})\big)$. The determinant in (140) is the multi-variable version of this.

(ii) Redundancy for Gauge Fields: What we now wish to do is to apply this observation to the problem of functional integration over gauge fields, where the key problem is that instead of a function $f(\mathbf{Q})$ with redundant variables in it, we deal with a functional $Z\left[\bar{\psi},\, \psi;\, \mathbf{A}^\mu\right]$ of a gauge field $\mathbf{A}^\mu(x)$ which also has redundant variables in it - the redundancy being generated from some given $\mathbf{A}^\mu(x)$ by making a gauge transformation. The generating functional Z should be invariant under any gauge transformation, since no physical quantity should depend on which gauge we choose; and indeed we know it is invariant, because the action is invariant.

What this means is that the obvious form for the generating functional,

$$Z\left[\bar{\psi},\, \psi;\, \mathbf{A}^\mu\right] \;=\; \int \mathcal{D}\bar{\psi}\,\mathcal{D}\psi\,\mathcal{D}\mathbf{A}^\mu\; e^{\frac{i}{\hbar}S\left[\bar{\psi},\,\psi;\,\mathbf{A}^\mu\right]} \qquad (??) \tag{141}$$

cannot be right, because it contains a "hidden infinity", coming from the integration over all gauge-transformed configuration of $\mathbf{A}^\mu$.

One might argue here that all one needs to do is to fix a gauge, and then calculate from there. In the old canonical formulation, this is what was done with QED, but it led to severe technical problems, which we will note in passing below. But in the path integral formulation, all that one has to do is to extract the Jacobian determinant in (14) (or rather, its generalization to functionals).

What happens if we do just apply (141) naively? We can do this most simply with QED, and so we are interested in the following integral:

$$I[J_\mu] \;=\; \int \mathcal{D}\mathcal{A}_\mu\; e^{\frac{i}{\hbar}\left\{S_0[\mathcal{A}_\mu]+\int d^4x\; J_\mu(x)\mathcal{A}^\mu(x)\right\}} \tag{142}$$

where $S_0[\mathcal{A}_\mu]$ is given by (62), which we rewrite as

$$\begin{aligned} S_0[\mathcal{A}_\mu] \;&=\; -\frac{1}{4\mu_0}\int d^4x\; \mathbf{F}_{\mu\nu}(x)\mathbf{F}^{\mu\nu}(x) \\ &=\; \frac{1}{2\mu_0}\int d^4x\;\; \mathcal{A}_\mu(x)\;\left[\eta^{\mu\nu}\partial^2 - \partial^\mu\partial^\nu\right]\;\mathcal{A}_\mu(x) \end{aligned} \tag{143}$$

which we can also write in k-space as

$$S_0[\mathcal{A}_\mu] \;=\; -\frac{1}{2\mu_0}\sum_q \;\mathcal{A}_\mu(q)\;\left[q^2\eta^{\mu\nu} + q^\mu q^\nu\right]\;\mathcal{A}_\nu(-q) \tag{144}$$

and you will notice the similarity of this result to that for the phonon system; we can in the same way divide this free field term into transverse and longitudinal parts, i.e., write

$$\mathcal{A}_\mu(q) \;=\; \mathcal{A}^{\parallel}_\nu(q)\hat{q}_\mu\hat{q}_\nu \;+\; \mathcal{A}^{\perp}_\nu(q)\left[\delta_{\mu\nu} - \hat{q}_\mu\hat{q}_\nu\right] \tag{145}$$

and we see that the operator $\hat{Q}^{\mu\nu}_0(\mathbf{q}) = q^2\eta^{\mu\nu} - \hat{q}_\mu\hat{q}_\nu$, or equivalently the operator $Q^{\mu\nu}_0(x) = \eta^{\mu\nu}\partial^2 - \partial^\mu\partial^\nu$, are both entirely transverse (which is what we would expect for photons, which are indeed transverse excitations).

The next move is then clearly supposed to be to do the functional integral in (142), following the usual line of development:

$$\int \mathcal{D}\mathcal{A}_\mu \, e^{\frac{i}{\hbar}\int d^Dx \left[\frac{1}{2\mu_0}\left(\mathcal{A}_\mu, Q_o^{\mu\nu}, \mathcal{A}_\nu\right) + \left(J_\mu \mathcal{A}^\mu\right)\right]} \sim \frac{1}{|Q_o^{\mu\nu}|} e^{-\frac{i\mu_o}{2\hbar}\int d^Dx \, (J Q_o^{-1} J)} \tag{146}$$

where the inverse operator is:

$$Q_0^{-1} = \left(\eta^{\mu\nu}\partial^2 - \partial^\mu\partial^\nu\right)^{-1} \tag{147}$$

