Notes for course on Physics of Particles and Fields, CMI, Autumn 2014

Govind S. Krishnaswami, May 2, 2015 (updated August 31, 2018)

This is an introductory course on particle physics, which does not assume prior knowledge of particle physics or much quantum field theory. These lecture notes (posted on the website http://www.cmi.ac.in/~govind/teaching/particle-phys-pg-o14) are very sketchy and are no substitute for attendance and taking notes at lectures or reading other texts and references. Please let me know (via govind@cmi.ac.in) of any comments or corrections.

Contents

1 Introduction

1.1 Some books on particle physics

The material in these rough notes is not original, it is based on what I have learned from my teachers and colleagues and material in several fine books (which contain a lot more!), e.g.,

1. F Halzen and A D Martin, Quarks and Leptons: An Introductory Course in Modern Particle Physics
2. D J Griffiths, Introduction to elementary particle physics
3. K Huang, Quarks, leptons and gauge fields
4. K Gottfried and V F Weisskopf, Concepts in Particle Physics.

5. D. H. Perkins: Introduction to High Energy Physics

6. Alessandro Bettini, Introduction to Elementary Particle Physics.

1.2 'Elementary particles'

• The physics of particles and fields deals with the fundamental constituents of matter and their interactions. This branch of physics is also called (elementary) particle physics or high energy physics and sometimes sub-atomic or sub-nuclear physics. Sub-atomic/nuclear means of size less than atomic/nuclear dimensions, rather than 'lying inside the atom/nucleus'.

• The word elementary particle has a usage depending on context and era. Around 1934, the photon, electron, proton, neutron, positron and neutrino were believed to be the only 'elementary' particles. It turned out that there are many more elementary particles like the soon-to-be-discovered muon. What is more, it turned out that the neutron and proton are not truly elementary, they have a size of about a fermi (10^{-15} m) and are composites of quarks and gluons.

• Roughly, an elementary particle is one that may be treated as a point particle (or at least one with fixed shape) with fixed properties and with no sub-structure for the purposes under consideration. Any system (like an atom, nucleus or perhaps even a black hole) in its ground state could serve as an elementary particle if the energy gap to the first excited state exceeds the thermal or other energies available, since then, the system will display a fixed shape, size, mass, angular momentum etc., associated to its ground state.

• In Newtonian and celestial mechanics we often treat billiard balls, the sun and planets as point particles. In much of biology and some of chemistry, atoms may be treated as elementary particles, as the energies exchanged are too small to excite atoms. In atomic physics, photons, electrons and nuclei can be treated as elementary; they are much smaller than atomic sizes of an angstrom. The eV energies of atomic physics are not sufficient to excite nuclei from their ground states (that would require 0.1-10 MeV). In nuclear physics, we may often treat pions, protons and neutrons as elementary, since the available energies of order MeV are too small to excite them (that would require about a 100 MeV). In fact, even an alpha particle (He-4 nucleus consisting of 2 protons and 2 neutrons) can be regarded as an elementary particle in much of low energy nuclear physics, since it is very tightly bound (binding energy of 28 MeV). In the physics of hadrons (hadrons comprise baryons like n, p and mesons like π, K), we deal with energies of the order of 100s of MeV to 100s of GeV at present. In hadronic physics, quarks, gluons, photons, electrons, neutrinos, W, Z etc are the relevant elementary particles. Hadrons are bound states of quarks and gluons, while e, ν, γ, W, Z etc are relevant when studying decays and collisions with hadrons. A peculiar feature of hadrons is that quarks and gluons seem to be permanently confined within them, we haven't been able to isolate a quark or a gluon from a hadron, unlike how we ionize atoms.

• The elementary particles we deal with in this course are primarily subatomic or subnuclear. Aside from its location or energy and momentum (which could change), an elementary particle is often characterized by its fixed mass, spin/angular momentum, intrinsic parity, charge, electric and magnetic moments and other quantum numbers like baryon and lepton number, strangeness, charm, color, weak isospin, weak hypercharge etc. Most of these attributes are associated with space-time or internal symmetries.

• For some purposes, Wigner's theorem (1939) gives a way of thinking about an elementary particle: one whose quantum states (labelled by momentum and spin projection) carry an irreducible unitary representation of the Poincare group of space-time symmetries. The relevant irreducible representations of the Poincare group are labelled by mass and spin. The group of space-time symmetries may be extended to include internal symmetries like electromagnetic gauge symmetry, leading to additional quantum numbers, some of which were mentioned above.

1.3 Cast of characters: Particles and fields of the standard model

• The standard model (SM) of particle physics is our current theory of elementary particles. It is a remarkably successful and elegant relativistic quantum field theory based on the 'gauge principle' and 'renormalizability'. The principle of local gauge invariance is a generalization of the gauge symmetry of electromagnetism. Renormalizability is essentially predictive power, i.e., that the theory must depend only on a finite number of free parameters, which once determined from experiment, allow definite predictions for all other physical quantities like scattering cross sections, decay rates and binding energies of bound states.

• To each elementary particle of the SM, there is associated a quantum field whose elementary excitation (produced by a creation operator acting on the vacuum) is the particle. Among other things, this makes it plausible why all photons (or all electrons, etc.) are identical: they are all produced from the same field. The particles are roughly divided into matter particles (fermionic spin-half quarks and leptons) and force carriers (spin one gauge bosons) and a spin zero Higgs boson. All these particles have been experimentally found and are point-like to current precision ($\sim 10^{-18} - 10^{-19}$ m).

1.3.1 Quarks and leptons

• The fundamental fermions (6 flavors of quarks and 6 leptons) may be arranged in three families of mass eigenstates. The quarks (named by Gell-Mann after the 3 quarks in J Joyce's Fennigan's Wake) are

$$\begin{pmatrix} up \\ down \end{pmatrix}, \begin{pmatrix} charm \\ strange \end{pmatrix}, \begin{pmatrix} top \\ bottom \end{pmatrix}, \quad \text{with charges} \quad \begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix} \quad \text{in units of the proton charge.} \tag{1}$$

u, c, t are called up-type quarks while d, s, b are down-type. Quarks feel the electromagnetic force since they are charged. Each quark comes in three colors (say red, green, blue, e.g. u_r, u_g, u_b). Color is to the strong force what charge is to the EM force. So quarks feel the strong force. In addition, the quarks carry weak-isospin and weak-hyper-charge quantum numbers which enable them to participate in the weak interactions. The up and down type quarks differ by one unit of electric charge, which permits weak transitions between them by emission or absorption of the $W^{\pm}$ particles of unit charge. We are over-simplifying a bit here, unlike EM and strong, the weak interactions are not parity invariant, they distinguish left from right-handed quarks[1]. A theory/phenomenon that distinguishes left from right (or mirror reflections) is called chiral. The quarks can be split into their left and right-handed components, which have different weak isospin T_3 and hyper-charge Y_W assignments. The up- and down-type left handed quarks[2] have $T_3 = \pm\frac{1}{2}$ while the right handed ones have $T_3 = 0$. Weak hyper-charges are given by $Y = 2(Q - T_3)$. Quarks also interact gravitationally since they carry energy. Quarks are unique in that they feel all the 4 forces.

• The leptons (Greek 'lepton' for thin/slender) comprise the electron, muon and tau as well as their neutrinos. They too come in 3 families of mass eigenstates

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}, \quad \text{with charges} \quad \begin{pmatrix} 0 \\ -1 \end{pmatrix} \quad \text{in units of the proton charge.} \tag{2}$$

Like the quarks, leptons carry weak isospin and weak-hyper-charge quantum numbers, but no color quantum numbers. Again, one needs to make a distinction between left handed and right handed leptons. The left-handed neutrinos and charged leptons have $T_3 = \pm\frac{1}{2}$ respectively. There are no right handed neutrinos in the standard model, while the right-handed charged leptons have $T_3 = 0$. As before the weak-hyper-charges are determined by $Y_W = 2(Q - T_3)$. As a consequence, leptons feel the weak and gravitational forces but not the strong force. The charged leptons (e, μ, τ) feel the electromagnetic force while the neutral neutrinos do not.

• The masses of the charged leptons are $m_e = 511$ keV/c^2, $m_\mu = 105.66$ MeV/c^2 and $m_\tau = 1.777$ GeV/c^2. Despite the historical name, lightness/slenderness is not a defining property of leptons. The tau lepton is almost twice as heavy as the proton ($m_p = 938$ MeV/c^2). The muon was initially called the μ-meson. That term is no longer appropriate. The word meson is now reserved for the subset of hadrons (strongly interacting particles) with integer spin (e.g. π, K, ρ mesons).

• Quark masses are only roughly determined since they have not been isolated from inside hadrons. However, using weak and electromagnetic probes/currents to look deep inside a proton, one may estimate their masses. The 'current' quark masses of the 'light' quarks are roughly

$$m_u \approx 0 - 5, \quad m_d \approx 6 - 10, \quad \text{and} \quad m_s \approx 130 - 200 \quad \text{MeV/c}^2. \tag{3}$$

The 'heavier' quarks have masses which are increasingly better determined (why?)

$$m_c \approx 1.3, \quad m_b \approx 4.2, \quad m_t \approx 173 \quad (\text{in GeV/c}^2). \tag{4}$$

[1] Quarks are spin half Dirac particles, associated to 4 component spinor fields. In terms of Dirac's γ matrices, $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ is hermitian traceless and squares to the identity. Its $+1$ eigenspace is the space of right handed spinors and its -1 eigenspace consists of LH spinors. Any spinor is uniquely the sum of LH and RH spinors. $P_{L,R} = \frac{1}{2}(1 \mp \gamma^5)$ project to the two subspaces. Helicity is the projection of angular momentum on the direction of momentum $h = \mathbf{J} \cdot \hat{p}$, eigenvalues of helicity are $\pm\frac{1}{2}\hbar$ for spin half particles. For massless particles, +/- helicity corresponds to right and left handedness.

[2] There is a complication due to quantum mechanical mixing. The weak interactions involve weak eigenstates rather than mass eigenstates. So it is not $(u, d)_L$ that forms a weak isospin doublet but $(u, d')_L$ where d' is a linear combination of the d-type quarks. But this does not affect the weak isospin T_3 assignments given here. In particular, a strange quark s has been seen to convert to an up quark in the weak decays of strange particles, this is allowed since there is a little bit of s inside d'.

In the standard model, the neutrinos were taken to be massless, though oscillation experiments show that neutrinos have masses, which however have not been determined beyond some upper limits. Cosmological constraints suggest that the sum of the 3 neutrino masses is less than 1 eV/c^2. The huge variation of quark and lepton masses by over 11 orders of magnitude, from less than an eV/c^2 to 173 GeV/c^2 is not understood.

• To each of these quarks and leptons, there is an anti-particle having opposite electric charge, color, magnetic moment, strangeness, isospin projection etc. Three is the smallest number of families that can accommodate CP violation, which is believed to be responsible for the preponderance of matter over anti-matter in the universe. The aesthetically pleasing quark-lepton symmetry with equal number of quark and lepton families ensures cancellation of possible anomalies that could render the quantum theory inconsistent. Check that the total charge of each family (quarks and leptons included) is zero.

• Naturally occurring matter is largely composed of particles from the first family u, d, e, (and the neutrinos $\bar{\nu}_e$ and ν_e from radioactive beta decay). Members of the 2nd and 3rd family decay to those of the first family. The muon decays to the electron (and 2 neutrinos) and the τ lepton decays to neutrinos and muon/electron or quarks in the first family. Particles (hadrons) containing the 2nd and 3rd families of quarks c, s, t, b are unstable and decay to hadrons containing u and d quarks.

• The second and third families are at first glance, more massive copies of the first family, though there is quantum mechanical mixing between the families to form three 'generations' of weak eigenstates e.g., $\begin{pmatrix} u \\ d' \end{pmatrix}, \begin{pmatrix} c \\ s' \end{pmatrix}, \begin{pmatrix} t \\ b' \end{pmatrix}$ when they participate in the weak interactions. (d', s', b') are linearly related to (d, s, b) via the 3×3 CKM (Cabbibo-Kobayashi-Maskawa) mixing matrix. $(u, d'), (c, s')$ and (t, b') each forms a doublet under weak isospin. Charge-changing weak interactions allow transitions between members of a generation (not a family). The fact that d' is a linear combination of d, s, b means that all three can decay to an up quark. Indeed $d \to u$ is seen in beta decay while $s \to u$ is seen in decay of strange particles. If there were no such mixing, the strange quark could not decay to a u. So mixing is both experimentally observed and permitted by the linearity of the quantum mechanical state space.

• Anti-quarks and anti-leptons have opposite handedness to their matter counterparts. So anti-quarks come in RH weak-isospin doublets and LH weak-isospin singlets. Similarly, $(\bar{\nu}_e, e)_R$ is a RH weak isospin doublet. LH charged anti-leptons are singlets under weak-isospin $\bar{e}_R, \bar{\mu}_R, \bar{\tau}_R$. Anti-neutrinos are always RH, there are no LH anti-neutrinos in the SM. The weak interactions treat the left and right-handed components differently (e.g. they have different T_3 and Y_W). We say that the weak interactions are chiral. Weak interactions to violate reflection (parity) symmetry, as we shall see later.

1.3.2 Hadrons

• Hadrons, though they are not point-like elementary particles like the quarks and leptons (at current precision), deserve special mention since most of the particles discovered so far are hadrons, there are hundreds of them. They were named (by L. Okun) after the Greek word 'adros' meaning bulky. Hadron is the collective name for strongly interacting particles that have been isolated. This would include nuclei since they too feel the strong nuclear force, though nuclei with more than one nucleon are more often referred to as nuclei rather than as hadrons.

• Hadrons are subdivided into integer spin bosonic mesons like $\pi, K, \eta, \rho, \omega, \phi$ and half-odd integer spin fermionic baryons (from Greek word 'barys' for heavy) like n, p, Δ, Λ, Σ, Ξ, Ω. The name meson (from Greek 'mesos' meaning middle or intermediate) seemed reasonable since the first discovered mesons (π, K) had masses ($m^{\pm}_{\pm} = 140$ MeV, $m_{K^{\pm}} = 494$ MeV) intermediate between that of the electron and of the heavier baryons (n,p with $m_n, m_p \approx 938 - 939 MeV$). However, mesons with mass more than that of the proton have been discovered.

• Hadrons are bound states of quarks and gluons. Mesons typically consist of a valence quark and anti-quark, while baryons consist of three valence quarks. Anti-baryons have three valence anti-quarks. Though quarks and gluons feel the strong interactions, they are not called hadrons, they have never been detected in isolation. In particle reactions we may produce quarks and gluons, but we only detect hadrons in detectors. Quarks and gluons either decay or 'hadronize' to form bound states, before we can detect them. What is more, it is empirically found (and theoretically suspected) that all hadrons are color neutral. It appears that color is confined within hadrons.

• The proton is made of two up and one down quark (its valence quarks), along with an infinite sea of virtual quark anti-quark pairs and gluons). For many purposes (determination of charge, baryon number, isospin, strangeness, magnetic moment etc., but not mass, momentum or structure) we may treat the proton as simply made of valence quarks. The neutron's valence quark content is $n = udd$. The proton and neutron form a strong isospin doublet with $I_3(p) = \frac{1}{2}, I_3(n) = -\frac{1}{2}$. The up and down quarks also form a strong isospin doublet with $I_3(u) = \frac{1}{2}, I_3(d) = -\frac{1}{2}$. Other quarks (c,s,t,b) do not carry strong isospin (i.e., they are singlets under strong isospin). The anti-proton and anti-neutron are $\bar{p} = \bar{u}\bar{u}\bar{d}, \bar{n} = \bar{u}\bar{d}\bar{d}$. $(\bar{n}, \bar{p})$ form an isospin doublet: $I_3(\bar{n}) = \frac{1}{2}, I_3(\bar{p}) = -\frac{1}{2}$. Baryons are assigned baryon number $B = +1$ while anti-baryons have $B = -1$. Though neutrons and anti-neutrons cannot be distinguished by their electric charge, they can be distinguished by their isospin, baryon number and magnetic moment!

• There are four kaons $K^+, K^-, K^0, \bar{K}^0$ (having masses around 500 MeV/c^2). They are strange mesons (contain a strange quark or strange anti-quark) made of the 'light' quarks u,d,s alone (charm quark with a mass of 1.3 GeV would be too heavy, as would b or t). $K^{\pm}$ are anti-particles of each other, as are K^0 and $\bar{K}^0$. K^+ and K^0 form an isospin doublet as do $\bar{K}^0$ and K^-. Based on this, try to figure out the valence quark content of the kaons. Do you get $K^+ = u\bar{s}$, $K^0 = d\bar{s}$, $K^- = d\bar{s}$ and $\bar{K}^0 = \bar{d}s$?

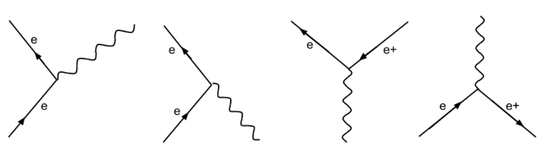
• The atoms of chemistry are bound states of electrons and nuclei, held together by electrostatic Coulomb forces between electrically charged electrons and nuclei. Similarly, hadrons are the atoms of hadronic physics, they are bound states of quarks and gluons, held together by strong forces between colored particles. Atoms are electrically neutral, and hadrons are color neutral. A major difference is that an electron can be isolated from an atom, while it has not been possible to isolate a quark or gluon from a hadron.

• Molecules of chemistry are bound states of atoms, held together by residual electromagnetic forces (van der Waal's forces including dipolar forces) between neutral atoms. Similarly, nuclei are the molecules of hadronic physics, they are bound states of several hadrons (in fact nucleons), held together by residual strong forces between the quarks and gluons inside different colorless hadrons. The simplest such hadronic 'molecule' (nucleus) is the isospin one deuteron, an np bound state. It is the nucleus of deuterium, a heavy isotope of hydrogen. It turns out there are no nn or pp bound states, the deuteron is the only nucleus with baryon number 2.

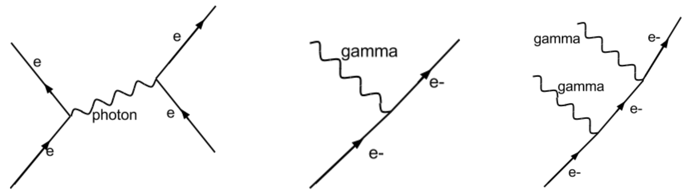
1.3.3 Force carriers, gauge and Higgs bosons, interactions and Feynman diagrams

• In the quantum field theoretic description, the force between a pair of electrons is (to first approximation) due to the exchange of a massless photon γ (emission and absorption of a photon), which is called the 'quantum' of the EM field. The photon is associated to the EM vector potential or gauge field $A_\mu(x)$. So the photon is called a vector boson or gauge boson. The cross section for the scattering of a pair of electrons is calculated using Feynman diagrams (FD). Each diagram gives an amplitude for a process. Amplitudes are added up and squared to get probabilities. The legs or lines describe free propagation of particles and vertices describe interactions. The simplest FD for the scattering of electrons is drawn in the figure. Time runs upwards. Incoming particles (incoming asymptotic states) are associated with external lines/legs at the bottom ($t \to -\infty$) of the FD. Outgoing particles are associated to external legs at the top of the diagram ($t \to \infty$). Incoming and outgoing particles (and anti-particles) are 'real' particles, they come from the source (like a hot filament or a radioactive nucleus or particle accelerator) and are received in the detector. They are characterized by being on mass shell, i.e. the square of their 4-momentum is equal to their mass: $p_\mu p^\mu = m^2$. The internal lines, like the wavy photon line correspond to virtual particles that are exchanged, they are intermediate states. Virtual particles need not be on mass shell and are not detected, they are for the purpose of calculating the amplitude. So $p_\mu p^\mu \neq 0$ in general, for the virtual photon that is exchanged, even though real photons are massless. In fact, to determine the amplitude, one needs to sum over all possible 4-momenta of the internal lines (virtual particles), subject to conservation of 4-momentum at each vertex where legs meet. This sum is much like the sum over intermediate states encountered in calculating energy shifts to second order in quantum mechanical perturbation theory:

$$H = H_0 + gH_1, \quad E_n = E_n^{(0)} + gE_n^{(1)} + g^2 E_n^{(2)} + \ldots \quad \text{where} \quad E_n^{(2)} = \sum_m \frac{\langle \psi_n | H_1 | \psi_m \rangle \langle \psi_m | H_1 | \psi_n \rangle}{E_n^{(0)} - E_m^{(0)}}. \tag{5}$$



• The arrows in the FD distinguish between particles and anti-particles. No arrow is placed on a line for a particle which is its own anti-particle, like the photon. For lines entering (ordered



by time) a vertex (as in the case of incoming states), the arrow points towards the vertex for electrons and away from the vertex for positrons. For lines leaving a vertex (as for the out-states), arrow points away from the vertex for electrons and towards the vertex for positrons. In the case of electrons and positrons, the arrow also indicates the direction of movement of negative electric charge.

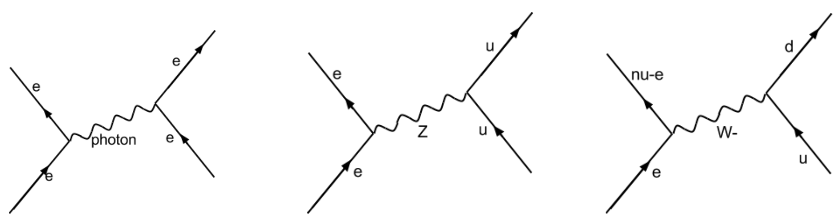
• The statement that a particle couples or interacts directly (or *at tree level*) with another particle is a consequence of the presence of interaction term(s) (cubic or higher order) in the SM Lagrangian that contains a product of the respective fields. Such an interaction may also be pictorially represented as a vertex in a Feynman diagram, as in the examples above. In the case of quantum electrodynamics, the interaction is $j^\mu A_\mu$ where $j^\mu = e\bar{\psi}\gamma^\mu\psi$. The photon field A is a linear combination of photon creation and photon annihilation operators $a^\dagger, a$ while ψ annihilates an electron or creates a positron. Meanwhile, $\bar{\psi}$ creates an electron or annihilates a positron. Thus the basic photon emission vertex $e^- \to e^-\gamma$ corresponds to destruction of an electron and creation of a photon and an electron $\bar{\psi}a^\dagger\psi$. However, it is important to recognize that this basic tri-linear interaction vertex does not describe a physical process, energy-momentum conservation would forbid the electrons and the photon to all be on mass shell – check this. The trilinear vertex is merely an ingredient that is used to construct diagrams describing real processes (i.e., those where all external legs are on-shell, as in the case of electron-electron scattering by one photon exchange). In particular $e^- \to e^-\gamma$ does not describe a real process. But electrons can radiate photons (as happens classically when they are accelerated). The simplest such process involves two photon emission $e^- \to e^-\gamma \to e^-\gamma\gamma$ where the electron in the intermediate state is a virtual electron. See the figure. An electron can also emit a single real photon in the Coulomb field of a nucleus, which provides the momentum balance. Draw the two leading order Feynman diagrams for elastic $e + e^- \to e^+e^-$ Bhabha scattering.

• Photons couple to all quarks and charge leptons in the same way as they couple to electrons, i.e., through the tri-linear vertex (weighted by the charge) $\gamma f\bar{f}$ where f is the charged fermion. So we also have scattering of muons via exchange of a photon.

• Feynman diagrams are a convenient way of depicting interactions and summarizing the procedure of determining amplitudes in perturbative quantum field theory. Perturbation theory applied to scattering processes is not the only method of deriving physical consequences from the SM Lagrangian. In particular, it is not very well-suited to the study of bound states. Nevertheless, a lot of what we know has been determined using perturbation theory.

• The other 'force carriers' include 8 massless colored gluons g for the strong interactions, and the massive weak gauge bosons $W^\pm$ and Z^0 ($m_{W^\pm} = 80$ GeV/c^2, $m_z = 91$ GeV/c^2) which mediate the charge-changing and charge-preserving weak interactions. All these are spin one gauge bosons associated to gauge fields. They are also called vector bosons since the corresponding gauge fields are vector fields like the electromagnetic vector potential $A_\mu(x)$. The numbers of these gauge bosons is equal to the dimensions of the Lie algebras of the various gauge groups. The EM (charge) gauge group $U(1)_Q$ is one dimensional, there is one type of photon. The color gauge group $SU(3)$ is 8 dimensional, resulting in 8 types of gluons. The electroweak interactions are governed by $SU(2)_L \times U(1)_Y$, corresponding to weak isospin and weak hypercharge. The cartesian product means the elements of weak isospin group commute with those of the hypercharge group. This is a 4 dimensional group, with generators corresponding to W^+, W^-, Z^0 and the photon. The generator corresponding to the third component of weak isospin 'T_3' and the generator corresponding to weak hypercharge 'Y_W' mix with each other (i.e. form 2 linear combinations) to give rise to the photon and Z^0 bosons.

• $U(1)$ is the 1-dimensional circle group of complex numbers of unit magnitude $e^{i\theta}$. SU(n) consists of $n \times n$ unitary matrices $g^\dagger g = I$ of determinant $\det g = 1$. Its Lie algebra consists of matrices that depart infinitesimally from the identity $g \approx I + A$. Then unitarity becomes anti-hermiticity $A + A^\dagger = 0$ and unimodularity becomes $1 \approx \det(1 + A) \approx 1 + \operatorname{tr} A$. So the Lie algebra of $SU(n)$ is the space of traceless anti-hermitian matrices. It is conventional to work with hermitian matrices by taking $A \to iA$. The space of such matrices is $n^2 - 1$ dimensional.



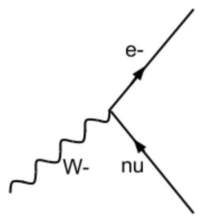
• Gauge fields at each space-time point are elements of the Lie algebra of the corresponding gauge group. So they can be written as a linear combination of a set of generators (basis elements of the Lie algebra). E.g. The Pauli matrices $\sigma_+, \sigma_-, \sigma_3$ are a basis for SU(2), so the weak isospin gauge field may be written as $W_\mu(x) = W_\mu^+(x)\sigma_+ + W_\mu^-(x)\sigma_- + W_\mu^3(x)\sigma(3)$. $W_\mu^\pm(x)$ are the gauge fields corresponding to the $W^\pm$ weak gauge bosons. In addition, we have the gauge field $B_\mu(x)$ for U$(1)_Y$. Then the photon and Z^0 fields are a pair of orthogonal linear combinations of W_μ^3 and B_μ:

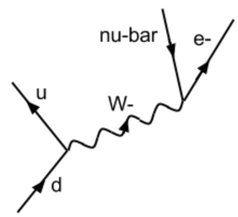
$$A_\mu = \cos\theta_w\, B_\mu + \sin\theta_w\, W_\mu^3 \quad \text{and} \quad Z_\mu = -\sin\theta_w\, B_\mu + \cos\theta_w\, W_\mu^3. \tag{6}$$

$\theta_w \approx 30$ degrees is called the weak mixing or Weinberg angle.

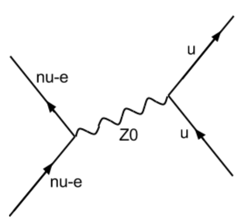
• The charge changing $W^\pm$ weak gauge bosons cause transitions between members of a lepton or quark generation, $(\nu_e, e), (\nu_\mu, \mu)(\nu_\tau, \tau), (u, d'), (c, s'), (t, b')$ e.g. the vertex $\bar\nu e W$ describes (for instance) transmutation of an electron into a neutrino while emitting a W^-. This is a lot like the vertex for photon emission by an electron in quantum electrodynamics. As before, this describes a virtual process, an electron is too light to decay into a W^-. But unlike in QED, the charge of the leptons changes since $W^\pm$ are charged. The same interaction term $\bar e \nu_e W$ also describes the decay of a W^- into an electron and anti-neutrino.

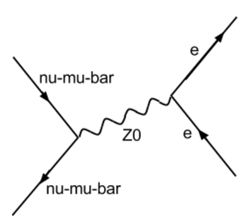
• Similarly we have the charge raising process $d' \to uW^-$. Since the weak eigenstate d' is a combination of d, s, b mass eigenstates, this means a d quark (mass eigenstate) can make a transition to a u quark by W^- emission. This is what happens in beta decay of a neutron. Indeed if we put together $d \to uW^-$ and $W^- \to e^-\bar{\nu}_e$ we get $d \to ue^-\bar{\nu}$.





• The photon couples to quarks and charged leptons. The Z^0 being another linear combination of the same gauge fields has somewhat similar properties. It is neutral and its own anti-particle. The Z^0 boson couples to all quarks and leptons (neutrinos included) through the $Zf\bar{f}$ vertex: $e^- \to e^-Z, \nu_e \to \nu_e Z$ or $Zu \to u$ or $d \to Zd$ etc[3]. Z^0 interactions are charge preserving weak interactions (at any given vertex), they are also called neutral current interactions, they do not change the flavor of quarks or leptons. One example is neutrino nucleon scattering. The basic process is neutrino quark scattering, e.g., $\nu_e u \to \nu_e u$. The up quark in the final state hadronizes and there is typically a shower of hadrons produced. This is seen in the detector as a sudden hadron shower without the track of any incoming particle (neutrinos do not leave tracks like charged particles). Another such process is (anti-) neutrino electron elastic scattering $\bar{\nu}_\mu e \to \bar{\nu}_\mu e$ by Z^0 exchange. This is a 'gold-plated' process since it cannot be mediated by anything but a Z and has minimal contamination from 'backgrounds' (other electron recoil processes that could mimic it). Since it is a muon-anti-neutrino, it cannot couple to an electron via $W^\pm$, and neutrinos do not couple to photons or gluons either. This is one of the processes by which neutral currents were first discovered in 1973 at the giant bubble chamber 'Gargamelle' at CERN. In the bubble chamber, one photographed the recoil of an electron when exposed to a beam of muon anti-neutrinos, with missing energy in the final state carried away by the $\bar{\nu}_\mu$.





• Photons couple *directly* to the charged leptons and quarks. They do not interact directly with other photons, or with gluons[4]. But due to electro-weak mixing, they do couple directly

[3]Unlike the electromagnetic vertex $Af\bar{f}$ which is always proportional to the electric charge of f, the $Zf\bar{f}$ vertex involves both the electromagnetic and weak charges and the weak mixing angle, due to EW mixing. (It is not just proportional to the weak hyper-charge or weak isospin). There is another difference between γ and Z^0 interactions, the latter are parity violating.

[4]But can you draw a higher order 'loop' diagram by which photons can scatter off each other?

to the $W^\pm$ and Z^0 weak gauge bosons via $AWW, AAWW, AZWW$. Similarly, the weak gauge bosons couple to each other $ZWW, WWWW$, $WWZZ$ too. These interactions are determined by the gauge principle along with the constraints of renormalizability. The field strength associated to a (generally non-abelian) gauge field is

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu, A_\nu] \tag{7}$$

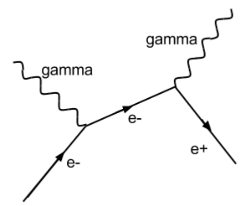
where g is called the gauge coupling constant. For an $SU(2)$ gauge theory, $A_\mu(x)$ is a 2×2 matrix in the SU(2) Lie algebra, and different Lie algebra elements do not commute in general. For an abelian gauge group like U(1) of EM, the commutator term vanishes. The field strength is also a 2×2 matrix in the Lie algebra. Now the Lagrangian density includes the term $-\frac{1}{4}$ tr $F_{\mu\nu}F^{\mu\nu}$ just as in E & M, except that we need to take the trace to get a gauge-invariant real function of x.

• The quartic interaction vertices among gauge bosons arise from the product of a pair of commutators, one from each field strength. For example, for the $SU(2) \times U(1)$ theory, we have 4 gauge fields W^+, W^-, Z, A leading to quartic terms such as $g^2[W^+, W^-][W^+, W^-]$ and $g^2[W^-, Z][W^+, A]$ etc. However, the $[A, Z]$ commutator vanishes since the third component of weak isospin commutes with the generator of hypercharge as well as with itself (recall that Z, A are both linear combinations of W^3_μ and B_μ), which explains the absence of an $AZAZ$ vertex.

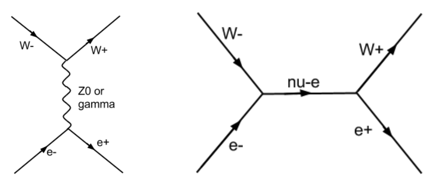
• The cubic interaction vertices AWW and ZWW come from the terms $(\partial A)[W^+, W^-]$ and $(\partial Z)[W^+, W^-]$ in the Lagrangian. In fact, it is via the AWW (i.e., $\gamma \to W^+W^-$) coupling that the W bosons were produced copiously at CERN and SLAC beginning in the 1990s (They had been discovered earlier in 1983 at CERN via proton anti-proton collisions). e^+e^- annihilation via head on collisions resulted in $e^+e^- \to \gamma \to W^+W^-$ or $e^+e^- \to Z^0 \to W^+W^-$. If one plots the total cross section for e^+e^- annihilation as a function of center of mass energy of the colliding electrons and positrons, there is a peak (called the Z-pole) around $m_Z = 91$ GeV. Such peaks in cross section are interpreted as due to the production of resonances, short-lived particles. In this case, the short-lived particle is the 'intermediate' Z boson. This is analogous to what happens in quantum mechanical potential scattering. If a particle scatters against an attractive potential well, then one finds a peak in the cross section associated to each of the bound states that the potential supports. However, this does not mean that Z^0 is a bound state of e^+ and e^-, that would have to be the case in non-relativistic qm, but relativistic QFT allows for the production of new particles which are not simply bound states of the colliding particles. Note that there is no W boson peak in the e^+e^- scattering cross section since there is no $e^+e^- \to W$ vertex. On the other hand, the W bosons were originally discovered in 1983 at CERN in $p\bar{p}$ collisions, where a quark and anti-quark collided to form a W, $u\bar{d} \to W^+ \to e^+\nu_e$. So one expects a peak in the $u\bar{d}$ collision cross section at a CM energy equal to the mass of the W boson. The Z boson was also discovered in the same $p\bar{p}$ experiments at CERN via processes such as $u\bar{u} \to Z^0 \to e^+e^-$ and $d\bar{d} \to Z^0 \to e^+e^-$.

• Gauge invariance alone would permit terms in the Lagrangian such as tr $F_{\mu\nu}F^{\mu\nu}F_{\rho\sigma}F^{\rho\sigma}$. However, the corresponding quantum theory is not renormalizable. Renormalizability precludes vertices where more than 4 gauge bosons meet.

• Matter particles aren't the only ones that can be detected as real particles. Photons too can be detected and counted. Conversely, the force carriers (gauge bosons) are not the only possible particles that can mediate processes. Matter particles too can feature as virtual intermediaries with real gauge bosons in initial/final states. E.g. The leading order FD for e^+e^- annihilation to produce two real photons is shown.

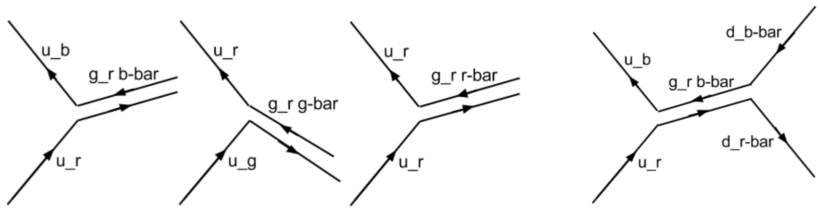


• Another example: at sufficiently high energies, the collision of an electron and positron can produce a real W^+W^- pair, the process being mediated by photon or Z exchange. A W boson, though it typically decays in less than 10^{-24}s, weighs nearly as much as a Rubidium Rb^{85}_{37} nucleus (80 GeV$/c^2$), yet it seems to be point-like. On the other hand, a lepton or quark could occur as an intermediary. Indeed the same process $e^+e^- \to W^+W^-$ could proceed through the exchange of a virtual neutrino.



• Since gluons carry color quantum numbers, they feel the strong force both among themselves via the 3 and 4 gluon vertices $ggg, gggg$ *and* while interacting with quarks $\bar{q}gq$. But gluons do not participate in electroweak interactions directly.

• The basic quark gluon vertex $\bar{q}gq$ is a tri-linear vertex just like the photon-electron vertex. Gluons carry color, so they could change the color of a quark. But gluons do not carry flavor and strong interactions do not convert d to u or c to d etc (only charge changing weak interactions can do this). If we label the three colors r, g, b, then a red up quark could convert to a green up quark by emitting a red-anti-green gluon. Similarly we could have $u_b \to u_g + g_{b\bar{g}}$ etc. For a given quark flavor, there are 8 different types of quark-gluon vertices, which are the independent terms in the double sum over colors $\bar{u}^a A^b_a u_b$. They correspond to the 8 types of gluons which form a basis for the space of traceless 3×3 hermitian matrices in color space. The gluon field being a matrix in color space has a row and a column index $(A_\mu)^a_b$, which may be regarded as a color and anti-color index. There are also two diagonal gluons, say $g_{r\bar{r}}$ and $g_{b\bar{b}}$ (with $g_{g\bar{g}} = -g_{r\bar{r}} - g_{b\bar{b}}$ being a linear combination due to tracelessness.) Gluon legs/propagators in Feynman diagrams are drawn either in the form of a coil/spring or with a double-line, one each for the color and anti-color index. In the FD shown, quarks have a color index (e.g. u_r) while anti-quarks (e.g., the down anti-quark $d_{\bar{r}}$) have an anti-color index. The direction of arrows distinguishes between quarks and anti-quarks.



• Due to their strong self-interactions, gluons alone (i.e. without any quark 'matter'), can combine to form color and charge neutral baryon number zero bound states (glueballs) that can even be heavier than a proton! However, glueballs have not yet been unambiguously identified in experimental data possibly due to mixing with mesons having the same quantum numbers, though there are several candidates.

• In addition, there is a spin zero massive Higgs particle ($m_H = 125$ GeV$/c^2$, discovered at CERN in 2012) associated to the Higgs field. In the SM, the Higgs field is responsible for the masses of the weak gauge bosons $W^{\pm}, Z^0$, the quarks and charged leptons. The Higgs particle couples to all these massive particles with a strength proportional to the corresponding masses.

• At present, there is no quantum theory of the gravitational force (though there are efforts to understand the issues involved) nor any experimental detection of quantum gravitational effects. The graviton is hypothesized to be the massless spin-2 quantum of the gravitational field.

1.3.4 General remarks on QCD and the Electroweak standard model

• The standard model is sub-divided into a gauge theory of the strong interactions based on the SU(3) group (Quantum Chromodynamics - QCD) and the SU$(2)_L\times$ U$(1)_Y$ electro-weak theory, sometimes called the Glashow-Weinberg-Salam model[5] which governs the electromagnetic and weak interactions. Sometimes this is called electroweak unification. Though the electroweak theory involves a unified treatment, it is not a unification in the technical sense since there are two independent coupling constants, one for SU$(2)_L$ and one for U$(1)_Y$ (in addition to the strong coupling constant for color SU(3))[6]. However, it involves quantum mechanical mixing between the weak isospin and hypercharge generators to result in gauge bosons with definite masses: the photon and Z^0. While the weak hypercharge group is abelian, the weak isospin and color groups are non-abelian. The electroweak theory includes Quantum Electrodynamics (QED).

• Electrons are familiar constituents of matter while muons are detected in cosmic ray showers

[5] Occasionally called Quantum Flavordynamics (QFD), though this can be confusing. Charge changing weak interactions can change the flavor of quarks, say from strange to up, which is the reason for the name. However, the gauge bosons do not carry flavor quantum numbers and flavor is not a gauge symmetry like color or electric charge. The gauge symmetries of the GWS model are weak isospin and weak hypercharge.

[6] There is one independent coupling constant for each simple group or $U(1)$ that appears as a factor in the gauge group. It appears as g in formulae such as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu, A_\nu]$ as well as at vertices (like electric charge in $e\bar{\psi} A \psi$). Grand unified theories seek to embed the SM gauge group in a single simple group resulting in a theory with a single coupling constant. The simplest GUT predicted proton decay with an estimated half-life of order 10^{31} years, but the experimental lower bound on proton half life is now about 10^{34} y.

and from particle (pion, kaon) decays in accelerators, τ leptons have been produced (briefly) in accelerator collisions. Neutrinos come to us from nuclear reactors, nuclear beta decay, thermonuclear fusion in the sun, cosmic ray showers etc. On the other hand, quarks and gluons have never been detected or isolated. For the first time in physics we have a situation where some of the fundamental constituents of matter (quarks and gluons) seem to be hidden from detection, permanently confined inside other particles (the hadrons)! The mechanism for this confinement is not well understood, though it is likely to be a consequence of QCD[7]. As mentioned, nuclei too are held together by strong forces between nucleons. Thus both hadronic and nuclear structure are, in principle, described by QCD.