However this operator inverse is formally infinite; writing

$$\left[\eta^{\mu\nu}\partial^2 - \partial^\mu\partial^\nu\right] \left(Q_0(x,x')\right)_{\mu\beta} = \delta^\mu_\beta \delta(x-x') \tag{148}$$

and multiplying to the left by ∂_μ, we get

$$0 \times Q_0 = \partial_\beta \delta(x-x') \tag{149}$$

which is a contradiction unless $\hat{Q}_0$ is infinite. This infinity is a reflection of the gauge invariance, because $\mathcal{A}^\mu(q)$ contains longitudinal (as well as transverse) degrees of freedom, which are untouched by $\hat{Q}_0$, i.e., they have zero eigenvalue when acted upon by $\hat{Q}_0$ (and hence infinite eigenvalue when operated on by $\hat{Q}_0^{-1}$). Any gauge transformation of the form (57), adding a term $\partial^\mu\psi(x)$ to $\mathcal{A}^\mu$, will thus also have zero eigenvalue when acted upon $\hat{Q}_0$, as we see easily:

$$Q_0^{\mu\nu}\partial^\mu\psi(x) = \left(\eta^{\mu\nu}\partial^2 - \partial^\mu\partial^\nu\right)\partial^\mu\psi(x) = 0 \tag{150}$$

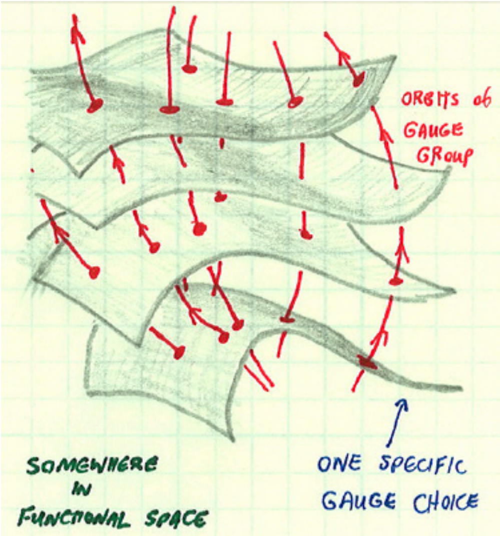
So how do we deal with this? The naive answer is to fix the gauge before doing the calculation, but this has the disadvantage that we immediately lose Lorentz invariance in the calculations. In the early days of QFT this was a big problem, actually first solved by Tomonaga. Other methods introduced a photon mass into the calculation - this removed the problem of zero eigenvalues, but destroyed gauge invariance (the mass was set to zero at the end of the calculation). However all such techniques were extremely messy to implement, and quite impossibly complicated for non-Abelian gauge theories.

B.4.3 (b) FADDEEV-POPOV TECHNIQUE

Let's first describe the solution of Fadeev and Popov in geometric terms, and then go on to see how it works in detail.

Imagine the space of all possible configurations of some gauge field $\mathbf{A}_\mu(x)$. This is of course a very large space, which we show here in the figure in the form of a 3-d caricature. Now suppose we are in some specific gauge, and we look at all such configurations in this gauge. We show all such configurations in this gauge on a hyper-sheet, depicted as a simple 2-d sheet in the figure. We then write all of these configurations, in this gauge, as

$$\mathbf{A}^\mu(x) \longrightarrow \bar{\mathbf{A}}^\mu(x) \qquad \text{(fixed gauge)} \tag{151}$$



and then consider the set of all possible gauge transformation on $\bar{\mathbf{A}}^\mu(x)$ that are required to reverse this many-one mapping, ie., to produce the full set of possible configurations; we write these as

$$\{A^\mu_a(x)\} \;=\; \{A^\mu_a(x) + \alpha^\mu_a(x)\} \;\equiv\; \{A^\mu_a(x) + \big(\partial^\mu \Lambda_a(x) + g_0 f_{abc} \Lambda^b(x) A^\mu_c(x)\big)\} \tag{152}$$

where it is understood that we deal here with infinite sets - the set of all gauge configurations $\{\mathbf{A}^\mu(x) \equiv \{A^\mu_a(x)\}$ is produced by starting with the set of all configurations possible in some fixed gauge $\{\bar{\mathbf{A}}^\mu(x)\}$, and then adding all possible gauge transformations $\{\boldsymbol{\alpha}^\mu(x)\} \equiv \{\alpha^\mu_a(x)\}$, for arbitrary differentiable functions $\boldsymbol{\alpha}^\mu(x)$. More precisely, we define the group G of all possible equivalence classes of a given $\tilde{A}_\mu(x)$, called in mathematics the "orbit" of the gauge group; the full set of functions $\mathbf{A}_\mu(x)$ is then produced by the set of all orbits of all possible configurations $\bar{\mathbf{A}}_\mu(x)$ in a fixed gauge. To show that this geometrical picture is meaningful is a job for mathematicians, which we will not enter into here.

Consider now the functional integral in (142) again - we will now write

$$\begin{aligned} I[J_\mu] &= \int \mathcal{D}\mathbf{A}_\mu \, e^{\frac{i}{\hbar}\left\{S_0[\mathbf{A}_\mu] + \int d^4x \, \mathbf{J}_\mu(x)\mathbf{A}^\mu(x)\right\}} \\ &= \int \mathcal{D}\alpha_\mu(x) \int \mathcal{D}\bar{\mathbf{A}}_\mu \, e^{\frac{i}{\hbar}\left\{S_0[\tilde{\mathbf{A}}_\mu + \alpha_\mu] + \int d^4x \, \mathbf{J}_\mu(x)[\tilde{\mathbf{A}}^\mu + \alpha^\mu]\right\}} \\ &= \int \mathcal{D}\alpha_\mu \int \mathcal{D}\bar{\mathbf{A}}_\mu(x) \, e^{\frac{i}{\hbar}\left\{S_0[\mathbf{A}^\mu] + \int d^4x \, \mathbf{J}_\mu(x)\mathbf{A}^\mu(x)\right\}} \end{aligned} \tag{153}$$

and we see that the problem in (153) is that integrand appearing in the functional integral, i.e., the function $\exp\left\{i/\hbar\big(S_0[\mathbf{A}^\mu] + \int d^4x \, \mathbf{J}_\mu(x)\mathbf{A}^\mu(x)\big)\right\}$, is invariant under changes in gauge, so that the functional integration $\int \mathcal{D}\boldsymbol{\alpha}^\mu(x)$ simply produces an infinite multiplication of the answer - what we would like to do is to get rid of this redundant multiplication factor.

To do so, let's recall the development in eqs. (138)-(140), and do the same now for gauge functionals, instead of just ordinary functions. Notice that we cannot just stick a factor like $\delta\big(\chi(x)\big)$ into the functional integral, for two reasons. First, we want to keep the theory gauge invariant, for many different reasons (most of which have not yet become apparent). Second, in the functional integral (142) and (153), we don't actually yet know how to properly "measure" the "volume", in the hyperspace of functionals, of the domain defined by a specific gauge choice. To see this we simply recall the determinant appearing in the finite-dimensional integral in (137), which acts as a Jacobian for the change of variable - it is not yet clear how to generalize this to cover for the much larger space of functions we are now dealing with.

Thus it is not enough to just stick a factor $\delta\Big(G\big(\bar{\mathbf{A}}^\mu(x)\big)\Big)$ into the functional integral (141), and integrate over the full $\mathbf{A}^\mu(x)$. What we want is something like

$$Z\left[\bar{\psi},\,\psi,\,\mathbf{A}_\mu\right] \;=\; \int \mathcal{D}\boldsymbol{\alpha}^\mu(x) \int \mathcal{D}\bar{\mathbf{A}}_\mu\;\boldsymbol{\Delta}_{FP}\;\delta\Big(G\big(\bar{\mathbf{A}}^\mu(x)\big)\Big) \int \mathcal{D}\bar{\psi}\mathcal{D}\psi\; e^{\frac{i}{\hbar}S[\bar{\psi},\,\psi,\,\mathbf{A}_\mu]} \tag{154a}$$

where $\boldsymbol{\Delta}_{FP}$ is the relevant determinant, now called the "Fadeev-Popov determinant". Our job is to find an expression for it, which is valid for any coneeivable form for the gauge transformation.

Formally this is easy. In exact analogy with (140), we have

$$\boldsymbol{\Delta}_{FP} \;=\; \det\left|\frac{\delta G\big(\mathbf{A}_\mu(x)\big)}{\delta\boldsymbol{\alpha}^\mu(x')}\right| \;\equiv\; \det\left|\frac{\delta G\big(A^a_\mu(x)\big)}{\delta\alpha^\mu_b(x')}\right| \;=\; \Delta^{ab}_{FP}(x,x')$$

$$\text{for}\;\; G\big(\mathbf{A}_\mu(x)\big) \;=\; 0 \qquad \text{(gauge constraint)} \tag{155a}$$

$$\text{so that} \qquad \int \mathcal{D}\boldsymbol{\alpha}^\mu \boldsymbol{\Delta}_{FP}[\boldsymbol{\alpha}] \;=\; 1 \tag{155b}$$

However these formal expressions are not terribly illuminating until we try to use them - this we will see in the next sub-section, for both QED and for non-Abelian gauge theories.