• Knowing the atomic hamiltonian is only a small step along the way to understanding the structure and interactions of atoms. Similarly, and especially due to confinement, knowing the Lagrangian of the standard model is but a small step towards understanding the structure, properties and phenomena associated with the particles of the sub-atomic world. Moreover, to appreciate why the standard model is as it is and to develop an intuition for subatomic phenomena, one benefits from a study of the evolution of the subject since the 19th century. What is more, one does not need to know either QCD or the electroweak theory to understand much of the physics of beta decay or pion nucleon interactions, since there are quantitative (low energy) effective theories for these phenomena. On the other hand, the consequences that follow from the Lagrangian of the standard model are only partly understood, especially in relation to the low energy behavior of the strong interactions and the high energy behavior of the electroweak interactions (since perturbation theory breaks down in these regimes). In particular, we do not know how to calculate the masses of pions or nucleons from QCD, nor the 'wave function' of the quarks inside. What is more, it is quite possible that the description of strong interactions in terms of (unisolated) quarks and gluons, may not be the only possible one. There may be other 'dual' formulations that are better suited to deriving the physical properties of the observed hadrons. Finally, there are many unexplained features and 17 input parameters (quark and lepton masses, mixing angles, coupling constants) that enter the standard model, which await a deeper understanding. Are the strong and electroweak interactions unified at some higher energy? Do quarks and leptons have sub-structure? Quite apart from all of this, there is a lot of new and deep mathematics and physics involved in understanding the standard model and related theories. Just as understanding atomic physics was stimulated and accompanied by many technological developments (semiconductors, lasers, microwave technology etc) there are numerous technological developments that are associated with particle physics (detector and accelerator technology, superconducting magnets, use of accelerators for condensed matter/materials research, medical diagnosis and treatment). All these features make the physics of particles and fields a very rich, deep and active subject.

1.4 Orders of magnitude and natural units

• In atomic physics the typical energies of radiative transitions are electron volts or tens of eV, as are ionization energies. Binding energies (energy required to dissociate an atom into electrons

[7]However, it is possible that at high temperatures and densities present in the early universe, quarks and gluons were in a different quark-gluon plasma phase, and not confined in hadrons.

and nucleus) range from 10-1000 eV with increasing atomic number Z. Atomic dimensions are of order Angstroms (10^{-8} cm) or tens of angstroms. Lifetimes of atoms in excited states are typically of order nanoseconds (for electric dipole 'E1' transitions), though they can be longer if forbidden by a selection rule for E1. At these energies and dimensions, nuclei (and electrons) may be treated as 'elementary' point particles, they are in their ground states since the energy is not adequate to excite nuclei. The electron is point-like to current precision ($10^{-18} - 10^{-19}$ m).

• When energy exchanges are of order $.1-10$ MeV nuclei may be excited and show a spectrum of excited states just like atoms. The energies encountered in particle physics range from eV to TeV at present. The lightest charged particle, the electron, has a mass of 511 keV/c^2. Protons and neutrons have a size of order of a fermi (femtometer, 1 fm = 10^{-15} m). Photons can have a variety of energies depending on their wavelength. For example, 800-1500 MHz photons from a mobile phone antenna have an energy of $h\nu \approx$ micro eV, substantially less than that of UV or visible light (eV) and insufficient to cause radiation damage by ionizing atoms in our body. [Recall that $h = 6.6. \times 10^{-34}$ Js while $\hbar = 10^{-34}$ Js $= 6.6 \times 10^{-16}$ eV s]. Photons in the Balmer emission series ($n_f = 2$) have eV energy corresponding to electronic transitions in hydrogen, and lie in the visible region of the spectrum (that is why Balmer discovered them first, before the $n_f = 1$ UV Lyman series). X-ray photons (discovered as radiation emitted by accelerated Cathode ray electrons) have energies of order keV or 10s of keV. Primary X-rays can ionize atoms, which then relax producing secondary X-rays. The name gamma rays is often used for photons of MeV energy and above, initially discovered in gamma radioactive decays of nuclei.

• The SI system takes as basic units a length (meter), a mass (kg) and a time (second). CGS works with cm, gram and second[8]. In particle physics, both relativity and quantum mechanics are relevant and $\hbar = 6.6 \times 10^{-22}$ MeV s and $c = 3 \times 10^8$ m/s are universal constants. So it is convenient to take the basic units to be a speed (in units of c), an action (in units of $\hbar$) and a mass (or energy). In these units the speed of light $c = 1$ and $\hbar = 1$. Measuring speeds in units of c is like Mach units in aerodynamics, where speeds are quoted in units of the sound speed. The unit of mass (or energy) could be the rest mass (or rest energy) of a particle playing an important role in the discussion or quite often GeV/c^2 (or GeV) which is quite close to the rest mass (energy) of a nucleon.

• In natural units ($c, \hbar$, mass), all physical quantities are expressed as a numerical factor times $c^\alpha \hbar^\beta M^p$. The powers are all determined by dimensional analysis, there is no ambiguity. In natural mass units, we say that all quantities (denoted by primed variables below) have dimensions of mass to some power M^p in units of c and $\hbar$ to some other powers. E.g., momenta have dimensions of mass in units of c, $p = p'c$. Energies have dimensions of mass in units of c^2, $E = E'c^2$. Lengths have dimensions of inverse mass in units of $\hbar/c$, $l = l'\hbar/c$ as is seen from the de Broglie wavelength $\lambda = h/p$. Times also have dimensions of M^{-1} in units of $\hbar/c^2$, $t = t'\hbar/c^2$. Velocities are dimensionless (i.e. M^0) when measured in units of c, $v = v'c$. Action and angular momenta are dimensionless (i.e. M^0) in units of $\hbar$: $L = L'\hbar$. Force has dimensions of M^2 in units of $c^3/\hbar$.

• If instead we use the basic units $c, \hbar$ and an energy, then all physical quantities (denoted

[8]The CGS unit of force is a dyne (1 Newton = 10^5 dyne) and energy is an erg (1 Joule = 10^7 ergs) since 1m $=$ 100 cm and 1 kg = 1000 g.

by tilde variables here) have dimensions of a power of energy in units of appropriate powers of $\hbar$ and c. E.g. $E = \tilde{E}$. Masses have dimensions of energy in units of c^{-2}, $m = \tilde{m}/c^2$. Momenta have dimensions of energy in units of $1/c$, $p = \tilde{p}/c$. Lengths have dimensions of inverse energy in units of $\hbar c$: $l = \tilde{l}\hbar c$. Time has dimensions of inverse energy in units of $\hbar$, $t = \tilde{t}\hbar$, etc. So quantities in natural energy units may be converted to natural mass units by multiplying by an appropriate power of c^2, e.g., $\tilde{t}c^2 = t'$. Most often we talk in energy units where masses, momenta and energies are all quoted in GeV, lengths and times in inverse GeV.

• What is a fermi in energy units? Ans: 5 GeV^{-1}. To see this note that $\tilde{l} = \frac{1\,fm}{\hbar c}$. Moreover,

$$\hbar c = 6.6 \times 10^{-22}\text{MeV s} \times 3 \times 10^8 m/s = 197\,\text{MeV fm} \tag{8}$$

Thus $\tilde{l} = 1/(197\,\text{MeV}) \approx 1/.2$ GeV$^{-1} \approx 5$ GeV^{-1}.

• In natural units, the reduced Compton wavelength of a particle is the reciprocal of its mass $\frac{\lambda}{2\pi} = \frac{\hbar}{mc} = \frac{1}{m}$ in units of $\hbar/c$ or equivalently the reciprocal of its rest energy in units of $\hbar c$. A particle of mass 100 GeV/c^2 (the Higgs particle has a rest mass energy of 125 GeV) has a reduced Compton wavelength of $\frac{\hbar}{100GeV/c^2 c} = \frac{\hbar c}{100GeV}$ which is 2×10^{-3} fm, which is about a thousandth the size of a proton. Equivalently, a reduced Compton wavelength of 10^{-18} m is (1/200) GeV^{-1} corresponding to a particle of mass 200 GeV/c^2.

• If τ is the mean life-time of an unstable particle or state, its energy width is defined as $\Gamma = \hbar/\tau$ or $\Gamma = \tau^{-1}$ in natural units. So an energy width of 1 MeV (or a life-time of 1 MeV^{-1}) corresponds to a mean life of 6.6×10^{-22} s. Indeed many hadronic resonances which decay via the strong interactions have a life-time of order 10^{-23} s.

• A kilogram is a huge mass of 5.6×10^{26} GeV/c^2. A Joule is 6.24×10^9 GeV. A meter is 5×10^{15} (GeV)$^{-1}$. A second is 1.5×10^{24} GeV^{-1}. Or a time of an inverse GeV is 6.7×10^{-25} s.

• The physical parameters defining the hydrogen atom are the electron mass and charge, treating the proton as infinitely massive. The mean speed of an electron in the g.s. of hydrogen is α in units of c, $v = \alpha c$. The fine structure constant $\alpha = \frac{e^2}{4\pi\hbar c}$ (in rationalized units) will be discussed shortly. The binding energy of the g.s. of the hydrogen atom (one Rydberg) is Ry $\frac{1}{2}mv^2 = \frac{1}{2}mc^2\alpha^2$ or $\frac{1}{2}m\alpha^2$ in units of c^2. Being a length, the Bohr radius is an inverse mass in natural units, so it must be inversely proportional to the only mass in the problem, that of the electron, the proportionality factor can be calculated, it is $1/\alpha$. So $a_0' = 1/(\alpha m)$ in natural units i.e., in units of $\hbar/c$. So the Bohr radius is $a_0 = a_0' \frac{\hbar}{c} = \frac{4\pi\hbar^2}{me^2}$. $a_0 \to 0$ as $\hbar \to 0$, indicating classical collapse of electron into nucleus. Also a_0 decreases with growing electric charge which increases the electrostatic attraction to the nucleus.

• In atomic physics, the binding energies of atoms (of order 10s or 100s of eV) are a very small fraction of the mass energies of constituents (GeV for the proton and half an MeV for the electron). So one talks of the energy of the ground state of hydrogen relative to the energy of far separated proton and electron. In particle physics, binding energies can be comparable or even more than mass energies of constituents (essentially because we deal with relativistic rather than non-relativistic bound states). So energies of states in particle physics are usually total energies E, including both binding BE and mass energies ME of constituents.

• Using dimensional analysis, factors of $\hbar$ and c can be re-introduced into formulas obtained in natural units. E.g. any length l' given as an inverse mass must simply be multiplied by the numerical value of $\hbar/c$ to get its SI or cgs value. Similarly, a time $\tilde{t}$ given as an inverse energy must be multiplied by $\hbar$.

• Prefixes for powers of 10 include (for small numbers, small letter) micro 10^{-6}, nano 10^{-9}, pico 10^{-12}, femto 10^{-15}, atto 10^{-18} and for large numbers (capital prefix) mega (million 10^6), giga (billion 10^9), tera 10^{12}, peta 10^{15}, exa 10^{18}. For example, PCs with a terabyte of hard disk storage are common. Collisions at the LHC produce peta bytes of data per year. The global internet traffic was of order 21 exabytes per month in 2010. The LHC is designed to collide protons of energy 7 tera eV. The fastest super computers reached 10-20 petaflops (floating point operations per second) in 2012. ISRO plans to make a 133 exaflops super computer by 2017. Electrons, photons etc seem to be point-like down to attometer precision. Pulses of light of femto (and even atto-second) duration have been produced and used to observe chemical reactions. Femtosecond lasers are used in eye surgery. Other small numbers arise as differences. Experimental tests of the equivalence principle show that the difference between inertial and gravitational masses is less than one part in 10^{12}. The difference between the measured value and QED prediction for the electron anomalous magnetic moment $(g-2)/2$ is less than one part in 10^8.

• In quantum field theory and particle physics, it is convenient to use rationalized Heaviside-Lorentz units for charge, In Gaussian units, factors of 4π appear in Maxwell's equations rather than in Coulomb's law, making them convenient in charged particle mechanics and atomic physics. In HL units (like SI), factors of 4π appear in Coulomb/Biot Savart laws rather than in Maxwell equations, which makes them convenient in field theory. HL and Gaussian units are both CGS systems, they only differ in the treatment of factors of 4π for charge. In HL units, Maxwell's equations are written as

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times E = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \cdot \mathbf{E} = \rho \quad \text{and} \quad \nabla \times \mathbf{B} = \frac{\mathbf{j}}{c} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}. \tag{9}$$

In HLU, charges are normalized so that Coulomb's law takes the form $\mathbf{F} = \frac{q_1 q_2 \hat{r}}{4\pi r^2}$. The electric field of a point charge is $\mathbf{E} = \frac{q}{4\pi r^2}\hat{r}$ and the Biot-Savart law reads $\mathbf{B} = \frac{1}{4\pi c}\oint \frac{I d\mathbf{l}\times\mathbf{r}}{r^3}$. Charges in HLU are related to Gaussian (electrostatic units esu or statcouloumb) units via $q_{hl} = \sqrt{4\pi} q_g$ while the fields and potentials are $E_{hl} = \frac{E_g}{\sqrt{4\pi}}$, $B_{hl} = \frac{B_g}{\sqrt{4\pi}}$ and $A_{hl} = \frac{A_g}{\sqrt{4\pi}}$. It follows that the expression for the conjugate momentum $\pi = \mathbf{p} - e\mathbf{A}/c$ and the Lorentz force law $\mathbf{F} = q(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B})$ and fields in terms of potentials $\mathbf{E} = -\nabla\phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}$, $\mathbf{B} = \nabla \times \mathbf{A}$ take the same form in both Gaussian and HL rationalized units. The magnitude of the charge of the electron in Gaussian units is $e_g = 4.8 \times 10^{-10}$ statcoul. In going from formulae in SI units to HLU, ϵ_0, μ_0 are put equal to one, since they have been absorbed into the definition of charge, and all times come with a factor of c. Sommerfeld's dimensionless fine structure constant in HLU is $\alpha = \frac{e^2}{4\pi\hbar c}$. Putting in the values $c = 3 \times 10^{10}$ cm/s and $\hbar = 1.05 \times 10^{-27}$ erg.s we find $\alpha = \frac{4.8^2 \times 10^{-20}}{1.05 \times 3 \times 10^{-17}} = 7.3 \times 10^{-3} \approx 1/137$. So in these units, electric charge has dimensions of M^0 in units of $\sqrt{\hbar c}$: $e = e'\sqrt{\hbar c}$ where $e' = \sqrt{4\pi\alpha}$.

• In Gaussian and HL units, E, B and A have different dimensions than in SI. Check that

$$[q] = \sqrt{\hbar c} = \sqrt{ML^3}/T, \quad [\mathbf{E}] = [\mathbf{B}] = [\mathbf{F}/q] = \frac{M^2 c^{5/2}}{\hbar^{3/2}} = \frac{M^{\frac{1}{2}}}{TL^{\frac{1}{2}}} \quad \text{and} \quad [\mathbf{A}] = [\phi] = \frac{Mc^{3/2}}{\hbar^{\frac{1}{2}}} = \frac{\sqrt{ML}}{T}.$$

So for instance, eA/c has dimensions of momentum as needed in $\pi = p - eA/c$. In HLU and Gaussian natural units, charge and current density have the dimensions

$$[\rho] = \frac{M^3 c^{7/2}}{\hbar^{5/2}} = \frac{\sqrt{ML}}{T^3} \quad \text{and} \quad [\mathbf{j}] = [\rho c] = \frac{M^3 c^{9/2}}{\hbar^{5/2}} = \frac{\sqrt{M}}{T^2\sqrt{L}} \tag{10}$$

Check that these are the dimensions of charge per unit volume and charge per unit area per unit time. A useful 'take away' is that in natural units, charges have dimensions of M^0, gauge potentials have dimensions of M, while electric and magnetic fields (or field strengths $F_{\mu\nu}$) have dimensions of M^2 and charge and current densities have dimensions of M^3.

• In Gaussian & HL units, the 4-vector gauge potential is $A^\mu = (\phi, \mathbf{A})$. Introduction of the gauge potential automatically solves the homogeneous Maxwell equations. Faraday's field strength is $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ while the 4-current is $j^\mu = (c\rho, \mathbf{j})$. The inhomogeneous Maxwell equations take the form $\partial_\mu F^{\mu\nu} = \frac{1}{c} j^\nu$ along with the consistency condition (since $F^{\mu\nu}$ is anti-symmetric) $\partial_\mu j^\mu = \frac{\partial\rho}{\partial t} + \nabla \cdot \mathbf{j} = 0$ for local charge conservation.

• Cross sections have dimensions of area. A barn $= 10^{-28}$ m^2 $= 100$ fm^2 is a unit of area used in particle physics, it is a rather large area. The cross sectional area over which the charge of the proton (charge radius $.8fm$) is distributed is about 2 fm^2 $= 20$ mb (millibarns). The word barn is based on a joke/code from the WW2 days of atom bomb research: the uranium nucleus looked as big as a *barn* to incoming neutrons. A uranium nucleus has a cross sectional area of about 200 fm^2 or 2 barns. Since 1 fm is 5 GeV^{-1} a cross section of 1 GeV^{-2} is 0.4 millibarns.

• When examining a process such as $e^+e^- \to Z^0$ at a collider experiment or Higgs production in pp collisions, one is interested in the expected number of interesting events (per unit time). Suppose the cross section for the specific process is predicted to be σ. Then the luminosity $\mathcal{L}$ of the collider is a quantity with dimensions of $L^{-2}T^{-1}$ in terms of which the event rate is given by $\mathcal{L}\sigma$. In the case of a collider, we have two counter rotating beams. The particles usually go round f times per unit time in n_b bunches, each of which consists of n_p particles and has a cross sectional area A. The beams intersect at detector locations where collisions occur. Now the number of clockwise rotating particles (say protons) that arrive at the detector per unit time per unit area normal to the beam is given by fn_bn_p/A, this is the beam intensity or flux. Now each of these particles could collide with any of the n_p particles (say protons) in a counter rotating bunch. So the maximum number of collisions per unit area per unit time is roughly $\mathcal{L} = fn_bn_p^2/A$, which is called the luminosity. Not all these collisions are expected to produce a Higgs, only a fraction determined by the cross section σ, so the expected rate of Higgs events is $\mathcal{L}\sigma$.

• For example, the design luminosity of the LHC is 10^{34} cm^{-2} s^{-1}, with 2808 bunches, each containing 1.15×10^{11} protons going round the ring at a frequency of 40 MHz. One can then estimate the beam cross sectional area. The cross section for Higgs production in pp collisions at a center of mass energy of 7 TeV at CERN is about 15 picobarns. How many Higgs events are expected per year? The integrated luminosity of an accelerator is the time integral of the luminosity it delivers to the detectors, this is what really matters, accounting for down-time. It has dimensions of an inverse cross section and is usually quoted in inverse femtobarns.

1.5 Comparison of strong, weak, electromagnetic and gravitational interactions

• We have already mentioned that among the matter particles of the standard model, only the quarks feel the strong force, which binds them into hadrons like the proton. Quarks and charged leptons feel the electromagnetic force. All quarks and leptons participate in the weak interactions. All particles feel the gravitational force. Let us get an idea of the relative strengths, ranges and time scales associated with these four forces. The forces cause scattering between particles, decays of unstable particles as well as formation of bound states.

1.5.1 Comparison of ranges of the 4 interactions

• In the quantum field theoretic description, a force between particles can be modeled as due to the exchange of a mediating particle (typically a boson), the quantum of the force. Photons mediate the EM force while pions mediate the strong force between nucleons and $W^{\pm}$ mediate the weak force responsible for beta decay. Heuristically, the range of the force transmitted by a carrier of mass m is of order the reduced Compton wavelength $\lambdabar = \hbar/mc$ of the carrier (setting aside surprises like confinement). This may be familiar from the form of the Yukawa potential $-\frac{g^2}{4\pi r}e^{-r/\lambdabar}$ between a pair of nucleons, which we will derive later (g is a strong analogue of electric charge, assigned to each of the nucleons). Roughly, a heavier intermediary exists for a shorter duration and can travel less distance. According to the energy-time relation, the intermediate state exists for a duration of order $(\hbar/mc^2)$[9]. In that time it can travel at most a distance $c \times \hbar/mc^2$. If there is more than one possible intermediary, then exchange of the lightest one dominates with growing separation.

• The gravitational force between masses has an infinite range (there is no characteristic distance over which it decays exponentially), as does the electromagnetic force between charges, both decreasing inversely with the square of separation $V(r) = -GM_1M_2/r$ and $V = q_1q_2/4\pi r$. This is associated with both being mediated by massless quanta, the graviton and the photon, with infinite Compton wavelengths.

• The strong force between nucleons (within nuclei or when scattered), has a range of about 1-2 fermi. Based on this, Yukawa predicted the existence of π-mesons with a mass in the 100-200 MeV range which would mediate the inter-nucleon force. Just such pions with masses $m_{\pi^\pm} = 140$ MeV were discovered in 1947. The range of the strong inter-nucleon force is of order the Compton wavelength of the pions (about 1.5 Fermi).

• Weak interactions (like those responsible for $n \to p^+e^-\bar{\nu}_e$) have the shortest range corresponding to the very massive weak gauge bosons $m_W = 81$ GeV, $m_Z = 90$ GeV (their Comp-

[9] Our estimate for the energy difference between initial and intermediate state as the rest energy mc^2 of the mediator is valid when the scattering particles have an energy small compared to the rest energy mc^2 of the intermediary and the momentum of the intermediary is small compared to its rest energy.

ton wavelengths are about 2×10^{-18} m). Weak interactions may be regarded as point/contact interactions at momentum transfers significantly lower than 80 GeV. As in an optical microscope, the momentum transfer is a measure of the reciprocal of wave length of the probe, as we will show. So when incoming and outgoing particles have low energies, the momentum transfer is small and we are probing the interaction at low resolution. Indeed, Fermi's original 1934 theory of beta decay involved a 4-fermion vertex $\frac{G_F}{\sqrt{2}}\bar{p}n\bar{e}\nu_e$ without mention of any intermediate W boson. The Fermi coupling $G_F \sim 1/M_w^2$.

• The strong force between quarks is due to exchange of gluons, which are massless. So one might expect it to correspond to a long range $1/r$ potential, like in EM and gravity. First, the idea of a potential or force is an idea from non-relativistic classical and quantum mechanics. The interactions between particles in relativistic QFT cannot always be exactly specified by a potential, though it is often a useful concept. Now consider a meson composed of a heavy quark and anti-quark, such as $J/\psi = c\bar{c}$ or $\Upsilon = b\bar{b}$. Such heavy quark mesons display a tower of excited energy levels, just as an $e^- p^+$ bound state does (hydrogen levels). What is more, the binding energy of these bound states is somewhat less than the rest energies of the constituents, unlike in hadrons composed of light quarks (e.g. proton or pion), so a non-relativistic treatment of the quarks is a reasonable first approximation. One finds that the observed spectrum of excited states can be theoretically obtained by assuming an inter-quark potential of the form $V(r) = -\frac{4\alpha_s}{3r} + kr$ where $\alpha_s = \frac{g_s^2}{4\pi\hbar c}$ is the strong version of the EM fine structure constant[10]. The attractive $-\alpha_s/r$ potential dominates at short distances, and is the strong analogue of the Coulomb potential energy between charges. It is due to 1-gluon exchange between quarks, just as the Coulomb law comes from 1-photon exchange in EM. But there is a surprise at larger separations, the potential grows linearly, corresponding to a constant force $F = -\nabla V$. In other words, it would cost a linearly increasing energy to pull a pair of heavy quarks apart. This is a symptom of the phenomenon of confinement, quarks have not been isolated from hadrons. When one tries to pull the quarks apart by supplying energy, at some point, it becomes energetically favorable for a quark-antiquark pair to form in between and the meson to break up into a pair of mesons as depicted in this cartoon

$$\bar{q}ggq \to \bar{q}ggggggq \to \bar{q}ggggggq\bar{q}ggggggq \to (\bar{q}ggggggq) \quad (\bar{q}ggggggq) \tag{11}$$

There is a further surprise as we probe a hadron at higher resolution (higher momentum transfers q, or loosely, higher energies E). First, the non-relativistic approximation breaks down and one cannot describe the situation in terms of the above potential, though the concept of coupling α_s continues to make sense. One finds that the strong coupling constant α_s decreases with increasing energy as $\alpha_s \sim 1/\log(E/\Lambda_{QCD})$ where $\Lambda_{QCD} \approx 200$ MeV is a constant 'scale parameter' of the strong interactions. So for momentum transfers much more than 200 MeV, (i.e. at short distances) the α_s vanishes logarithmically. Asymptotically (at high momentum transfers), quarks behave as free particles. This is the famous phenomenon of asymptotic freedom.

• Despite the quarks being confined within hadrons, it is possible to 'see' them using weak and electromagnetic probes and measure their (fractional) electric charges, spin and weak isospin

[10] g_s appears at the quark gluon vertex $g_s\bar{q}Aq$, in the covariant derivative and in $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu, A_\nu]$ etc.

quantum numbers. This is achieved for instance via $e^- p^+$ scattering or νp^+ scattering. $ep \to ep$ elastic scattering proceeds through exchange of a photon between the electron and say an up quark. Draw a Feynman diagram. The quark vertex is proportional to the charge of the up quark, which can be inferred from measurements of the scattering cross section. The exchanged photon plays the same role as light in a microscope to used to peer into an amoeba. Similarly, we have the deep inelastic scattering process $\nu_\mu p^+(uud) \to \mu^- p^+(uud)\pi^+(u\bar{d})$. Draw the leading order FD for this process, it involves both W exchange and gluon exchange. It must involve the W boson since neutrinos only interact weakly and can convert to a muon only via a W.

1.5.2 Comparison of decay rates due to the three interactions

• It has not been possible to predict when a given particle will decay, statements about decay are statistical statements about large populations. The decay of a particle refers to spontaneous decay of an isolated particle at rest (it does not refer to conversion of the particle to different particle(s) as a result of a collision). Decays of particles are found to follow the exponential depletion law $N(t) = N(0)e^{-t/\tau}$. The mean life-time τ is the time it takes for a large population of N particles to reduce to N/e particles.

• The same particle when it is in a bound state may or may not decay. Free neutrons decay with a mean lifetime of 15 minutes but neutrons bound in a stable nucleus or a neutron star can live essentially for ever. A free neutron moving at speed v relative to the lab lives on average a time of $\gamma\tau$ where $\gamma = (1 - v^2/c^2)^{1/2}$ as viewed from the lab.

• Weak interactions are responsible for nuclear beta decay (decay of a neutron $n \to p^+ e^- \bar{\nu}_e$). It is a very slow process compared to electromagnetic radiative decay (through photon emission) or decay of unstable hadrons through strong interactions. The shortest nuclear β decay mean-lives are of the order of milli-seconds though they can stretch to millions or even billions of years. The muon also decays weakly via $\mu^- \to e^- \nu_\mu \bar{\nu}_e$ in 2.2 μs[11]. The large variation in weak decay life-times is because of the mass difference between parent particle and decay products. If the mass difference is larger (relative to the parent mass), then the decay can proceed faster, it is like sliding down a steeper gradient[12] On the other hand, electromagnetic radiative decays of atoms typically take a few nano-seconds. This is one reason beta decay is referred to as a weak interaction, compared to electromagnetism. In fact, weak interactions are manifested primarily in decay or scattering processes forbidden by conservation laws to occur through the strong or EM interactions.

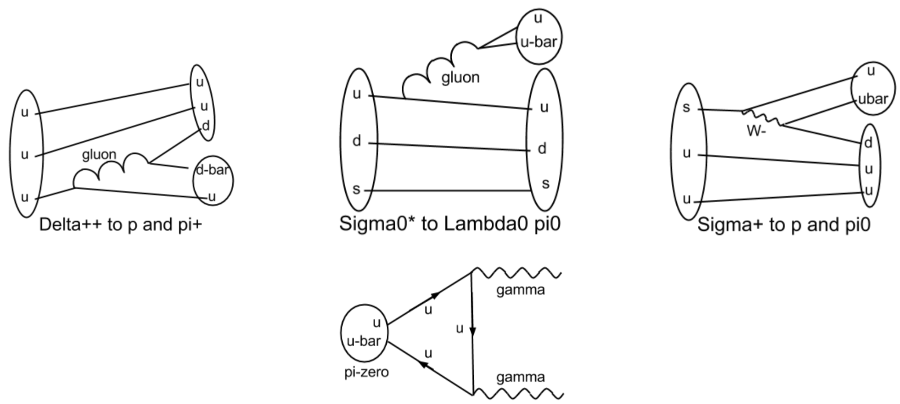
[11]The EM decay $\mu^- \to e^-\gamma$ has not been seen so far, it violates electron L_e and muon L_μ lepton number conservation.

[12]This is a general rule, not special to weak interactions. A nice illustration is the Geiger-Nuttal law found empirically in 1911 (and explained in 1928 by Condon and Gurney and independently by Gamow using tunneling) which states that α decay (which proceeds through strong interactions!) half lives of nuclei satisfy $t_{\frac{1}{2}} \propto e^{aZ/\sqrt{E}}$ where E is the energy of the α particle emitted and Z is the atomic number. So short-lived nuclei emit more energetic alpha particles as they 'slide down a steeper slope'! The exponential dependence on energy also explains how small differences in mass defects between parent and daughter nuclei can lead to huge differences in decay half lives. For example, Thorium 232 alpha decay has a half life of 14 billion years while Radium has an alpha decay half life of 1590y.

• All hadrons (baryons and mesons) other than the proton are unstable in isolation. Hadrons that decay through the strong interactions typically do so with a life time of 10^{-23} s (e.g. $\Delta^{++}(uuu) \to p^+(uud) + \pi^+(u\bar{d})$, a quark-level Feynman diagram is shown) or

$$\Sigma^{0*}(uds) \to \Lambda^0(uds)\pi^0(u\bar{u}), \qquad \tau_{\Sigma^{0*}} = 10^{-23}s, \quad \text{(strong decay)} \tag{12}$$

Sometimes, hadrons that do not decay strongly (e.g. because gluon interactions cannot change quark flavor) are called ‘stable hadrons’.



• Electromagnetic decays of hadrons occur more slowly, the neutral pion, which primarily decays to 2 photons has a mean life of 10^{-16} s. The spin half neutral Σ baryon decays electromagnetically to a photon and the Lambda baryon in 10^{-19} s.

$$\Sigma^0(uds) \to \Lambda^0(uds)\,\gamma \quad \text{with} \quad \tau_{\Sigma^0} = 10^{-19}s, \quad \text{(EM decay)}. \tag{13}$$

• Weak decays of hadrons tend to take a lot longer. For instance $\pi^+ \to \mu^+\bar{\nu}_\mu$ has a life-time of 26 ns[13] (another weak decay mode is $\pi^+ \to \pi^0 e^+ \nu_e$). Free neutrons beta decay via the weak interaction $n \to p^+ e^- \bar{\nu}$ with a mean life of about 15 minutes. And the spin half Σ^+ baryon decays weakly via

$$\Sigma^+(uus) \to p^+(uud)\pi^0(u\bar{u}) \quad \text{with} \quad \tau_{\Sigma^+} = 10^{-10}s \quad \text{(weak decay)} \tag{14}$$

• The upshot is that decay rates generally increase as we go from weak through electromagnetic to strong interactions. Analogously, the production rates, or cross sections increase from weak through EM to strong interactions. For example, Halzen and Martin quote the following typical cross sections. Weak cross sections are $\sim 10^{-11}$ mb (e.g. neutrino proton elastic scattering $\nu p \to \nu p$ or muon production in neutrino proton inelastic scattering $\nu p^+ \to \mu^- p^+ \pi^+$). In general, neutrino cross sections are roughly of order $\sigma_\nu \sim E \times 10^{-11}$ mb where E is the neutrino energy in GeV. On the other hand, electromagnetic scattering processes such as $\gamma p \to p\pi^0$ have cross sections of order 10^{-3} mb. And typical strong interaction cross sections (e.g. $\pi p \to \pi p$) are of order 20-40 mb.

[13] Cosmic ray protons in collisions with molecules in the upper atmosphere produce particle showers which include pions. These pions decay to muons, many of which live long enough to reach the Earth’s surface.

1.5.3 Comparison of strengths of forces

• We have seen that decay rates generally increase as we go from weak through electromagnetic to strong interactions. This reflects the increasing strength of the interactions. In fact, we can get a very rough estimate (ignoring the significant effects of kinematic factors etc.) of the relative strengths of the three interactions by comparing the decay rates of $\pi^+, \pi^0, \Delta^{++}$, which decay via the weak, em and strong interactions respectively with life times $\tau^w_{\pi^+} = 2.6 \times 10^{-8}$ s, $\tau^{em}_{\pi^0} = 10^{-16}$ s and $\tau^s_{\Delta^{++}} = 10^{-23}$ seconds. Draw the leading order FD for each decay. Each is a second order process[14] (involving one W, one γ and one u-quark exchange at leading order) with decay rate proportional to the square of the corresponding 'fine structure' constant $\alpha_w, \alpha_{em}, \alpha_s$. Thus we find

$$\frac{\alpha_s}{\alpha_{em}} \approx \sqrt{\frac{10^{-16}}{10^{-23}}} \approx 3000 \quad \text{and} \quad \frac{\alpha_{em}}{\alpha_w} \approx \sqrt{\frac{10^{-8}}{10^{-16}}} \approx 10^4. \qquad (15)$$

These are *very crude* estimates, even their orders of magnitude cannot be trusted. But they do indicate that strong interactions are significantly stronger than EM which are in turn stronger than the weak interactions at the energy scales relevant to these decays.

• To compare the strength of gravity with other forces between particles, consider a pair of protons (treated as point particles) separated by a distance given by a natural length associated with the proton, its reduced Compton wavelength $\lambda\!\!\!^{-}_p = \frac{\hbar}{m_p c} = 1.32$ Fermi. The gravitational potential energy is of magnitude $G_N m_p^2/\lambda\!\!\!^{-}_p$ where $G_N = 6.67 \times 10^{-11}$ Nm^2/kg^2 (or 6.71×10^{-39} GeV^{-2} in natural units). To get a dimensionless measure of the gravitational potential energy, let us consider it in units of the rest energy of the proton (use $\hbar c \approx 10^{-34} \times 3 \times 10^8$ Jm)

$$\frac{\text{grav. potn. egy}}{\text{rest energy}} = \frac{G_N m_p^2}{\hbar c} = \frac{6.7 \times 10^{-11} \times (1.6 \times 10^{-27})^2}{3 \times 10^{-26}} = 5.4 \times 10^{-39}. \qquad (16)$$

So the gravitational potential energy is 39 orders of magnitude less than the rest energy of the proton and can be ignored at energies of order the rest energy of the proton (1 GeV). On the other hand, we may ask at what energy E the gravitational force becomes comparable to the rest energy of the proton. To find out, we note that gravity couples to energy, not just mass, so the ratio of gravitational to rest energy is of order $G_N E^2$ in natural units. For this to be of order one, we need energies of order $E_P \sim G_N^{-1/2} \sim 1.2 \times 10^{19}$ GeV. This is called the Planck energy and is significantly higher than the center of mass energy available at the most energetic accelerator (LHC, 7000 GeV) or even the most energetic cosmic ray particles detected (10^{12} GeV). So gravitational forces are negligible in present day particle physics.

• Alternatively, we may ask what the mass of a point particle M_{Pl} must be so that the gravitational potential energy between two such particles separated by their reduced Compton wavelength equals the rest energy of either. So $G_N M_{Pl}^2/\hbar c = 1$ or $M_{Pl} = \sqrt{\frac{\hbar c}{G_N}} = 1.22 \times 10^{19}$ GeV/c^2, this is the Planck mass. The Planck length is defined as the reduced Compton wavelength of a Planck mass particle, i.e., $l_p = \frac{\hbar}{M_{Pl} c} = 1.6 \times 10^{-35}$ m. The Planck energy is the rest

[14] For e.g., at each of the two EM vertices of $\pi^0 \to 2\gamma$, there is a factor of u-quark charge $2e/3$. So the amplitude is proportional to $\alpha = e^2/4\pi$ and the decay rate ($\propto$ square of amplitude) is proportional to α^2.

energy of a Planck mass particle, $E_{Pl} = M_{Pl}c^2$ which is seen to be the same as the gravitational potential energy of a pair of point-like Planck mass objects separated by a Planck length $E_p = \frac{G_N M_p^2}{l_p}$.

• To compare gravity with EM, let us find the electrostatic energy between idealized point-like protons separated by their Compton wavelength, in units of the proton rest energy

$$\frac{\text{elec. potn. egy.}}{\text{rest energy}} = \frac{e^2}{4\pi \frac{\hbar}{mc} mc^2} = \frac{e^2}{4\pi\hbar c} = \alpha. \tag{17}$$

We notice that this ratio is independent of the proton mass and is in fact the fine structure constant α. It turns out that α is not quite a constant, but increases very slowly (logarithmically) from $1/137 = 7.3 \times 10^{-3}$ at eV energies relevant to atomic physics to about $1/129$ at the rest energy 90 GeV of the Z^0. In any case, the electric force at current energies is about 10^{36} times as strong as the gravitational force.

• The ordering of strengths of the three forces can be seen from other phenomena too. The existence of stable nuclei implies that the attractive strong force among nucleons outweighs the electric repulsion between protons in a nucleus. So though strong interactions operate only on nuclear and sub-nuclear scales, they must be much stronger than electric forces at lengths of a fermi. This is borne out by measurements.

• The most familiar manifestations of the weak interactions are in beta decay and muon decay $n \to pe\bar{\nu}_e$ and $\mu \to e\bar{\nu}_e\nu_\mu$. These are charge changing weak interactions, mediated by $W^\pm$ exchange. They also lead to a small parity violating inter-nucleon force (say $ud \to ud$ via W^+ exchange), which has been measured, but is vastly smaller than the inter-nucleon strong force. There is also the charge preserving ('neutral current') weak interaction mediated by the Z^0, which leads to small parity violating effects in atomic spectra. Again, these effects are much smaller than those due to the electric forces, as is clear from the success of non-relativistic QM in the Coulomb potential in reproducing the hydrogen spectrum.

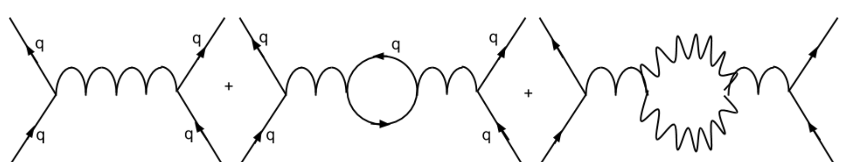
• As found earlier, weak interactions are very short ranged. In 1934, Fermi developed a 4-fermion (n, p, e, ν) contact interaction to describe beta decay (it is a low-energy approximation to the electroweak theory that took shape in the 1960s and 1970s).

$$S = \int \left[\bar{p}(i\partial\!\!\!/ - m)p + \bar{n}(i\partial\!\!\!/ - m)n + \bar{e}(i\partial\!\!\!/ - m)e + \bar{\nu}_e(i\partial\!\!\!/ - m)\nu_e + G_F \bar{p}n\bar{e}\nu_e + \text{h.c.}\right] d^4x \tag{18}$$

From the kinetic terms, we infer that the fermion fields have dimensions $M^{3/2}$ in natural units, to ensure the action have dimensions of $\hbar$. Unlike α_{em}, Fermi's coupling constant G_F is not dimensionless, but (like G_N of gravity) has dimensions of inverse mass-squared. From beta decay rates, one finds $G_F \approx 10^{-5}$ GeV^{-2}. Fermi's theory (at tree level) was very successful in explaining beta decays of all nuclei, muon decay and other (low energy) weak processes known at the time. However, the effective strength of the weak interaction $G_F E^2$ increases quadratically with energy of particles involved and becomes comparable to the electromagnetic coupling at energies of tens of GeV. By then, Fermi's theory ceases to be valid. In fact, it is not renormalizable and loses predictive power when one tries to calculate quantum corrections perturbatively beyond tree level. A similar problem afflicts perturbative quantum gravity. In the case of weak interactions, Fermi's theory was replaced with the Glashow-Weinberg-Salam standard model, in which weak and electromagnetic interactions get mixed and the contact 4-fermi interaction is replaced by exchange of W bosons.