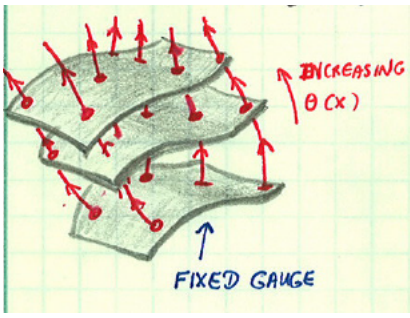
In preparation for this, let's first rewrite (155a) for a general non-Abelian gauge theory. If we assume that the non-Abelian gauge transformation can always be characterized by an "angle" in some space of gauge transformations (this angle being $\theta(x)$ for $U(1)$ transformations and the hyperangle $\boldsymbol{\Lambda}(x) = \Lambda_a(x)$ for YM theories), then we can rewrite the Fadeev-Popov factor in (154) as

$$\int \mathcal{D}\boldsymbol{\alpha}^\mu(x)\;\boldsymbol{\Delta}_{FP}\;\delta G\big(\mathbf{A}_\mu(x)\big) \;\;\rightarrow\;\; \int \mathcal{D}\boldsymbol{\Lambda}(x)\;\boldsymbol{\Delta}_{FP}(x,x')\;\delta G\big(\mathbf{A}_\mu(x')\big) \tag{156}$$

where now

$$\boldsymbol{\Delta}_{FP}(x,x') \;\equiv\; \Delta^{ab}_{FP}(x,x') \;\;=\;\; \det\left|\frac{\delta G\big(A^a_\mu(x)\big)}{\delta\Lambda_b(x')}\right| \;\equiv\; \det\left|M^{ab}(x,x')\right| \tag{157}$$

This is a much more transparent formula since the functional integral is over hyperangles $\Lambda^a(x)$ in some finite-dimensional space, something we know how to do. Note that the Fadeev-Popov determinant has a physical meaning that is evident from (157). If we make an



infinitesimal change in the gauge by effecting a change $\delta\mathbf{\Lambda}(x')$ in the gauge angle, then $\mathbf{\Delta}_{FP}(x,x')$ measures the "response" of the gauge-fixing function $G\big(\mathbf{A}_\mu(x)\big)$ to this change. In other words, if we write

$$\delta G\big(A_a^\mu(x)\big) \;=\; \int d^4x' \; M_{ab}(x,x')\;\delta\Lambda^b(x') \tag{158}$$

then $M^{ab}(x,x')$ measures the response of the a-th component of the gauge constraint function $G\big(A_a^\mu(x)\big)$ to the change $\delta\Lambda_b(x')$ in the gauge angle.

B.4.4: GAUGE FIELDS in HIGH-ENERGY PHYSICS

We will not go into too much detail here. The case of QED is relatively easy to understand, as we will see. Non-Abelian gauge theories are more messy because the number of degrees of freedom is large. It would take us too far afield to discuss the application of the results to either QED or to the electroweak theory (and the latter requires, in any case, an appeal to a spontaneous symmetry-breaking, i.e., to the Anderson-Higgs mechanism).

B.4.4 (a) QUANTUM ELECTRODYNAMICS

This is of course the simplest theory. We have parametrized gauge transformations in this theory by the shifty $A^\mu(x) \longrightarrow A^\mu(x) + \partial^\mu\chi(x)$, but it is now time to formulate this a little more generally. We assume the gauge constraint in the form

$$\mathbf{G}\Big(A^\mu\big(x,\theta(x)\big)\Big) \;=\; 0 \qquad \left\{\begin{array}{ll} \mathbf{G}\big(A^\mu(x)\big) = \partial_\mu A^\mu(x) & \text{(Lorentz)} \\ \mathbf{G}\big(A^\mu(x)\big) = \nabla\cdot\mathbf{A}(x) & \text{(Coulomb)} \\ \mathbf{G}\big(A^\mu(x)\big) = \quad A^z(x) & \text{(Axial)} \end{array}\right\} \tag{159}$$

when on the RHS of (159) we give 3 common examples of fixed gauge choices in QED. In what follows we will use a slight variation on the Lorentz gauge, by adding an extra function $\chi(x)$; this can simply be viewed as part of the gauge transformation, and its usefulness will become apparent later. We then write

$$\mathbf{G}(A^\mu) \;=\; \partial^\mu A_\mu(x) - \chi(x) \;\equiv\; \big(\partial^\mu \bar{A}_\mu(x) + \partial^2\theta(x)\big) - \chi(x) \tag{160}$$

where $\bar{A}_\mu(x)$ is some fixed gauge satisfying $\mathbf{G}(A^\mu) = 0$. Now eqtn (155b) implies that $\int \mathcal{D}\theta(x)\, \delta\big(\mathbf{G}(A^\mu)\big) = \Delta_{FP}^{-1}$; thus we have [2]

$$\int \mathcal{D}\theta(x)\, \delta\big(\mathbf{G}(A^\mu)\big) \;=\; \big(\det|\partial^2|\big)^{-1} \tag{161}$$

so that the FP determinant is

$$\mathbf{\Delta}_{FP}(x,x') \;=\; \det|(\partial^\mu \partial_\mu)_{x,x'}| \;\equiv\; \det|\partial^2| \tag{162}$$

which is a constant $\mathbf{\Delta}_{PF}$; recall that the determinant $\det|\partial^2| \equiv \det|\partial^2_{xx'}|$, where $\partial_{xx'} \equiv \langle x|\partial^2|x'\rangle = \partial_x \delta(x-x')$, should be interpreted as a matrix, which is diagonal in x-space (compare notes on path integrals in section A). Thus we get

$$\begin{aligned} I[A^\mu] &= \int \mathcal{D}A^\mu(x)\, e^{\frac{i}{\hbar}S[A^\mu]} \\ &= \mathbf{\Delta}_{FP} \int \mathcal{D}A^\mu(x)\, \delta\big(\partial^\mu A_\mu - \chi(x)\big)\, e^{\frac{i}{\hbar}S[A]} \end{aligned} \tag{163}$$

with the constant outside the integration. However we still have to deal with the δ-functional $\delta(\partial^\mu A - \chi)$, and this is where the 't Hooft trick comes in handy. Suppose we functionally integrate now over $\chi(x)$, but now inserting some functions $H_t(x)$ in the integral, i.e., we multiply I by the factor $\int \mathcal{D}\chi(x)\, H_t(x)$, (where we if we choose $H_t = 1$, then we get rid of the χ-function as though it had never been there at all). We then have

$$I[A^\mu] \;=\; \frac{1}{\mathcal{N}} \int \mathcal{D}A^\mu\; H_t(\partial^\mu A_\mu)\, e^{\frac{i}{\hbar}S[A]} \tag{164}$$

where $\mathcal{N}$, the normalizing factor, is just a constant:

$$\mathcal{N} \;=\; \frac{\int \mathcal{D}\chi\; H_t(\chi)}{\mathbf{\Delta}_{FP}} \tag{165}$$

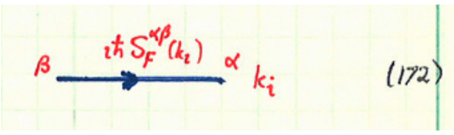
The nice thing about this trick is that we can now make $H_t(\chi)$ the exponential of something - this way everything is now in the exponential, and we can read off the Feynman rules. The choice made by 't Hooft was

$$H_t(\chi) \;=\; \exp \frac{-i}{2\alpha} \int d^4x \frac{1}{\mu_0 \hbar} \chi^2(x) \tag{166}$$

where α is just a number; we then finally get (ignoring the factor $1/\mathcal{N}$):

$$Z_{QED}[J_\mu] \;=\; \int \mathcal{D}\bar{\psi}\mathcal{D}\psi \int \mathcal{D}A^\mu\; e^{\frac{i}{\hbar}\left\{S_0[A] - \frac{1}{2\alpha}\int d^4x \frac{1}{\mu_0}\, (\partial_\mu A^\mu)^2 + \int d^4x\; J_\mu(x) A^\mu(x)\right\}} \tag{167}$$

[2]Recall the footnote to eqtn. (137). We are just using the fundamental generalization of the usual formula that $\int dx\, \delta(ax-b) = 1/a$ to $\int \mathcal{D}\theta(x)\, \delta\big(\partial^2\theta(x) - f(x)\big) = 1/\det(\partial^2)$.



with $S_0[A^\mu]$ given by (143) or (144). We may now do the usual functional integration that led to (146) and (147), but now we can write

$$Z_{QED}[J_\mu] = \int \mathcal{D}\bar{\psi}\mathcal{D}\psi \int \mathcal{D}A^\mu \; e^{-\frac{i}{\hbar}\int d^4x \left[\frac{1}{2\mu_0}A^\mu(x)D^0_{\mu\nu}(x,x')A^\nu(x')\right] - \int d^4x \; J_\mu(x)A^\mu(x)} \tag{168}$$

where

$$D^0_{\mu\nu}(x,x') = \left(\eta_{\mu\nu}\partial^2 + (\alpha^{-1}-1)\partial_\mu\partial_\nu\right)^{-1} \tag{169}$$

and this operator does not have the pathology of $Q_0(x,x')$ in (147); it has an equivalent representation in momentum space as

$$D^0_{\mu\nu}(q) = -\frac{1}{q^2+i\epsilon}\left[\eta_{\mu\nu} + (\alpha-1)\hat{q}_\mu\hat{q}_\nu\right] \tag{170}$$

where as usual, $\hat{q}^\mu = q^\mu/|q|$. The parameter α now acts as a "regularizer", getting rid of the "zero mode" problem we had before. Different values of α give different gauges that had been used in QED long before Fadeev and Popov; for example