• Let us briefly consider the coupling constants in the standard model with gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$. There are three coupling constants one for each simple group[15] or $\mathrm{U}(1)$ factor in the gauge group: g_1 for weak hypercharge $\mathrm{U}(1)_Y$, g_2 for weak isospin $\mathrm{SU}(2)$ and g_3 for color $\mathrm{SU}(3)$. Corresponding to these three coupling constants are three 'fine structure' constants $\alpha_i = \frac{g_i^2}{4\pi\hbar c}$. $\alpha_3 = \alpha_s$ is the strong one. But there is mixing among the weak isospin and weak hypercharge generators. The weak mixing angle is given by $\tan\theta_w = g_1/g_2$ and the fine structure constant of QED is $\alpha_{em} = \alpha_2 \sin^2\theta_w$. The charged weak $g_w = \frac{g_{em}}{\sin\theta_w}$ and neutral weak $g_z = \frac{g_{em}}{\sin\theta_w \cos\theta_w}$ couplings can be expressed in terms of the EM coupling g_{em} and weak mixing angle.

• It turns out that α_{em} as well as the other α_i are not constant, but change with energy scale (or distance scale) associated to the probe. They are called 'running' coupling constants. The charge of an electron (which measures the strength of its coupling to photons or the strength of its scattering from other electrons) increases as one probes it more closely (smaller wavelength or higher energy/momentum carried by the photon probe). At large distances, the charge is screened due to 'polarization of the vacuum'. Analogous effects are familiar from atomic physics (screening of nuclear charge by inner shell electrons) and electrostatics in a polarizable medium (dielectric like a plasma). The vacuum itself behaves as a polarizable medium due to the effects of virtual electron positron pairs. It is found that the fine structure constant increases logarithmically with the energy scale (or reciprocal of distance scale). On the other hand the coupling constants for the color and weak isospin groups decrease with energy scale. This anti-screening of color and weak isospin is due to the non-abelian gauge groups involved. For example, in QCD, there is competition between the screening due to virtual $q\bar{q}$ pairs and anti-screening due to virtual gluons (see figure) and the latter dominate in practice, so $\alpha_s = \alpha_3$ decreases (logarithmically) with energy.



Scattering of quarks: Vacuum polarization due to virtual quark and gluon loops

• The running of coupling constants (when they are small) is given to leading order in perturbation theory by a renormalization group equation (RGE) for $\alpha(\mu)$. μ is called the sliding

[15] If a group G has a non-trivial normal subgroup H, then we can form the quotient group G/H. If the subgroup H is not normal in G, then G/H does not have the structure of a group, it is just a coset space (cosets cannot be multiplied to get cosets in general). So if H is normal, G may be 'broken down' to H and G/H. The same process can be applied to H and G/H, thus breaking a group into smaller groups. The process terminates at simple groups, which are those that do not have any non-trivial normal subgroups. $U(1)$ is not simple, since N^{th} roots of unity form normal subgroups for every N. But $SU(n)$ is simple for $n = 2, 3, \ldots$

energy/momentum scale.

$$\mu \frac{d\alpha(\mu)}{d\mu} = \beta(\alpha) \tag{19}$$

The β function is evaluated by using Feynman diagrams of the sort shown. One finds for small α that $\beta(\alpha) = \beta_0 \alpha^2 + \mathcal{O}(\alpha^3)$. We say that the beta function has a double zero at the origin ($\alpha = 0$), this means $\alpha = 0$ is a fixed point of the RG flow. The sign of β_0 determines whether the coupling grows ($\beta_0 > 0$) or decreases ($\beta_0 < 0$) with energy μ. Indeed, the solution of this ODE is

$$\alpha(\mu)^{-1} = \alpha(\mu_0)^{-1} - \beta_0 \log \frac{\mu}{\mu_0} \quad \text{or} \quad \alpha(\mu) = \frac{\alpha(\mu_0)}{1 - \beta_0 \alpha(\mu_0) \log(\mu/\mu_0)} \tag{20}$$

The value of α at a reference scale μ_0 needs to be fixed experimentally. Once this is done, the theory predicts the values of the running couplings at other energy scales (where α is still small). In particular, the formula for $\alpha(\mu)$ can only be trusted when α is small. The parameter $\beta_0 = \frac{1}{12\pi}(4N_g - 11N_c)$ where N_c is the number of colors in the case of QCD (or the N in an $SU(N)$ gauge group) and N_g is the number of fermion generations (3 generations of quarks in the case of QCD). The first term $4N_g$ in β_0 arises from fermion loops[16] while the second term $11N_c$ comes from gauge boson loops[17]. In QED, there are no gauge boson (photon) self-couplings and no gauge boson loops, so in effect, $N_c = 0$. So in QED $\beta_0 > 0$ and Sommerfeld's fine structure constant grows logarithmically with energy: $\alpha = 1/137 = 7.3 \times 10^{-3}$ at energies relevant to atomic physics to about $1/129$ at the rest energy 90 GeV of the Z^0. QCD displays the opposite behavior since $\beta_0 < 0$ for sufficiently many colors $\frac{N_c}{N_g} > \frac{4}{11}$, which is the case in nature $N_c = 3$ and $N_g = 3$. So $\alpha_s \to 0$ at high energies and quarks become asymptotically free. At an energy of 3 GeV α_s is about $.2 - .3$ while at 90 GeV it decreases to about 0.12. Over this entire range of energies, the strong force between quarks is significantly stronger than electric forces between quarks. QCD becomes strongly coupled at relatively low energies, i.e., $\mu \lesssim 1$ GeV and QED becomes strongly coupled at high energies. In these regimes, perturbation theory cannot be trusted. A naive application of the above formula for α_{em} would suggest that it has a simple pole (blows up) at an energy of $\mu = \mu_0 e^{1/\beta_0 \alpha(\mu_0)} \sim e^{137\pi}$ eV. This is called the Landau pole and led to a temporary loss of confidence in the validity of QED and QFT in the 1950s. However, this energy is way above the Planck energy and there is no reason to trust perturbative QED when α grows big, other methods to study the theory are called for. On the other hand, the largeness of α_s at low energies (sometimes dubbed infrared slavery in contrast to asymptotic freedom), necessitates the development of non-perturbative methods to study QCD in the infra-red, quarks are strongly bound together by gluons leading to phenomena like confinement when one tries to extract them from hadrons. The UV behavior of QCD is perturbative due to asymptotic freedom.

• Using the above RGE, the scale dependence of the three running couplings $\alpha_1, \alpha_2, \alpha_3$ can be found. At leading order in perturbation theory, they only differ in the values of β_0, which is

[16] It is proportional to the number of fermion-anti-fermion flavors that could circulate in the loop, i.e, u, d, s, c, b, t

[17] If we use the double-line notation for gluon lines and fix the colors of the asymptotic quark states, there remains a color loop whose possible colors we must sum over in computing the Feynman amplitude.

negative for α_2 and α_3 and positive for α_1. In a plot of $\alpha_i(\mu)^{-1}$ against $\log(\mu/\mu_0)$, we will get straight lines, with positive slopes for the color and weak isospin groups and negative slope for the weak hypercharge group. The predicted running of these three couplings has been measured carefully and confirmed upto μ over a 100 GeV. Interestingly the three straight lines seem to get rather close at an energy of 10^{14} GeV. The 3 gauge couplings seem to meet at this scale dubbed the scale of grand unification (GUT scale). Attempts have been made to find a grand unified simple gauge group which would govern interactions at the GUT scale. The simplest grand unified models predict the correct weak mixing angle (which is a free parameter in the SM), they also predict proton decay, which has not yet been seen.

2 Concepts from mechanics, quantum theory and relativity

2.1 Rutherford scattering cross section

• Scattering of alpha particles against gold atoms (by Geiger, Marsden and Rutherford in 1909-1911) was instrumental in identifying the nuclear model of the atom where a heavy positive charge is concentrated at a point-like nucleus with light electrons surrounding it.

• Suppose the beam of incoming particles has an intensity/flux of $\mathcal{F}$ particles per unit time per unit area normal to the beam. After scattering, we want to find dN, the number of particles entering solid angle $d\Omega$ per unit time. dN is the rate at which particles should be detected by a detector covering solid angle $d\Omega$. This should be proportional to the incoming flux, so $dN = \mathcal{F} d\sigma$ where $d\sigma$ is an area. $d\sigma$ may be interpreted as the area normal to the beam through which the particles must pass so that they are scattered into $d\Omega$. The ratio $\frac{d\sigma}{d\Omega}$ is called the differential scattering cross section.

• Suppose charge q_1 of mass m and energy E scatters off point charge q_2 fixed at the origin with repulsive Coulomb potential $V = \frac{q_1 q_2}{4\pi r}$. Rutherford's differential scattering cross section is

$$\frac{d\sigma}{d\Omega} = \left(\frac{q_1 q_2}{16\pi E \sin^2 \frac{\theta}{2}} \right)^2 \tag{21}$$

where θ is the (polar) scattering angle $\theta = 0, \pi$ for forward/back scattering.

• Let us recall how this formula is obtained. Consider particles coming in through an annulus with impact parameters between b and $b+db$ and any azimuthal angle ϕ. They will be scattered by angles between θ and $\theta + d\theta$ and therefore correspond to $d\Omega = 2\pi \sin\theta d\theta$. Now $dN = 2\pi b db \mathcal{F}$ is the product of the area of the annulus and the incident flux. Divide by $\mathcal{F}$ to get $d\sigma$ and divide by $d\Omega$ to get $\frac{d\sigma}{d\Omega} = -\frac{b}{\sin\theta}\frac{db}{d\theta}$. The minus sign is because larger impact parameters imply smaller scattering angles.

• By integrating the energy equation $E = \frac{1}{2}mv^2 = \frac{1}{2}m\dot{r}^2 + \frac{L^2}{2mr^2} + V(r)$ in a repulsive Coulomb potential, one obtains the following relation between impact parameter and scattering angle $1 + \frac{4b^2}{a^2} = \text{cosec}^2(\theta/2)$. Differentiate to get the Rutherford differential cross section

$$\frac{8b}{a^2} db = -2\sin^{-3}(\theta/2)\cos(\theta/2)\frac{1}{2}d\theta \quad \Rightarrow \quad -\frac{db}{d\theta} = \frac{\cos(\theta/2)}{\sin^3(\theta/2)}\frac{a^2}{8b}. \tag{22}$$

• To derive the relation between b and θ we use Newton's 2nd law. In fact we only need to integrate once if we use conservation of energy $E = \frac{1}{2}m\dot{r}^2 + \frac{l^2}{2mr^2} + \frac{q_1 q_2}{4\pi r}$. If v is the initial speed of the projectile (say coming in from the left), then its (conserved) angular momentum and energy are $l = mvb$ and $E = \frac{1}{2}mv^2$. So we may eliminate l in favor of b: $l^2/2m = Eb^2$. Since we are interested in the angular deviation of the orbit rather than its time dependence, we parametrize the orbit by $r(\phi)$ instead of $\mathbf{r}(t)$, where ϕ is the polar angle in the plane of the trajectory measured with respect to the direction of the projectile source. The orbit equation simplifies if we use $u = 1/r$ in place of r. Thus $\dot{r} = r'(\phi)\dot{\phi}$ and $l = mr^2\dot{\phi}$ give

$$\dot{r} = -u'(\phi)\frac{l}{m} \quad \Rightarrow \quad \frac{1}{2}m\dot{r}^2 = \frac{l^2}{2m}u'(\phi)^2 = Eb^2 u'(\phi)^2. \tag{23}$$

and the conservation of energy becomes $Eb^2 u'(\phi)^2 = E - Eb^2u^2 - \frac{q_1 q_2}{4\pi}u$. Defining the length $a = \frac{q_1 q_2}{4\pi E}$ which is the (hypothetical) radial distance at which the Coulomb energy equals the total projectile energy[18]

$$d\phi = \frac{b du}{\sqrt{1 - b^2u^2 - au}} \tag{24}$$

Now the projectile comes in from $r = \infty$ with $\phi = 0$, reaches a point of closest approach where $r = r_{min} = 1/u_{max}$ [$\dot{r} = 0$, the mid-point of the trajectory, with $\phi = \phi_0$. u_{max} is the positive root of the quadratic under the square-root sign, $u_{max} = \left(\sqrt{a^2 + 4b^2} - a\right)/2b^2$] and eventually scatters off to $r = \infty$ reaching an asymptotic $\phi = 2\phi_0$. Draw a diagram! The scattering angle is $\theta = \pi - 2\phi_0$. The integral can be done by completing the square and a trigonometric substitution using $\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x$,

$$\phi_0 = \int_0^{u_{max}} \frac{b\, du}{\sqrt{1 - au - b^2u^2}} = \frac{\pi}{2} - \arcsin\frac{1}{\sqrt{1 + 4b^2/a^2}} \quad \Rightarrow \quad \sin^2(\theta/2) = \frac{1}{1 + 4b^2/a^2}. \tag{25}$$

This is the advertised relation between impact parameter and scattering angle for Coulomb scattering, which leads to the Rutherford cross section.

• Physical remarks on Rutherford cross section formula.

• (1) The Coulomb field has a long range (though the scatterer is point-like). Particles with large impact parameters (small scattering angles) contribute significantly to the total cross section $\sigma = \int (d\sigma/d\Omega) 2\pi \sin\theta d\theta$. In fact, strictly speaking, the total cross section for Coulomb scattering is infinite due to a divergence at $\theta = 0$. This means there is scattering even for large impact parameters b, though the scattering angle decreases with b.

• (2) The charges need not have the same sign, the potential could be attractive or repulsive, since the cross section only depends on the square of each of the charges.

• (3) The cross section falls off as $1/E^2$ with increase in energy. This is a characteristic feature of scattering of point charges at high energies, since the energy is the only dimensional parameter in the problem[19] with which to construct a quantity with dimensions of area (particle masses are small compared to sufficiently large E and there is no length scale from the

[18]Except in the case of zero impact parameter, the projectile does not actually get as close as a since it never comes to rest if the angular momentum is non-zero

[19]A 'scale invariant' potential like the Coulomb potential $q_1 q_2/4\pi r$ does not introduce any length scale of

dimensions of the point particles). For instance, the cross section for e^+e^- annihilation to any final state has been measured from MeV to 100s of GeV center of mass energies (see fig. 5.3, p.145 in Perkins 4th Ed.). It shows an overall $1/E^2$ fall off with localized peaks corresponding to resonances (e.g. the $q\bar{q}$ vector mesons) like $\rho(d\bar{d} - u\bar{u}, 776MeV), \omega(d\bar{d} + u\bar{u}, 783MeV), \phi(s\bar{s}, 1019MeV), J/\psi(c\bar{c}, 3.1GeV), \Upsilon(b\bar{b}, 9.5GeV)$ and the weak gauge bosons $Z^0(91GeV), W^+W^-(160GeV)$ etc. The scalar mesons π^0 and η^0 can also appear as resonances in e^+e^- annihilation, but they are suppressed due to the need for two photon processes unlike the single photon intermediary which is adequate in the above cases.

• (4) Rutherford scattering cross section depends on θ, it is not isotropic. There is non-trivial scattering for large impart parameters. In other words, different incoming angular momenta $L = mvb$ of the projectile scatter differently but non-trivially. This is to be contrasted with a very short range potential (like a delta function or a Yukawa potential $e^{-r/\xi}/r$ with ξ much less than the impact parameter or de Broglie wavelength of incoming matter waves), where S-wave $(L = 0)$ scattering dominates, and there is no scattering for larger impact parameters $b \gg \xi$, and $d\sigma/d\Omega$ is independent of θ.

• (5) Though the differential cross section decays monotonically as θ goes from 0 to π (back scatter), there is still significant scattering through large angles. The fall-off would be much faster if the repulsive charge isn't concentrated at a point, but spread all over the atom. This was the experimental signal favoring Rutherford's nuclear model of the atom over J J Thomson's 'plum pudding' model.

• (6) The same Rutherford formula arises in non-relativistic QM in the first Born approximation (which is valid at high (non-relativistic) energies). If $E = \hbar^2 k^2/2m$ is the initial and final kinetic energy (elastic scattering, the electron just changes direction), then the magnitude of transferred momentum $\mathbf{q} = \vec{k}_f - \vec{k}_i = k\hat{r} - k\hat{z}$ is $q = 2k\sin(\theta/2)$ since

$$q^2 = (k\hat{r} - k\hat{z})^2 = 2k^2(1 - \cos\theta) = 4k^2\sin^2(\theta/2) \tag{26}$$

The Born scattering amplitude $f(\theta,\phi)$ is proportional to the Fourier transform[20] of the potential, $f(\theta,\phi) \approx -\frac{2m}{4\pi\hbar^2}\tilde{V}(q)$ where

$$\tilde{V}(\mathbf{q}) = \int V(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \quad \text{and} \quad V(r) = \frac{q_1 q_2}{4\pi r} \quad \Rightarrow \quad \tilde{V}(q) = \frac{q_1 q_2}{q^2} \tag{27}$$

And the cross section is the square of the scattering amplitude

$$f(\theta,\phi) \approx -\frac{2m}{4\pi\hbar^2}\tilde{V}(q) \approx -\frac{2m}{4\pi\hbar^2}\frac{q_1 q_2}{q^2} = -\frac{q_1 q_2}{16\pi E}\frac{1}{\sin^2\theta/2} \quad \Rightarrow \quad \frac{d\sigma}{d\Omega} = |f|^2 \approx \left(\frac{q_1 q_2}{16\pi E \sin^2\theta/2}\right)^2. \tag{28}$$

its own. Other potentials may have a length scale ξ associated with them, in which case the formula for the cross section could depend on both E and ξ. An example of non-point-like particle scattering is hadron-hadron scattering, hadrons have a size a of order a Fermi and at high energies, the cross section approaches a constant, roughly the classically expected value of $\pi(2a)^2$.

[20]The FT can be got from knowledge that the electrostatic potential for a point charge e, $\phi = e/4\pi r$ satisfies Poisson's equation $-\nabla^2\phi = \rho = e\delta^3(\mathbf{r})$ since $E = -\nabla\phi$ and $\nabla\cdot E = \rho$. Fourier expanding $\phi(r) = \int \tilde{\phi}(q) e^{i\mathbf{q}\cdot\mathbf{r}}[dq]$ and $\nabla^2\phi(r) = -\int q^2\tilde{\phi}(q)e^{i\mathbf{q}\cdot\mathbf{r}}[dq]$. Using $\delta^3(\mathbf{r}) = \int e^{i\mathbf{q}\cdot\mathbf{r}}[dq]$ Poison's theorem gives $\tilde{\phi}(q) = -e/q^2$ if $\phi(r) = -e/4\pi r$.

• (7) In a relativistic treatment using Feynman diagrams in QED, scattering between a pair of charges is (at leading order) due to exchange of a virtual photon. For example, one may consider electron muon scattering or electron proton scattering (so that one of the particles is much heavier than the other and we may ignore its recoil, just as in the α-gold nucleus case.) The calculation is more involved (dealing with Dirac spinors), but the answer, Mott's formula (see for e.g. Griffiths 2nd ed., p.255), bears a resemblance to Rutherford's formula (and reduces to it in the non-relativistic limit $p \ll mc$). Assuming both particles have charge of magnitude e,

$$\left[\frac{d\sigma}{d\Omega}\right]_{\text{Mott}} = \left(\frac{e^2}{16\pi(p^2/2m)\sin^2(\theta/2)}\right)^2 \left[1 + \frac{p^2}{m^2c^2}\cos^2(\theta/2)\right]. \tag{29}$$

p is the magnitude of the lab frame momentum of the electron, m its mass and θ the scattering angle. In the non-relativistic approximation $p^2/2m$ is simply the kinetic energy E of the incoming electron.

2.2 Need for quantum treatment and high momentum transfers

• Quantum effects become important when we are interested in phenomena at a length scale comparable to or smaller than the typical de Broglie wavelength $\lambda = h/p$ of the particles involved. Atoms are about an angstrom in size. The typical electron kinetic energies KE (by kinetic energy we mean $\sqrt{p^2c^2 + m^2c^4} - mc^2$, which reduces to $p^2/2m$ in the non-relativistic limit) are of order eV or 10 eV (binding energy is 13.6 eV, and kinetic and potential energies are comparable) and the electron mass is half an MeV, so a non-relativistic treatment suffices, though relativistic effects are manifested in the fine structure of atomic spectra. Typical atomic electron momenta are $p = \sqrt{2m\text{KE}} \sim 1$ keV$/c$ implying de Broglie wavelengths $\lambda\!\!\!^- = \frac{\hbar c}{pc} = \frac{200MeV fm}{1keV} \approx 2$ Angstroms, comparable to the size of atoms. Atomic electrons require a quantum mechanical treatment.

• Nuclei are of size 1-10 fm as we go from deuterium to uranium. The binding energy of a helium nucleus is 28 MeV while the rest energy of its constituents is 3755.67 MeV. More generally, the binding energy per nucleon in a nucleus ranges from 1-2 MeV to 10 MeV, which is much smaller than the rest mass 938 MeV/c^2 of nucleons. So nuclear structure may be treated non-relativistically. The momentum of a proton with kinetic energy 2 MeV is about 60 MeV/c. It follows that the reduced de Broglie wave length is about $\lambda\!\!\!^- = \frac{\hbar c}{pc} \sim \frac{200\,\text{MeV fm}}{60MeV} \sim 3.3$ fm which is comparable to nuclear dimensions. So like in atomic structure, non-relativistic qm provides a good first approximation to nuclear structure.

• In most of sub-nuclear particle physics, we require both quantum mechanics and special relativity since particle kinetic energies often significantly exceed their rest energies (10 GeV electrons or 1 TeV protons!) and we probe subnuclear dimensions comparable or smaller than relevant de Broglie wavelengths. Moreover, number of particles is typically not conserved (e.g. due to pair production), which is forbidden in non-relativistic qm.

• To resolve structures at small length scales we need probes with high energy (or momentum transfer). To see why, note that the diffraction limit of resolution d of an optical microscope is given by Abbe's formula

$$d \approx \frac{\lambda}{2n \sin\alpha} \tag{30}$$

where $\alpha = \theta$ is the aperture angle of the objective lens (half the opening angle of the cone of light it receives from the source via the specimen) and λ the wavelength of light used and n the refractive index of the medium between specimen and objective. So shorter wavelengths of light (or higher energies $E = hc/\lambda$) resolve better. For a general probe (including matter particles), we have by de Broglie's relation $\lambda = h/p$

$$d \approx \frac{h}{2p \sin\theta} \tag{31}$$

Now θ is the maximum angle through which the incoming particles have been scattered from initial momentum $\mathbf{p}_i = p\hat{z}$ to final momentum $\mathbf{p}_f = p\hat{r}$, assuming elastic scattering so that $|\mathbf{p}_i| = |\mathbf{p}_f| = p$. It follows that the momentum transfer $\mathbf{q} = \mathbf{p}_f - \mathbf{p}_i$ has magnitude $q^2 = 2p^2 - 2p^2\cos\theta$ or $q = 2p\sin(\theta/2)$. So $d \approx \frac{h}{2q\cos(\theta/2)}$ and we need large momentum transfers to probe short distances by scattering through any particular angle. By using virtual photons and weak gauge bosons with large momentum transfers in ep and νp scattering, the deep inelastic scattering experiments of the late 1960s and early 1970s discovered the quark structure of the proton.

• By 2014, momentum transfers of order $q = 200$ GeV/c and more are available. This corresponds to a resolution of $d = \hbar c/(200\ GeV) \approx 10^{-18}$m. Quarks, (and leptons and gauge bosons) behave like point-like particles down to this resolution.

• High energies are also needed to produce the particles encountered in particle physics, since they have high masses like $m_{\text{Higgs}} = 125$ or $m_{\text{top}} = 175$ GeV$/c^2$. In an e^+e^- colliding beam experiment, the cm energy is available for production of new particles.

2.3 Resonances and Breit-Wigner line shape

• Most of the particles discovered are unstable, they decay to other particles through one or more 'decay channels'. E.g. $Z^0 \to \nu\bar{\nu}$ or $Z^0 \to e^+e^-$ or $Z^0 \to q\bar{q}$. They often do not live long enough to leave tracks in detectors, but are manifested as peaks (resonances) in scattering cross sections. Resonances are real ('on-shell') particles, they are not virtual particles (of course, a Z can also appear as a virtual intermediary in a calculation). Hadrons are bound states of quarks and gluons held together by the QCD interactions. However, some hadrons (like the Δ) can decay via the strong interactions to lighter hadrons. Hadrons that decay through the strong interactions are called hadronic resonances, typically they decay in 10^{-23}s. Those hadrons that do not decay strongly may decay weakly or electromagnetically, and typically live much longer, so they are called stable hadrons (e.g., nucleons, Σ, Λ^0, Ξ among baryons (and their anti-particles) and pions, kaons and η^0 among mesons – they are usually not called resonances). The neutron is a stable hadron, it decays weakly with a mean life of about 15 minutes. The proton is the only hadron that has not been seen to decay. The first baryonic resonance identified was the Δ resonance, which decays strongly to a pion and a nucleon (it comes in 4 charge states, which form an isospin $3/2$ quartet Δ^- Δ^0, Δ^+ Δ^{++}). Conversely, it is found that the cross section for π-N scattering goes through a local maximum ('resonance') when the invariant mass[21] of the $\pi - N$ colliding system crosses the 'mass' of the Δ (1232 MeV). Hadronic resonances typically decay to the stable hadrons mentioned above via the strong interactions. Some of the initially discovered hadronic resonances were denoted with an asterisk, e.g. K^*, σ^*, Ξ^* N^*, Δ^* etc. In more recent notation, different resonances with the same isospin are denoted with the same letter and mass in parentheses (e.g. N(1440), N(1520) etc. and $\Delta(1232), \Delta(1600)$ etc., see the particle data booklet).

• The scattering cross section in the neighborhood of a resonance has a characteristic shape called the Breit-Wigner resonance curve. It is a Lorentzian line shape characterized by an energy width Γ and a peak value $\sigma_{\rm peak}$. A heuristic derivation follows.

• As in radioactive decay, the rate of depletion of a sample of N unstable particles A is proportional to the number of particles present

$$\frac{dN}{dt} = -WN \tag{32}$$

where W is the total decay rate. It is the sum of partial decay rates for the various channels (labelled by i, $A \to a + b, A \to c + d$ etc.) $W = \sum_i W_i$. Thus $N(t) = N(0)e^{-Wt}$. The time $\tau = 1/W$ is called the mean life time of the particle, the time after which on average $N(0)/e$ particles survive. $\Gamma = \hbar W$ is called the energy width of the resonance ($\hbar W_i = \Gamma_i$ are the partial widths). By the energy-time relation, we would expect $\Gamma = \hbar/\tau$ to be the spread in (rest) energies of the unstable state. The point is that an unstable state does not have a well-defined (real) energy since it does not correspond to a stationary state (a stationary state has wave function with purely harmonic time depenedance $\psi(0)e^{-iEt/\hbar}$). Suppose a population of resonances are produced at $t = 0$ or observed to exist at $t = 0$. Then for $t \geq 0$, we could model the amplitude (whose square should be the number of surviving particles) as,

$$\psi(t) = \psi(0)e^{-iMc^2t/\hbar}e^{-t/2\tau} \quad \text{where} \quad M \quad \text{is the central value of the rest mass of the resonance,} \tag{33}$$

so that $N(t) = |\psi(t)|^2 = |\psi(0)|^2 e^{-t/\tau}$ where $N(0) = |\psi(0)|^2$. In natural units, the amplitude in the energy basis is given by the Fourier transform of ψ:

$$\chi(E) = \int_0^\infty \psi(t)e^{iEt} = \frac{i\psi(0)}{E - M + \frac{i\Gamma}{2}} \quad \Rightarrow \quad |\chi(E)|^2 = \frac{|\psi(0)|^2}{(E-M)^2 + \frac{\Gamma^2}{4}}. \tag{34}$$

In a collision of the decay products a, b at a center of momentum energy E, the cross section for producing the resonant state A would be expected to be proportional to $|\chi(E)|^2$. Thus the Breit-Wigner resonance curve is

$$\sigma(E) = \sigma_{\rm peak}\frac{\Gamma^2/4}{(E-M)^2 + \Gamma^2/4} \quad \text{where} \quad \sigma_{\rm peak} = 4|\psi(0)|^2/\Gamma^2. \tag{35}$$

[21]The invariant mass is the Lorentz invariant length of the total initial 4-momentum $\sqrt{(p_\pi + p_N)^2}$. It equals the total energy of the colliding particles in the center of momentum frame, which is the frame in which the total 3-momentum is zero.

This is a bell-shaped curve ('Lorentzian') centered at $E = M$. Γ is the full width of the resonance curve at half maximum ($\sigma = \frac{1}{2}\sigma_{\text{peak}}$ when $E = M \pm \Gamma/2$) where σ_{peak} is the peak value of the cross-section. This is of course a simple model, but it does remarkably well in fitting resonance cross sections. In practice cross sections may not quite be symmetric about $E = M$ and this can be understood using scattering theory. Moreover, the Breit-Wigner cross section decreases only quadratically and the tail of one resonance can affect the cross section due to a neighboring resonance.

• The first hadronic resonance to be discovered (by Fermi and Anderson in 1952 using the Chicago cyclotron; the idea that they were seeing a resonance was suggested by K Brueckner) was $p^+\pi^+ \to \Delta^{++} \to p^+\pi^+$. $\Delta^{++} = uuu$ is an isospin $3/2$ baryon, an excited state of a nucleon (nucleons have angular momentum half, while Δ has $J = 3/2$, the quarks are going round with more angular momentum, a little like the difference between the 1S and 2P states of hydrogen). It is a broad resonance with a width of $\Gamma = 120$ MeV around a central mass of 1232 MeV. Being a broad resonance, Δ^{++} typically decays very quickly with a mean life time $\tau = \frac{\hbar}{\Gamma} \approx 10^{-23}\,$s.

• The Z^0 boson arises as a resonance in e^+e^- annihilation (to any final state), at $M_Z = 91$ GeV with a full-width $\Gamma = 2.5$ GeV. What is its life time? Since Γ is the sum of partial widths for the various decay modes, the more the number of contributing decay channels, the wider the resonance. Z can decay to any $f\bar{f}$ pair where f is a quark or lepton whose mass is less than half that of the Z^0 (so Z^0 cannot decay to $t\bar{t}$ or W^+W^-, though a virtual (far off-shell) Z can produce a real W^+W^- pair in high energy e^+e^- collisions). In fact, this can be used to bound the number of neutrinos to which the Z^0 could decay. The experimentally determined Z-width ($\Gamma = 2.5$ GeV) can be used to show that there can be at most 3 types of neutrinos of mass less than $M_Z/2$ that the Z couples to, indeed one finds $N_\nu = 2.99 \pm 0.01$. In particular, there cannot be a 4th generation of leptons with a light neutrino that couples in the same way as the first 3 generations.

• The Hoyle state is a narrow long-lived ($\Gamma = 10\text{eV}$) resonance of the Carbon nucleus with a mass 7.654 MeV above the C_6^{12} ground state. It was predicted to exist by F Hoyle in 1953 and experimentally found subsequently by W Fowler. This resonance is the key to carbon synthesis in helium burning stars via the so-called triple alpha process. The existence of this resonance allows (by increasing the cross section to a level that is adequate to explain the observed abundance of Carbon in the universe) three alpha particles to fuse (first two αs fuse to form beryllium-8, which can capture a third α particle.). The C_6^{12} resonance then decays by photon emission to arrive at the carbon ground state.

• The lowest lying baryons (spin half baryon octet ($n, p, \Sigma^\pm, \Sigma^0,\ \Xi^-, \Xi^0$)) and lowest lying mesons (spin zero meson nonet ($\pi^\pm, \pi^0, K^\pm, K^0, \bar{K}^0, \eta, \eta'$)) are the stable hadrons built from light quarks (u, d, s). They are all stable against strong decays. With the exception of the proton, all of them decay via the weak or EM interactions, which makes them quite long lived on the 10^{-23} s time scale of the strong interactions. Above these (in mass) we have baryonic and mesonic resonances with spin more than half and more than zero respectively. These resonances decay through the strong interactions. The lightest resonances are the spin 3/2 baryon decuplet (10 of them the $\Delta, \Sigma^*, \Xi^*, \Omega$ quartet, triplet, doublet and singlet) and the spin one meson nonet

$(\rho, K^*, \phi, \omega)$. Look up the quark content of the baryon decuplet and arrange them in an I_3 vs S plot. Baryonic resonances with higher spin (5/2, 7/2,9/2,11/2) have been found, as have mesonic resonances with spin $2, 3, 4$.

• When the squares of the masses M^2 of hadronic resonances are plotted against their spin (total angular momentum J) one often finds collections of resonances with a linear relation between M^2 and L. These collections are called Regge 'trajectories'. For example, we have the Regge trajectory of rotational excitations of the spin 1 ρ meson, spin 2 A_2 meson, spin 3 g meson and spin 4 δ meson. Similarly there is a Regge trajectory of rotational excitations of the $J = 3/2$ Δ. The trajectories seem to have roughly the same slope but different intercepts. We will discuss Regge trajectories later.

2.4 Size and binding energy of non-relativistic bound states

A non-relativistic bound state system is one where the speeds v of constituents (in the rest frame of the system) are small compared to c. Examples include atoms, the solar system, molecules, nuclei (with more than one nucleon) and mesons with two heavy quarks ($J/\psi = c\bar{c}$ and $\Upsilon = b\bar{b}$). We may use the uncertainty principle to argue that the size R of such a system is large compared to the Compton wavelengths $\lambda = \frac{h}{mc}$ of the constituents. Let us estimate $\Delta x \sim R, \Delta p \sim mv$. Then $Rmv \sim \hbar$ if we take the uncertainty bound as a reasonable estimate, which is usually reasonable for low lying states. It follows that $\frac{v}{c} \sim \frac{\hbar}{mc}\frac{1}{R} = \frac{\lambda}{2\pi R}$. So $v \ll c$ implies $\lambda \ll R$, i.e. the Compton wavelengths of constituents are small compared to the spatial extent of the bound state. An atom is a good example (electron Compton wavelength 2.4×10^{-12} m is less than the Bohr radius). Such a non-relativistic system also has small binding energies compared to rest masses of constituents. Indeed, typically mean kinetic and potential energies are of the same order (the virial theorem relates them), so we may take the binding energy $BE \sim mv^2$. Then $\frac{BE}{mc^2} \sim \frac{v^2}{c^2} \ll 1$.

• *Estimate the binding energy of the earth in its bound orbit around the sun. How does it compare with the sum of rest mass energies of the sun and earth? Compare with the situation in the Hydrogen atom.*

• Relativistic bound states include protons, pions and indeed any hadron made from the light quarks. They have binding energies comparable to or exceeding the rest energies of constituents, which move at speeds comparable to that of light. Typically, they do not consist of a definite number of particles. Relativistic bound states are poorly understood.

2.5 Angular momentum

2.5.1 Representations of angular momentum Lie algebra

• In non-relativistic QM, angular momentum observables, irrespective of whether they are orbital, spin or combinations of both satisfy the same angular momentum (SU(2)) Lie algebra commutation relations[22]

$$[J_i, J_j] = i\hbar\epsilon_{ijk}J_k \tag{36}$$

Check that $J^2 = J_1^2 + J_2^2 + J_3^2$ commutes with all three, it is the only independent Casimir of the algebra and spans its center. In qm, angular momentum observables are represented as hermitian operators. Irreducible representations are labelled by the eigenvalues of J^2, which is a multiple ($\hbar^2 j(j+1)$) of the identity in any such representation. It is conventional to choose a basis for a representation where J^2 and J_z are diagonal, there are at most two such independent operators that are simultaneously diagonalizable. An angular momentum j irreducible unitary representation is $2j+1$ dimensional, the representation space is $\mathbb{C}^{2j+1}$. The $2j+1$ simultaneous eigenstates of J^2 and J_z are $|jm\rangle$ with eigenvalues $\hbar j(j+1)$ and $\hbar m$ respectively. Here $m = -j, -j+1, \ldots, j-1, j$ and j is either a non-negative integer or half an odd positive integer. This is obtained by defining the raising and lowering operators $J_\pm = J_x \pm J_-$ which satisfy the commutators $[J^2, J_\pm] = 0$ and $[J_z, J_\pm] = \pm\hbar J_\pm$. These imply that $J_\pm$ raise and lower the value of m without affecting j. The boundedness (above) of $J_z^2 = J^2 - J_x^2 - J_y^2$ can be used to argue that J_- and J_+ cannot indefinitely lower and raise the value of m in a *given* irreducible representation. One finds that there are states of highest and lowest angular momentum projection $J_+|jj\rangle = J_-|j,-j\rangle = 0$. Moreover,

$$J_\pm|jm\rangle = \hbar\sqrt{j(j+1) - m(m\pm 1)}|j, m\pm 1\rangle. \tag{37}$$

The factor on the right is obtained from the identities $J^2 = J_+J_- + J_z^2 - \hbar J_z = J_-J_+ + J_z^2 + \hbar J_z$. It gives us the matrix elements of $J_\pm$ (in this basis) in the representation labelled j.

• For $j = 0$ we get a 1d irreducible representation with J_i represented by the 1×1 zero matrix for all components i. This is the 'trivial' representation, but nature has chosen it for all the spin zero particles including the Higgs, pions, kaons, η, η'. The spin state of any such particle is given by a single complex number, the corresponding field is just a 1-component function of location (scalar field).

• By contrast, in the $j = \frac{1}{2}$ representation, the spin wave function is a 2 component complex vector ('spinor'). All quarks and leptons have spin half as do the octet of spin half baryons (including the nucleons). For $j = \frac{1}{2}$, in the basis in which J_z is diagonal, $J_i = \frac{1}{2}\hbar\sigma_i$ in terms of the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}; \quad \sigma_+ = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}, \quad \sigma_- = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} \tag{38}$$

σ_i span the space of 2×2 hermitian traceless matrices and satisfy the Lie algebra $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$. They also satisfy the identity (special to $j = \frac{1}{2}$) $\sigma_i\sigma_j = \delta_{ij}I + i\epsilon_{ijk}\sigma_k$ which implies the anti-commutation relation $\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}$.

• For $j = 1$ the matrix elements in the $|11\rangle, |10\rangle, |1-1\rangle$ ordered basis are

$$J_+ = \sqrt{2}\hbar \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad J_- = \sqrt{2}\hbar \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \text{and} \quad J_z = \hbar \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{39}$$

[22] The same SU(2) Lie algebra applies to the spin of relativistic particles, as long as they are not massless. Spin of a massless particle is *not* described by SU(2), and will be discussed later.

Massive spin one particles include the $W^{\pm}$, Z^0 weak gauge bosons and vector mesons ρ, K^*, ϕ, ω. The spin of a massless particle is not described by the $SU(2)$ Lie algebra. So the photon and gluon, which have spin 1, are *not* described by the above states, in particular, they have only two, not three spin states ('polarizations').

• The next possibility is $j = 3/2$. The Δ baryons ($\Delta^{++}(uuu), \Delta^{+}(uud), \Delta^{0}(udd), \Delta^{-}(ddd)$) have angular momentum (spin) 3/2 and their spin wave functions can be thought of as 4-component vectors with angular momenta represented by 4×4 matrices.

• An infinite dimensional (reducible) representation of the angular momentum algebra is provided by the orbital angular momentum differential operators $L_i = (\mathbf{r} \times \mathbf{p})_i = \epsilon_{ijk} r_j (-i\hbar \frac{\partial}{\partial r_k})$ acting on wavefunctions $\psi(r, \theta, \phi)$. Despite appearances, L_i don't involve the radial coordinate r (check that $L_i f(r) = 0$). In fact,

$$L_z = -i\hbar \frac{\partial}{\partial \phi}, \quad L_x = i\hbar \left(\sin\phi \frac{\partial}{\partial \theta} + \cot\theta \cos\phi \frac{\partial}{\partial \phi} \right), \quad L_y = i\hbar \left(-\cos\phi \frac{\partial}{\partial \theta} + \cot\theta \sin\phi \frac{\partial}{\partial \phi} \right). \tag{40}$$

So they take functions on the unit sphere $f(\theta, \phi)$ to functions on the unit sphere. A convenient basis for square-integrable complex-valued functions on S^2 is provided by the spherical harmonics $Y_{lm}(\theta, \phi)$. We may expand any such function uniquely as

$$\psi(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm} Y_{lm}(\theta, \phi), \qquad c_{lm} \in \mathbb{C}. \tag{41}$$

Each angular momentum l representation for $l = 0, 1, 2 \ldots$, appears as a sub-representation of this infinite dimensional representation. Each eigen-space (with eigenvalue $\hbar^2 l(l+1)$) of the Casimir

$$L^2 = -\hbar^2 \left[\frac{1}{\sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial \phi^2} \right]. \tag{42}$$

is an invariant subspace and carries the irreducible angular momentum l representation.

2.5.2 Angular momentum and rotations

• Angular momenta are known to generate rotations in classical mechanics (infinitesimal canonical transformations, say via the Poisson bracket). In qm, orbital angular momentum components $L_i = \epsilon_{ijk} r_j (-i\hbar \frac{\partial}{\partial r_k})$ generate rotations in the sense that the action of the infinite dimensional unitary operator $U(\hat{n}, \theta) = \exp(i\theta \mathbf{L} \cdot \hat{n}/\hbar)$ on the coordinate space wave function gives

$$e^{i\theta \mathbf{L} \cdot \hat{n}/\hbar} \psi(\mathbf{r}) = \psi(R\mathbf{r}). \tag{43}$$

where $R = e^{i\theta \mathbf{L} \cdot \hat{n}/\hbar}$ is the 3×3 SO(3) rotation matrix (obtained using the $j = 1$ representation of L_i) corresponding to the rotation of Euclidean space by angle θ counterclockwise about $\hat{n}$. For an infinitesimal rotation, $R\mathbf{r} = \mathbf{r} + (\delta\theta)\hat{n} \times \mathbf{r}$. An infinitesimal rotation corresponds to $U \approx I + \frac{i\delta\theta}{\hbar}(n_x L_x + n_y L_y + n_z L_z)$ where $\hat{n} = (n_x, n_y, n_z)$.

• If $\psi(\mathbf{r})$ is the coordinate space wave function with respect to a coordinate frame $\mathcal{F}$, then $\psi' = U(\hat{n}, \theta)\psi$ is the wave function with respect to a rotated frame $\mathcal{F}'$. In general, the expectation values and matrix elements of observables ($\langle\phi|A|\psi\rangle$) depend on the coordinate frame used, they are related by $\langle\phi'|A|\psi'\rangle = \langle\phi|U^\dagger AU|\psi\rangle$. A scalar observable is one that has the same matrix elements in all frames related by rotations. Applying the condition to an infinitesimal rotation we find that a scalar operator must commute with the angular momentum generators $[A, L_i] = 0$. The hamiltonian $H = p^2/2m - e^2/4\pi r$ of the hydrogen atom is an example of a scalar observable, as is the Casimir L^2.

Transformation of a $j = 1$ triplet under rotations: vectors

• The states of a $j = 1$ multiplet transform like an ordinary three vector under rotations. To see this, consider a $j = 1$ triplet $|m\rangle, m = 1, 0, -1$. Under an infinitesimal rotation $U(\hat{z}, \theta) \approx I + i\hbar^{-1}\theta L_z$ about the z axis, we find

$$\delta|m\rangle = i\hbar^{-1}\theta L_z|m\rangle = i\theta m|m\rangle. \tag{44}$$

So $|0\rangle$ is unchanged while the $|\pm 1\rangle$ states transform into multiples of themselves,

$$\delta|\pm 1\rangle = \pm i\theta|\pm 1\rangle. \tag{45}$$

Moreover, check that a rotation by $\theta = 2\pi$ leaves the state unchanged, $U(\hat{z}, 2\pi) = I$.

• To see the connection to rotation of a vector, recall that under an infinitesimal counter-clockwise rotation of axes by angle θ, the components of a vector v_1, v_2, v_3 transform into

$$\begin{pmatrix} v_1' \\ v_2' \\ v_3' \end{pmatrix} = \begin{pmatrix} 1 & -\sin\theta & 0 \\ \sin\theta & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \approx \begin{pmatrix} 1 & -\theta & 0 \\ \theta & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \tag{46}$$

So $\delta v_1 = -\theta v_2$, $\delta v_2 = \theta v_1$ and $\delta v_3 = 0$. In this basis, v_1 and v_2 are mixed under a rotation. To 'disentangle' them we can go to the 'helicity basis' by defining the complex combinations $v_\pm = (v_1 \pm iv_2)$ so that (upper signs are read together and lower signs give a separate equation)

$$\delta v_\pm = \delta v_1 \pm i\delta v_2 = -\theta v_2 \pm i\theta v_1 = \pm i\theta v_\pm, \quad \delta v_3 = 0. \tag{47}$$

This is precisely how the states of an angular momentum $j = 1$ triplet transform under rotations. Partly for this reason, a massive spin-1 particle is called a vector particle. Taking the z direction along its momentum, the $m = \pm 1, 0$ spin-projection states are called the positive, negative and zero helicity states. This explains the use of the term 'helicity basis' above. As shown, they transform under rotations in the same way as components of a 3-vector. The W and Z weak gauge bosons are examples of spin-1 vector bosons. Photons and gluons are also spin-1 vector bosons, but they are massless and lack the zero helicity state, they only come in positive and negative helicity versions. We are aware of this from the purely transverse polarization of EM waves. W and Z particles can have longitudinal as well as transverse polarizations. Spin for a massless particle is not described in terms of the SU(2) Lie algebra commutation relations, it is more subtle, we will study it later.

Transformation of a $j = \frac{1}{2}$ doublet under rotations: spinors

• Transformation of $j = \frac{1}{2}$ spinors under rotations. $\mathbb{C}^2$ is the space of states of a spin half system. Any such state (spinor $|\psi\rangle$) is a linear combination of $m = \pm 1$ eigenstates of σ_3,

$|\uparrow\rangle = \begin{pmatrix}1\\0\end{pmatrix}$ and $|\downarrow\rangle = \begin{pmatrix}0\\1\end{pmatrix}$, $|\psi\rangle = c_\uparrow|\uparrow\rangle + c_\downarrow|\downarrow\rangle$ with $|c_\uparrow|^2 + |c_\downarrow|^2 = 1$. Under a rotation, a spinor transforms by the unitary matrix $U(\hat n, \theta) = \exp \frac{1}{2} i\theta\vec\sigma \cdot \hat n$. Using $\sigma_i^2 = I$ we get

$$U(\hat n, \theta)|\psi\rangle = e^{\frac{1}{2}i\theta\vec\sigma\cdot\hat n}|\psi\rangle = \cos(\theta/2)I|\psi\rangle + i\sigma\cdot\hat n\sin(\theta/2)|\psi\rangle \tag{48}$$

For example,

$$\begin{aligned} U|\uparrow\rangle &= \cos(\theta/2)|\uparrow\rangle + i\sin(\theta/2)(n_x|\downarrow\rangle + in_y|\downarrow\rangle + n_z|\uparrow\rangle) \\ &= [\cos(\theta/2) + in_z\sin(\theta/2)]\,|\uparrow\rangle + [\sin(\theta/2)(in_x - n_y)]\,|\downarrow\rangle. \end{aligned} \tag{49}$$

Starting from, say, the $|\uparrow\rangle$ spinor, by suitable choice of $\hat n$ and θ, $U|\uparrow\rangle$ can be made equal to any $|\psi\rangle \in \mathbb{C}^2$. Indeed the set of such unitary transformations is a three parameter family (parametrized by the direction $\hat n$ and angle θ) as is the space of unit spinors in $\mathbb{C}^2$. So there is always a rotated coordinate frame in which a given spinor is just $|\uparrow\rangle$. The analogous statement is not true for $j > \frac{1}{2}$ since rotations are still a 3 parameter family but the unit vectors in $\mathbb{C}^{2j+1}$ are a larger ($4j+1$ real parameter) family. So not every unit vector in a $j = 1$ multiplet ($\mathbb{C}^3$) can be rotated to the $|j = 1, m = 1\rangle$ state. (Rotations of 3d Euclidean space, which are represented by the unitary operators $U(\hat n, \theta)$ on C^{2j+1} do not exhaust all unitary transformations of the Hilbert space for $j > \frac{1}{2}$.) An interesting physical consequence (according to T D Lee, as I learned from H S Mani) is that spin half particles cannot have a quadrupole moment while higher spin particles can.

• Moreover, to any unitary transformation of $\mathbb{C}^2$, there corresponds a rotation of 3d Euclidean space $\mathbb{R}^3$, this is not true for $j > \frac{1}{2}$. If the $\mathbb{C}^2$ is the space of states of a spin half system, then the corresponding Euclidean space is just ordinary coordinate space. But $\mathbb{C}^2$ could also be the space of isospin states, then the corresponding $\mathbb{R}^3$ is not ordinary coordinate space but an 'internal' isospin space.

• The observables of any two state system (hermitian 2×2 matrices) can be written as a linear combination of Pauli matrices and the identity matrix $A = a_0 I + a_i\sigma_i$. And the states of a 2 state system can always be written as a linear combination of $|\uparrow\rangle$ and $|\downarrow\rangle$, where $\uparrow, \downarrow$ are the eigenstates of σ_3. This analogy between two state systems and the spin $j = \frac{1}{2}$ system is often exploited in the treatment of two level systems (atoms, particles etc).

• Interestingly, a rotation by 2π does not return a non-zero spinor to its initial state since $U(\hat n, 2\pi) = -I$. But a rotation by 4π does, $U(\hat n, 4\pi) = I$. We say that a spinor (state vector of a spin half system) is double valued in ordinary 3d coordinate space. This double-valued nature has been experimentally detected through interference of beams of spin-half particles (neutrons) whose spins have been rotated using magnetic fields.

2.5.3 Comment on helicity and spin

• For a massive particle, its intrinsic angular momentum (spin) could be defined as its angular momentum in its rest frame. More generally, spin can be defined via the helicity $h = \mathbf{J}\cdot\hat p$, which is the projection of total angular momentum on the momentum of the particle. The advantage of this concept is that it applies to massless particles like the photon, that cannot be brought to rest. A massive particle can viewed in a frame in which it is going slowly, where non-relativistic QM applies, and then by taking $\hat{z}$ along $\vec{p}$, it is clear that $h = -s, s+1, \ldots, s-1, s$ where s is its spin. More generally, we can define the spin of a particle to be the maximum value of helicity. We will find that $h = \pm 1$ for a photon, so it is a spin one particle. For an electron $h = \pm\frac{1}{2}$, it is a spin half particle. $W^\pm$ and Z^0 have $h = 0, \pm 1$, they are spin one particles. Unlike a massive particle, a massless particle of spin s has only two possible helicities $h = \pm s$. This is familiar to us from the lack of longitudinal polarisation for EM waves.

• The Higgs is the only spin zero (i.e. scalar) elementary particle in the SM. Quarks and leptons have spin half. Photons, gluons, $W^\pm$ and Z^0 have spin one (i.e. vector particles). Gravitons would have spin 2. Composite particles like hadrons, nuclei and atoms too can be characterized by an angular momentum J. This angular momentum is due to a combination of the spins of the constituents as well as the orbital angular momenta of the constituents. When we speak of the spin of a hadron or nucleus or atom, we mean its total angular momentum J.

• Nature seems not to have used elementary particles of spin higher than two. However, there are composite particles with higher spin. Hadrons furnish many examples. Pions and kaons, η, η' have spin zero. The baryons $p, n, \Sigma, \Lambda, \Xi$ have spin half. The vector meson resonances ρ, ω, ϕ, K^* have spin one. The baryon resonances $\Delta, \Sigma^*, \Xi^*, \Omega$ have spin $3/2$. Meson resonances with spin as high as $2, 3$ and 4 have been found. Baryon resonances with spins as high as $11/2$ have been found. For example, among the excited states of the nucleon, the isospin $3/2$ $\Delta(2400)$ and $\Delta(1920)$ baryons have spin $11/2$ and $7/2$ respectively, while the isospin half baryons $N(2190)$ and $N(1688)$ have spin $7/2$ and $5/2$ respectively.

• The hydrogen atom (in any state) has integer angular momentum. Indeed, any bound state of an even number of spin half particles (e.g. proton and electron) must have integer spin since $\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$, and the orbital angular momentum of any state of hydrogen is always an integer $l = 0, 1, 2 \ldots$. Thus the total angular momentum $J = L + S$ of an H-atom must correspond to integer j. The deuteron has total angular momentum one, it is a spin one particle.

• The concept of helicity originally arose as a conserved quantity in inviscid fluid flow. $h = \mathbf{w} \cdot \mathbf{v}$ where $\mathbf{v}$ is the velocity vector field and $\mathbf{w} = \nabla \times \mathbf{v}$ is the vorticity field. So spin of a particle is a bit like vorticity of a fluid element.

2.5.4 Clebsch Gordan coefficients

• Consider a combination of two systems described by the angular momentum observables $\vec{J}_1$, $\vec{J}_2$ which commute with each other. Moreover suppose the individual systems have fixed angular momentum quantum numbers $j = j_1$ and $j = j_2$ respectively (corresponding to their individual Hilbert spaces $\mathbb{C}^{2j_i+1}$ carrying the relevant irreducible representations). Their respective Hilbert spaces have bases $|j_1, m_1\rangle$ and $|j_2, m_2\rangle$ with appropriate ranges of m_1, m_2. E.g., we may be combining the spin and orbital angular momenta of an e^- in an atom, or the spins of the two e^-s in a He-atom, the isospins of a π and a nucleon or p and n in a deuteron. The combined Hilbert space is the tensor product $\mathbb{C}^{2j_i+1} \otimes \mathbb{C}^{2j_i+1}$, which is the span of the tensor products of the basis vectors of the individual Hilbert spaces $\{|j_1 m_1\rangle \otimes |j_2 m_2\rangle\}$. Its dimension is the product of dimensions. The total angular momentum $\vec{J} = \vec{J_1} \otimes I + I \otimes \vec{J_2} \equiv J_1 + J_2$ again satisfies the angular momentum algebra and the eigenvalues of J^2 are $\hbar^2 j(j+1)$ where the possible values of j are (recall angular momentum addition or see below)

$$|j_1 - j_2|, |j_1 - j_2| + 1, \cdots, |j_1 - j_2| - 1, |j_1 + j_2|. \tag{50}$$

In other words, the combined system does not (in general) carry an irreducible representation of angular momentum due to the presence of more than one highest weight state (annihilated by $J_+ = J_{1+} + J_{2+}$). Instead we have a direct sum decomposition of the tensor product Hilbert space into irreducible multiplets

$$\mathbb{C}^{2j_1+1} \otimes \mathbb{C}^{2j_2+1} \equiv j_1 \otimes j_2 = |j_1 - j_2| \oplus (|j_1 - j_2| + 1) \oplus \ldots \oplus (|j_1 - j_2| - 1) \oplus |j_1 + j_2|. \tag{51}$$

The total state space $\mathbb{C}^{2j_1+1} \otimes \mathbb{C}^{2j_2+1}$ has the uncoupled basis of tensor product states $|j_1 m_1\rangle |j_2 m_2\rangle$ in which $J_1^2, J_{1z}, J_2^2, J_{2z}$ are all diagonal. It also has a coupled basis labelled by $|j m j_1 j_2\rangle$ abbreviated $|jm\rangle$ (where $\hbar^2 j(j+1), \hbar m$ are the eigenvalues of the total J and total J_z) in which J^2, J_z, J_1^2, J_2^2 are diagonal. Both the uncoupled and coupled bases are orthonormal bases for the total state space, so one may write

$$|jm\rangle = \sum_{m_1 + m_2 = m} C_{jm;m_1 m_2} |m_1\rangle |m_2\rangle \tag{52}$$

with the restriction of the range of summation arising from $J_z = J_{1z} + J_{2z}$. The CG coefficients $C_{\cdot;\cdot}$ are the matrix elements of a square $(2j_1+1) \times (2j_2+1)$ dimensional orthogonal matrix. Let us find them in some important examples.

• Clebsch-Gordan coefficients for addition of $j_1 = \frac{1}{2}$ and $j_2 = \frac{1}{2}$ is important in a 2-electron system like the He-atom or a 2 nucleon system like the deuteron. Here $j = 1, 0$ so $\frac{1}{2} \otimes \frac{1}{2}$ decomposes into the direct sum of a triplet $\{|jm\rangle = |11\rangle, |10\rangle, |1-1\rangle\}$ and a singlet $|jm\rangle = |00\rangle$. Now $|\frac{1}{2}\rangle|\frac{1}{2}\rangle = \uparrow\uparrow$ is the only tensor product state with $m = m_1 + m_2 = 1$, so we must have (with suitable choice of phase)

$$|11\rangle = |\frac{1}{2}\rangle|\frac{1}{2}\rangle \quad \text{and similarly} \quad |1-1\rangle = |-\frac{1}{2}\rangle|-\frac{1}{2}\rangle. \tag{53}$$

To find the CG coefficients for $|10\rangle$ we apply J_- to $|11\rangle$ or J_+ to $|1-1\rangle$ and recall that $J_\pm |jm\rangle = \sqrt{j(j+1) - m(m \pm 1)}\hbar |j, m \pm 1\rangle$. In natural units ($\hbar$ doesn't appear in the CG coefficients anyway),

$$J_-|11\rangle = \sqrt{2}|10\rangle = J_{1-}|\uparrow\rangle|\uparrow\rangle + J_{2-}|\uparrow\rangle|\uparrow\rangle = |\downarrow\rangle|\uparrow\rangle + |\uparrow\rangle|\downarrow\rangle. \tag{54}$$

Thus $|10\rangle = 2^{-\frac{1}{2}}(\uparrow\uparrow + \downarrow\downarrow)$. The singlet state $|00\rangle$ has $m = 0$ and so must be a linear combination $a \uparrow\downarrow + b \downarrow\uparrow$. The coefficients must be determined (up to normalization) by the condition that $|00\rangle$ must be annihilated by $J_\pm$, or equivalently that it must be orthogonal to $|10\rangle$. The latter gives $|00\rangle = 2^{-\frac{1}{2}}(\uparrow\uparrow - \downarrow\downarrow)$. The CG coefficients may be summarized in a 4×4 matrix

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} & 0 \\ 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{55}$$

with columns labelled by the ordered uncoupled basis states $|\uparrow\rangle|\uparrow\rangle, |\uparrow\rangle|\downarrow\rangle, |\downarrow\rangle|\uparrow\rangle, |\downarrow\rangle|\downarrow\rangle$ and the rows labelled by the ordered coupled basis states $|11\rangle, |10\rangle, |00\rangle, |1-1\rangle$ in decreasing order of m. The CG coefficient $C_{jm;m_1m_2}$ is given by the entry in the jm row and the m_1m_2 column. Check that $CC^t = I$. Our choice of ordered bases and phases ensures that C is both orthogonal and symmetric as a matrix.

• Find the Clebsch-Gordan coefficients for addition of $j = 1$ and $j = \frac{1}{2}$. The combination decomposes as $\frac{3}{2} \oplus \frac{1}{2}$. The uncoupled basis is ordered as

$$|1\rangle|\uparrow\rangle,\ |1\rangle|\downarrow\rangle,\ |0\rangle|\uparrow\rangle,\ |0\rangle|\downarrow\rangle,\ |-1\rangle|\uparrow\rangle,\ |-1\rangle|\downarrow\rangle. \tag{56}$$

and the coupled basis is ordered with decreasing $m = m_1 + m_2$

$$|3/2,3/2\rangle,\ |3/2,1/2\rangle,\ |1/2,1/2\rangle,\ |3/2,-1/2\rangle,\ |1/2,-1/2\rangle,\ |3/2,-3/2\rangle. \tag{57}$$

• The matrix of CG coefficients for $1\otimes\frac{1}{2} = \frac{3}{2}\oplus\frac{1}{2}$ with rows and columns labelled by uncoupled and coupled bases (in the order given above) is (by choosing phases in a suitable way so that it is a symmetric matrix)

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & \sqrt{\frac{2}{3}} & 0 & 0 & 0 \\ 0 & \sqrt{\frac{2}{3}} & -\frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{3}} & -\sqrt{\frac{2}{3}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{58}$$

2.6 Resonant scattering from partial wave expansion

• As an application of angular momentum and potential scattering theory, we re-derive the Breit-Wigner formula for the elastic cross section $a + b \to c \to a + b$ (in the vicinity of a resonance c) using the expansion in angular momentum partial cross sections. Suppose the resonant sate c occurs for center of momentum energies E in the neighborhood of the central mass M then the Breit-Wigner formula is

$$\sigma_{\text{elastic}} \approx \sigma_{\text{peak}} \frac{\Gamma^2/4}{(E-M)^2 + \Gamma^2/4}. \tag{59}$$

We will relate the energy width Γ and the peak cross section σ_{peak} to the phase shift and incoming wave number k appearing in the partial wave expansion of potential scattering. Now the Breit-Wigner formula for the cross section must be considered as a local addition (around $E = M$) to the background cross section when the particles are off resonance. For example, the πN cross section is typically around 20-30 mb, but near the resonance (i.e. around a centre of momentum energy 1232 MeV of the πN system -or- pion kinetic energy 195 MeV in the rest frame of the nucleon), the cross section goes up to 200 millibarns! Similarly, the e^+e^- annihilation cross section which is around 1 nb at a CM energy of 86 GeV, suddenly goes up to 30 nb at E = 91 GeV around the Z^0 resonance.

• Recall that the differential x-section $\frac{d\sigma}{d\Omega}$ is the absolute square $|f(\theta,\phi)|^2$ of the scattering amplitude. For a spherically symmetric scattering potential $V(r)$ f is independent of the azimuthal angle ϕ. Legendre polynomials $\left(\int_{-1}^{1} P_l(x)P_{l'}(x)dx = \frac{2}{2l+1}\delta_{ll'}\right)$ provide a basis for functions of $x = \cos\theta$, so the scattering amplitude may be expressed as a sum of partial wave amplitudes

$$f(\theta) = \sum_l (2l+1)\, a_l\, P_l(\cos\theta). \tag{60}$$

The partial wave amplitudes may be expressed in terms of the phase shifts δ_l (one uses unitarity to get this formula)

$$a_l = \frac{e^{2i\delta_l} - 1}{2ik} = \frac{e^{i\delta_l}\sin\delta_l}{k}, \tag{61}$$

where k is the angular wave number ($\hbar k$ is the momentum) of the incoming particles. The incoming particles have kinetic energy $KE = \hbar^2k^2/2m$. Note that this kinetic energy is not the same as the energy E that appears in the Breit-Wigner formula. The latter is the centre of momentum energy of the colliding particles a and b, or equivalently, the Lorentz invariant quantity $E = \sqrt{(p_a^\mu + p_b^\mu)^2}$. Using the orthogonality of Legendre polynomials, the total elastic cross section is expressed as a sum of partial cross sections

$$\sigma = 4\pi\sum_{l=0}^{\infty}(2l+1)|a_l|^2 \quad \text{or} \quad \sigma = \sum_l \sigma_l \quad \text{where} \quad \sigma_l = \frac{4\pi}{k^2}(2l+1)\sin^2\delta_l. \tag{62}$$

The incoming plane wave

$$e^{ikz} = e^{ikr\cos\theta} = \sum_{l=0}^{\infty} i^l(2l+1)j_l(kr)P_l(\cos\theta) \tag{63}$$

may be resolved into spherical waves of definite angular momentum l. Each scatters with a partial cross section σ_l. Off resonance, various angular momentum channels contribute. But near a resonance, a particular angular momentum channel tends to dominate the cross section $\sigma_{l_0} \gg \sigma_l$ for $l \neq l_0$. This may be visible in the angular distribution of scattered particles: near a resonance, they would be distributed roughly like $P_{l_0}(\cos\theta)^2$ (over and above a non-resonant background cross section). Based on this experimental finding, we will say that if $|a_l|^2$ passes through a maximum at some value of $l = l_0$ for some particular wave number k (or CM energy E), then the partial x-section σ_{l_0} will pass through a maximum and the two colliding particles are said to resonate. The resonant state is characterized by the the angular momentum l_0 and mass M corresponding to the center of momentum energy of the two particles at resonance. For hadronic resonances, one may also assign an isospin and parity to the resonant state. Our treatment here assumes the scattering particles are spin-less. More generally, the spin of the scattering particles also contribute, and the resonant state is characterized by a total angular momentum J which receives contribution both from l_0 and from the spins s_a and s_b. For the

$J = 3/2$ Δ resonance, since $s_p = \frac{1}{2}$ and $s_\pi = 0$, we would expect that the resonance is in the $l = 1$ channel.

• The condition for a resonance in the angular momentum l channel is that $\delta_l(k) = \pi/2$. In the neighborhood of a resonance, we will assume that $\sigma_{\rm tot}$ (after subtracting the non-resonant background cross section) is well approximated by the partial cross section σ_l, which has reached a maximum. Then

$$\sigma \approx \sigma_l = 4\pi(2l+1)|a_l|^2 \tag{64}$$

Now let us write the partial wave amplitude in a manner suggestive of the Breit-Wigner formula which involves a quotient,

$$ka_l = \frac{\sin\delta_l}{e^{-i\delta_l}} = \frac{1}{\cot\delta - i} \tag{65}$$

The resonant energy $E = M$ is defined as the one at which $\delta_l = \pi/2$ or $\cot\delta_l = 0$. Expanding in a Taylor series around $E = M$ we have

$$\cot\delta_l(E) = (E-M)\frac{d}{dE}\cot\delta_l(E)_{|E=M} + \mathcal{O}(E-M)^2. \tag{66}$$

The derivative has dimensions of inverse energy, and we define Γ by

$$\frac{d}{dE}\cot\delta_l(E)_{|E=M} = -\frac{2}{\Gamma} \tag{67}$$

The factors are chosen so that Γ coincides with the full width at half maximum appearing in the Breit-Wigner line shape. Then near the resonance, the partial wave amplitude is

$$ka_l \approx \frac{1}{\frac{2}{\Gamma}(M-E) - i} = \frac{\Gamma/2}{(M-E) - i\Gamma/2}. \tag{68}$$

Thus the partial cross section becomes

$$\sigma_{\rm elastic} \approx \sigma_l \approx \frac{4\pi}{k^2}(2l+1)\left|\frac{\Gamma/2}{(M-E)-i\Gamma/2}\right|^2 = \frac{4\pi}{k^2}(2l+1)\frac{\Gamma^2/4}{(E-M)^2+\Gamma^2/4}. \tag{69}$$

This is the Breit-Wigner formula with $\sigma_{\rm peak} = 4\pi(2l+1)/k^2$ where $l = J$ is the angular momentum of the resonant state. This applies to collision of spin-less particles $a + b \to c \to a + b$. More generally, if the colliding particles have spins s_a, s_b, then the Breit-Wigner cross section, averaged over the spin states of a and b is

$$\sigma_{\rm elastic} \approx \frac{4\pi}{k^2}\frac{(2J+1)}{(2s_a+1)(2s_b+1)}\frac{\Gamma^2/4}{(E-M)^2+\Gamma^2/4}. \tag{70}$$

2.7 Introduction to concept of parity

• Parity or reflection through the origin $\Pi(x,y,z) = (-x,-y,-z)$ is a discrete transformation of Euclidean space. In spherical polar coordinates $\Pi(r,\theta,\phi) = (r, \pi-\theta, \pi+\phi)$. Π may also be regarded as a composition of reflection in the $x-y$ plane (for instance) $(x,y,z) \to (x,y,-z)$ followed by rotation by π about the z-axis. So in 3 spatial dimensions, reflection through the origin and in a plane are both discrete transformations which cannot be continuously deformed to the identity[23]. If rotations are a symmetry of the system, then both definitions of parity may be used. Parity is often called mirror reflection and satisfies $\Pi^2 = I$.

• It has been found experimentally that parity is a symmetry of gravitational, EM and strong interactions, but not the weak interactions. As with other symmetries, if a Hamiltonian/Lagrangian or law of nature is parity-invariant, then parity transforms solutions of the eq. of motion or allowed phenomena into other solutions or allowed phenomena. EM and strong interactions are parity invariant in the sense that the probability for any EM/strong process is equal to the probability for the mirror-reflected process. Gottfried and Weisskopf give an example from atomic physics. Consider the angular distribution of radiation emitted by an ensemble of hydrogen atoms in a definite excited state. If the laws of atomic physics and atom-radiation interaction are parity invariant, then the angular distribution should be unchanged upon reflection in any plane that leaves the initial sate probability distribution $|\psi(\mathbf{r})|^2$ invariant.

• Parity is a discrete transformation. When it is a symmetry, it leads to a conserved quantity, though not via Noether's theorem. Dynamics specified by a hamiltonian H is parity invariant if $[\Pi, H] = 0$. It follows that the expectation value of Π in any state is conserved in time and an eigenstate of parity remains one of the same parity. In a scattering or decay that does not involve the weak interactions, the parity of the initial state (assuming it is one of definite parity) must equal the parity of the final state. Moreover, eigenstates of the hamiltonian $H|E\rangle = E|E\rangle$ can be chosen to have definite parity (even or odd, if necessary by forming the combinations $|E\rangle \pm \Pi|E\rangle$) and the parity of a stationary state is constant in time. For example, spherical harmonics $Y_{lm}(\theta,\phi)$ have parity $(-1)^l$, as may be seen from the Rodriguez formula for associated Legendre polynomials $Y_{lm} \propto P_{lm}(\cos\theta)e^{im\phi}$ where

$$P_{lm}(x) = (1-x^2)^{|m|/2} d_x^{|m|} \frac{1}{(2l)!!} d_x^l (x^2-1)^l \quad \text{and} \quad x = \cos\theta, \quad \Pi(\theta,\phi) = (\pi-\theta, \pi+\phi). \quad (71)$$

• Parity of a system composed of two sub-systems is multiplicative, provided their interactions are parity invariant. If the two sub-systems are far separated, the wave function $\Psi = \psi_1\psi_2$ is a product and $\Pi = \Pi_1\Pi_2$. Processes that bring the sub-systems closer will preserve parity by assumption.

2.7.1 Transformation of scalars, vectors and tensors under reflection

• A vector is a quantity that transforms like position under rotations. Among vectors we distinguish between polar vectors that change sign under reflections, and axial vectors that are parity even. For example, position $\mathbf{r}$, momentum $\mathbf{p}$, velocity $\mathbf{v}$, acceleration $\mathbf{a}$, force $\mathbf{F}$, electric field $\mathbf{E}$, current density $\mathbf{j}$ and polarization $\vec{\epsilon}$ are polar vectors. A cross product of polar vectors is an

[23]By contrast, in two spatial dimensions, reflection in the origin $(x,y) \to (-x,-y)$ is simply rotation by π in the $x-y$ plane and cannot be used to define parity. Instead define parity in 2D as reflection in a line through the origin, say the y-axis: $(x,y) \to (-x,y)$. The choice of the line does not matter as long as rotations are a symmetry of the system

axial vector. For example angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, torque $\mathbf{r} \times \mathbf{F}$, and the magnetic field $\mathbf{B}$ ($d\mathbf{B} \propto d\mathbf{l} \times \mathbf{r}$) are axial vectors. It follows that the vector potential ($\mathbf{A}$ in $\mathbf{B} = \nabla \times \mathbf{A}$) is a polar vector. The cross product of polar and axial vectors is a polar vector, an example is the Poynting vector $\mathbf{S} \propto \mathbf{E} \times \mathbf{B}$. A linear combination of polar and axial vectors does not have definite parity, though it transforms as a vector under rotations. Such a quantity appears in the parity violating $V - A$ theory of weak interactions where the difference between a vector and an axial vector current appears in the ineteraction.

• A scalar (under rotations) is a quantity that is independent of the orientation of the coordinate frame. Examples include constants like the rest mass or charge of a particle and the speed of light or fields like density, pressure, energy and temperature of a fluid. In fact, we identify two types of scalars. True scalars (simply called scalars) are invariant under reflections while pseudoscalars are odd under reflections. All the quantities mentioned above are true scalars. The dot product of two polar vectors or two axial vectors is a scalar. For example, power $= \mathbf{F} \cdot \mathbf{v}$ is a scalar, as is electromagnetic energy density $\frac{1}{2}(\mathbf{E} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{B})$. The dot product of a polar and axial vector (e.g., $\mathbf{E} \cdot \mathbf{B}$) is a pseudo scalar. The triple product of three polar vectors is a pseudo scalar. Check that Maxwell's equations are invariant under parity. Show that the magnetic dipole energy $\propto \mathbf{L} \cdot \mathbf{B}$ and electric dipole energy $\propto \mathbf{r} \cdot \mathbf{E}$ (or $\propto \mathbf{p} \cdot \mathbf{A}$ in radiation gauge) are scalars.

• While strong and EM processes have been found to conserve parity, weak interactions violate parity (maximally, in a sense). Angular momentum (both orbital and spin), being an axial vector, does not change sign under parity, while momentum is a polar vector and does change sign. It follows that helicity $h = J \cdot \hat{p}$ is a pseudo-scalar. So parity reverses the helicity of a particle. In particular, parity takes a $h = -1$ (negative helicity or left handed) neutrino to a $h = 1$ (positive helicity or right handed) neutrino (draw a figure). RH neutrinos have not been observed and are not included in the SM. This indicates that parity is not a symmetry of the *structure* of neutrinos. We will discuss parity violation in the weak *interactions* later.

• Under an orthogonal transformation, the Levi-Civita symbol transforms as $\epsilon'_{ijk} = R_{ii'} R_{jj'} R_{kk'} \epsilon_{i'j'k'}$. It follows that $\epsilon'_{123} = R_{1i} R_{2j} R_{3k} \epsilon_{ijk} = \det R$ etc. For a reflection $\det R = -1$: the ϵ symbol changes sign under parity, it is a pseudo-tensor.

2.7.2 Intrinsic parity

• In addition to the parity of the wavefunction of a system of particles, we may assign an intrinsic parity to the constituent particles. Gottfried & Weisskopf give a nice motivating example. Consider two $J = 0$ states $|\psi_s\rangle$ and $|\psi_p\rangle$ of a He atom. In $|\psi_s\rangle$, the electron spins are in a singlet state and both electrons are in S-wave states of zero angular momentum, so $\Pi|\psi_s\rangle = |\psi_s\rangle$ and ψ_s has even parity, it is called a scalar. In $|\psi_p\rangle$, one electron is in a P-wave state, the other in an S-wave state, giving $L = 1$ and $\Pi_p = -1$. The electron spins are in one of the triplet states with $S = 1$. Orbital and spin angular momentum are combined in such a way that $\mathbf{J} = \mathbf{L} + \mathbf{S} = 0$. Since ψ_p is has odd parity, it is called a pseudo-scalar. Now, each of these Helium states ψ_s, ψ_p may be considered as particles in their own right and perhaps even with the same center of mass wave function $\chi(R)$. Nevertheless, they behave differently under mirror reflection, and are assigned even and odd 'intrinsic' parities. Of course, we could assign these parities because we knew the substructure of these He states.

• Parity in atomic physics. In atomic processes, the number of electrons and nucleons is conserved, while the number of photons may change. Consider 1-photon emission. The parity of initial or final state is the product of (1) the intrinsic parities of the electrons and nucleons (2) the parity of the orbital wave function of the electrons Π_i, Π_f (sometimes called extrinsic parity) and (3) the parity of the emitted photon state Π_γ. Since the numbers of nucleons and electrons are conserved, the intrinsic parities of electrons and nucleons drop out and conservation of parity implies $\Pi_i = \Pi_f \Pi_\gamma$. It is possible to assign a parity to the emitted photon state. 1-photon states of definite angular momentum and parity are in 1-1 correspondence with certain solutions of Maxwell's equations, the electric and magnetic multipole fields. For instance $\Pi_\gamma = -1$ in electric dipole radiation while $\Pi_\gamma = 1$ for magnetic dipole radiation In particular, the parity of the electron wave function must change in an electric dipole transition. So an electric dipole transition from 2S to 1S is forbidden by parity conservation while 2P to 1S is allowed. The possibility of emitting single photons makes it possible to measure the parity of photon states, something that is not possible for nucleons or electrons. In strong and EM interactions, electrons or nucleons cannot be produced singly (unlike photons). An electron can only be produced in association with a positron. It is possible to find the parity of the e^+e^- pair, but not of an electron. It is possible to produce electrons singly, as in a beta decay, but then such a process is parity violating and cannot be used to find the parity of the electron.

• In strong interactions pions (and other mesons like kaons) play a similar role to that played by photons in EM. An accelerated charge radiates photons while an accelerated neutron can radiate pions. An excited hadron may decay by pion emission just as an excited atom may decay by photon emission. In strong interactions, such as $p + p \to d + \pi^+$ or $p + p \to \pi^+ + p + n$, pions can be produced or annihilated singly, and the parity of pions can be measured. Pions have negative intrinsic parity, they are called pseudo-scalars since they have zero spin in addition. Since isospin (which is a symmetry of the strong interactions) transforms pions into each other, they must all have the same negative parity, as must the kaons and η, η', using the SU(3) transformations among u, d, s quarks.

• The negative parity of π^- was experimentally established by studying the decay of an 'atom' made of a deuteron $d^+ = np$ and a π^-. It was in its ground state (l=0) and it decayed to two neutrons via the strong force. The outgoing neutrons were found to be in an $l = 1$ state. In the reaction $np\pi^- \to nn$, the intrinsic parities of the nucleons cancel out. So the parity of the π^- should equal the extrinsic parity of the final state, $\Pi_{\pi^-} = (-1)^l = -1$. The negative parity of pion can also be explained using the quark model.

• While the intrinsic parity of a fermion is a matter of convention, fermion-anti-fermion pairs such as e^+e^- can be produced singly, and their parity measured (the $\pi^0 = \bar{d}d - \bar{u}u$ is another example). From the theory of the Dirac field (to be discussed later) one can show that spin-half fermion and anti-fermion have opposite intrinsic parities. By convention, electrons, muons and tau leptons and all quarks are assigned even parity. It follows that a $q\bar{q}$ bound state with zero orbital angular momentum must have negative parity, as is the case for pions and kaons, they are pseudoscalars (zero spin and odd parity). On the other hand, the ρ meson resonances (and the rest of the spin 1 meson nonet K^*, ω, ϕ) are quark-anti-quark bound states with orbital angular momentum $l = 0, j = 1$. They have odd parity and are called vector mesons (states of a $j = 1$ multiplet transform as a 3d vector as shown previously). Unlike a fermion and its anti-particle, a boson and its anti-particle are not required to have opposite parities, for example, $\pi^\pm$ both have odd parity and the odd parity π^0 is its own anti-particle.

• The number of nucleons is conserved in strong interactions, so their intrinsic parities cancel out in reactions. By convention, nucleons are assigned positive parity. Since the spin-half baryons $n, p, \Lambda, \Sigma, \Xi$ transform into each other under flavor $\mathrm{SU}(3)$ symmetry of the light u, d, s quarks, all these baryons have the same even parity. This is consistent with the assignment of even parities to all the quarks and the fact that the quarks are in an $l = 0$ state in these lowest lying stable baryons. Strange particles may be produced in pairs (of opposite strangeness, in collisions of non-strange hadrons) in strong interactions, e.g. $p + p \to K^+ + \Lambda^0 + p$. So the parity of the $\Lambda^0 K^+$ pair can be measured relative to that of the proton, it is negative. Since kaons transform into pions under flavour $\mathrm{SU}(3)$ symmetry, they are assigned negative parity while $\Pi_{\Lambda^0} = +1$ consistent with what we said about the spin half baryon octet. The spin $3/2$ baryons $\Delta, \Sigma^*, \Xi^*, \Omega$ are composed of quarks which are again in orbital angular momentum $l = 0$ state, they too have even parity.

• It is possible to assign negative intrinsic parity to photons (this is different from the parities of the multipole radiation fields) and gluons, essentially because the vector potential $\mathbf{A}$ is a polar vector. As we will see, parity is not conserved in the weak interactions, so we cannot consistently assign parities to particles undergoing weak interactions (in particular we do not assign intrinsic parities to $W^\pm, Z^0, \nu, \bar\nu$). To summarize, some even parity particles are e^-, μ^-, τ^-, quarks (by convention), octet of spin-half baryons, decuplet of spin-$3/2$ baryons, nonet of vector mesons and Higgs scalar. Some odd parity particles are the nonet of spin zero pseudoscalar mesons, photons, gluons, e^+, μ^+, τ^+ and the anti-quarks.