$\alpha = 1$ Feynman gauge
$\alpha = 0$ Landau gauge

We may now read off the Feynman rules for QED (compare the rules already derived for Dirac fermions in section B.3, eqs. (44)-(48)). We have, for the Lagrangian

$$\mathcal{L} \;=\; \bar{\psi}_\alpha\big(i\gamma^\mu D_\mu - m\big)\psi_\alpha \;+\; J_\mu A^\mu \;-\; \frac{1}{2\alpha\mu_0}(\partial_\mu A^\mu)^2 - \frac{1}{2\mu_0}F_{\mu\nu}F^{\mu\nu} \tag{171}$$

the following rules:

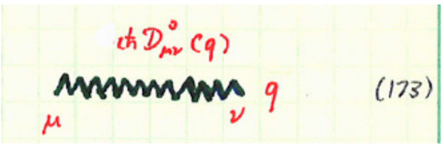
1. The fermion propagator is $G_0^{\alpha\beta}(k_j) = i\hbar S_F^{\alpha\beta}(k_j)$, for a fermion line carrying momentum k_j, as shown in the figure (172);

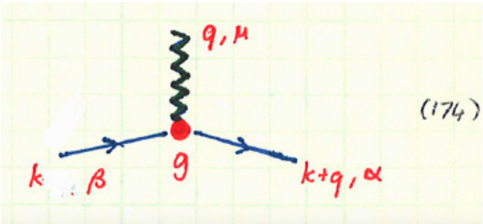
2.The photon propagator is given by the factor, as shown in the figure (173),

$$i\hbar D^0_{\mu\nu}(q) \;=\; \frac{-i\hbar}{q^2+i\epsilon}\left[\eta_{\mu\nu} + (\alpha-1)\hat{q}_\mu\hat{q}_\nu\right]$$

3. Each interaction vertex contributes the factor, as shown in the figure (174)

$$-\frac{i}{\hbar}g \;=\; -\frac{i}{\hbar}e\gamma^\mu_{\alpha\beta}\,\delta\Big(\sum_i k_i + \sum_j q_j\Big)$$





where the γ-matrix now contains spin indices α, β acting between $\bar{\psi}_\alpha$ and ψ_β, and e is the charge of the particle.

Then, as usual, we sum over all spinor indices, integrate over all moments k_i and q_j, and multiply by a symmetry factor; and the momentum integrations appear in the form

$$(-1)^L \int \frac{dk_1}{(2\pi)^D} \cdots \int \frac{dk_n}{(2\pi)^D} \int \frac{dq_1}{(2\pi)^D} \cdots \int \frac{dq_m}{(2\pi)^D} \tag{175}$$

where L is the number of fermion loops.

Actually, the rules for QED are almost identical in form to those for the electron-phonon problem, covered in section B.3; and the topology of the diagrams is identical to that for the coupled field problem of section B.3.

B.4.4 (b) YANG-MILLS SU(N) GAUGE THEORY

Let us go back to our key formula in (157), for the FP determinant. To make things clear, we will simply go through the same manoeuvres as we did for the $U(1)$ gauge field, we will pick the same generalized Lorentz gauge,

$$\begin{aligned} G\big(A_a^\mu(x)\big) &= \partial_\mu A_a^\mu - \chi_a(x) \\ &= \partial_\mu \bar{A}_a^\mu + \partial_\mu\big(\partial^\mu \Lambda_a + g_0 f_{abc} \Lambda^b \bar{A}_c^\mu\big) - \chi_a(x) \end{aligned} \tag{176}$$

Following through the steps as for the QED calculation, we then find that

$$\Delta_{FP}^{ab}(x,x') = \det|\delta^{ab}\,\partial^2 + g_0 f^{abc} \partial_\mu A_c^\mu|\;\delta(x-x') \tag{177}$$