2.7.3 Parity violation in the weak interaction

• To explain some puzzling features of weak decays of strange mesons, Lee and Yang in 1956 suggested that parity may not be conserved in the weak interactions. During 1954-56 a conundrum called the $\tau - \theta$ puzzle unfolded. Strange mesons decaying (weakly) to 2 and 3 pions $\theta^+ \to \pi^+\pi^0$ and $\tau^+ \to \pi^+\pi^+\pi^-$ were observed. However τ and θ were found to have the same masses and mean lifetimes and they seemed to be identical in all other respects (strangeness, isospin). However, they would have to have positive and negative parities if parity were conserved in the above reactions (pions have negative parity). So it was reasoned that they could not be the same particle. Lee and Yang studied the data available and pointed out that parity conservation had simply been assumed, and had not been experimentally established in the weak interactions (while it had been tested in EM and strong processes). They suggested that τ and θ were simply two different decay modes of the same particle (now called the K^+) and that the weak interactions did not conserve parity.

• Recall that if parity is a symmetry, then a mirror reflected process must have the same frequency of occurrence as the original process. To test this, one may set up two experiments which are mirror images of each other (and not identical) and see whether they give the same output. In 1956, C S Wu and collaborators (of Columbia Univ, though the experiment was done at the National Bureau of Standards (now NIST), Maryland due to the need for very low temperatures) prepared a sample of Co-60 nuclei with their spins (total angular momenta) $\mathbf{J}$ polarized in a common direction (say $\hat{z}$). The direction of out-going beta electrons was measured. Let us consider the effect of reflections in the $x-y$ plane, under which the position polar vector $(x,y,z) \to (x,y,-z)$. The e^- momentum $\mathbf{p}$ is also a polar vector. Under a reflection in the xy plane, $\mathbf{p} = (p_x, p_y, p_z) \mapsto (p_x, p_y, -p_z)$. On the other hand, the nuclear angular momentum $\mathbf{J}$ is an axial vector, so under a mirror reflection $J_z \to J_z$. So the mirror reflected version of a beta decay process would have the Cobalt nucleus polarized along $\hat{z}$ but with beta particle emerging with momentum reflected in the $x-y$ plane. If parity were a symmetry, these two processes must have equal probability. In other words, the angular distribution of electrons must be up-down symmetric or $\mathrm{Prob}(\theta) = \mathrm{Prob}(\pi - \theta)$ where θ is the polar angle between $\hat{z}$ and the direction of electron momentum. However, the experimental measurements were significantly up-down asymmetric. This showed that parity is not conserved in the weak interactions. If the Cobalt nuclei weren't spin polarized (by maintaining low thermal fluctuations), then the up-down symmetry would be hidden. The reaction was ${}^{60}\mathrm{Co} \to {}^{60}\,\mathrm{Ni}^{**} + e^- + \bar{\nu}_e$. The spin of Cobalt is 5 while that of the excited Nickel nucleus is 4. Using the fact that anti-neutrinos are exclusively RH, we can understand why most of the electrons came out in the direction opposite to the spin of the Cobalt nuclei. To first approximation we ignore the recoil of the Nickel nucleus so that the electron and anti-neutrino moment must be back to back and the Ni spin is is in the same direction as the Co spin (i.e. along $\hat{z}$). The spins of the e and $\bar{\nu}_e$ must add to one and point upwards to make up for the difference between Co and Ni angular momenta. Thus the electron momentum must be along $-\hat{z}$ and the $\bar{\nu}_e$ must travel upwards. We also see that the electron emerges with negative helicity in this example. This is generally the case, relativistic charged leptons that participate in the charge changing weak interactions predominantly have negative helicity (see below).

• At the same time, L Lederman and R L Garvin (at Columbia) found evidence for parity violation in the $\pi \to \mu \to e$ decay chain. Another group of V. Telegdi et. al. also experimentally demonstrated parity violation.

2.7.4 Helicity of ultra-relativistic fermions in charged weak processes and $\pi^{\pm}$ decay

• Another interesting experimental fact about weak interactions concerns the helicity of the fermions and anti-fermions (leptons or quarks and their anti-particles) produced in charged weak processes. Charged weak processes involve the $W^{\pm}$, e.g. $W^- \to \mu^- \bar{\nu}_\mu$, $W^- \to e^- \bar{\nu}_e$, $W^+ \to e^+ \nu_e$ etc. All neutrinos produced are found to have negative helicity while all anti-neutrinos have positive helicity. But it is not just the neutrinos that obey this helicity selection rule, so do ultra-relativistic (i.e. produced with energy much more than rest mass) quarks and leptons. It is found that ultra-relativistic fermions (e^-, μ^-, τ^- and the quarks) are produced predominantly with negative helicity in charged weak interactions. A small fraction ($\sim m/E$) of the fermions come out with positive helicity, also called the 'wrong' helicity. On the other hand, the anti-fermions (e^+, μ^+, τ^+ and anti-quarks) are predominantly produced with positive helicity.

• Let us apply this helicity selection rule to understand the peculiar case of charged pion decay. Based on the 140 MeV pion mass, two weak decay modes are allowed $\pi^- \to e^- \bar{\nu}_e$ and $\pi^- \to \mu^- \bar{\nu}_\mu$. One might expect the former to be the dominant decay mode since $m_\pi - m_e \gg m_\pi - m_\mu$ and it would be like sliding down a steeper slope. However, the decay to an electron is greatly suppressed due to the helicity selection rule and charged pions predominantly decay to muons. To see this, consider the π^- decay in its rest frame where e^- and $\bar{\nu}_e$ would emerge 'back-to-back' with zero total momentum, to conserve momentum. Since the $\bar{\nu}_e$ has positive helicity, its spin must point away from the π^- decay vertex. The electron (which typically emerges with a speed close to c due to $m_\pi >> m_e$) must have negative helicity, and so its spin must point towards the interaction vertex. So the total angular momentum of the final state is non-zero and points in the direction of the momentum of the $\bar{\nu}_e$. However, the initial state has zero angular momentum since pions have spin zero. So this decay is forbidden by the conservation of angular momentum. The same decay with muons replacing electrons would also be forbidden, except that the muons are typically not ultra-relativistic since $m_\pi - m_\mu < m_\mu$. It is through the 'wrong' helicity $(h = +1)$ μ^- that π^- decay proceeds to a large extent.

2.8 Identical particles: Fermions and Bosons

• All electrons (protons, photons, hydrogen atoms in their ground state etc) have been found to be identical in their intrinsic physical characteristics (mass, charge, spin, magnetic moment). Moreover, electrons cannot be unambiguously distinguished by their trajectories due to Heisenberg uncertainty. This experimental fact is implemented in non-relativistic QM by requiring (1) that the hamiltonian (which specifies the dynamics) is invariant under a permutation of identical particles and (2) that the permutation of identical particles results in a wave function that represents the same physical state $P_{ab}\psi(\ldots q_a \ldots q_b \ldots) = e^{i\theta}\psi(\ldots q_b \ldots q_a \ldots)$ where q denotes all the degrees of freedom of each particle (position, spin etc). The permutation/exchange operator is hermitian, $P^2 = I$, so its eigenvalues are ± 1 and it commutes with the hamiltonian. Thus energy eigenstates may be chosen to be simultaneously eigenstates of the permutation operators. So the phases $e^{i\theta} = \pm 1$ and the corresponding eigenspaces are the bosonic and fermionic multi-particle states. [There are other possibilities. For instance, particles moving on a plane may in a sense be attached to magnetic flux tubes extending into the third dimension. Then the particles are no longer strictly localized around a point and require a different treatment, leading to 'anyonic' statistics.]

• In a second quantized framework the symmetry and anti-symmetry of state functions is a consequence of the corresponding creation and annihilation operators satisfying commutation or anti-commutation relations. It is a remarkable experimental fact that all fermions have half odd integer spin and all bosons have integer spin. This is called the spin-statistics connection. It can be shown to hold in relativistic quantum field theory using causality.

• Anti-symmetry and Pauli principle: Suppose the number of identical fermions is approximately constant and the multi-particle wave function can be approximately built from single particle wave functions by anti-symmetrization (say, when each moves almost independently of the others). Then the requirement of anti-symmetry implies that no more than one identical fermion can occupy the same single particle state. This is Pauli's principle. It can be usefully applied to multi-electron atoms, nuclei and even to the valence quarks in baryons. In particular, two electrons with the same spin cannot occupy the same spatial location or same atomic orbital. This leads to the 'degeneracy pressure' of fermions responsible for the stability of ordinary matter and white dwarfs. When the gravitational attraction can no longer be withstood, the electrons in a white dwarf may be captured by the protons via the weak interaction $e^- + p \to n + \nu_e$. There is a transition to a neutron star (Pauli's principle is not violated). On the other hand, identical bosons can occupy the same single particle state and indeed tend to condense to the lowest such state at low temperatures.

• A composite particle that is a bound state of n identical fermions behaves as a fermion or boson according as n is odd or even. To understand this, note that the addition of an even number of half-odd integer spins always gives an integer spin while addition of an odd number always gives a half-odd integer spin (and orbital angular momentum is always an integer multiple of $\hbar$). For example, a system of identical (say, all in their ground state) nuclei of baryon number B behaves as a system of fermions or bosons according as B is odd or even. This statement assumes that the neutron and proton can be treated as different isospin states of the same particle (nucleon). This is a good approximation if their mass difference and electromagnetic effects can be ignored. The fermionic and bosonic nature of baryons and mesons (3 and 2 valence quark bound states) is another illustration of this.

• For a two body system, the spatial part of the wave function $\psi(\mathbf{r})$ of a state of definite $\mathbf{L}^2$ and L_z is the product of a radial part and a spherical harmonic Y_{lm}. Here $\mathbf{r}$ is the relative coordinate, it changes sign under exchange of particles $\mathbf{r} \to -\mathbf{r}$. From our study of parity we may conclude that under particle exchange, this orbital wave function is symmetric or anti-symmetric, according as l is even or odd. This fact is often quite useful.

• Let us apply the ideas of identical particles and angular momentum conservation to the dominant decay of the neutral pion $\pi^0 \to 2\gamma$, to argue (the original argument is due to L D Landau Doklady Akad. Nauk (USSR) 60, 207 (1948) and C N Yang Phys Rev 77, 242 (1950)) that the spin of the π^0 cannot be one. Working in the π^0 rest frame, the final state depends on three vectors, the relative momentum $\mathbf{k}$ of the two photons and their polarizations $\vec{\epsilon}_1, \vec{\epsilon}_2$. $\pm\mathbf{k}$ are the directions in which the photons emerge. The final state must be symmetric under exchange of photons on account of their Bose statistics. It must be linear in each of the polarizations (as it lies in the tensor product Hilbert space) and must transform as a vector under rotations if the pion has $J = 1$ and angular momentum is conserved. In momentum space, the wave function would then be some scalar function of $\mathbf{k} \cdot \mathbf{k}$ times this vector. There are several vectors that one can construct. $\epsilon_1 \times \epsilon_2$ is inadmissible since it is anti-symmetric under exchange. $(\epsilon_1 \cdot \epsilon_2)\mathbf{k}$ is also anti-symmetric since $\mathbf{k} \to -\mathbf{k}$ under exchange of the photons. The only symmetric vector is proportional to $\mathbf{k} \times (\epsilon_\mathbf{1} \times \epsilon_\mathbf{2}) = \epsilon_\mathbf{1}(\mathbf{k} \cdot \epsilon_\mathbf{2}) - \epsilon_\mathbf{2}(\mathbf{k} \cdot \epsilon_\mathbf{1})$, but it is identically zero since photons are transversely polarized. Thus the amplitude for two photon emission by a spin-1 π^0 vanishes. Since the process does occur, π^0 must have spin 0 or $2, 3 \ldots$. Additional arguments show that pions are spin less. [Note also that if the π^0 spin were non-zero, then the angular distribution of photons emitted by a polarized π^0 sample would be anisotropic, which would be another way to investigate their spin.]

2.9 Lorentz symmetry and relativistic mechanics

2.9.1 Introduction

• The idea of an inertial frame arose in Newtonian mechanics, i.e., one in which Newton's law of inertia (1st law) holds. Newton thought there was an absolute rest frame and his laws held in frames in uniform motion with respect to this absolute frame. Maxwell thought light travels in an ether medium and that his equations held in a frame at rest relative to the ether. It was assumed that the ether is at rest in Newton's absolute frame. However, the Michelson-Morley experiment failed to detect any relative motion between the earth and the ether. The speed of light was the same in the direction of Earth's motion around the sun and opposite to it as well as in other directions. Due to the apparent absence of any preferred frame, Einstein concluded that there is no absolute rest frame or ether medium and that the speed of light in vacuum c is the same for all observers. He also postulated that the laws of physics in special relativity (as in Galilean relativity), are the same in all inertial frames (i.e., frames related by rotation or uniform motion ['boosts'] (and translations)): this is Lorentz invariance (Poincare invariance, if we include translations). A difference between Galilean and Lorentz invariance is the formula relating space-time coordinates under a boost.

• The space-time coordinates of an event $(\mathbf{r}', t')$ in frame S' which moves at $v\hat{x}$ relative to frame S are

$$x' = \gamma(x - vt), \quad y' = y, \quad z' = z, \quad ct' = \gamma(ct - xv/c), \quad \text{with} \quad \gamma = \left(1 - \frac{v^2}{c^2}\right)^{-1/2}, \tag{72}$$

where the two frames coincide at $t = 0$. This formula for a Lorentz boost is a symmetry of Maxwell's equations but not of Newton's equations of mechanics. Maxwell's equations are valid in all frames in uniform relative motion with the same constant speed of light c. Einstein modified Newton's equations so that they respect the constancy of the speed of light, and are invariant under Lorentz transformations.

• In particular, simultaneity is frame-dependent in general. Two simultaneous events in S with coordinates (x_a, t) and (x_b, t) occur at different times $t'_{a,b} = \gamma(t - x_{a,b}v/c^2)$ in a boosted frame.

• Moving rods are Lorentz contracted. A rod of length l' lying along the x'-axis in frame S' has a length $l = l'/\gamma$ when viewed from from S. To see this suppose the rod extends from the origin of S' to x' so that $l' = x' - 0$. Then $x' = \gamma(x - vt)$ and $0 = \gamma(x_0 - vt)$ where $x_0(t)$ is the coordinate of the origin of S' (or left end of rod) when viewed from S and $x(t)$ is the coordinate of the right end of the rod when viewed from S, so that $l = x - x_0$. Subtracting, $l' = \gamma(x - x_0) = \gamma l$.

• Time dilation: 'moving clocks run slower'. A time interval τ' according to a clock in S' is assigned a time interval $\tau = \gamma\tau'$ by an observer in S. To see this, suppose τ is the time interval between when the origins of S and S' coincide and when the origin of S' is at location x_0

viewed from S. Then $t_1 = t'_1 = 0$ while $t_2 = x_0/v$ and $ct'_2 = \gamma(ct_2 - x_0 v/c) = \gamma x_0(\beta^{-1} - \beta)$. With $\tau = t_2 - t_1$ and $\tau' = t'_2 - t'_1$ we find $c\tau' = \gamma\tau(c^2 - v^2)/c$ or $\tau' = \tau/\gamma$.

• Time dilation implies that moving unstable particles viewed from the lab will appear on average to live longer than their (rest-frame) mean life time. E.g. muons produced in cosmic ray showers in the upper atmosphere (several km above sea level) have a mean life time of $\tau = 2.2\mu s$ in their rest frame. They would be expected to cover at most $c\tau = 660$m at ultra-relativistic energies, but in fact some make it well below the Earth's surface! On the other hand, if a particle does not decay at rest, it cannot be made to decay by moving it at speed v.

• Velocity transformation formula. Let $\mathbf{u} = \frac{d\mathbf{r}}{dt}$ and $\mathbf{u}' = \frac{d\mathbf{r}'}{dt'}$ be the velocities of a particle according to two observers S and S' (boosted by velocity $v\hat{x}$ relative to S). How are they related. To find out we write the Lorentz boost in infinitesimal form

$$dx' = \gamma(dx - vdt), \quad dy' = dy, \quad dz' = dz, \quad cdt' = \gamma(cdt - \beta dx) \tag{73}$$

Then we find

$$u'_x = \frac{dx'}{dt'} = \frac{dx - vdt}{dt - \beta dx/c} = \frac{u_x - v}{1 - vu_x/c^2}, \quad u'_y = \frac{u_y}{\gamma(1 - u_x v/c^2)}, \quad u'_z = \frac{u_z}{\gamma(1 - uu_x/c^2)}. \tag{74}$$

Inverting, $u_x = \frac{u'_x + v}{1 + vu'_x/c^2}$ is the x-component of the velocity of the particle as viewed from S when it has velocity component u'_x in S' while S' moves with velocity $v\hat{x}$ relative to S. In particular, if the velocity of the particle is in the same direction as the boost (i.e. x), then $u_y = u'_y = u_z = u'_z = 0$ and we get the velocity composition law for two collinear velocities: if $\beta_i = v_i/c$ then $\beta = \frac{\beta_1 + \beta_2}{1 + \beta_1\beta_2}$. In other words, the composition of two Lorentz boosts with parameters $\beta_{1,2}$ is a boost with parameter β. This is also the rule for addition of hyperbolic tangents $\tanh(\theta_1 + \theta_2) = \frac{\tanh\theta_1 + \tanh\theta_2}{1 + \tanh\theta_1 \tanh\theta_2}$. We define the *rapidities* θ_i by $\beta_i = \tanh\theta_i$. Speed or v/c is sometimes an awkward variable in relativity since it is ≈ 1 for all relativistic particles. Rapidity is more convenient since it stretches the interval $0 \le v/c \le 1$ to the infinite range $0 \le \theta \le \infty$. Rapidity is particularly natural since rapidities simply add under composition $\theta = \theta_1 + \theta_2$ while velocities 'add' in the more complicated manner given above. Check that c composed with any velocity v is again c, ensuring the constancy of the speed of light in all inertial frames. In particular, the wave fronts of light emitted from a point are spherical in all such frames, unlike the wave fronts of sound which look flattened in the direction of motion of an observer. If light is emitted from $\mathbf{r} = 0$ at $t = 0$ in frame S, the wave front at time t is the sphere $r(t)^2 = c^2t^2$. The graph of an expanding spherical wavefront takes the shape of a light-cone in a space-time diagram.

• In terms of rapidity $\theta = \operatorname{arctanh}(v/c)$ we have $\sinh\theta = \gamma v/c$ and $\cosh\theta = \gamma$. Thus the above Lorentz boost takes the form of a hyperbolic rotation

$$x' = x\cosh\theta - ct\sinh\theta, \quad y' = y, \quad z' = z, \quad ct' = -x\sinh\theta + ct\cosh\theta. \tag{75}$$

• When a particle's speed $|\mathbf{v}|$ approaches that of light, relativistic effects become significant. Acceleration of a massive particle can (in principle) increase its momentum or kinetic energy without bound, but it would take an infinite amount of energy to reach the speed of light. The

Newtonian relations $E = p^2/2m$ and $\mathbf{p} = m\mathbf{v}$ between energy or momentum and velocity are modified (as explained later) to $E = \gamma mc^2$ and $\mathbf{p} = \gamma m\mathbf{v}$ where $\gamma = (1 - v^2/c^2)^{-1/2}$. The relativistic expression includes the rest energy $E_0 = mc^2$ of the particle whose (rest) mass is m (sometimes γm is called the relativistically increased mass). Conservation of energy holds in relativistic processes only if we include the rest energy. For example, a neutron at rest can beta decay producing a proton, and fast moving $\bar{\nu}_e$ and electron. *Estimate the average distance (viewed from the Earth) that a 1.057 GeV ($= 10m_\mu$) muon travels before decaying.*

• It follows that $E = \frac{pc^2}{v}$. So as the speed $v \to c$, energy becomes linear in momentum $E \gtrsim pc$. For a massless particle like the photon, $E = pc$. So for ultra-relativistic particles, energy and momentum are equal in natural units (GeV). More generally, the formulae for E and $\mathbf{p}$ imply the 'mass-shell' relation $E^2 = \mathbf{p}^2c^2 + m^2c^4$, which defines a two-sheeted hyperboloid in the 4d energy-momentum space with coordinates $(E/c, p_1, p_2, p_3)$. Real massive particles have an energy-momentum 4-vector that lies on the positive energy sheet/shell. When viewed from different reference frames, a particle can have different energy-momentum 4-vectors (though all observers agree on the sign of E and the value of m). The totality of the tips of all these 4-vectors is the mass shell, it is a Lorentz-invariant construct. 'Virtual' particles (which appear in intermediate stages of calculations) can be off mass-shell. For massless particles (like photons, gluons and gravitons) the mass shell becomes the positive energy part of the light cone $E^2 = \mathbf{p}^2c^2$.

• E or p are more convenient ways of specifying the motion than speed. For example a 1.022 MeV (twice the rest energy) electron has speed $v = (\sqrt{3}/2)c = 0.87c$. A 10.22 MeV electron has speed $v = .9987c$ while a 1.022 GeV electron has speed $v = 0.99999987$c. While the speeds may not seem very different, the energies are vastly different and can be used in e^+e^- collisions to produce new particles with higher masses. Annihilation of 1 GeV electrons and positrons is capable of producing (say) charged pions or kaons with masses of 140 and 494 MeV/c^2 while 10 MeV electrons which have a similar speed, would be incapable of doing so.

2.9.2 Minkowski space and Lorentz group

• Space-time coordinates of an event in frame S are specified by (contravaraint) position 4-vector $x^\mu = (x^0, x^1, x^2, x^3) = (ct, x, y, z)$ (regarded as a column vector below). Coordinates x' in a Lorentz boosted or rotated frame S' are given by $x'^\mu = \Lambda^\mu_\nu x^\nu$. If S' moves at speed v in x direction relative to S, then Λ is block diagonal $\begin{pmatrix} \gamma & -\gamma\beta \\ -\gamma\beta & \gamma \end{pmatrix}$ in the 0-1 subspace and identity in the 2-3 subspace. If S, S' are related by a rotation/reflection, then $\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix}$ is block diagonal where $R^tR = I$ is an $O(3)$ matrix.

• Minkowski introduced his inner product so that boosts could be regarded as hyperbolic/imaginary rotations. The Minkowski inner product between 4-vectors is

$$x \cdot y = (x, y) = x^0y^0 - x^1y^1 - x^2y^2 - x^3y^3 = x^\mu \eta_{\mu\nu} y^\nu = x^t \eta y \quad \text{where} \quad \eta_{\mu\nu} = \text{diag}(1, -1, -1, -1), \tag{76}$$

$\eta_{\mu\nu}x^\nu$ are called the covariant components of the position 4-vector, so $x \cdot y = x^\mu y_\mu$. The metric and its inverse $\eta^{\mu\nu}$ ($\eta^{\mu\nu}\eta_{\nu\rho} = \delta^\mu_\rho$) may be used to lower and raise indices and define new tensors.

• Boosts and orthogonal transformations leave the Minkowski inner product invariant in the sense $(x', y') = (x, y)$ since $\Lambda^t \eta \Lambda = \eta$. So the inner products of 4-vectors $x \cdot y$ and the $(length)^2$ of a 4-vector x^2 are Lorentz invariant, same for all inertial observers.

• The set of all Λ leaving η invariant form the homogeneous Lorentz group $O(1,3)$ (they are the linear transformations of Minkowski space that preserve the inner products of 4-vectors, in particular, they must preserve the origin and not involve translations). Since it preserves the inner product, a Lorentz transformation (LT) necessarily takes time-like vectors to time-like vectors, space-like vectors to space-like vectors and null vectors to null vectors. The Lorentz group is a 6 dimensional group. In other words, the constraints $\Lambda^t \eta \Lambda = \eta$ admit a 6 parameter family of solutions Λ. The Lie algebra of the Lorentz group has 6 independent generators, which can be taken as three infinitesimal rotations (L_x, L_y, L_z) and three infinitesimal boosts (K_x, K_y, K_z).

• In addition to boosts and rotations (and compositions thereof) the Lorentz group includes discrete transformations: parity (reflections such as $x \to -x, y \to y, z \to z, t \to t$, a reflection in any other plane or through the origin can be got from this one by a suitable rotation) and time-reversal $t \to -t$ (and compositions of all these). In fact, the Lorentz group has 4 connected components: the connected component of the identity is the group of proper orthochronous Lorentz transformations $SO(3,1)$. Proper means determinant 1, e.g., not involving reflection in an odd number of spatial directions if the sense of time is not reversed. Orthochronous means maintaining the sense of time. The others are proper non-orthochronous, improper orthochronous and improper non-orthochronous. They are labelled by the sign of $\det \Lambda$ and Λ^0_0. Indeed, from $\Lambda^t \eta \Lambda = \eta$ we see that $\det \Lambda = \pm 1$. If $\det \Lambda = 1$ we call it a proper LT and improper otherwise. From the $(0,0)$ component of $\Lambda^t \eta \Lambda = \eta$, we find that the first column parametrizes a 2-sheeted hyperboloid $(\Lambda^0_0)^2 - (\Lambda^1_0)^2 - (\Lambda^2_0)^2 - (\Lambda^3_0)^2 = 1$, with sheets labelled by the sign of Λ^0_0. Though parity and time reversal are symmetries of the EM and strong interactions, it was found in 1956 and 1964 (indirectly) that neither parity nor time reversal is a symmetry of weak interactions. In the standard model, the time reversal invariance is broken by a phase in the CKM mass matrix for quarks and parity is violated by the vector minus axial vector structure of weak interactions. Only the proper orthochronous Lorentz transformations (i.e. the connected component of the identity) seem to be symmetries of all the interactions.

• Minkowski space is the disjoint union of time-like $x^2 > 0$, space like $x^2 < 0$ and light-like $x^2 = 0$ (null) 4-vectors. The separation vector between the ends of a rod is space-like. Tangent vectors to the trajectory of a massive particle are time-like, as we will see. Tangent vectors to a light ray are null vectors. The Lorentz group acts on Minkowski space via $x \mapsto \Lambda x$. The orbit of a point x in Minkwoski space is the set of all 4- vectors Λx as Λ ranges over the group. The orbit of the origin $x = 0$ is the origin. The orbit of a light-like vector is the light cone. The orbit of a time-like vector is the two sheeted hyperboloid on which it lies. The orbit of a space like vector is the 1-sheeted hyperboloid that it lies on. We say that Minkowski space is foliated by the orbits of the Lorentz group. Each orbit other than the origin is a 3-dimensional manifold (i.e. a ‘hypersurface’).

2.9.3 Relativistic momentum, energy and particle motion

• Consider the trajectory of a massive particle moving around the lab. It is a curve $x^\mu(\tau)$ in Minkowski space. τ is proper-time, time as measured by a clock attached to the particle. τ is the arc length along the curve in units of c. If $d\tau$ is a proper-time interval around an instant when the particle's speed is v, then the corresponding time interval measured by a stationary observer in the lab is $dt = \gamma d\tau$. The proper-time interval when the position vector of a particle is incremented by dx^μ is the Minkowski-length of the tangent vector in units of c: $c^2(d\tau)^2 = \eta_{\mu\nu}dx^\mu dx^\nu = c^2dt^2 - d\mathbf{r}^2$; it is Lorentz invariant. The Newtonian velocity $v^i = \frac{dx^i}{dt}$ of the particle, where x^i are its lab frame coordinates and t lab time does not transform as the spatial components of a 4-vector since t transforms as the zeroth component of a 4-vector. On the other hand, the 4-velocity $u^\mu = \frac{dx^\mu}{d\tau}$ transforms as a contravariant 4-vector just like the position 4-vector since $d\tau$ is Lorentz invariant. The Newtonian and proper velocities are related: $u^i = \frac{dx^i}{d\tau} = \gamma\frac{dx^i}{dt} = \gamma v^i$ and the zeroth component $u^0 = \gamma c$. So $u^\mu = \gamma(c, \frac{d\mathbf{r}}{dt}) = \gamma(c, \mathbf{v})$. The 4-velocity is a time-like vector of constant length: $u^2 = \gamma^2(c^2 - v^2) = c^2$.

• The contravariant components of the momentum 4-vector of a massive particle are defined as $p^\mu = mu^\mu = m\gamma(c, \mathbf{v}) = (E/c, \mathbf{p})$. This serves as the relativistic definition of energy E. We see that $E = \gamma mc^2$ while $\mathbf{p} = \gamma m\mathbf{v}$ as mentioned earlier. Like position x^μ, p^μ transforms as a 4-vector under Lorentz boosts, $E' = \gamma(E - p_x v)$ and $p_x' = \gamma(p_x - \frac{vE}{c^2})$. Its covariant components are $p_\mu = \eta_{\mu\nu}p^\nu = (E/c, -\mathbf{p})$. It is time-like $p^2 = m^2c^2$. The possible 4-momenta of a massive particle lie on the 2 sheeted hyperboloid $E^2/c^2 - \mathbf{p}^2c^2 = m^2c^2$.

• The relativistic energy E includes the rest energy (mc^2), non-relativistic kinetic energy $\mathbf{p}^2/2m$ and further corrections, but not any potential energy:

$$E = \sqrt{\mathbf{p}^2c^2 + m^2c^4} = mc^2 + \frac{\mathbf{p}^2}{2m} + \frac{3}{8}\frac{\mathbf{p}^4}{m^3c^2} + \cdots \tag{77}$$

Sometimes E is called relativistic kinetic energy to distinguish it from potential energy. More often, $E - mc^2$ is referred to as relativistic kinetic energy.

• The above considerations do not apply to massless particles. They travel at the speed of light and follow trajectories whose tangent vectors are always null $d\tau^2 = 0$. For a massless particle $E = |\mathbf{p}|c$. The 4-momentum of a massless particle $(|\mathbf{p}|, \mathbf{p})$ is a null vector. The possible 4-momenta of a massless particle lie on a light cone. For a photon, the energy is determined by its frequency $E = h\nu$.

• Tachyons, particles with space-like momenta $p^2 < 0$ do not exist. Note that for a tachyon, there would be observers who would disagree as to the sign of its energy, its energy can be made arbitrarily negative in a suitable frame since its 4-momentum lies on a 1-sheeted hyperboloid. However, there are 'virtual' (off mass shell) particles with space-like momenta that are exchanged between real (onshell) particles. A famous example of this is the space-like $(q^\mu q_\mu < 0)$ virtual photon exchanged between an electron and a quark in deep inelastic scattering, the experiment in which quarks were discovered. However, such virtual particles are not detected, they only enter into calculations, though in an essential way.

2.9.4 Relativistic Lorentz force equation

• Newton's law of inertia (1st law) continues to hold in special relativity. A particle travels in a straight line with respect to the Minkowski metric when it is not acted on by any force. The relativistically covariant generalization of Newton's second law is

$$\frac{dp^\mu}{d\tau} = (\text{force})^\mu. \tag{78}$$

For example, a particle of charge e in an electromagnetic field feels the Lorentz force

$$\frac{dp^\mu}{d\tau} = eF^{\mu\nu}\frac{u_\nu}{c}. \tag{79}$$

where u^μ is the 4-velocity and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is Faraday's anti-symmetric covariant electromagnetic field strength tensor. Here $\partial_\mu = (\frac{1}{c}\partial_t, \nabla)$. The gauge potential $A^\mu = (\phi, \mathbf{A})$ includes the scalar and vector potentials. The electric and magnetic fields $\mathbf{B} = \nabla \times \mathbf{A}$ and $\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \nabla\phi$ with cartesian components $E^i, B^i = \epsilon^{ijk}\partial_j A^k$ are then the components of the field strength $F^{0i} = \partial^0 A^i - \partial^i A^0 = -E^i$ and

$$F^{ij} = \partial^i A^j - \partial^j A^i = -\partial_i A^j + \partial_j A^i = -\epsilon^{ijk}B^k \quad \Rightarrow \quad F^{\mu\nu} = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{pmatrix}. \tag{80}$$

The covariant components are given by $F_{\mu\nu} = \eta_{\mu\rho}\eta_{\nu\sigma}F^{\rho\sigma}$. So $F_{0i} = -F^{0i}$ and $F_{ij} = F^{ij}$.

• If the speed of the charge $v \ll c$, then show that the spatial components of the above equation reduces to the non-relativistic Lorentz force equation in the lab frame

$$\frac{d\mathbf{p}}{dt} = e\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}\right) \tag{81}$$

Show that the time component reduces to the equation for evolution of energy $\frac{dE}{dt} = e\mathbf{E}\cdot\mathbf{v}$, the rhs is the rate at which the Lorentz force does work on the particle, the magnetic field does not contribute.

2.9.5 Momentum conservation, (in)elastic collisions and forbidden processes

• 4-momentum is conserved in particle interactions. So the initial energy (sum of rest and kinetic energy) and 3-momentum must equal those in the final state in any collision/interaction/decay. Rest energy and kinetic energy are typically not conserved, though the sum is. A collision is elastic if kinetic energy (or equivalently rest energy) is conserved, e.g. $A + B \to C + D + E$ where $m_A + m_B = m_C + m_D + m_E$. Since there are very few such relations among masses of distinct particles, an elastic collision usually implies that the particles in the initial state are the same as the particles in the final state, e.g. Moller scattering $e^-e^- \to e^-e^-$ or Compton scattering $\gamma e \to \gamma e$ or Rutherford scattering $\alpha N \to \alpha N$. In the 'charge exchange' strong process $\pi^- p \to \pi^0 n$ kinetic energy is conserved (ignoring the mass splittings within an isospin multiplet) even though the final state particles are not the same as those in the initial state. A famous inelastic scattering process is deep inelastic scattering ('DIS', first carried out at Stanford by the MIT-SLAC team lead by J Friedman, H Kendall and R Taylor beginning 1967 (Nobel prize 1990)), which was instrumental in the discovery of point-like quarks inside the proton, e.g. $ep \to eX$ uses the electromagnetic force to probe proton structure, X is a state containing several hadrons with total baryon number one and electric charge one (there is a kinematic parameter that allows one to quantify the degree of inelasticity). Similarly, the weak interaction was also used to probe the structure of the proton via neutrino nucleon deep inelastic scattering, e.g., $\nu_e p \to e^- X$ where X typically contains several hadrons with total baryon number one and charge two.

• These DIS experiments succeeded the elastic $ep \to ep$ scattering experiments conducted at Stanford by R Hofstadter (Nobel Prize 1961), as the electron beam energy was raised above 1 GeV. Elastic ep and electron-deuteron scattering gave a lot of information on nucleons, including their charge distribution, magnetic moments (elastic electric and magnetic form factors) and size (charge radius). These elastic scattering experiments indicated that the proton is a soft sphere of charge. Thus, it came as a surprise when the DIS experiments observed 'hard' scattering of electrons off protons by decreasing the wavelength of the exchanged virtual photon. The charge inside the proton, when probed at short distances, was concentrated in point-like 'partons', later identified with quarks.

• As a consequence, spontaneous decay $A \to D_1 + D_2 + \cdots$ is forbidden if the rest mass m_A is less than the sum of the rest masses of the daughter particles $\sum_i m_i$. This is because the initial 4-momentum (in the center of momentum (CM) frame, where the 3 momentum is zero) $p_i^\mu = (m_A, \vec{0})$ must equal the final $p_f^\mu = (\sum m_i + \text{k.e.}, \sum_i \mathbf{p}_i)$, and the kinetic energy of decay products is non-negative. $M_A = \sum_i m_i$ is called the threshold for the decay. For instance, the hydrogen atom in its ground state ($m_H = m_e + m_p - 13.6eV$) is stable against decay to an electron and proton due to the binding energy. The hydrogen g.s. is also stable against decay via K-electron capture $H = e^- p^+ \to n\nu_e$ since the mass of the neutron is significantly (≈ 782 keV) more than that of the hydrogen atom. K-electron capture does happen in some proton-rich nuclei which (due to the nuclear binding energy) can reduce their mass by replacing a proton with a neutron. On the other hand, a free neutron ($m_N = 939.565$ MeV) does decay to $pe\bar{\nu}_e$ ($m_p = 938.272$ MeV, $m_e = .511$ MeV, $m_{\bar{\nu}} < m_e$). The mass difference of ≈ 782 keV (assuming $m_\nu < 1$ keV) is manifested in the kinetic energy of decay products.

• Let us give some more applications of energy-momentum conservation. Radiation of a (real) photon by an isolated (real) electron (or positron) $e^- \to e^- + \gamma$ is forbidden by energy conservation. In the rest frame of the initial electron $E_i = m_e < E_f \geq m_e + E_\gamma$. Electrons can radiate photons when they are accelerated in the Coulomb field of a nucleus, the nucleus provides the momentum balance.

• The reverse reaction of photon absorption by free electrons $\gamma e \to e$ is similarly forbidden by energy conservation. On the other hand, photons do scatter off electrons via Compton scattering $e\gamma \to e\gamma$.

• Free photons cannot turn into e^+e^- pairs. We may always go to a frame where the photon's energy is less than $2m_e$, while the final state has energy at least $2m_e$. On the other hand, a photon in the Coulomb field of a nucleus can pair produce.

• Pair annihilation to a single photon $e^+e^- \to \gamma$ is forbidden by momentum conservation. In the CM frame of the initial state $\mathbf{p}_{\rm tot} = 0$, but the final momentum cannot be zero as the photon moves at the speed of light in all frames. Pair annihilation to two photons has been observed.

2.9.6 Conservation of energy-momentum in 2 body decay

• Two body decays (i.e. spontaneous decay of a parent to two decay products) are more common (e.g. α and γ decay of nuclei $N(A,Z) \to N(A-4, Z-2) + \alpha$, $N^* \to N\gamma$, $\Lambda^0 \to p\pi^-$ etc.) than 3 body decay (e.g. β decay). The π^+ decays after coming to rest in photographic emulsion. In the decay of a positive pion π^+ (at rest) to μ^+ the μ^+ always emerges with the same kinetic energy of 4.1 MeV and travels roughly the same distance of about 600 microns. As we will show, a mono-energetic spectrum of energies of decay products is characteristic of a 2-body decay. Indeed, we now know that $\pi^+ \to \mu^+\nu_\mu$.

• The kinematics of two body decay (as opposed to 3 body decay) is strongly constrained by energy and momentum conservation. Consider the decay $a \to b + c$ in the rest frame of parent a. Let the 4-momenta of a, b, c be p, q, r. Then in natural units, $p = (m_a, \vec{0})$ and momentum conservation implies $q = (E_b, \vec{p}_b)$ and $r = (E_c, -\vec{p}_b)$. We will show that the energies of the daughter particles E_b, E_c are fixed by the masses! By 4-momentum conservation, $p^\mu - r^\mu = q^\mu$. Squaring,

$$p^2 + r^2 - 2p\cdot r = q^2 \quad \text{or} \quad m_a^2 + m_c^2 - 2m_a E_c = m_b^2 \quad \Rightarrow \quad E_c = \frac{m_a^2 + m_c^2 - m_b^2}{2m_a}. \qquad (82)$$

And by $b \leftrightarrow c$ symmetry, $E_b = \frac{m_a^2 + m_b^2 - m_c^2}{2m_a}$. In other words, the daughter particles are mono-energetic, they do not come out with a range of energies. In 3 body decay, energy-momentum conservation allows for a range of daughter particle energies, as was found in beta decay. Initially, it was mistakenly thought that beta decay must also be a 2 body decay, by analogy with alpha and gamma decays of radioactive nuclei. But the continuous spectrum of beta particles from a given radionuclide invalidated this hypothesis.

• Furthermore, since the energy of a particle is bounded below by its rest energy, $E_b \geq m_b$ which means $m_a^2 + m_b^2 - m_c^2 \geq 2m_a m_b$ or $(m_a - m_b)^2 \geq m_c^2$. From $p = q + r$, it is clear that $m_a \geq E_b \geq m_b$, so we must have $m_a \geq m_b + m_c$. As one would expect, the daughters must be less massive than the parent for the decay to conserve energy. In particular, a proton cannot decay to a neutron and anything else. And a photon cannot decay to massive particles. In particular, for a photon to produce a real e^+e^- pair, there must be some other particle (usually a nucleus) to conserve energy-momentum.

• Remark: The Q-value of a decay is the energy released as kinetic energy of the decay products, e.g. in the beta decay of a nucleus, $N(A,Z) \to N(A, Z+1) + e^- + \bar{\nu}_e$. The huge range in beta decay lifetimes (or alpha decay lifetimes) is due to the differences in Q-values, despite all beta decays being governed by an interaction of the same strength (Fermi's 4 fermion coupling).

The term Q-value is also used in other areas of chemistry and physics to quantify the energy released in a reaction and perhaps traces its origin to the energy released as heat in an exothermic reaction. It is to be distinguished from the Q-factor (quality factor) of a resonance which is the ratio of the energy of the resonant state divided by the energy width of the resonance. A narrow resonance has a high Q-factor, it is more stable and lasts longer since dissipative effects that lead to its decay are weaker. One speaks of the Q-factor of a resonant cavity or LCR circuit.

• The Geiger-Nuttal law found empirically in 1911 states that α decay half lives of nuclei satisfy $t_{\frac{1}{2}} \propto e^{aZ/\sqrt{E}}$ where E is the energy of the α particle emitted and Z is the atomic number. So short-lived nuclei emit more energetic alpha particles as they 'slide down a steeper slope'! The exponential dependence on energy also explains how small differences in mass defects between parent and daughter nuclei can lead to large differences in decay half lives. For example, Thorium 232 alpha decay has a half life of 14 billion years while Radium has an alpha decay half life of 1590 years. The Geiger-Nuttal law was explained in 1928 by Condon and Gurney and independently by Gamow. The alpha particle tunnels across a barrier between an attractive nuclear potential well (due to the strong interactions) and the outside of a nucleus where there is electrostatic repulsion.

2.9.7 Threshold energy for particle production via scattering on a fixed target

When two particles scatter at sufficient energies, the scattering may not be elastic, and the final state may involve one or more new particles, e.g. production of the Δ resonance ($\pi^+ p \to \Delta^{++}$) when a pion beam is incident on a hydrogen target or strange baryon production when a pion or kaon beam strikes a hydrogen target ($\pi^- p \to \Lambda^0 K^0$, $K^- p \to \Lambda^0 \pi^0$, $K^- p \to \Sigma^- \pi^+$ etc., these are strong interactions that conserve strangeness. The subsequent decay of a strange hadron may proceed via a strangeness changing charged weak interaction). It is found that there is a minimal (threshold) energy of the pion or kaon beam for the desired final state to be produced. Moreover, this threshold energy must be more than the difference between the invariant mass of the final state (M_{final}) and the target, $M_{\text{final}} - M_p$. For, if the incident meson beam has just so much energy, then the final state would have to be produced at rest, and then 3-momentum would not be conserved.

• We can find the threshold energy using 4-momentum conservation. For definiteness, consider $a + b \to c + d$ where the 4-momenta are p, q, r, s and particle a (proton) is the target at rest in the lab and E_b^{lab} is the lab frame energy of the projectile beam (meson) b ($q = (E_b^{\text{lab}}, \mathbf{p}_b^{\text{lab}})$). Then $p + q = r + s$ or squaring, in the lab frame

$$m_a^2 + m_b^2 + 2m_a E_b^{\text{lab}} = m_c^2 + m_d^2 + 2r \cdot s \tag{83}$$

So the projectile energy E_b^{lab} is minimal when $r \cdot s$ is minimal. Since $r \cdot s$ is Lorentz invariant, we may choose to evaluate it, say, in the CM frame of the c-d system. In this frame $r \cdot s = E_r^{\text{cm}} E_s^{\text{cm}} - \mathbf{p}_r \cdot \mathbf{p}_s = E_r^{\text{cm}} E_s^{\text{cm}} + |\mathbf{p}_r^{\text{cm}}|^2$ since the 3-momenta of c and d are opposite. This is minimal when the final state particles are produced at rest in the CM frame, i.e., $r \cdot s \geq m_c m_d$. Thus we find

$$E_b^{\text{lab}} \geq E_b^{\text{threshold}} \equiv \frac{(m_c + m_d)^2 - m_a^2 - m_b^2}{2m_a}. \tag{84}$$

The threshold energy is the minimum energy of the incident b-particle beam needed to produce the final state (with rest mass $m_c + m_d$) when incident on a stationary target of a. More generally, if the final state is a single particle or one with three or more particles, then we would replace $m_c + m_d$ above with the sum of the rest masses of the desired final state particles. At threshold, the particles in the final state are produced at rest in the CM frame of the system. Equivalently, in the lab frame (rest frame of target) at threshold, the out-going particles all move ('together') in the same direction as the incoming b-particle beam; for them to fly in different directions the beam energy must be above threshold.

• Find the threshold pion beam energy for production of the Δ resonance and compare it with the πN resonance cross section data. Find the threshold charged pion and kaon beam energies for production of the Λ^0, Σ^- strange baryons in collisions with a stationary proton target.

• Note that the threshold energy is just a kinematic minimum to produce the desired final state. The event rate depends on the probability of producing the final state (which depends on the dynamics – forces of interaction between particles) as well as on the phase space available for the final state, which are in general, energy dependent. Thus one may require higher energies than the threshold to obtain a desired event rate.

• Instead of a fixed target, one may also consider a colliding beam set up where a and b collide with equal and opposite 3-momenta. Find the threshold energy $E_b^{\text{min, cm}}$ for production of a final state with invariant mass M.

2.9.8 Mandelstam variables and s, t, u channel scattering

• Mandelstam variables (S Mandelstam 1958) s, t, u are Lorentz invariant variables associated with $2 \to 2$ scattering $a + b \to c + d$. They have dimensions of mass-squared and encode the energy, momentum, and angles between particle momenta in a frame-independent manner. If the 4-momenta are p_a, p_b, p_c, p_d then momentum conservation implies the relation $p_a + p_b = p_c + p_d$ while each incoming/outgoing particle is on mass shell $p_a^2 = m_a^2$ etc. We may construct interesting Lorentz scalars by squaring the sums or differences of momenta. It is conventional to define (in units where $c = 1$)

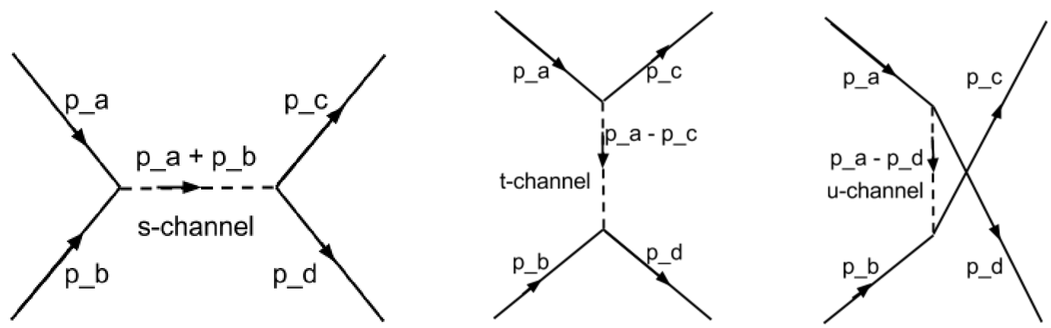
$$s = (p_a + p_b)^2 = (p_c + p_d)^2, \quad t = (p_a - p_c)^2 = (p_b - p_d)^2, \quad \text{and} \quad u = (p_a - p_d)^2 = (p_b - p_c)^2. \tag{85}$$

s is called the square of center of mass energy. This is because in the center of momentum frame, the 3-momenta of the incoming particles $\mathbf{p}_a = -\mathbf{p}_b$ are equal and opposite so that $p_a + p_b = (E_a + E_b, \mathbf{0})$. Thus $E_{CM} = E_a + E_b = \sqrt{s}$. For example the highest energy $p\bar{p}$ collisions at the Tevatron at Fermilab were at $\sqrt{s} \approx 1.8$ TeV.

• One could construct other Lorentz invariants like the scalar products $p_a \cdot p_b$, but they are not independent, $2 p_a \cdot p_b = s - m_a^2 - m_b^2$ etc. In fact, only two among s, t and u are independent. One checks that

$$s + t + u = m_a^2 + m_b^2 + m_c^2 + m_d^2. \tag{86}$$

• An interpretation of s, t, u is that they are the squares of the 4-momenta carried by a single particle that mediates the interaction among a, b, c and d. One speaks of s, t and u channel scattering as shown in these three Feynman diagrams (arrows indicate flow of momentum and time runs to the right).



• For example, $\bar{\nu}_\mu e \to \bar{\nu}_\mu e$ elastic scattering via Z exchange proceeds through the t-channel if we agree to label the particles $\bar{\nu}_\mu = a, c$ and $e^- = b, d$. With this convention, Z exchange cannot occur in the s or u channels.

• What is the CM energy in head-on collisions of 9 GeV (total energy) electrons with 3.1 GeV (total energy) positrons?

3 Acceleration of particles

3.1 Nature's particle accelerators

• Natural radioactive decay of unstable nuclei (uranium salts, radium, polonium, radon etc.) provided a reliable source of α, β and γ ray particles. This was the main source of particles in the first third of the 20th century and was widely exploited by Rutherford, the Curies and others leading to many discoveries including nuclei, proton, neutron, artificial radioactivity, fission etc. The energy of α, β, γ rays from radioactive decay is limited to a few MeV by the nuclear mass difference between parent and daughter nuclei. For example, one gets a mono-energetic spectrum of 5 MeV alpha particles from Radon alpha decay (2 body decay), which was used in the Geiger-Marsden gold foil experiment. Rutherford (1919) discovered that the proton is a constituent of the nucleus by ejecting protons from a Nitrogen nucleus that was irradiated with alpha particles from a radioactive source, $\alpha_2^4 + N_7^{14} \to O_8^{17} + p_1^1$.

• Cosmic rays are another natural source of particles with a very wide range of energies (many have energies of 100s of MeV and several GeV while the most energetic (but very rare) ones detected so far have 10^{12} GeV). Primary cosmic rays are mostly protons (and some alpha particles, heavier nuclei and a few electrons) coming from largely unknown sources (perhaps some astrophysical analogue particle accelerators are involved, in the 1950s, the Crab nebula was identified as one possible source, other potential sources include supernovae, gamma ray bursts, active galactic nuclei, quasars etc.). Neutral particles like photons and neutrinos also come to us, but are usually not called cosmic rays. Collisions of primary cosmic ray particles with O_2, N_2 gas molecules in the atmosphere produces showers of secondary cosmic rays which reach the earth. The secondary particles include protons, photons, pions, kaons, muons, electrons, alpha particles, neutrinos and their anti-particles. Pions and kaons (via pions) decay in the atmosphere to muons and neutrinos which reach ground level. Cosmic rays arrive at random, the flux is low and they have not been controlled to form beams. The primary cosmic ray flux decreases with energy, it is about a 1000 per m^2 per second with GeV energies, about 1 per m^2 per second with TeV energies etc. Nevertheless, detection of cosmic rays using cloud chambers, photographic emulsions and scintillation counters led to the discovery of positrons (C D Anderson (using cloud chambers in 1932 as a student of Millikan at Caltech) and Blackett & Occhialini at Cambridge in 1932-33), muons (C D Anderson & graduate student S Neddermeyer, 1936-37), pions (Powell, Occhialini, Muirhead & Lattes 1947 in Bristol using photographic emulsions) and kaons (Rochester and Butler, 1947 at Manchester using cloud chambers) and the study of particle interactions during the first half of the 20th century. They are still of much interest as a free source of particles for detector testing/calibration, as the source of the most energetic particles and as a window into astrophysics and cosmology.

• Nature also sends us neutrinos, mainly from fusion in the solar core ($4p \to \alpha + 2e^+ + 2\nu_e$ which releases 27 MeV of energy, the neutrinos alone can escape from the solar core) and occasionally from super novae. Solar neutrino flux is very large, 10^{10} per square cm per second. But neutrinos interact very weakly (cross sections roughly of order $\sigma_\nu \sim E \times 10^{-11}$ mb where E is the neutrino energy in GeV below the scale of electroweak mixing (100 GeV); the cross section then decreases), and most pass through the Earth without interacting. Atmospheric neutrinos are those produced in cosmic ray showers. Atmospheric neutrinos were first detected in 1965 at the Kolar Gold Fields mine in India.

• Certain reactions do not require acceleration of particles. Grand unified theories which seek to go beyond the standard model predict proton decay via reactions such as $p \to e^+\pi^0 \to e^+2\gamma$. The simplest such $\mathrm{SU}(5)$ model of Georgi and Glashow suggested a half life of order 10^{31} years. Experiments to check this (like Kamioka nucleon decay experiment begun in 1983 in Japan - 'Kamiokande' the last three letters 'nde' stand for nucleon decay experiment) looked for proton decay in a large tank of water, containing of order 10^{32} protons (3 kilotons of water surrounded by 1000 photo multiplier tubes), so that one may expect to see a few events in a year. No proton decay event has been seen so far, the current (2014) lower bound on the proton mean life is of order 10^{34} years. But in a stroke of luck, the same experiment unexpectedly detected neutrinos from supernova 1987a, a supernova that was visible to the naked eye and occurred in a nearby galaxy, the large Magellanic cloud.

3.2 Artificial particle accelerators

• The need for controlled high flux beams of high energy particles to explore sub-atomic physics led to the development of particle accelerators. The basic principle is to use an electric field to accelerate charged particles. We have no direct way of accelerating neutral particles (beams of neutral particles like neutrinos, $K^0, \bar{K}^0$ and neutrons can be got indirectly through decays of previously accelerated charged particles, or their production in high energy collisions of charged particles with a target). Electrons and protons (and their anti-particles) are the ones most frequently accelerated in accelerators.

• To accelerate particles, one needs a high vacuum environment. Otherwise, there is a lot of energy loss due to collisions and scattering off atoms in the gas. So the development of acceleration technology went hand in hand with the development of vacuum technology, though we do not discuss it here.

• The particle accelerator most of us are familiar with is the picture tube of an old TV or cathode ray tube of an old computer monitor. Here electrons are boiled off a cathode by heating it and then accelerated by applying an electric field of about 10 kilo volts. The 10 KeV electron beam is then deflected as needed by applying a magnetic field generated by current carrying coils. The electrons produce flashes of light when they strike the fluorescent screen which is coated with a phosphor.

• Protons for acceleration are obtained by ionizing a gas of hydrogen. By applying large electric fields, electric discharges are produced in the gas chamber, producing electrons and protons, which are pulled towards oppositely charged electrodes and injected into the accelerator. The same procedure is also used to obtain other ion beams.

3.3 Van de Graaff generator

Van de Graaf at Princeton and then MIT developed the static voltage generator named after him beginning in 1929. By 1931 he had achieved a 150 kilo volt potential difference. This was done by transferring charge from one electrode to another on a silk belt running between two pulleys. It was like a water wheel that moves water up from a tank. Van de Graaf generators can reach about 5 mega volt potential difference which can be used to accelerate electrons and protons etc. By using a pair in *tandem*, it is possible to reach 10-25 Mega Volt potential differences.

3.4 Cockroft Walton generator

Cockroft and Walton (1932) working at the Cavendish Lab, Cambridge under Rutherford developed a voltage multiplier ('cascade generator') that bears their name. It converted 200 kilo volts AC current from a transformer into a DC voltage of 700 kilo volts using capacitors and rectifiers (AC to DC converter). They were motivated by Rutherford's desire to have high energy beams of electrons, atoms etc that could not be obtained from natural radioactive decay. By colliding the 700 KeV proton beam with a Lithium target, they artificially split an atom (nucleus) for the first time ($p+$ $\mathrm{Li}_3^7 \to 2\alpha^4$), creating an international sensation. G. Gamow had told them that protons of a few 100 KeV could penetrate the nuclear barrier (a sort of reverse alpha decay) and induce nuclear reactions. Cockroft Walton generators are still used in the initial stage of larger particle accelerators today. They are also used to generate the 10s of KeV voltages needed in TV sets.

3.5 Linear particle accelerator

• The idea of particle accelerators go back to the Swede Gustav Ising (Stockholm 1924, different from the German Ernst Ising who solved the 1d Ising model as a student of Wilhelm Lenz of the Runge-Lenz vector fame), Hungarain Leo Szilard (Berlin 1928, who also had the idea of the nuclear chain reaction) and Norwegian Rolf Wideroe (Aachen, Germany 1928). The first working model was constructed by R Wideroe who achieved 'energy doubling' in 1928.

• The basic idea of a linear particle accelerator is to accelerate charged particles (say electrons) to a high energy by a succession of small accelerations while passing through an evacuated cylindrical pipe. Traditionally, the electrons were initially produced by boiling them off a metal filament or other cathode, as in a TV picture tube. Inside the pipe is a collinear sequence of cylinders (metal electrodes) separated by gaps. When the particles are inside the tubes, they feel no electric field. Each acceleration is due to a relatively small voltage difference across the gap between two successive electrodes. If the voltage on each electrode was fixed, then successive electrodes would have to be maintained at increasingly high voltages which would not be feasible. Instead, the voltages on the electrodes are set using a radio frequency power source. So a given electrode periodically switches from positive to negative voltage. The lengths of the electrodes and radio frequency are adjusted so that when the electrons are between a pair of electrodes, they always feel an accelerating, rather than decelerating voltage. In other words, the voltage on a given electrode must switch sign between when the electron enters and exits the electrode. This is called the method of resonance acceleration, patented and implemented by Wideroe. Since the electrons are speeding up, the lengths of the electrodes must increase along the length of the accelerator. Such an accelerator is called a linear accelerator or linac. The longest linac (2 miles) is at the Stanford Linear Accelerator Center (SLAC), it accelerates electrons to about 50 GeV. Linacs are also used in the initial stages of larger circular accelerators like the Tevatron at Fermilab. The next big accelerator after the LHC is likely to be an e^+e^- linear collider.

• Typically, the electron beam consists of bunches of electrons separated by gaps, rather than a steady stream. Since electrons in the beam repel, the beam tends to spread out and periodically needs to be focussed, as we shall describe soon.

3.6 The Cyclotron

• Upon reading about Wideroe's resonance accelerator, E O Lawrence at Berkeley (1929 onwards) had the idea of bending the charged particle trajectories using a magnetic field, so that the same accelerating voltage difference could be traversed several times. This led to a sequence of increasingly energetic cyclotrons built by Lawrence, his student Livingstone and collaborators. In a vacuum chamber, two hollow semicircular electrodes (called 'dees' for their D shape) were aligned to form a disk with a gap in between. Charged particles (say from a radioactive source) entered the cyclotron at its center. The disk (originally about 10 cm in diameter) was placed between the poles of an electromagnet producing a roughly constant magnetic field (12.7 kilo Gauss, 1 Tesla $= 10^4$ Gauss) normal to the Ds. So charged particles would be bent in a semi-circular path inside the Ds with an angular frequency of revolution ω.

$$mr\omega^2 = \frac{ev \times B}{c} \quad \text{and} \quad v = r\omega \quad \Rightarrow \quad \omega = \frac{eB}{mc}. \tag{87}$$

This formula for the Larmor or cyclotron frequency is valid for non-relativistic speeds. It has the important consequence that the frequency is independent of the radius of the orbit. In the presence of a voltage difference between the D's, a charged particle would feel an electric force when traversing the gap. The voltages on the two electrodes had to be switched periodically (by way of an alternating voltage of frequency equalling ω from a radio frequency oscillator) to ensure that particles were accelerated each time they crossed the gap. Thus the particles spiral outwards, but the time between successive gap traversals remains fixed. The particle beam is extracted from a hole in the rim of one of the Ds.

• By 1931 the Berkeley cyclotron could accelerate protons to 1.22 MeV using a 11 inch magnet. By the end of the 1930s, cyclotrons had grown to 5 feet in diameter, accelerating protons to 8 MeV. By the late 1940s they had reached a few hundred MeV.

3.7 Synchro-cyclotron

• The constancy of the cyclotron frequency $\omega = \frac{eB}{mc}$ with increasing charged particle speed is violated by relativistic effects. As the speed of particles approach c a greater force is needed to produce the same acceleration. The relativistic cyclotron frequency is $\omega = \frac{eB}{\gamma mc}$ where $\gamma = (1 - v^2/c^2)^{-1/2}$. So as the electrons speed up, they take longer to reach the gap between D's during the accelerating part of the RF cycle, and go off resonance. This problem was addressed by decreasing the frequency of the RF voltage applied to the electrodes to keep them in step with the faster particles going round larger spirals. However, the lower energy particles in the center of the cyclotrons would then go off resonance. So a variable frequency RF voltage could not accelerate a continuous beam of particles. So the particles were sent in bunches from the center. A new bunch was injected once the previous one had been accelerated and ejected from the periphery. The resulting device is called the synchro-cyclotron (SC), patented by E McMillan (1945) who worked with Lawrence at Berkeley and constructed the first electron synchro-cyclotron (the Russian V Veksler had published the idea a little earlier). It was used to produce proton beams with kinetic energy of a few hundred MeV. Colliding protons from the Berkeley cyclotron with nuclear targets resulted in production of pions, though these went undetected. Lawrence and his group at Berkeley had a special expertise in building accelerators, but had not devoted as much effort to develop the detectors needed to study particle interactions. Pions were eventually discovered in cosmic ray showers in 1947 by Powell, Lattes, Occhialini and Muirhead at Bristol. They were subsequently identified by the Berkeley group when Lattes joined them. The Berkeley synchrocyclotron was 5m in diameter requiring a very large, heavy and expensive magnet. It was impractical to make synchrocyclotrons with larger magnets needed to go to higher energies.

3.8 Synchrotrons

• The successor to synchrocyclotrons are synchrotrons. The aim was to reach higher energies without the need for magnets of ever increasing diameter. This was achieved by steadily increasing the magnetic field as the particles were accelerated, so that the orbit radius remained fixed with bunches of particles going round in circles rather than spiraling out. Instead of a single big magnet, a ring of magnets surrounded the evacuated beam pipe in a toroidal arrangement. In essence, a synchrotron is a linear accelerator that is bent into a ring using a sequence of bending magnets. An alternating RF voltage is applied at several places around the ring to accelerate the particles. Among the early synchrotrons was the Berkeley Bevatron (1950) which produced 6.5 billion eV protons and the Cosmotron at Brookhaven, New York (3.3 GeV protons in 1952). The first major synchrotron at CERN was the proton synchrotron (PS) which was contemporaneous with the Alternating Gradient Synchrotron at Brookhaven. These accelerated protons to about 30 GeV by 1960 using rings of 200m diameter. In the mid 1970s the super proton synchrotron at CERN and the machine at Fermilab could accelerate protons to 400-500 GeV. The magnetic fields employed in the bending magnets (which produce a uniform vertical field over the width of the beam pipe) are of the order of a Tesla or more, and in each gap between RF cavities, the proton energy is increased by a few MeV. The Fermilab 'Tevatron' accelerated protons to 800 GeV in 1984 and to 980 GeV in 2001.

• Synchrotrons have grown larger in diameter to achieve larger energies and remain the main proton accelerators today. The synchrotron at the Large Hadron Collider in CERN has a 27 km circumference and produced 4 TeV protons in 2012.

• Sometimes, the intensity of the beam produced by an accelerator is not sufficient though the energy may be adequate. A storage ring is just like a synchrotron in that electrons or protons or their anti-particles are kept circulating for several hours using bending and focusing magnets, but with out accelerating them. However, RF accelerating cavities are used to replenish any energy lost to synchrotron radiation. As more particles are fed into the storage ring, the intensity goes up. When the beam flux is adequate, the particles are extracted for collisions. For example at the ISR (Intersecting Storage Rings) at CERN (1971-84), very high flux beams of counter rotating protons were brought to collide at specific locations at a CM energy of 62 GeV. The two rings can contain different/same particles, they need not be anti-particles of each others.

3.9 Power radiated by an accelerated non-relativistic charged particle

• The Lorentz force law governs the non-relativistic motion of a particle of mass m and charge e in an EM field. Let $\mathbf{p} = m\mathbf{v}$ be its momentum, then (in cgs/Gaussian or HL units)

$$\frac{d\mathbf{p}}{dt} = e\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}\right) \tag{88}$$

We may determine the rate of change of kinetic energy $T = \frac{\mathbf{p}^2}{2m}$

$$\dot{T} = \frac{1}{m}\mathbf{p}\cdot\dot{\mathbf{p}} = \frac{e}{m}\mathbf{E}\cdot\mathbf{p} = e\,\mathbf{v}\cdot\mathbf{E}. \tag{89}$$

Electric fields are used to speed up (accelerate) charged particles. The magnetic field does no work, but it can bend the particle trajectory, which is very useful, though it too counts as acceleration!

• A charged particle that is accelerated at $\mathbf{a}$ emits EM radiation. Larmor's non-relativistic formula for the power radiated is (in rationalized HL units, check the dimensions!)

$$P = \frac{2}{3}\frac{e^2}{4\pi}\frac{\mathbf{a}^2}{c^3} = \frac{2}{3}\frac{e^2}{4\pi}\frac{\dot{\mathbf{p}}^2}{m^2c^3}. \tag{90}$$

where $\mathbf{a}$ is the acceleration and $\dot{\mathbf{p}}$ is the rate of change of momentum. For example, electrons oscillate back and forth in an antenna and emit radio waves. There is no power loss in the absence of acceleration. The relativistically covariant generalization is $P = -\frac{2}{3}\frac{e^2}{4\pi m^2 c^3}\frac{dp^\mu}{d\tau}\frac{dp_\mu}{d\tau}$ where τ is the proper time along the trajectory of a massive particle whose 4-momentum is p^μ. We will work with the non-relativistic formula here.

• A charged particle moving in a circular orbit will lose energy due to 'synchrotron' radiation. Let us find the energy loss for a given particle energy E and ring radius r. If the radius of the orbit is r and speed is v, then $a = \frac{v^2}{r}$ and $E = \frac{1}{2}mv^2$

$$P = \frac{8}{3}\frac{e^2}{4\pi}\frac{E^2}{m^2r^2c^3} \tag{91}$$

The power lost in a ring of radius r for fixed particle energy decreases as the square of the particle mass. So synchrotron radiation losses are much more for electrons than for protons of the same energy and ring radius. This is why electron accelerators (e.g. Stanford Linear Accelerator Center, though LEP at CERN is an exception) tend to be linear while proton accelerators are circular (e.g. CERN Proton Synchrotron, Alternating gradient synchrotron at Brookhaven, Fermilab Tevatron and the LHC at CERN). As a consequence, the most advanced synchrotrons built so far (2014) allow us to accelerate protons to 4 TeV (LHC), but electrons only to about 50 GeV (Large Electron Positron Collider, CERN).

• To reduce the losses for a given desired particle energy E, the ring must have as large a radius as possible. The LHC tunnel has a circumference of 27 km. *From these data, estimate the power loss for 7 TeV protons at the LHC. Of course, the Larmor formula is only approximately valid for relativistic protons.*

• If the centripetal acceleration for circular motion is due to the Lorentz force in a constant magnetic field B perpendicular to the plane of motion, then $mv^2/r = evB/c$. The cyclotron frequency $\omega = \frac{v}{r} = \frac{eB}{mc}$ a constant (within the non-relativistic approximation) though r decreases. Then the radiation energy loss is exponential

$$\frac{dE}{dt} = -\frac{4}{3}\frac{e^2}{4\pi}\frac{\omega^2 E}{mc^3}. \tag{92}$$

• The EM radiation produced by charged particles (especially electrons) going round a synchrotron is called synchrotron radiation. Though it results in loss of beam energy, it can also be used as an intense source of synchrotron radiation (X-rays) for further research. Applications include use of the radiation to study the structure of materials and condensed matter as well as medical imaging and X-ray crystallography. Old particle accelerators are often reincarnated at synchrotron light sources. Storage rings also serve as synchrotron light sources. There are several such synchrotron light sources around the world, such as the Australian Synchrotron Facility, Melbourne, Advanced Photon Source at Argonne Lab near Chicago and the European Synchrotron Radiation Facility at Grenoble in France.

3.10 Colliding beam versus fixed-target experiments

• Suppose we wish to produce new particle C of mass m_c through collisions of available particles such as electrons and protons (or secondary beams of neutrinos or pions). One possibility is to collide a beam of particle A of mass m_A and energy E (lab 4-momentum $p_A^\mu = (E_A, \mathbf{p})$) with a stationary target of particles of mass m_B ($p_B^\mu = (m_B, \vec{0})$). To what energy must A be accelerated to be able to create the new particle (or set of particles including the new one) of invariant mass m_C? Conservation of 4-momentum implies $p_C^\mu = p_A^\mu + p_B^\mu = (E_A + m_B, \mathbf{p})$ while $p_C^2 = m_C^2$. Thus we must have

$$p_A^2 + p_B^2 + 2p_A \cdot p_B = m_C^2 \quad \text{or} \quad m_A^2 + m_B^2 + 2E_A m_B = m_C^2. \tag{93}$$

which implies that the projectile energy must be at least

$$E_A = \frac{m_C^2 - m_A^2 - m_B^2}{2m_B} \tag{94}$$

So we see that the required energy grows quadratically with the mass of the new particle. Furthermore, a significant portion of the kinetic energy of the projectile A is simply carried away by C by conservation of momentum, and does not contribute to the mass-energy of C. Typically, C is a heavy new particle and we are limited by the energy E_A to which we can accelerate A. A more efficient use of the beam energy to create new particles is provided by a colliding beam experiment to be described below.

• However, fixed target experiments have played a very important role. They were the only experiments up to the 1960s since it is very difficult to aim two beams at each other and produce a significant number of collisions, it is a lot easier to aim a beam at a large fixed target like a block of iron or liquid hydrogen or heavywater! E.g. the $\bar{p}$ was discovered (by O. Chamberlain and E. Segre, Nobel prize 1959) via pair production at the Berkeley Bevatron in 1955 by colliding 6.5 GeV protons on a stationary liquid hydrogen target via the reaction $p + p \to p + p + p + \bar{p}$. Let us find the minimal energy E of protons to create an antiproton of mass 938 MeV. At the threshold energy E, in the center of momentum frame, the initial momentum is zero and the four particles in the final state must be produced at rest, so that all the projectile energy is used for particle production. So the total CM 4-momentum in the final state is $p_{cm}^\mu = (4m, \vec{0})$, which must equal the initial momentum by momentum conservation. The total lab frame 4-momentum in the initial state is $p_{lab}^\mu = (E + m, \mathbf{p})$. The initial momenta in the two frames are related by a Lorentz boost, but must have the same invariant length-squared $(E + m)^2 - \mathbf{p}^2 = 16m^2$. Moreover $E^2 - \mathbf{p}^2 = m^2$. We find $E = 7m \approx 6.5$ GeV. (Actually, the protons in the target are not quite at rest, they have some so-called Fermi motion in the nuclear environment, so an energy of slightly less than $7m$ is adequate for producing anti-protons. However, this is only the minimum/threshold energy, one may require higher energies to achieve a desired event rate.)

• There are several reasons for the importance of fixed target experiments. Cosmic rays are freely available highly energetic particles raining down on the Earth. Cosmic rays colliding with a fixed target in a cloud or bubble chamber were used to discover the e^+, charged π, K mesons and other strange particles in the 1950s. But we have no way of colliding cosmic rays with each other. The Pierre Auger project is a 'fixed target' cosmic ray 'observatory' currently operating over 3000 square km in Argentina! Neutrinos interact only weakly, they have a very low cross section for interacting with other particles. All neutrino experiments (e.g. Borexino currently detecting solar neutrinos in Italy) have used large fixed targets of liquid, iron etc.

• The first colliders were built only in the 1960s, but they have largely replaced fixed target experiments outside of cosmic ray and neutrino experiments. Rolf Wideroe who developed the first linear accelerator also had the idea for a collider in 1943. His collaborator Bruno Touschek (who survived being shot and left to die during the second world war!) was the first to propose a collider that used a single storage ring for two counter propagating beams of oppositely charged particles. He built the first collider in Frascati near Rome, in 1961, colliding 250 MeV electrons with positrons.

• In the simplest colliding beam experiments, identical particles (or anti-particles, $e{+}e^-, p\bar{p}, pp$) with the same energy E are made to collide head-on, so that the center of momentum frame coincides with the lab frame. The total initial 4-momentum is $p^\mu = (2E, \vec{0})$. The final state as a whole is at rest in the lab frame. The particles produced in the final state are not required to carry away any momentum (though they often do, flying off in all directions). So the entire center of mass energy $2E$ is in principle available for producing new particles.

• Consider the general case of a colliding beam experiment, where particle a with 4-momentum $p_a = (E_a, \mathbf{p}_a)$ collides with particle b with $p_b = (E_b, \mathbf{p}_b)$. Let us work in the center of momentum frame, in which the total 3-momentum is zero (both before and after the collision, indeed at all times). The final state is at rest in the CM frame, so the final 4-momentum is $p_f = (E, 0)$. The energy E of the final state in the center of momentum frame is the maximum energy that is available for production of new particles. Thus we have

$$p_f^2 = E^2 = p_i^2 = (p_a + p_b)^2 = p_a^2 + p_b^2 + 2p_a \cdot p_b = m_a^2 + m_b^2 + 2(E_a E_b - \mathbf{p}_a \cdot \mathbf{p}_b). \quad (95)$$

Suppose the initial 3-momenta $\mathbf{p}_a$ and $\mathbf{p}_b$ are at an angle θ and that the energies $E_{a,b} \gg m_{a,b}$ are significantly more than rest masses. Then $|\mathbf{p}_{a,b}| \approx E_{a,b}$. Then the CM energy is

$$E^2 \approx 2E_a E_b (1 - \cos\theta) \quad (96)$$

So to maximize the energy available for creation of new particles, we must have $\theta = \pi$ (head on collision), in which case, $E_{cm} \approx \sqrt{4E_a E_b}$. In particular, if the colliding particles have the same energies $E_a = E_b$, then $E_{cm} \sim 2E_a$ (as obtained previously). E_{cm} is the energy available for production of new particles in the final state, which is produced at rest in the CM frame.

• On the other hand, in the case of a fixed target experiment, with particle b at rest in the lab ($\mathbf{p}_b = 0$), the CM energy is

$$E_{cm}^2 = m_a^2 + m_b^2 + 2m_b E_a. \quad (97)$$

And again, if the rest masses are small compared to E_a, then the CM energy available in a fixed target experiment is $E_{cm} \approx \sqrt{2m_b E_a}$ (this is the invariant mass of the final state, which is the maximum possible mass of a particle created in the collision). So the available cm energy in a colliding beam experiment grows linearly with the beam energy while it grows as the square root of the beam energy in a fixed target experiment.

• The (design) CM energy in pp collisions (7 TeV on 7 TeV) at the LHC is 14 TeV. If we wanted this CM energy to be achieved via fixed target collisions with a proton, then what beam energy would we need? From the answer, it should be clear that with current beam energies, the only way to achieve a CM energy of 14 TeV is via colliding beams. The drawback of colliding beam experiments is that beam flux (and so event rates) are much lower than in fixed target experiments (with the same CM energy), the latter benefit from the availability of dense target materials.

3.11 Asymmetric colliders

• The advantage of a symmetric collider (say head-on collision of e^+e^- beams of equal energy) over a fixed-target experiment is that it reduces the beam energy required to produce a new particle of given mass. This is because no energy is unnecessarily 'wasted' in kinetic energy of the final state particles: they can even be produced at rest in the lab since the initial total 3-momentum is zero. Typically, in a colliding beam experiment, the final state particles fly out in all directions from the collision vertex. On the other hand, in a fixed target experiment, the final products typically move in the direction of the original beam with significant momentum. This can be advantageous in special circumstances. If the particles produced are very short-lived, then they may decay before they are detected. On the other hand, the lab-frame mean lifetime of a particle grows with its speed $\tau' = \gamma\tau$. So we may lengthen the tracks of short-lived particles by producing them with significant momentum. This is what is done at asymmetric colliders, which are a cross between symmetric colliding beam and fixed target set-ups. For example, at beauty-factories (BaBar at PEP2 SLAC Stanford during 1999-2008 and Belle at KEKB, Tsukuba Japan), B mesons $\Upsilon(4S)$ containing $b, \bar{b}$ quarks are produced in $e^+ \; e^-$ annihilations. One requires about 10 GeV CM energy to produce these mesons since $m_b \sim 4.2$ GeV. However, instead of colliding electrons and positrons of equal and opposite momenta, one collides 9 GeV electrons of momentum $\mathbf{p}$ (what is $|\mathbf{p}|$?) with 3.1 GeV positrons of momentum $\mathbf{p}'$ (what is $|\mathbf{p}'|$?) where $\mathbf{p}$ and $\mathbf{p}'$ are oppositely directed but of different magnitudes What is the center of momentum energy? As a consequence the final products as a whole move in the direction of $\mathbf{p} - \mathbf{p}'$. The B^0 and $\bar{B}^0$ mesons produced in the decay of Υ live longer than if they were produced at rest. Using the accurate silicon vertex detector it was possible to study the weak decays of B^0 and $\bar{B}^0$ and learn about CP violation in the weak interactions. The results confirm the CKM theory of mixing of quark families.

• Another example of an asymmetric collider is the electron proton collider at HERA (Hadron Elektron Ring Anlage) at DESY (Deutsches Elektronen-Synchrotron) in Hamburg. ep deep inelastic scattering experiments have been used to probe distances of a tenth of an attometer deep inside the proton. The measurements confirm perturbative QCD calculations and also determine non-perturbative aspects of the distribution of quarks and gluons in a nucleon that have not yet been theoretically understood.

3.12 Focusing the beam using quadrupole magnets

• Charged particles in a synchrotron beam repel each other causing the beam to spread out. They are also subject to imperfections in the bending magnets etc so the particles do not travel in circular orbits of a common radius, but suffer so-called betatron oscillations about a circular path. A typical proton may make a hundred thousand revolutions around the ring, covering nearly a million kilometers, so beam stability is essential. The beam consists of bunches of, say 10^12 protons as at the LHC. The bunches are separated and do not significantly affect each other. However, there is a tendency for a given bunch to spread out both laterally (in the plane perpendicular to the beam pipe) and longitudinally.

• To get a high intensity proton beam, it is focussed laterally, essentially by using magnetic analogues of converging and diverging lenses. These are ‘focusing magnets’, with appropriate magnetic field gradients. The method is called strong focusing, it has been an important advance in the development of particle accelerators. Suppose the beam travels into the plane of the paper. An F-type quadrupole magnet focuses the beam in the horizontal direction while defocusing in the vertical direction. A D-type quadrupole magnet defocuses in the horizontal direction while focusing in the vertical direction. By placing F- and D-type quadrupoles at suitable intervals, it is possible to produce net focusing both horizontally and vertically. A quadrupole magnet may be visualized as a system of 4 bar-magnets (coil electromagnets with an iron core) lying in the plane of the paper, at right angles to each other, with S and N poles alternating $S(NorthEast)N(SE)S(SouthWest)N(NW)$ as one goes around the beam pipe. The magnetic field is zero at the center but increases outwards, resulting in a gradient. A D-type quadrupole is an F-type quadrupole rotated by 90 degrees in the plane of the paper. The arrangement results in an alternating gradient synchrotron (AGS). Draw the field lines of such a quadrupole magnet and demonstrate its focusing and defocusing properties. The focusing and defocusing effects are greater in the periphery of the beam pipe (near the pole pieces of the magnets) than at the center, due to the field gradient. So having been focussed horizontally by an F quadrupole, the horizontal defocusing due to the next D quadrupole does not entirely undo the effect of the previous F magnet. Thus it is possible to obtain focusing in both directions using a succession of alternating gradient quadrupole magnets!

• The resonance acceleration principle using RF cavities has an inbuilt tendency towards compressing bunches longitudinally. The synchrotron is set up to optimally accelerate particles that emerge from an RF cavity when the oscillating voltage has grown, say, half-way to its maximum. Such particles receive the optimal kick and stay on resonance. Particles that are going too fast arrive early, see a smaller voltage, and receive a smaller kick. Particles that are going a bit too slow arrive late, by which time the voltage has grown. They get a greater kick which pushes them towards the optimal position in the middle of a bunch.