We may also carry out the integration using the 't Hooft trick, introducing the obvious generalization of (166) as

$$H_t(\chi_a) = \exp \frac{-i}{2\alpha} \int d^4x \left(\frac{1}{\hbar} \chi_a(x) \chi^a(x)\right) \tag{178}$$

and so we can write the partition function or generating functional as

$$\begin{aligned} Z_{YM}[J_\mu^a] &= \int \mathcal{D}\mathbf{A}^\mu(x)\ \mathbf{\Delta}_{FP}\ e^{\frac{i}{\hbar}\left(S^0_{YM}[\mathbf{A}]+\int d^4x\left[\mathbf{J}_\mu(x)\mathbf{A}^\mu(x)-\frac{1}{2\alpha}(\partial_\mu\mathbf{A}^\mu)^2\right]\right)} \\ &= \int \mathcal{D}A^\mu_a(x)\ \Delta^{ab}_{FP}(x)\ e^{\frac{i}{\hbar}\left(S^0_{YM}[A^\mu_b(x)]+\int d^4x\left[J^b_\mu(x)A^\mu_b(x)-\frac{1}{2\alpha}(\partial_\mu\mathbf{A}^\mu)^2\right]\right)} \end{aligned} \tag{179}$$

with $S^0_{YM}[\mathbf{A}^\mu]$ given by:

$$S^0_{YM}[\mathbf{A}^\mu] = -\frac{1}{\hbar}\int d^4x\ \mathbf{F}_{\mu\nu}(x)\mathbf{F}^{\mu\nu}(x) \tag{180}$$

However there is now a big difference. In the case of QED, the Faddeev-Popov determinant Δ_{FP} in (162) was just a constant, independent of the field $\mathbf{A}^\mu(x)$. This is no longer the case - the determinant in (177) is clearly dependent on $\mathbf{A}^\mu(x)$, and so we cannot take it outside the functional integral in (179), as we did for QED in (163). This makes the non-Abelian case much more difficult.

At this point the founders of modern QFT introduced a trick that had been invented by Feynman, during his earlier research into quantum gravity. He wrote the determinant Δ^{ab}_{FP} as a functional integration over a set of fake fermion fields - this takes us back to the result we found in section B.2 for integration over Fermion fields, that they give a determinant in the numerator (cf. eqs. (25) and (28) in section B.2). Thus we write

$$\begin{aligned} \Delta^{ab}_{FP}(x,x') &= \det|M^{ab}(x,x')| = \int \mathcal{D}\bar{\mathbf{c}}(x)\int \mathcal{D}\mathbf{c}(x)\ e^{iS_{GH}[\bar{\mathbf{c}},\mathbf{c}]} \\ S_{GH}[\bar{\mathbf{c}},\mathbf{c}] &= \int d^4x \int d^4x'\ \bar{\mathbf{c}}_a(x)M^{ab}(x,x')\mathbf{c}_b(x) \end{aligned} \tag{181}$$

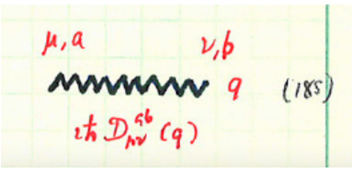
where the fermion fields $\mathbf{c}(x) \equiv c^a(x)$ and $\bar{\mathbf{c}}(x)$ obey all the Grassmann rules discussed before. The fields were called "ghost fields" by Feynman, and they created much confusion at the time (Feynman introduced them to prevent the theory from losing unitarity - it was only later that their role as a determinant was understood).

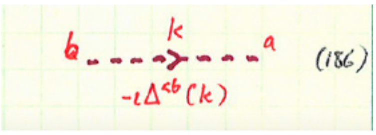
Thus we can finally write that

$$Z_{YM}[J^a_\mu] = \int \mathcal{D}\bar{\mathbf{c}}\mathcal{D}\mathbf{c}\int \mathcal{D}\mathbf{A}^\mu\ e^{\frac{i}{\hbar}\left(S^0_{YM}[\mathbf{A}^\mu]+S_{GH}[\bar{\mathbf{c}},\mathbf{c}]+\int d^4x\ \left[\mathbf{J}_\mu(x)\mathbf{A}^\mu(x)-\frac{1}{2\alpha}(\partial_\mu\mathbf{A}^\mu)^2\right]\right)} \tag{182}$$