3.13 Secondary and tertiary beams from fixed targets

• Bunches of protons accelerated at a synchrotron can be extracted and aimed at a target, often beryllium, copper or carbon. Collisions with nuclei produce many particles, of which pions and kaons are plentiful since they are the lightest hadrons; they are of particular interest. For example, at the LHC in 2012 there were about 10^{11} 4 TeV protons in each bunch and 10^7 bunches per second. At the proton synchrotron at CERN in 1963, there were 10^{12} 25 GeV protons every three seconds. The point is that there are sufficiently many protons to produce large numbers of pions and kaons (more pions than kaons). However, pions and kaons are unstable. Can we make secondary beams of them? If they are sufficiently energetic and moving at relativistic speeds, then time dilation allows us to do so.

$$m_{\pi^\pm} = 139.6\, MeV, \quad m_{\pi^0} = 135\, MeV, \quad m_{K^\pm} = 493.7\, MeV, \quad m_{K^0} = 497.6\, MeV. \quad (98)$$

• The mean lifetime of charged pions is about $\tau_{\pi^\pm} = 26$ns, with the primary decay mode being $\pi^+ \to \mu^+ + \bar{\nu}_\mu$. A charged pion moving at speed v has an energy $E = \gamma mc^2$. Viewed from the lab frame, it has a mean life time of $\gamma\tau$. Thus a relativistic pion of energy say 14 GeV has $\gamma = 100$ and lives on average 2.6μs. In this time it can travel on average $2.6 \times 10^{-6} \times 3 \times 10^8 = 780$m. This is enough to make secondary beams of pions. Magnets may be used (like a prism for bending light) to bend the charged pion trajectories. More energetic pions bend less so we may make secondary beams with specified energy (like selecting light of specific wavelength using a prism).

• Neutral pions (which are a linear combination of $u\bar{u}$ and $d\bar{d}$) have a mean life-time of 10^{-17}s ($\pi^0 \to 2\gamma$), which is too short for producing a secondary beam.

• Charged kaons $K^\pm$ have a mean life-time 12ns, roughly half that of charged pions. $K^+ \to \mu^+\bar{\nu}_\mu$. Thus it is possible to prepare secondary beams of charged kaons as well. A 5 GeV beam of kaons has $\gamma \approx 10$ and will travel on average 36m. This is adequate for secondary beams, as were produced using the main injector at Fermilab with 120 GeV protons.

• Due to mixing, the neutral kaons K^0 and $\bar{K}^0$ do not have definite lifetimes. The point is that they are strangeness eigenstates, being produced via the strong interactions in proton nucleus or pion nucleon collisions. But they decay via the weak interactions. The states with definite life-times are a long lived K-long K_L ($\tau_{K_L} = 50$ns. $K_L \to 3\pi$) and a short lived K-short K_S ($\tau_{K_S} = 90$ps, $K_s \to 2\pi$). So K-long lives longer than charged pions and one can form secondary beams of K_L, though one cannot control them with magnetic fields as they are neutral. Remarkably, it is possible to 'generate' K_S in a beam of K_L by passing it through matter, as we shall see shortly.

• Neutral kaons are of interest in studying CKM mixing and CP violation. Charged kaon beams have been used to produce other hadrons (especially strange particles) in collisions with nuclei in bubble chambers in the 1960s and subsequently. The KTeV experiment at Fermilab used secondary beams of kaons.

• Charged pions and kaons decay producing neutrinos (more muon neutrinos than electron neutrinos). Neutrinos interact very weakly and pass through shielding that absorbs all the other decay products. Thus, one obtains tertiary beams of mostly $\nu_\mu, \bar{\nu}_\mu$. It is possible to select neutrinos from anti-neutrinos by selecting the charge of the parent pions and kaons using bending magnets. By colliding the neutrinos with a fixed target detector, one obtains neutrino-nucleon deep inelastic scattering events, which probe the structure of the proton and also study the weak interactions. Nuclear fission reactors are sources of $\bar{\nu}_e$ from beta decay of neutrons.

4 Neutral Kaon oscillations/mixing and regeneration

• We have mentioned 2 body scattering and decays to two bodies. Do 1-body decays exist? If so, the parent and daughter must have the same mass, so one must look for particles with the same mass that can make a transition ('oscillation') without the net emission/absorption of any other particle. Particle-anti-particle pairs have degenerate masses and are obvious candidates for oscillations. But for this to happen, they must also have the same electric charge and baryon number and other exactly conserved quantum numbers. Examples include the neutral kaons $K^0 = d\bar{s}, \bar{K}^0 = s\bar{d}$, neutral charmed mesons $D^0 = u\bar{c}, \bar{D}^0 = c\bar{u}$, neutral bottom mesons $B^0 = d\bar{b}, \bar{B}^0 = b\bar{d}$ and neutral B_s-mesons $B_s^0 = s\bar{b}, \bar{B}_s^0 = b\bar{s}$. On the other hand, we cannot have oscillations between the degenerate but oppositely charged kaons $K^{\pm}$, neutron and anti-neutron or Λ^0 and $\bar{\Lambda}^0$. To study the $K^0 - \bar{K}^0$ system, it helps to have the concept of charge conjugation at our disposal.

4.1 Charge conjugation and C-invariance

• The operation of replacing a particle with its anti-particle is called charge conjugation C. Acting on a one-particle state, it reverses all the internal quantum numbers, both those associated to gauge symmetries and global symmetries. In particular, the electric charge Q, magnetic moment, color, weak hypercharge Y_w and weak isospin T_3 are reversed. So are the baryon number, lepton number, isospin projection I_3, strangeness, charm, beauty, topness etc. Space-time coordinates, spin, angular momentum and helicity are unchanged. E.g. $\pi^{\pm}$ are related by charge conjugation, as are $K^{\pm}$.

• Maxwell's equations of classical EM are C-invariant in the sense that they are unchanged if we reverse the signs of $\rho, \vec{j}, \vec{E}, \vec{B}, A_\mu$. EM and Strong interactions are empirically found to be symmetric under C (Charge conjugation invariance was first proposed by Kramers.). This means the rate of a (strong/EM) reaction is is same as the rate of the reaction that results from replacing all particles by their anti-particles. The conjugate reaction may happen to be the same reaction. For instance, the rate of π^+ and π^- production in $p\bar{p}$ annihilation $p + \bar{p} \to \pi^+ + \pi^- + \cdots$ are equal to better than a percent, the same applies to kaons in $p\bar{p} \to K^+K^- + \cdots$ (see Perkins). There are more stringent tests of C-invariance.

• $C^2 = I$, so the eigenvalues of C are ± 1, called C-parity or charge conjugation. The charge conjugation operator C, when acting on a $|\pi^+\rangle$ state produces a state of a negative pion: $C|\pi^+\rangle \propto |\pi^-\rangle$. In fact, one may choose phases so that

$$C|\pi^+\rangle = |\pi^-\rangle \quad \text{and} \quad C|\pi^-\rangle = |\pi^+\rangle. \tag{99}$$

Similarly, we may take $C|K^{\pm}\rangle = |K^{\mp}\rangle$, $C|K^0\rangle = |\bar{K}^0\rangle$ and $C|\bar{K}^0\rangle = |K^0\rangle$. A 1-particle state can have definite C-parity only if the particle is its own anti-particle, like the photon, Z^0, π^0, $\eta, \eta', J/\psi, \Upsilon$ etc. Reversal of sign of A_μ under charge conjugation means that a photon has negative C-parity $C|\gamma\rangle = -|\gamma\rangle$. Like parity, C-parity is a multiplicative quantum number so for a multi-photon state $C|n\gamma\rangle = (-1)^n|n\gamma\rangle$. The C-conserving EM decay of the neutral pion to two photons $\pi^0 \to 2\gamma$ implies that it has even C-parity: $C|\pi^0\rangle = |\pi^0\rangle$. C-invariance of EM would then forbid $\pi^0 \to 3\gamma$, and indeed this decay has not been seen. The upper limit on the branching ratio is 3.1×10^{-8}. Another even C-parity particle is the neutral η meson $(u\bar{u} - d\bar{d} - 2s\bar{s})$. Its decays too preserve C-parity (e.g. $\eta \to 2\gamma$, $\eta \to 3\pi^0$).

• C-invariance is broken (in a sense maximally) by the weak interactions. Under C, a left-handed neutrino is mapped to a left-handed anti-neutrino. The latter does not exist in the standard model and has not been seen. So it is not even possible to compare the rate of LH neutrino production to LH anti-neutrino production in weak interaction reactions that would otherwise be charge conjugates of each other. Similarly, the charge conjugate of a LH electron is a LH positron. Though both exist, they have different weak interactions: the LH electron transforms as a doublet under weak isospin while the LH positron is a singlet under weak isospin.

4.2 Neutral Kaons: production and strong interactions

• Recall that the kaons form two $I = \frac{1}{2}$ doublets, the $S = 1$ doublet $(K^+ = u\bar{s}, K^0 = d\bar{s})$ and the $S = -1$ pair $(\bar{K}^0 = \bar{d}s, K^- = \bar{u}s)$. The neutral kaons are anti-particles of each other and form a very interesting '2-state' system first studied by Gell-Mann and Pais in the mid 1950s. Feynman includes a discussion in his lectures on physics. Neutral kaons are produced in strong interactions.

• E.g., K^0 are produced *in association* with Λ^0 baryons ($S = -1$) when a pion beam strikes a proton target

$$\pi^-(d\bar{u}) + p(uud) \to \Lambda^0(dus) + K^0(\bar{s}d) \tag{100}$$

or in association with a Σ^+ baryon ($S = -1$) when two protons collide

$$p + p \to p(uud) + \Sigma^+(uus) + K^0(\bar{s}d). \tag{101}$$

A point to notice here is that to conserve baryon number, a strange baryon (which necessarily has negative strangeness), is produced in association with a K^0 and never a $\bar{K}^0$. So if a Λ^0 is produced in πp collisions at not too high energies, we can be sure that the resulting neutral kaon is not a $\bar{K}^0$.

• However $\bar{K}^0$ can be produced in association with K^+ mesons in $\pi^+ p$ collisions:

$$\pi^+(u\bar{d}) + p(duu) \to K^+(u\bar{s}) + \bar{K}^0(s\bar{d}) + p(duu) \tag{102}$$

The threshold pion energy for this reaction is 1.5 GeV while that for production of K^0 in association with Λ^0 above is $.91$ GeV. So one may produce a pure K^0 beam by choosing the charged pion beam energy appropriately. The point is that K^0 and $\bar{K}^0$ are not the same particle, and it is possible to choose reactions where only one is produced, and make a beam of those neutral kaons.

• It is also possible to produce $\bar{K}^0$ in association with K^+ $(S = +1)$ by scattering protons off neutrons $p(uud) + n(udd) \to n(udd) + n(udd) + \bar{K}^0(\bar{d}s) + K^+(\bar{s}u)$.

• Another way of making neutral kaons is by aiming a secondary beam of charged kaons at a target containing nucleons:

$$K^+ + n \to K^0 + p, \quad \text{or} \quad K^-(s\bar{u}) + p(uud) \to \bar{K}^0(s\bar{d}) + n(duu). \tag{103}$$

By choosing the sign of the charged kaon beam, we can select neutral kaons with a specified strangeness.

• Neutral kaons can also be produced in association with charged kaons in $p\bar{p}$ annihilation reactions

$$p\bar{p} \to K^-\pi^+K^0 \quad \text{and} \quad p\bar{p} \to K^+\pi^-\bar{K}^0 \tag{104}$$

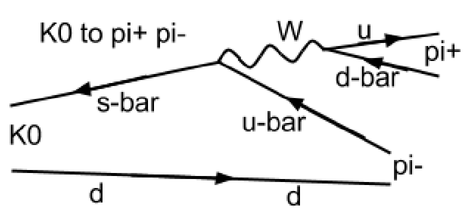
Which neutral kaon is produced can be determined from the charge of the pion/kaon co-produced.

• Neutral kaons interact differently with matter, depending on their strangeness. $\bar{K}^0$ are seen to produce the $S = -1$ Λ^0 baryons via $\bar{K}^0 p \to \Lambda^0\pi^+$. But K^0 does not produce Λ^0 upon striking protons unless the energy is so high as to allow reactions such as $K^0 p \to \Lambda^0 K^+ K^+ \pi^-$.

• A K^0 can 'exchange charge' with a proton, $K^0 + p \to K^+ + n$ while $\bar{K}^0$ cannot do so. On the other hand, $\bar{K}^0$ can produce the strange baryon $\Sigma^+(uus)$ upon colliding with a proton, $\bar{K}^0 + p \to \Sigma^+ + \pi^0$ but the $S = 1$ K^0 cannot produce a strange baryon in such a collision. Thus it is possible to detect the presence of neutral kaons of specified strangeness by examining how a beam of neutral kaons interacts with matter.

4.3 Neutral kaons: weak interactions, decays and CP eigenstates

• K^0 and $\bar{K}^0$ are stationary states of the strong interaction (QCD) hamiltonian H_{strong}, they are states of definite strangeness and mass (same mass 497.7 MeV) if we ignore the weak interactions. Such an approximation holds for 10^{-23} s during the strangeness-conserving production process via the strong interactions (due to the proximity of other strongly interacting particles p, π, Λ). When the kaons propagate after they are produced, the hamiltonian that governs their time evolution is in principle the sum of the strong, weak and EM hamiltonians. The strong interactions are responsible for binding the kaons while the weak interactions are responsible for their non-leptonic decay to pions or semi-leptonic decay to pions, leptons and neutrinos. Unlike neutral pions, neutral kaons cannot decay to photons or pions electromagnetically since EM conserves strangeness, so we can ignore EM. If we include the weak interactions, $K^0, \bar{K}^0$ are not stationary states, they do not have definite masses or life-times. They decay in strangeness changing weak interactions after a long time (10^{-8} - 10^{-10} s), so we can make secondary beams of neutral kaons. A Feynman diagram for the decay $K^0 \to \pi^+\pi^-$ is shown[24]. Draw similar diagrams for $\bar{K}^0 \to \pi^+\pi^-, \pi^0\pi^0$.



• It was discovered (theoretically predicted in 1955 and then experimentally discovered in 1960) that a pure K^0 beam, upon propagating in vacuum could turn into a mixed beam of K^0 and $\bar{K}^0$ with their numbers oscillating in time. This is called $K^0\bar{K}^0$ mixing or strangeness

[24]Though we can draw such diagrams, the particle that decays to $\pi^+\pi^-$ with a definite lifetime is neither K^0 not $\bar{K}^0$, but a certain linear combination that is to be determined.

oscillations. This could be detected since K^0 and $\bar{K}^0$ react differently when passing through matter, as explained above.

• K^0 and $\bar{K}^0$ are not stationary states (i.e. of definite mass and life-time; the combination of mass and lifetime $m - \frac{i}{2\tau}$ can be regarded as a complex energy eigenvalue corresponding to a stationary state of an unstable particle in its rest frame) if one includes the weak interactions $H_{\text{total}} = H_{\text{strong}} + H_{\text{weak}}$. Usually, the weak interactions may be ignored in comparison to the strong interactions. But when there are degeneracies, a small perturbation can alter the stationary states significantly (as we will see in the example below). Since charge changing weak interactions can change strangeness, the new states of definite mass (and life-time) need not have definite strangeness. Moreover, weak interactions can cause transitions (or oscillations) between K^0 and $\bar{K}^0$, which are forbidden by S-conserving strong interactions. From a study of weak decays, it is found that there are two neutral kaons with definite masses and life-times, K_{short} and K_{long} with $\tau_S = .9 \times 10^{-10}$ s and $\tau_L = .5 \times 10^{-7}$ s. They decay primarily to two and three pions, $K_S \to \pi^+\pi^-, \pi^0\pi^0$ and $K_L \to 3\pi^0, \pi^+\pi^-\pi^0$. The phase space for three body decay is smaller than for two body decay; it has a smaller Q value, which explains why K_L lives longer.

• To begin to understand all this, following Gell-Mann and Pais (1955), let us model the $K^0 - \bar{K}^0$ Hilbert space as $\mathbb{C}^2 = \text{span}(|K^0\rangle, |\bar{K}^0\rangle)$ and represent $H_{\text{strong}} = \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}$ where $E = 497.7$ MeV. In other words, we are ignoring the translational degrees of freedom and focusing on the two dimensional space labelled by strangeness of the neutral kaons.

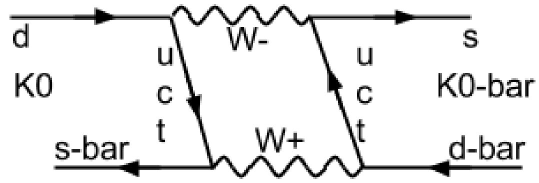
• For simplicity, we model the perturbing hamiltonian due to the weak interaction by the purely off-diagonal (off-diagonal matrix elements are the ones that cause transitions) real matrix $H_{\text{weak}} = \begin{pmatrix} 0 & w \\ w & 0 \end{pmatrix}$ with $w \ll E$. It follows that the perturbed eigenstates are

$$K_1 = \frac{1}{\sqrt{2}}\left(K^0 + \bar{K}^0\right) \quad \text{and} \quad K_2 = \frac{1}{\sqrt{2}}\left(K^0 - \bar{K}^0\right) \quad \text{with energies} \quad E \pm w. \tag{105}$$

If $H_{\text{weak}} = \begin{pmatrix} 0 & w \\ w^* & 0 \end{pmatrix}$ were hermitian, off-diagonal but not real, the perturbed energies would be $E \pm |w|$, corresponding to the eigenstates $\frac{1}{\sqrt{2}}(e^{i\arg(w)}K^0 \pm \bar{K}^0)$.

• By transitions between K^0 and $\bar{K}^0$ we mean reactions of the sort $K^0 \leftrightarrow \bar{K}^0$. This could happen, for instance, through an intermediate state that both K^0 and $\bar{K}^0$ can decay to, such as $\pi^+\pi^-$ or $\pi^+\pi^-\pi^0$. Thus mixing can take place through virtual intermediate pion states like $K^0 \to \pi^+\pi^- \to \bar{K}^0$. Such a $|\Delta S| = 2$ transition must involve two charge changing weak currents, i.e. two W bosons exchanged (neutral current Z^0 exchange does not change strangeness, nor would gluon or photon exchange). An example is a 1-loop 'box' Feynman diagram with virtual up type quarks and anti-quarks shown below.

• The true neutral kaon eigenstates K_1, K_2 of H_{total} are some linear combinations of K^0 and $\bar{K}^0$. We can find them without knowing the explicit form of the hamiltonian if we assume that it commutes with CP. Unlike P and C which are individually *not conserved* in weak interactions, CP is a symmetry of the strong interactions and very nearly (to about one part in a



thousand) a symmetry of the weak interactions. We will work in the approximation where CP is conserved, so the eigenstates of H are also CP eigenstates. The hamiltonian also possesses other symmetries like rotations and translations, but these act trivially within the two dimensional subspace under consideration, the corresponding symmetry generators are multiples of the identity and do not allow us to identify the states with definite lifetime.

• Now, being anti-particles, charge conjugation takes a one-K^0 state to a one-$\bar{K}^0$ state and vice-versa. For some phase ($|\alpha| = 1$),

$$C|K^0\rangle = \alpha|\bar{K}^0\rangle \quad \text{and} \quad C|\bar{K}^0\rangle = (1/\alpha)|K^0\rangle \quad \text{since} \quad C^2 = I. \tag{106}$$

It is possible to absorb the phase α into the state vector $|\bar{K}^0\rangle$ so that $C|K^0\rangle = |\bar{K}^0\rangle$ and $C|\bar{K}^0\rangle = |K^0\rangle$. Now kaons all have negative intrinsic parity (they are part of the pseudo-scalar meson octet with the pions) $P|K\rangle = -|K\rangle$. So $PC|K^0\rangle = -|\bar{K}^0\rangle$ and $PC|\bar{K}^0\rangle = -|K^0\rangle$. It is conventional to absorb the minus sign into the state vector $|\bar{K}^0\rangle$ to get (note that $CP = PC$ since C does not affect the spatial or spin degrees of freedom upon which P acts)

$$CP|K^0\rangle = |\bar{K}^0\rangle \quad \text{and} \quad CP|\bar{K}^0\rangle = |K^0\rangle. \tag{107}$$

So the CP matrix is simply the Pauli matrix σ_1 within this subspace, and its eigenstates are the previously introduced combinations K^0 and $\bar{K}^0$. Thus the strangeness eigenstates are not CP eigenstates, but the sum and difference $K_{1,2}$ are CP even and odd

$$|K_1\rangle = \frac{1}{\sqrt{2}}(|K^0\rangle + |\bar{K}^0\rangle), \;\; CP = +1, \quad \text{and} \quad |K_2\rangle = \frac{1}{\sqrt{2}}(|K^0\rangle - |\bar{K}^0\rangle), \;\; CP = -1. \tag{108}$$

To the extent that CP is a symmetry of the weak interactions K_1 and K_2 should be the states with definite masses and life-times.

• CP conserving weak interactions would allow $K_{1,2}$ to decay only into CP-even and CP-odd final states. Neutral kaons are seen to decay weakly to two pion ($\pi^0\pi^0$ or $\pi^+\pi^-$) and three-pion ($\pi^0\pi^0\pi^0$ or $\pi^+\pi^-\pi^0$) final states.

• Let us argue that the two pion final states are CP even[25]. Since kaons have no angular momentum (spin zero) the 2π final state does not have any orbital angular momentum, so its extrinsic parity $(-1)^l = 1$. The odd intrinsic parities of the pions combine to make $\pi\pi$ a parity even

[25]More generally, consider the two pion final state $|\pi^+\pi^-\rangle$. Pions have spin zero each so we may ignore spin. Suppose the pair has orbital angular momentum l. The intrinsic parities are both odd and multiply to give one. The extrinsic parity is $(-1)^l$. The effect of C is simply to exchange the pions, which is the same as reflection through the origin, under which $|\pi^+\pi^-\rangle = (-1)^l|\pi^+\pi^-\rangle$. So the C-parity is $(-1)^l$. Thus the CP eigenvalue is $(-1)^{2l} = 1$ irrespective of the value of l. Of course $l = 0$ since the decaying kaon was spin less.

state. C does not affect the $2\pi^0$ state while it exchanges the $\pi^\pm$, though this leaves an S-wave state unchanged. Thus the $\pi\pi$ final states have $CP = +1$ and only K_1 can decay to $\pi\pi$. Note that the Q value for 2-pion decays is $498 - 140 - 140 \approx 218$ MeV.

• The $\pi\pi\pi$ final states relevant to neutral kaon decay can be shown to be CP odd, so only K_2 can decay 3π while conserving CP[26]. Irrespective of this, the Q value ≈ 78 MeV for 3 pion decay is significantly less than for 2π decay. So K_1 which decays to two pions has a much shorter lifetime $\tau_1 \approx .9 \times 10^{-10}$s than (K_2) that decays to 3π in $\tau_2 \approx .5 \times 10^{-7}$s. $\tau_2/\tau_1 \approx 570$. The energy widths are $\Gamma_1 = 7.4\ \mu$eV and $\Gamma_2 = 0.013\ \mu$eV. These widths are very narrow compared to the masses (≈ 497.6 MeV), so the neutral kaons are much more stable compared to the hadronic resonances like the Δ that decay via the strong interactions.

• It turns out that CP isn't exactly conserved in weak decays[27], so the states of definite mass and lifetime aren't exactly CP eigenstates but a pair of short and long-lived neutral kaons K_S and K_L. However, $K_S \approx K_1$ and $K_L \approx K_2$, so we shall ignore the effects of CP violation in this discussion and use the symbols K_S and K_1 interchangeably as also K_L and K_2.

4.4 Neutral kaons: strangeness oscillations

• H_{weak} breaks the degeneracy in masses, so $K_{1,2}$ have slightly different masses $m_{1,2}$ ($\Delta m/m_{K^0} \approx 7 \times 10^{-15}$); they also have significantly different mean life times $\tau_{1,2}$ (or energy widths $\Gamma_{1,2} = \hbar/\tau_{1,2}$). The fact that the strangeness eigenstates $K^0, \bar{K}^0$ differ from the mass eigenstates leads to the phenomenon of strangeness oscillations, predicted by Gell-Mann and Pais in 1955 and experimentally detected in 1960.

• Suppose we start off with a neutral kaon produced in a strong interaction, it must have definite strangeness. For definiteness, if it was produced in $\pi^- p$ collisions in association with Λ^0, then it must have $S = 1$ and be a K^0. However, K^0 is not an eigenstate of H_{tot} and to examine its time evolution as it propagates (as part of a beam) we decompose it in terms of the CP eigenstates $K_{1,2}$. At $t = 0$ we have

$$|K^0(0)\rangle = \frac{1}{\sqrt{2}}\left(|K_1(0)\rangle + |K_2(0)\rangle\right) \quad \text{where} \quad K_1(0) \equiv K_1 \quad \text{and} \quad K_2(0) \equiv K_2. \qquad (109)$$

[26] Consider the $3\pi^0$ final state from neutral kaon decay, we will show that it is CP odd. First, C leaves each π^0 invariant, so we only need to determine the parity of this $3\pi^0$ state. Each pion has odd intrinsic parity, so the product of intrinsic parities is -1. It remains to consider the extrinsic parity from orbital motion. The $3\pi^0$ state must have zero total angular momentum J since the original kaon is spin-less (pseudo scalar). Now suppose the orbital angular momentum of one pair of pions about their CM is $\mathbf{l}$ and that of the remaining pion about the CM of the other two is $\mathbf{l}'$. Then $\mathbf{J} = \mathbf{l} + \mathbf{l}'$. But since $J = 0$, $l = l'$. What is more, since the pions are identical bosons, any two pions must be in a symmetric state under exchange, which means l is even. So both l and l' are even, resulting in even extrinsic parities $(-1)^l, (-1)^{l'}$ (in any case, the product of the extrinsic parities will be $+1$ since $l = l'$). In summary, the CP eigenvalue of a $J = 0$ $3\pi^0$ state is simply the product of intrinsic parities, and therefore $CP = -1$. The $\pi^+\pi^-\pi^0$ state can be CP odd or even, depending on the relative angular momentum l of the $\pi^+\pi^-$ pair. However, the small Q value ≈ 80 MeV of the decay, suggests $l = 0$. Now the $\pi^+\pi^-$ pair is CP even, as argued earlier for $K_S \to 2\pi$. Combining this with the CP odd π^0 $(C = 1, P = -1)$ we conclude that the $\pi^+\pi^-\pi^0$ system is CP odd.

[27] A manifestation of CP violation is the very rare weak decay of K_L to two pions, the amplitude for this is about 500 times less than the amplitude for $K_S \to \pi\pi$

In their rest frame, the amplitudes for the CP eigenstates have a time dependence determined by their masses and life-times

$$|K_1(t)\rangle = e^{-im_1 t - t/2\tau_1}|K_1(0)\rangle \quad \text{and} \quad |K_2(t)\rangle = e^{-im_2 t - t/2\tau_2}|K_2(0)\rangle. \tag{110}$$

Thus the time evolution of our initial K^0 is a superposition of a pair of oscillatory amplitudes, each with its own exponential damping.

$$|K^0(t)\rangle = \frac{1}{\sqrt{2}}\left(e^{-im_1 t - t/2\tau_1}|K_1(0)\rangle + e^{-im_2 t - t/2\tau_2}|K_2(0)\rangle\right) \tag{111}$$

Production of K^0 via the strong interactions happens in 10^{-23} seconds[28]. For a long time after this, i.e. till $t \lesssim \tau_1 \approx .9 \times 10^{-10}$s both K_1 and K_2 are present in the beam. As t approaches τ_1 one observes decays to two pions arising from the CP even K_1 component. When $\tau_2 \gtrsim t \gg \tau_1$, the K_1 component would have decayed and one would primarily see decays to three pions. This is in the rest frame of the K^0. If it were moving at a speed v, relative to the lab frame, then the masses and life-times must be scaled up by γ. In practice, the 2π decays are seen closer to the K^0 production region and the 3π decays at greater distances downstream since τ_2 is about 570 times larger than τ_1. In fact, $c\tau_1 = 2.67$ cm while $c\tau_2 = 15.5$m. A neutral kaon beam of energy 5 GeV, has $\gamma \approx 10$ and $\gamma c\tau_1 = 26.7$cm while $\gamma c\tau_2 = 155$m. So a few meters from the production region, all the short-lived K_1's would have decayed.

• In $|K^0(t)\rangle$, the amplitudes of K_1 and K_2 are unequal at $t > 0$. What this means is that an initially pure K^0 beam morphs into one containing neutral kaons of either strangeness[29]. To study this, we may allow the neutral kaon beam to pass through matter (e.g. a hydrogen target). The kaons will interact strongly with the nucleons, but unlike the weak decays, these are strangeness preserving reactions. While the ($S = -1$) $\bar{K}^0$ interacts strongly with protons, producing $S = -1$ strange baryons (like Λ), the K^0 does not interact as much and cannot produce Λ^0. Thus, starting with a pure K^0 beam, one can measure the intensity of K^0 at time t. The probability of finding a K^0 in $|K_1(t)\rangle$ is the absolute-square of the amplitude

$$\begin{aligned} \text{Ampl}(K^0,t) &= \langle K^0|K^0(t)\rangle = \frac{1}{\sqrt{2}}\left(\langle K_1| + \langle K_2|\right)|K^0(t)\rangle = \frac{1}{2}(e^{-im_1 t - t/2\tau_1} + e^{-im_2 t - t/2\tau_2}) \\ \Rightarrow \text{Prob}(K^0,t) &= \frac{1}{4}\left[e^{-t/\tau_1} + e^{-t/\tau_2} + 2e^{-\frac{t}{2}\left(\frac{1}{\tau_1}+\frac{1}{\tau_2}\right)}\cos(\Delta m\, t)\right] \quad \text{where } \Delta m = |m_2 - m_1| \end{aligned} \tag{112}$$

So the K^0 probability is expected to decrease exponentially from 1 at $t = 0$, but should show an oscillatory behavior superimposed on the exponential decay. The angular frequency of the oscillations is Δm, it has been measured (and calculated approximately using the 1-loop box Feynman diagram).

$$\Delta m = 3.5 \times 10^{-6}\,\text{eV} \approx \frac{1}{2\tau_1}, \quad m_2 > m_1. \tag{113}$$

[28] 10^{-23} s is the natural time-scale for the physics of hadrons. It is simply the time it takes light to cross a hadron, which is about a fermi in size.

[29] Note that kaons of opposite strangeness, being anti-particles, could annihilate if they collide. However, it is more likely for a single kaon to propagate or decay than for two of them to annihilate in such a beam. Moreover, the interference phenomena that lead to oscillations are due to the superposition principle of QM and apply to single kaons, just as the double slit interference pattern with photons may be developed with very low intensity light (single photons that do not interact with each other).

Though this mass difference is very small compared to the masses of other elementary particles, it is measurable since the corresponding oscillations have a time period of order $2\pi/\Delta m \approx 4\pi\tau_1 \approx 1$ ns. Due to time-dilation, this is manifested over a macroscopic length. Similarly, the probability of finding a $\bar{K}^0$ in the beam at time t is (plot these probabilities!)

$$\text{Prob}(\bar{K}^0, t) = |\langle \bar{K}^0 | K^0(t)\rangle|^2 = \frac{1}{4}\left[e^{-t/\tau_1} + e^{-t/\tau_2} - 2e^{-\frac{t}{2}\left(\frac{1}{\tau_1}+\frac{1}{\tau_2}\right)}\cos(\Delta m\, t)\right]. \qquad (114)$$

The probability of finding a $\bar{K}^0$ in an initially pure K^0 beam grows from 0 at $t = 0$, reaches a maximum at a time of order τ_1, oscillates and eventually decays to zero. Of course, after a time long compared to τ_2, there are no neutral kaons left, they have all decayed to two or three pions.

• The rate of $K^0 - \bar{K}^0$ strangeness oscillations that can be calculated from the 1-loop box Feynman diagram depends on the mass of the charm quark, which can appear as a virtual particle in some Feynman diagrams. By comparing the calculations with experimental measurement of the rate of strangeness oscillations, it was possible to predict the mass of the charm quark before it was discovered as a constituent of the J/ψ in 1974!

4.5 K_S regeneration

• A particularly striking manifestation of strangeness oscillations is in the phenomenon of K_S regeneration, predicted in 1955 by A Pais and O Pancini and detected by Pancini and collaborators in 1960. Suppose we begin with a beam (traveling through vacuum) of pure K^0 produced in $\pi^- p$ collisions in association with Λ^0 baryons (originally called hyperons). K^0 being an equal superposition of the short lived K_1 and long-lived K_2, we will initially see two pion decays by which the $K_1 = K_S$ component is depleted once $t \gtrsim \tau_1$ (in the rest frame), leaving a less intense beam of long lived $K_2 = K_L = (K^0 - \bar{K}^0)/\sqrt{2}$ (as long as $t \ll \tau_2$). Now, if we let the K_L beam pass through a layer of matter, the K^0 and $\bar{K}^0$ components will react differently. The $S = -1$ $\bar{K}^0$ are more easily absorbed (producing Λ^0 baryons) leaving a beam of more K^0's than $\bar{K}^0$s, so that (up to overall factors) the sequence of amplitudes may be written

$$K^0 \to K_L = \frac{1}{\sqrt{2}}(K^0 - \bar{K}^0) \to \frac{1}{\sqrt{2}}(fK^0 - \bar{f}\bar{K}^0) = \frac{1}{2}(f+\bar{f})K_L + \frac{1}{2}(f-\bar{f})K_S, \qquad (115)$$

where $f > \bar{f}$ are the amplitudes for K^0 and $\bar{K}^0$ to emerge from the layer of matter (Perkins' notation). Having emerged from the slab of matter the beam now has a K_S component which will decay to two pions! So by passing a neutral kaon beam (from which all the K_S had been depleted) through matter, K_S has been regenerated. This process can be repeated several times till all the kaons have decayed.

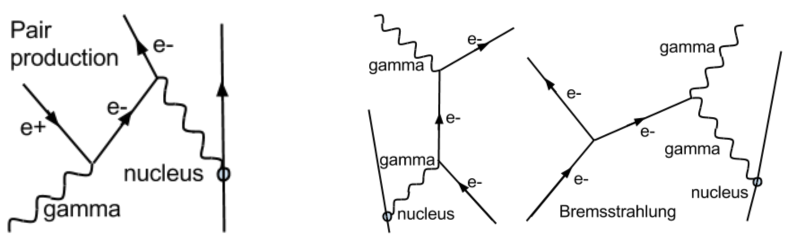
• These strange features of neutral kaons are partly a consequence of the superposition principle of quantum mechanics (and serve as a test of it).

5 Detection of particles: basic ideas

• The screen of a cathode ray tube (TV screen) serves as a detector of electrons, via the flashes of light they emit when hitting the scintillating material. Rutherford, Geiger and Marsden manually counted such flashes from α particles incident on a ZnS screen as a function of location using a microscope, in their alpha scattering experiments.

• Charged particles like electrons, muons, alpha particles etc are often detected by the tracks they leave. When a charged particle passes through a gas, it ionizes atoms along its path. The ions (due to their inertia) do not move much, while the liberated electrons tend to drift away. Many detectors like cloud, bubble, spark and proportional wire chambers work by detecting the ions (or electrons) left in the wake of the charged particle.

• Photons are neutral and do not ordinarily leave ion trails like electrons. However, high energy photons (gamma rays) can ionise atoms in a gas and the resulting ions/electrons can be collected. More spectacularly, photons of sufficient energy (more than $2m_e$) pair produce while passing through matter. Though pair production by real photons in vacuum $\gamma \to e^+e^-$ is not allowed due to momentum conservation, it can happen in the Coulomb field of a nucleus. The tracks of the e^+e^- pair can then be observed. The leading order Feynman diagram is drawn with time increasing upwards. There are three vertices (including the one at the nucleus) each of which contributes a factor of electric charge e. So the amplitude is $\propto e^3$ and the cross section to α^3, so pair production is called a third order process.



• Charged particles radiate photons when accelerated. Classically, the power radiated (say in a circular trajectory of radius r) grows with square of energy and decreases as the square of the mass $P = \frac{2e^2\mathbf{a}^2}{3\cdot 4\pi c^3} = \frac{2e^2E^2}{3\cdot 4\pi m^2r^2c^3}$. So this is particularly important for high energy electrons. Momentum conservation does not permit a free electron to radiate a single photon $e \to e\gamma$ in vacuum. However, an electron passing through matter is accelerated in the coulomb field of a nucleus and does radiate photons (the momentum being balanced by the nucleus). This is called bremsstrahlung, German for breaking radiation. There are two leading order FD for bremsstrahlung, which show that it is a third order process[30] If the photons produced are sufficiently energetic, they can pair produce in the vicinity of other nuclei. This produces an electromagnetic shower or avalanche of characteristic shape, and the charged particle tracks can be observed.

[30]Note that one needs to add the amplitudes represented by each of these diagrams and square the sum to get the physical rate. The photons that emerge cannot be identified as having been radiated by the electron or by the nucleus, it is the system that radiates, though we say that the electron radiates photons.

• Muons being 200 times heavier than electrons radiate less and typically do not produce electromagnetic showers. But they leave long tell-tale ion-tracks in detectors. Muon trackers are usually on the outer part of collider detectors, since muons make it quite far, unlike electrons. The tau lepton ($m_\tau = 1.777$ GeV) is even heavier than the muon, but it does not leave tracks, because it decays rapidly via the weak interactions (lifetime of 10^{-13} s) to charged leptons/pions and neutrinos.

• The neutron is a strongly interacting electrically neutral particle that lives long enough to be detected. Its EM interactions due to magnetic dipole moment are rather weak to be used as a method of detection. Beta decay to charged particles (electron and proton) is not a very practical means of detection since the rate is very low and the neutron may leave the lab before it decays. When passing through matter (especially hydrogen rich matter so that the target nuclei have the same mass as the projectile), neutrons can elastically scatter off protons (or other nuclei), and the recoiling protons/ions, being charged, can be more easily detected. When passing through matter, energetic neutrons can also collide inelastically with nuclei and produce several charged particles and daughter nuclei in a hadron shower which can be detected. Low energy neutrons can be absorbed by appropriate nuclei, which then undergo nuclear reactions (including fission) emitting gamma rays and other ionizing charged particles like beta and alpha rays, which can be detected.

• The neutral pion also interacts strongly, but is very short-lived. It decays in 10^{-17}s electromagnetically ($\pi^0 \to 2\gamma$) and the photons can be detected through pair production or otherwise (e.g. via the photo electric effect).

• We now introduce some specific devices/methods of particle detection.

5.1 Geiger counter

• In its simplest form, a Geiger counter consists of a hollow metal cylinder filled with a (typically inert noble) gas at low pressure. A wire along the axis was maintained at 1 kilo volt potential difference relative to the cylinder. When a charged particle (say an alpha or beta particle) passed through the cylinder, it left a trail of ionized atoms which were attracted to the wire due to the electric field. These ions produced more ions as they were accelerated towards the wire, resulting in a small electric pulse ('discharge') in the wire, which could be detected using an electrometer. In more sophisticated versions, the small current is amplified using amplifiers so that it can be heard on a speaker. Diodes and subsequently triodes, which were developed in the first decade of the 20th century were used to amplify the electrical pulse.

• The Geiger counter was developed by Geiger (and Rutherford) by 1913 and was improved upon by Geiger and Müller in 1928.

• Though it did not provide information on the track followed by the charged particle, the Geiger counter could detect the passage of individual alpha or beta particles, and hence count them. Clicks of a Geiger counter were used to measure radioactivity from uranic salts. However, they do not provide tracks of charged particles nor do they measure the energy of charged particles. Geiger counters can serve as triggers to switch on more sophisticated tracking detectors (like cloud chambers) when a charged particle passes by. A more sophisticated version of the Geiger counter, a proportional counter can indicate the energy of the charged particle by producing a current whose total charge is proportional to the energy. The point is that the number of ion-electron pairs produced by an ionising charged particle is roughly proportional to its energy. Each of these ion pairs is used to produce a small avalanche close to the anode wire, by carefully adjusting the voltage in the chamber, the gas pressure and composition.

5.2 Electroscope

• The electroscope as an instrument to detect electric charge goes back to the 1600s. It was used to discover cosmic rays in the early 20th century. Theodor Wulf constructed a sensitive electroscope in 1910 in the Netherlands. It consists of a pair of conducting leaves, say gold leaves that hang downwards and are joined at a vertex to a conducting central holder. When it is charged, the leaves repel each other. If a charged particle passes by, it ionizes the surrounding gas and electrons or ions drift to the electroscope discharging it. As a consequence, the leaves cease to repel. Natural radioactivity from the earth could discharge an electroscope, but this effect should decrease with height (distance from radioactive sources). What Wulf and later Victor Hess (Vienna, Nobel prize 1936) in more systematic studies found, was that the rate of discharge did not decrease with height, and in fact increased. This was discovered via Eiffel tower-top and balloon-borne experiments! It was deduced that the source of the discharging radiation was from space, leading to the discovery of cosmic rays.