The Feynman rules for this functional are extracted by writing the action in terms of a non-interacting "free field" part

$$S_0[\bar{\mathbf{c}},\mathbf{c};\ \mathbf{A}^\mu] = \int d^4x\ \left[\bar{\mathbf{c}}(x)\partial^2\mathbf{c}(x) - \frac{1}{4}\mathbf{F}_{\mu\nu}(x)\mathbf{F}^{\mu\nu}(x) - \frac{1}{2\alpha}(\partial_\mu\mathbf{A}^\mu)^2\right] \tag{183}$$





and an interacting part

$$\begin{aligned} S_{int}[\bar{\mathbf{c}}, \mathbf{c};\ \mathbf{A}^\mu] &= -\int d^4x \left\{ ig_0(\bar{\mathbf{c}}\mathbf{f}\mathbf{c}\ \partial_\mu \mathbf{A}^\mu) + \frac{1}{2} g_0 \mathbf{F}_{\mu\nu}\mathbf{f}\ \mathbf{A}^\mu \mathbf{A}^\nu - \frac{1}{4} g_0^2\ (\mathbf{f}\ \mathbf{f})(\mathbf{A}_\mu \mathbf{A}^\mu)^2 \right\} \\ &= -\int d^4x \left\{ ig_0(\bar{c}_a f^{abc} c_b \partial_\mu A^\mu_c) + \frac{g_0}{2} F^a_{\mu\nu} f^{bc}_a A^\mu_b A^\nu_c - \frac{g_0^2}{4} f^a_{bc} f^{de}_a A^b_\mu A^c_\mu A^\mu_d A^\nu_e \right\} \end{aligned} \tag{184}$$

and this allows us to read off the Feynman rules for the Yang-Mills gauge field (which, I emphasize, is still not coupled to the real world fermions that one might expect to exist in a real theory like the electroweak theory). We then have

1. A vector boson propagator given by the YM generalization of (173), as shown in the diagram (185)

$$i\hbar D^0_{\mu\nu}(q) = \frac{-i\hbar}{q^2 + i\epsilon} \delta^{ab} \left[\eta_{\mu\nu} + (\alpha - 1)\hat{q}^\mu \hat{q}^\nu \right]$$

2. A ghost propagator given by the diagram (186)

$$-i\Delta^{ab}(k) = -i\delta^{ab} \frac{1}{k^2 + i\epsilon}$$

where there is no $\hbar$ because of the definition (181); we notice that the ghost propagator is massless, as is obvious from (184), and is a scalar field as well.

3. Finally, from the interaction in (184) we get a whole variety of vertices, of the following form:

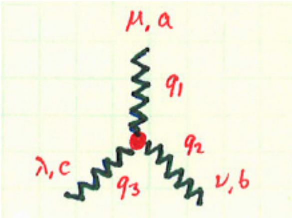
(a) 3-boson vertex, given by the 2nd term in (184), as

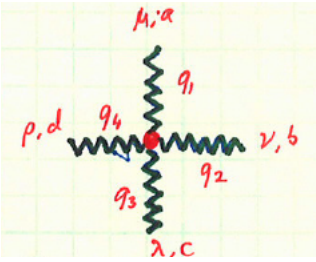
$$-\frac{i}{\hbar}\lambda_3 = \frac{i}{\hbar} g_0 f^{abc} \left[\eta_{\mu\nu}(q_1 - q_2)_\lambda + \eta_{\nu\lambda}(q_2 - q_3)_\mu + \eta_{\lambda\mu}(q_3 - q_1)_\nu \right] \delta(q_1 + q_2 + q_3)$$

in which 3 vector bosons $\mathbf{A}^\mu(q)$ interact at a point in spacetime, each scattering off their mutual "curvature";

(b) A 4-boson vertex, given by the 3rd term in (184), as

$$\begin{aligned} \tfrac{i}{\hbar}\lambda_4 = -\tfrac{i}{\hbar} g_0^2 [& f^{ab}_e f^{cd}_e (\eta_{\mu\lambda}\eta_{\nu\rho} - \eta_{\mu\rho}\eta_{\nu\lambda}) \\ & + f^{ac}_e f^{bd}_e (\eta_{\mu\nu}\eta_{\lambda\rho} - \eta_{\mu\rho}\eta_{\lambda\nu}) \\ & + f^{ad}_e f^{cb}_e (\eta_{\mu\lambda}\eta_{\nu\rho} - \eta_{\mu\nu}\eta_{\lambda\rho})] \delta(q_1 + q_2 + q_3 + q_4) \end{aligned}$$





with 4 bosons coupled at a point;

(c) A boson-ghost fermion interaction, coming from the 1st term in (184), given by

$$\frac{i}{\hbar} g_3 \;=\; g_0 f^{abc} q_\mu$$

which is produced by the absorption of a gauge boson by the ghost fermion.

Finally, we integrate over momenta as before - however, there are no external ghost fermion lines, and we still have a loop factor $(-1)^L$ for L ghost fermion loops.

This concludes this introduction to gauge fields in QFT. One of the most interesting topics that is usefully examined in this formalism, outside of particle physics, is the variety of non-relativistic fermionic and bosonic superfluid in Nature, as well as spin fluids. And of course, if we go to spin-2 gauge fields, we can also discuss quantum gravity.