5.3 Wilson cloud chamber

• The Scottish physicist C T R Wilson (Nobel prize 1927) developed the cloud chamber that bears his name. This is an example of a serendipitous development, Wilson originally wanted to reproduce the optical glory effect. In a failed attempt to do so, he was led to the path of developing the cloud chamber by 1911. He found in 1895 that water vapor condenses around charged particles to form droplets (drops of oil with electrons stuck to them were used later by Millikan to measure electron charge). The device consisted of a chamber filled with pure water vapor fitted with a piston. When the vapor is expanded using the piston, it adiabatically cools leading to supersaturation (the vapor pressure of water is exceeded). This is an unstable state, passage of a charged particle leaves ions which act as nucleation sites. Water vapor condenses around ions revealing the trails of charged particles. To work efficiently, one needs to know when to expand the vapor. The cloud chamber was used to detect cosmic rays by using Geiger counters as triggers for the expansion. Geiger counters were placed above and below the chamber (counter controlled cloud chamber, Blackett and Occhialini 1937, Nobel prize to Blackett in 1948). If both counters detected a pulse in quick succession, it signaled passage of a charged cosmic ray and the expansion was triggered. The resulting water droplets revealed the path which was photographed after activating a flash bulb. Tracks of α and β particles from natural radioactive decay were detected using the Wilson cloud chamber, emphasizing their particle-like nature. The α particle tracks are thicker than those of electrons, α's have higher ionizing ability due to higher charge. Later, particles produced in artificial nuclear disintegration and cosmic rays were also detected. Positrons (1932) muons (1936) and kaons (1947) were discovered in cosmic ray showers using cloud chamber photographs. By applying a magnetic field across the cloud chamber, information on the charge and momentum of charged particles could be obtained from their tracks. In a non-relativistic approximation, assuming velocity orthogonal to magnetic field,

$$\frac{qvB}{c} = \frac{mv^2}{r} = \frac{pv}{r} \quad \Rightarrow \quad p = \frac{qrB}{c} \tag{116}$$

High momentum electrons bend far less than low momentum ones. So the radius of curvature gives an estimate of momentum if the charge and applied B are known. The sign of charge is determined by whether the particle bends clockwise or anti-clockwise; e^+ and e^- bend in opposite directions. The curvature of an electron track increases with time as the electron loses energy to the material in the chamber, so cloud chamber photos show spirals.

• Moreover, particles of higher charge $|q|$ produce thicker tracks, since more ions are ionized. The chamber can be calibrated by using test particles of known charge. The thickness of the track is proportional to the square of the charge and inversely proportional to the square of the speed of the particle. So all particles of unit charge moving nearly at the speed of light leave tracks of the same width. For fast particles $v \simeq c$, one cannot tell the direction of motion as the track thickness does not appreciably increase within the dimensions of the chamber, given the limited energy loss. To find the direction one simply inserted a thin lead plate in the middle of the chamber. After passing through the plate, an electron/positron has less energy and its track is more curved in a magnetic field.

• Cloud chambers, while being relatively easy to construct, have the drawback of not containing much material to function as target. Glaser's bubble chamber solved this problem by using denser liquid instead of water vapor, which could act as target while continuing to reveal ion trails.

5.4 Photographic emulsions

• Like intense light, high energy charged particles too can be used to take photographs. Darkening of photographic plates (precursors to photographic film, glass coated with light sensitive chemical emulsions) was crucial to the discovery of X-rays and radioactivity. Hess confirmed the electroscopic detection of cosmic rays by the traces they left on photographic plates left on mountains and taken up in balloons.

• Powell, Occhialini, Muirhead and Lattes discovered the charged pions in 1947 using photographic emulsions to record tracks of cosmic rays at high altitude. Powell had been developing high quality photographic emulsions from the late 1930s. The thickness of the track left by a charged particle in photographic emulsion decreases with the speed of the particle and reaches a minimum as $v \to c$. Powell (Nobel prize 1950) and coworkers made essential improvements in photographic emulsion technology to record tracks of highly relativistic weakly ionizing particles, leading to the discovery of charged pions. A stack of emulsion covered plates (Silver Bromide crystals suspended in gelatin) produced images of cosmic ray tracks (black tracks of Silver grains), decays and interactions. Unlike cloud chambers which had to be expanded and contracted after each particle passage, the plates could be left to take photographs and many tracks could be recorded on the same plate. Photographic emulsions have much greater stopping power than air or vapor, so particle tracks are much shorter. The direction of motion of a charged particle through a photographic plate is given by direction of increase of black grains. A slower particle has greater ionizing ability. See the photographs in Perkins' book. A charged pion comes to rest in the emulsion and decays to a muon. The charged pion track ends and the muon tracks emerges at an angle, the missing momentum being carried by the uncharged muon neutrino. The resulting muon tracks were all of roughly the same length indicating that the decay is a 2 body decay and the muon is mono-energetic and decays in about 2.2 microseconds. At the end of the muon track is an electron track which again emerges at an angle, the missing momentum being carried by a pair of neutrinos.

• Photographic emulsions have good spatial resolution (half a micron) and are still in use in studies of cosmic rays and neutrinos as well as in astronomy. By very refined methods, distances of 1 micron between the point of production of a π^0 (via the strong interactions) and its decay (to 2γ via EM) can be measured, even though it does not leave a track.

5.5 Glaser bubble chamber

• The bubble chamber was invented by Donald Glaser (Nobel prize 1960) at Michigan in 1952. They were very widely used in tracking particles produced at accelerator experiments for the next three decades. The bubble chamber was based on and replaced Wilson's cloud chambers. Bubble chambers used liquid rather than water vapour used in cloud chambers. The greater density produced more target material for energetic particles (including neutrinos!) to interact with. While liquid hydrogen produced clear tracks, denser liquids like freon were used when more target mass was needed and also to ensure that photons and electrons produce EM showers. Cloud chambers also take longer (1 minute) than bubble chambers (1 second) for compression after expansion and track photography.

• Bubble chambers work in a similar manner as cloud chambers. The boiling temperature of a liquid decreases with decreasing pressure. A super heated liquid is one that is at a temperature above boiling point due to a sudden drop of pressure. The basic principle of operation of a bubble chamber is that a super-heated liquid is unstable and will begin to boil around perturbations such as the ion trails left by a charged particle. They are filled with liquid (liquid hydrogen requires a temperature of 20K; denser liquids like propane and freon have also been used. Glaser is said to have tried out beer without success!) and fitted with a movable piston to adjust the pressure. When the piston is pushed in, the liquid is at a high pressure of several atmospheres and at a temperature just below boiling. The piston is suddenly withdrawn leading to a drop in pressure and creating an unstable superheated liquid. If charged particles pass through in interesting directions (as determined by scintillation counters placed around the chamber) then, the compression of the chamber is triggered. Prior to this, the liquid has started to boil by forming bubbles around the ions left by charged particles. The compression happens at just the right time so that the bubbles are big enough to be photographed. Then the piston is pushed back in to prevent the whole liquid from boiling and to return it to its original high pressure state. The whole cycle can be completed in about a second in time for the next bunch of particles from an accelerator. A magnetic field (of order a Tesla) is applied across the bubble chamber to enable determination of particle charge and momentum. Particles with higher charge leave thicker tracks in bubble chambers. Muons leave long tracks since they radiate much less than electrons of the same energy and do not interact strongly.

• Glaser's first bubble chamber (1952) was only 3 cm in width. In the 1950s, L Alvarez (Berkeley), one of the champions of large-scale experiments, built along with coworkers, a 10 inch bubble chamber and also a 72 inch bubble chamber filled with deuterium exposed to pions and protons from the 6 GeV Berkeley Bevatron. Analysis of millions of bubble chamber photographs led to the discovery of numerous hadrons and resonances for which he received the 1968 Nobel prize. The Gargamelle bubble chamber used to discover neutral current neutrino interactions via elastic neutrino electron scattering $\bar{\nu}_\mu e \to \bar{\nu}_\mu e$ via Z^0 exchange at CERN in 1973 was 4.8 m long.

5.6 Spark chambers

• Spark chambers are made of several parallel metal plates with inert gas (He or Ne) in between successive plates. When a charged particle passes through, it leaves an ion trail in the gas. By applying a carefully chosen high voltage to alternate plates, one induces sparks along the ion trails between plates (like in a gas discharge tube or lightning) which are photographed. The timing of sparks can even be recorded using microphones. Spark chambers could be triggered with greater precision than bubble chambers. The material in the plates also functions as a target for particle interactions and increases the cross section for rare events like neutrino scattering.

• The basic idea of the spark chamber goes back to 1950 but it took some years to implement the idea and find just the right voltage. The development of the spark chamber by Fukui and Miyamoto in the late 1950s led almost immediately (in 1962) to the discovery that there were two types of neutrinos, muon and electron. This was achieved by the team of L Lederman, J Steinberger, M Schwarz et. al. using tertiary neutrino beams from the Alternating Gradient Synchrotron (AGS) at Brookhaven, Long Island, NY.

• In 1956 Cowan and Reines had discovered the electron type anti-neutrino coming from beta decay in nuclear reactors. Their 'signature' was that $\bar{\nu}_e$ produced positrons when interacting with matter. Lederman et. al. found that the neutrinos from the AGS produced largely muons rather than positrons when they interacted with the material in the spark chamber (see below). Positrons annihilate readily and produce an EM shower with a characteristic shape in spark chamber photographs. On the other hand, muons which radiate much less leave long tracks in the spark chamber, they do not produce EM showers. They concluded that there was a second kind of neutrino associated to the muon. This completed the second lepton family.

• Protons accelerated in the AGS were extracted, made to hit a target (e.g. Berillium) producing a shower of hadrons, including pions and kaons. The resulting particles were passed through a pile of mud. The pions and kaons decayed primarily to $\mu^-\bar{\nu}_\mu$ or $\mu^+\nu_\mu$ (with a lesser fraction of $e\bar{\nu}_e$ or $e^+\nu_e$ - recall the helicity rule). At the end of the mud, one was left with a beam of mostly $\nu_\mu, \bar{\nu}_\mu$ and fewer $\nu_e, \bar{\nu}_e$. The neutrino beam was directed at a spark chamber. A small fraction of the neutrinos then underwent neutrino nucleon deep inelastic scattering ($\nu_\mu p \to \mu^+ X$ where

X may include several hadrons, and similarly for $\nu_e, \bar{\nu}_\mu, \bar{\nu}_e$) Draw the Feynman diagram with $W^\pm$ exchange.

5.7 Wire chambers

• Analyzing millions of photographs from bubble and spark chambers was cumbersome. Beginning in the 1960s it became feasible to directly digitize the positions and times of charged particle tracks using electronics attached to wire chambers. Wire chambers are in a sense supped up versions of the Geiger counter.

• In essence, the metal plates of spark chambers are replaced with closely spaced (millimeter) parallel wires to which voltages are applied. The resulting sparks along ion trails, when charged particles pass through, are received by the nearest wire. This information (including the timing) is passed to a computer. By using several layers of wires oriented at different angles, accurate particle tracking is achieved using wire spark chambers.

• Multiwire proportional[31] chambers developed by G. Charpak (1968-70) provide even greater position (sub mm) and 10 ns time resolution for charged particle tracks. Here the free electrons produced along the particle track drift to the wires giving rise to small currents that are accurately measured and timed (typically, there are three layers of wires with the central layer held at high voltage being the one that received the electrons). This information is fed to a computer to reconstruct 3d tracks and determine particle momenta (using curvature of their trajectories in a magnetic field). Charpak received the Nobel prize in 1992. A multiwire proportional chamber requires elaborate electronics as each wire needs to be connected to electronic circuits to amplify the signals.

• Drift chambers work in a similar way as wire chambers but focus on measuring the 'drift time' taken for the electron signal to reach the 'sense' wire, to estimate distances of 10s of micrometers.

5.8 Scintillation counters

• Scintillation materials (crystals like CsI, NaI or organic materials (plastics containing anthracene dissolved in a liquid) and placed in a layer within a larger detector) give off flashes of light when charged particles pass through. The charged particle produces an electronic excitation in a molecule in the material. When an electron returns to the lower energy level, a photon is emitted, typically in the visible (often blue) or UV region. There are scintillation materials that give off flashes of light even when gamma rays pass through (not just charged particles). Geiger, Marsden and Rutherford used a Zinc Sulphide screen which produced flashes of light when struck by alpha particles that passed through their gold foil.

[31] An electron released when a charged particle ionizes a gas atom is accelerated towards the high voltage wire. If it picks up enough energy before reaching the wire, it ionizes other atoms producing more electrons, creating an avalanche. The number of secondary electrons is *proportional* to the number of primary ions by an amplification factor of order 10^5.

• The photon signal from scintillation counters can be very weak (1 or a few photons). It needs to be amplified. This is done by photo multiplier tubes PMTs.

• Scintillation detectors do not provide tracks of charged particles. However, a pulse recorded by a scintillation counter can be used as a trigger to activate other detectors like bubble chambers or spark chambers . They are also used to estimate the energy of charged particles and EM and hadron showers, as we shall see.

5.9 Cherenkov radiation

• Cherenkov radiation is emitted by charged particles moving in a polarizable medium (usually water or heavy water) faster (βc) than the speed (c/n) of light in the medium (which is always less than c). The phenomenon had been noticed before Pavel Cherenkov and his advisor Vavilov (1934) but had been mistaken for fluorescence[32]. When a plane flies at above the speed of sound in air, a shockwave forms between regions of sub-sonic and super-sonic flow. This wave can be heard by ground based-observers in a sonic boom. Cherenkov radiation is an optical analogue, when a charged particle (electron or muon, pion, kaon etc) moves faster than the local speed of light. There results an expanding cone of (usually bluish) light that follows the particle. The wave front of Cherenkov radiation is shaped like the wake of a ship and is at a fixed angle relative to the particle trajectory. The opening angle of the cone is given by

$$\sin\theta = \frac{(c/n)t}{\beta ct} = \frac{1}{\beta n}. \tag{117}$$

The opening angle of this cone decreases with the speed of the charged particle and may be used to determine its speed. Cherenkov radiation emitted in a tank forms ring-like patterns on the walls of the tank, which can be detected by photo multiplier tubes that line the walls. There is a photo you may find on the internet of technicians in a boat repairing the PMTs on walls of a Cherenkov tank. The Cherenkov effect was theoretically explained by Tamm and Frank, colleagues of Cherenkov in the same department. The three of them shared the 1958 Nobel prize.

5.10 Photo Multiplier Tubes

• PMTs are used to amplify the photon signal from scintillation counters or Cherenkov counters. So a scintillation detector is often fitted with a PMT. And hundreds of PMTs line the walls of huge tanks of water in order to detect Cherenkov radiation from fast moving charged particles.

• PMTs work on the basis of the photo electric effect. Visible and UV photons incident on a metallic cathode can cause photo electrons to be ejected. An electric field guides these photo electrons to an anode. Upon striking the anode, several more electrons are ejected, which are then directed at a second anode, generating yet more electrons. Thus, a single photon can generate an cascade of electrons, enough to produce a detectable current.

[32]Fluorescence is the emission of characteristic light by excited molecules, as they relax to a lower energy level after having been excited by radiation.

5.11 Silicon vertex detectors

• Highly sensitive doped semiconductor strips (10-20 μm width, made of Si/Ge) may be used to detect the passage of charged particles. The semiconductor strip functions as a diode which is reverse biased and does not conduct under normal circumstances. The charged particles ionize silicon atoms producing conduction electrons whose current is then detected. It is possible to get position resolution of less than 10 microns this way. This is needed to study weak decays of heavy quarks (b,c), which live for only $10^{-13}s$ and travel a fraction of a millimeter at energies of order 10-100 GeV. SVX detectors have enabled us to study the weak decay chain $b \to c \to s$ (of hadrons containing these quarks) from very close to the collision vertex. Current day collider detectors contain silicon strip detectors as their innermost component surrounding the collision vertex. The top quark, which decays to b before it can hadronize to form mesons/baryons, was discovered in 1995 at the Fermilab Tevatron (CDF and D0 detectors) by observing the b mesons produced in top decays.

5.12 Sampling Calorimeters

• Often, the energy of a charged particle is determined by measuring the curvature of its track in a fixed magnetic field. However, very high energy particles do not bend much and this method is not always feasible. The other problem is that of determining the energy of a shower of particles (electromagnetic or hadron showers and jets of hadrons), where there may be hundreds of densely packed short tracks rather than a single curved track.

• A sampling calorimeter is used to determine the energy of single charged particles or showers. A calorimeter typically consists of a sequence of heavy metal plates interspersed with scintillation counters or spark chambers. The metal plates (e.g. iron slabs) serve as target material to induce particles (like electrons, muons, nucleons and neutrinos) to interact, slow down/produce showers and deposit their energy. The light emitted in the interspersed scintillation counters and spark chambers is measured. By calibrating the calorimeter using test beams of known energy particles, one can arrive at accurate measurements of unknown particle energies and shower energies. Sampling calorimeters essentially use an analogue of Simpson's rule to sample the energy deposition function and 'integrate' it to estimate the total energy of a particle or shower.

• The recently sanctioned India Based Neutrino Observatory will use a sampling calorimeter to measure the energies of (especially) ν_μ and $\bar{\nu}_\mu$ that undergo deep inelastic scattering with nuclei in the detector producing $\mu^\pm$ and hadron showers. 30,000 RPCs (resistive plate chambers) will be used to detect muon tracks and measure their energies from the ionization produced in the gas, as in a wire chamber. A magnetic field will also be used to bend muon tracks to estimate their momenta.

6 Nuclear physics and isospin with a view towards particle physics

6.1 α, N, p and n particles and the discovery of the strong nuclear force

• Rutherford discovered the alpha particle in natural radioactive emission from uranium and radium in 1899. It was the least penetrative (through matter) of the three types of radioactive radiation α, β, γ. Now we know why: α particles, unlike β electrons and γ ray photons interact both strongly and electromagnetically. However (and partly for these reasons) the alpha particle was a work-horse of sub-atomic physics in the first third of the 20th century. Now we know that an α particle is a stable bound state of two neutrons and 2 protons (He_2^4 nucleus) with a large binding energy of 28 MeV. Bombarding atoms and nuclei with alpha particles led to many discoveries.

• During 1908-1913 Geiger and Marsden under the supervision of Rutherford at Manchester scattered alpha particles against thin metal foils (esp. gold foils which can be made very thin) and detected the alpha particles through flashes on a fluorescent screen beyond. The observation of wide angle scattering was used to deduce the concentration of positive charge in a point-like atomic nucleus. Rutherford worked out the differential scattering cross section for Coulomb scattering of point charges in 1911 and the experimental results of 1913 agreed with Rutherford's formula. Thus atomic nuclei were discovered.

• In 1919 Rutherford discovered the proton by bombarding hydrogen gas with 5 MeV alpha particles from Radium decay. αs being much heavier, knock protons (called H particles then) out of the atoms and they could be detected by counting scintillations they produced or via their tracks in a cloud chamber.

• Interestingly, for α particles of higher energy, the measured number and angular distribution of scattered protons did not match the Coulomb scattering formula, even after accounting for recoil of the target. There was an anomalously large scattered proton current produced. The experiments showed that the $\alpha - H$ force at short distances could not be purely electromagnetic and must be of a great intensity. Rutherford also estimated the distance at which the simple Coulomb repulsion picture must break down: around 3.5 Fermi. While the earlier Geiger-Marsden experiments involved high Z targets like gold, the H target was at the opposite extreme. 5 MeV was adequate to overcome the repulsive $\alpha - H$ Coulomb barrier and provide (in retrospect) the first experimental observation of the strong nuclear force. However, it was not initially interpreted as such (the deviations were initially attributed to the possible non-point-like nature of the alpha particle) and it took more than a decade and a half to *begin* understanding the nature of the strong nuclear force. However, Rutherford's discovery of the proton paved the way for the identification of the first constituent of the atomic nucleus, the proton.

• The neutron was discovered by Chadwick at the Cavendish lab in 1932 by bombarding Beryllium with α particles from a radioactive source via the reaction $\alpha_2^4 + Be_4^9 \to C_6^{12} + n$. While the alpha particle, nucleus and proton had been discovered essentially by chance, the existence of the neutron as a constituent of the nucleus had long been suspected.

• Electromagnetism was still the only force known and it was impossible to explain the stability of heavier nuclei if they consisted only of protons. The electron was the only other matter particle known at the time and electrons emerged from the nucleus in beta decay. So it was tempting to think that nuclei were made of protons and electrons. But there are many difficulties with this. If electrons were bound to protons by the electrostatic Coulomb potential, then the size of a nucleus could not be smaller than the Bohr radius, which is much larger than nuclear dimensions of a few Fermi. If some other force confined an electron within a fermi, then the uncertainty in its momentum would be at least of order $\hbar/1fm$ which is about 197 MeV/c. To confine an electron with such high (relativistic) momenta would require nuclear potentials of comparable size, but experimental observations indicate nuclear potentials of a few MeV at most. If we try to build the nucleus from protons and electrons alone, we also run into trouble with measured nuclear angular momenta. N-14 nucleus is problematic (as pointed out by Heitler and Herzberg in 1929) since it would have to contain 14 protons and 7 electrons based on mass number and charge. The spin of 21 spin half particles cannot be an integer, yet the measured spin of N-14 is one. Moreover, the measured mass numbers A of some nuclei were twice or even more than the atomic number Z (nuclear charge). In the absence of electrons, the nuclear mass could not come from protons alone. It was natural to conjecture (as Rutherford did in his Bakerian lecture to the Royal Society in 1920) that the nucleus consisted of protons as well as neutral neutrons[33]. Rutherford and his associates at the Cavendish desperately searched for the neutron from 1920 onwards. The breakthrough came in 1932, the year in which both the positron and neutron were discovered.

• The exposure of beryllium to alpha particles produced from Polonium radioactive decay was found (by Bother and Becker (1930) and by the Joliot-Curies (1932)) to produce neutral radiation of very high energy. It had to be of nuclear origin, rather than, say, from atomic electron transitions. These experimenters thought these were gamma rays and, if so, they had to have energy ~ 50 MeV, based on the manner in which they reacted with matter. But Rutherford and Chadwick did not believe the gamma ray hypothesis. Chadwick estimated the mass of the neutral particles by studying the recoil of hydrogen and nitrogen gas exposed to the radiation. He found a mass of order of the proton mass. Moreover, if it was gamma emission, then the reaction would be $\alpha_2^4 + Be_4^9 \to C_6^{13} + \gamma$. From the mass defect, the resulting gamma rays could have energies of at most 14 MeV. So the neutral radiation could not be gamma rays. Thus was discovered the neutron, as a constituent of the nucleus. Chadwick got the 1935 Nobel prize for the discovery of the neutron.

• One can infer the existence of neutrons from the missing momentum in proton nucleus collisions. In cloud chamber/photographic emulsions, one sees the track of the incoming p and of a recoiling nucleus, but there is missing momentum which is ascribed to an unseen neutral particle.

• The discovery of the neutron paved the way to understanding the nucleus as a bound state of protons and neutrons, as proposed by Heisenberg. It was also crucial for Fermi's theory of beta decay $n \to p + e^- + \bar{\nu}_e$.

[33]Neutrons were initially hypothesized to be proton-electron bound states, though this idea was problematic for similar reasons. The only known pe bound states, were the electromagnetically bound levels of the hydrogen atom, which is much larger than the nucleus. The spin of Nitrogen is also inconsistent with the neutron being an ep bound state with integer spin.

• Confining a neutron in a nucleus of linear dimension 1 fm corresponds to a momentum uncertainty of order $\Delta p \approx \hbar/1fm$. Putting $(\Delta p)^2 = \langle p^2 \rangle - \langle p \rangle^2$ and taking $\langle p \rangle = 0$, this corresponds to a non-relativistic kinetic energy $\langle \frac{p^2}{2m_n} \rangle \approx \frac{\hbar^2}{2 \cdot 1 GeV/c^2} \approx 20$ MeV. This energy is significantly less than the rest energy of neutrons and comparable to nuclear potential energies. So nucleons in a nucleus can be treated within non-relativistic quantum mechanics.

• On the experimental side, neutron scattering became a key technique for studying the structure of bulk matter (neutrons being neutral penetrate further than charged particles through solids and liquids). Fermi in Rome and others (including the Hahn group in Berlin, Rutherford's group in Cambridge and Jolio-Curie in Paris) reasoned that neutrons being neutral could penetrate nuclei better than alpha particles (which were the projectiles of choice till then) and led to the synthesis of new radioactive isotopes. In most cases the nucleus absorbed the neutron to form an unstable isotope which then suffered beta decay. There was one case which Fermi and collaborators did not understand properly: neutron scattering off Uranium where there seemed to be several radioactive products. Their experiments were stopped due to fascist political developments preceding WWII: he went to Stockholm for the Nobel prize and then sailed to the US. Meanwhile (in 1938) the master chemist Otto Hahn working with Fritz Strassmann in Berlin discovered nuclear fission (breakup of Uranium$_{92}$ into two nuclei of roughly half the size, Barium$_{56}$ and Krypton$_{36}$ (Technitium$_{43}$ and Lanthanum$_{57}$ etc were also produced in some fission reactions)) by slow neutron scattering off Uranium. Conventional wisdom was that scattering neutrons off Uranium would produce neighbors or isotopes of Uranium, but not an element roughly half its size. This was further investigated and explained in Stockholm by Lise Meitner and her nephew Otto Frisch (Meitner was Hahn's collaborator but had to leave Germany in 1938) who showed that a lot of energy was released. Slow neutrons were used by Fermi, L. Szilard and H. Anderson to produce the first nuclear chain reaction, working first at Columbia Univ and then in Chicago.

6.2 Strong force and nuclear spectra

• By scattering fast neutrons off nuclei, nuclear radii could be measured. Fast neutrons have a de Broglie wave length much smaller than nuclear dimensions and classical hard sphere scattering approximation applies. One finds that lead and uranium have 10 fm radii, while it is about 6 fm for nuclei in the middle of the periodic table and decreases further to about a fermi for light nuclei.

• Protons and neutrons are held together in atomic nuclei by the strong nuclear force which outweighs the electric repulsion between protons. Potentials[34] used to model nucleon scattering show that the force has a finite range of about 2-3 fm and is attractive for inter-nucleon separation above about .7 fm corresponding roughly to a Yukawa potential $-ge^{-r/\lambda}/r$. At smaller separations, it is repulsive (see Fig 2 on p.47 of Gottfried and Weisskopf). At separations of a Fermi, the strong force between protons is about 10 times their electrostatic repulsion. G & W use this to get a crude estimate of nuclear excitation energies and nuclear dimensions. If the

[34] Potentials can only be approximate since the inter-nucleon force also depends on the spin and isospin states of the nucleons, it is not simply a central force.

attractive part of the nuclear potential is approximated by a Coulomb potential, then the nuclear analogue of the fine stricture constant is about $\alpha_N \sim 1/10$. It follows that the nuclear Rydberg $\mathrm{Ry}_N = \frac{1}{2} m_N \alpha_N^2 \sim 938/200 \sim 5$ MeV and the nuclear Bohr radius $a_N = \frac{1}{m_N \alpha_N} \sim \frac{10}{938}$ MeV^{-1} ~ 2 fm. Thus one expects nuclear excitation energies to be a few MeV and nuclear dimensions to be a few fm, as is observed.

• As already discussed, the motion of nucleons in a nucleus may be treated within non-relativistic QM to good approximation: binding energies of a few MeV are small compared to rest energies of constituents (938 MeV) while de Broglie wavelengths are comparable to nuclear dimensions. Like atoms, nuclei have a spectrum of discrete excited states. In an atom the electrostatic potential due to the nucleus, along with Pauli's exclusion principle gives a first approximation to energy levels and wave functions; inter-electron repulsion introduces important corrections to the independent electron approximation. In the absence of a large central potential in a nucleus, many-body effects are more pronounced in nuclear structure, making it more challenging to determine theoretically. Nevertheless on account of rotation invariance, nuclear levels (like atomic levels) may be labelled by an angular momentum quantum number J, which is an integer or half-odd-integer, according as the baryon number ($=$ mass number A) is even or odd.

• Nuclei with large numbers of protons ($Z \gg 1$) tend to have more neutrons than protons ($A > 2Z$) to counter inter-proton electric repulsion; U-238 has only 92 protons and is quite stable (alpha decay with half life of 4.5 billion years). If an isotope has too few neutrons, then it is susceptible to alpha decay (e.g. Radon-222). In general, α decay increases the neutron fraction since the alpha particle has only 50% neutrons whereas the parent nucleus typically has $A/Z > 50\%$. A few nuclei are fissile (U-235, U-233 and Plutonium-239), susceptible to fission induced by a slow neutron to daughter nuclei of roughly half the size. These are used as fuel in nuclear reactors and in fission bombs. Th-232 is not fissile but available in beach sands in India. However, Th-232 can be converted into fissile U-233 by bombarding with thermal neutrons, followed by a couple of beta decays. Fission of U-233 produces energy as well as neutrons which can be used to convert some more Th-232 into fissile U-233. So Th-232 can play a useful role in the nuclear fuel cycle.

• On the other hand, if a nucleus has too many neutrons, then it is susceptible to beta decay converting a neutron to a proton. E.g. the stable isotopes with $Z = 1$ are hydrogen H_1^1 and deuterium H_1^2 while tritium H_1^3 undergoes beta decay.

6.3 Yukawa's π mesons

• Yukawa originally proposed the charged pions $\pi^\pm$ to mediate the strong force between proton and neutron, which would involve a charge exchange $p \leftrightarrow \pi^+ n, n \leftrightarrow p\pi^-$. Of course, it was natural to postulate a third neutral pion to mediate the strong force between a pair of protons or a pair of neutrons. In fact, Kemmer showed later that there had to be a third neutral pion by extending heisenberg's isospin symmetry to pions postulating that they were in an isospin triplet representation of SU(2).

• Yukawa thought that charged pions would decay via $\pi^- \to e^- \bar{\nu}_e$ and this could be used to explain neutron beta decay via the dissociation of a neutron $n \to p\pi^-$ followed by $\pi^- \to e^- \bar{\nu}_e$.

But this is not correct, neutron beta decay is a single three body decay $n \to pe\bar{\nu}$ and charged pions predominantly decay to muons $\pi^- \to \mu^- \bar{\nu}_\mu$. The μ^- subsequently decays to an electron and two neutrinos $\mu^- \to e^- \bar{\nu}_e \nu_\mu$.

• The production of pions in the upper atmosphere due to collisions of cosmic ray protons with nucleons proceeds through reactions such as

$$p + p \to n + \pi^+ + p \quad \text{and} \quad p + n \to n + \pi^+ + n. \tag{118}$$

With more energy more pions can be produced, including π^0s. But the pions can be captured/absorbed by a nucleus. And charged pions decay with a mean life of 10^{-8}s. For these reasons, pions do not easily reach the Earth's surface, and had to be found using high altitude/balloon borne experiments (Powell et. al. using photographic plates). This explains why the muon was discovered before the pion and mistaken for it.

6.4 Static Yukawa potential from massive Klein-Gordon field equation

• A scalar particle like the pion is associated to a scalar field, it transforms as a scalar under the Poincare group. Ignoring self-interactions, the simplest scalar field is the Klein-Gordon field, one that satisfies the KG equation $\left(\Box + \frac{m^2c^2}{\hbar^2}\right)\phi(\mathbf{r}, t) = 0$ where $\Box = \frac{1}{c^2}\partial_t^2 - \hbar^2\nabla^2$ is the d'Alembertian. The pion field was introduced to explain the strong force between nucleons. Yukawa argued that it must be massive since the force is short-ranged. Recall that the electric field encodes the force between static electric charges. The Coulomb potential arises as a particular static solution of Maxwell's equations (Gauss' Law) of electromagnetism in the presence of a point electric charge.

• So we seek a static solution of the KG equation in the presence of a source (say a neutron). In the presence of sources, there is an inhomogeneous term on the RHS of KG equation. For simplicity, consider an attractive point source of strength g_0, and let us look for a time-independent solution

$$\left(-\nabla^2 + \frac{m^2c^2}{\hbar^2}\right)\phi = -g_0\delta^3(\mathbf{r}). \tag{119}$$

This may be done by going to momentum space

$$\phi(\mathbf{r}) = \int \tilde{\phi}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}[d\mathbf{k}] \quad \text{and} \quad \delta^3(\mathbf{r}) = \int e^{i\mathbf{k}\cdot\mathbf{r}}[d\mathbf{k}] \quad \text{where} \quad [d\mathbf{k}] = \frac{d^3k}{(2\pi)^3}. \tag{120}$$

The equation becomes (in natural units)

$$(\mathbf{k}^2 + m^2)\tilde{\phi}(\mathbf{k}) = -g_0 \quad \text{or} \quad \tilde{\phi} = -\frac{g_0}{\mathbf{k}^2 + m^2} \quad \Rightarrow \quad \phi(\mathbf{r}) = -g_0 \int \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\mathbf{k}^2 + m^2}[d\mathbf{k}] = -g_0 \frac{e^{-mr}}{4\pi r}. \tag{121}$$

Thus the scalar field in the presence of a source is given by the Fourier transform of a 3d Lorentzian. The Fourier transform may be evaluated using contour integration to obtain the screened Coulomb or Yukawa potential (see problem set 7).

• The general solution of $(-\nabla^2 + \frac{m^2c^2}{\hbar^2})\phi = -g_0\, \delta^3(\mathbf{r})$ is the sum of a particular solution and the general solution of the homogeneous equation $-\nabla^2\phi = -m^2\phi$, which is the same as the non-relativistic Schrödinger eigenvalue problem for a free particle with negative energy. So the general solution to the homogeneous equation is $\phi(\mathbf{r}) = \sum_{\vec\kappa} A_{\vec\kappa} e^{\vec\kappa\cdot\mathbf{x}}$ where $\vec\kappa$ is any vector with $\vec\kappa^2 = m^2$. However, these solutions all blow up at large spatial distances, so we do not add them to the particular solution obtained by Fourier analysis.

• As an alternative to Fourier analysis, one may solve the differential equation in position space by looking for a spherically symmetric ϕ, which must satisfy

$$(-\nabla^2 + m^2)\phi = -\frac{1}{r}(r\phi)'' + m^2\phi = -g_0\delta^3(\mathbf{r}) \tag{122}$$

For $r > 0$ we have $(r\phi)'' = m^2 r\phi$ whose linearly independent solutions are $r\phi = e^{\pm mr}$. Keeping the decaying solution, $\phi = A\frac{e^{-mr}}{r}$. The proportionality constant A is fixed by examining the behavior near $r = 0$, where $\phi \to A/r$. Recalling $\nabla^2(1/r) = -4\pi\delta^3(\mathbf{r})$ and ignoring the sub-leading divergence of $m^2\phi$ at the origin, we get $\phi = -g_0\frac{e^{-mr}}{4\pi r}$.

• The quantity $(\mathbf{k}^2 + m^2)^{-1}$, which is the inverse of the operator $(-\nabla^2 + m^2)$ is an example of a propagator, it propagates the influence of the source at the origin $-g_0\delta^3(\mathbf{r})$ to determine the value of of the pion field at $\mathbf{r}$, $\phi(\mathbf{r})$. More generally, in a non-static situation, the propagator for a massive scalar (spin zero) field is $(\Box + m^2)^{-1}$ or $(-p^2 + m^2)^{-1}$ in momentum space, where $p^2 = E^2 - \mathbf{p}^2$ is the square of the 4-momentum $p^\mu = i\partial^\mu$. The propagator is the contribution of the exchange of a virtual scalar particle to the amplitude for a scattering or decay process. The intermediary need not be on mass shell, so $p^2 \neq m^2$ in general. Rather, one must integrate over all possible values of p^μ consistent with momentum conservation to find the amplitude.

6.5 Isospin

• The proton and neutron masses are the same to about one part in a thousand ($m_n = 939.565$, $m_p = 938.272$ MeV). Even with the accuracy available in 1932, Heisenberg suggested that they are nearly degenerate energy states of the same system, the nucleon. Moreover, weak interactions (like beta decay) transform nucleons into each other though the emission of other particles, just as happens between two levels of an atom in a radiative transition. Nucleons also feel essentially the same strong forces, which are (electric) charge-independent. In general, degeneracies in spectra are indicative of a symmetry (recall that degeneracy with respect to magnetic quantum number $E_{nlm} = E_{nlm'}$ in the hydrogen atom is due to rotation invariance). All this can be explained by postulating that the strong interactions possess a new internal isospin symmetry. Isospin (‘iso’ means ‘like’) was somewhat misleadingly named isotopic spin (by E P Wigner); it does not have much to do with isotopes. The more accurate term isobaric spin is sometimes used. Mass differences between the nucleons are attributed to electromagnetic corrections to strong forces.

• In so far as the isospin degree of freedom is concerned, the nucleon is modeled as a 2-state system with Hilbert space $\mathbb{C}^2$ spanned by $|p\rangle = (1,0)^t$ and $|n\rangle = (0,1)^t$. Hermitian observables can be expressed as linear combinations of the Pauli matrices $\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\tau_2 =$

$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and the identity. Including the ordinary spin-half of nucleons, the wave function must lie in the tensor product $\mathbb{C}^2_{\text{isospin}} \otimes \mathbb{C}^2_{\text{spin}}$ with the isospin/spin observables acting on the first/second factors. If we included translational degrees of freedom there would be a third tensor factor $L^2(\mathbb{R}^3)$ in the Hilbert space.

• By analogy with spin angular momentum $\vec{S}$, isospin is defined as a vector observable $\vec{I}$. $\vec{I} = (I_1, I_2, I_3)$ are the components of the isospin vector observable in a 3d 'internal' isospin space $\mathbb{R}^3_{\text{iso}}$, whose analogue in the case of spin (or angular mom) is ordinary 3d xyz-coordinate space. While angular momentum $\vec{J}$ generates rotations in ordinary xyz Euclidean space, $\vec{I}$ generates rotations in the internal $\mathbb{R}^3_{\text{iso}}$. The components of isospin are postulated to satisfy the SU(2) Lie algebra $[I_\alpha, I_\beta] = i\epsilon_{\alpha\beta\gamma} I_\gamma$. Now this Lie algebra possesses irreducible unitary representations of dimension $1, 2, 3, \cdots$. The nucleon doublet carries the 2d irreducible representation of isospin (denoted $\mathbf{2}$) where I_α are represented by 2×2 hermitian matrices. In the neutron-proton basis for $\mathbb{C}^2$, $I = \frac{1}{2}(\tau_1, \tau_2, \tau_3)$. This representation is also called the fundamental or defining representation of SU(2) Lie algebra. The proton and neutron are the isospin $I_3 = \pm\frac{1}{2}$ states (up and down relative to the 3rd direction of the internal $\mathbb{R}^3_{\text{iso}}$).

• Under a finite rotation by angle $|\vec{\theta}|$ about the axis $\vec{\theta}$ in isospin space, the nucleon doublet transforms as

$$N \to UN \quad \text{where} \quad U = \exp\left(-\frac{1}{2} i\vec{\tau} \cdot \vec{\theta}\right) = \left[\cos(\theta/2) - i\vec{\tau} \cdot \hat{\theta} \sin(\theta/2)\right] \in SU(2). \quad (123)$$

For example, under a rotation about the second axis [$\hat{n} = (0, 1, 0)$],

$$N = \begin{pmatrix} p \\ n \end{pmatrix} \to \begin{pmatrix} \cos\theta/2 & -\sin\theta/2 \\ \sin\theta/2 & \cos\theta/2 \end{pmatrix} \begin{pmatrix} p \\ n \end{pmatrix}. \quad (124)$$

Under an infinitesimal isospin transformation, the nucleon doublet $N \to N + \delta N$ where $\delta N = -\frac{1}{2} i\vec{\theta} \cdot \vec{\tau} N$. Here $\vec{\theta} = (\theta_1, \theta_2, \theta_3)$ are infinitesimal real parameters of the transformation specifying the axis and angle of rotation in isospin space.

• Isospin symmetry is the approximate (in retrospect because u and d are not exactly equal in masses) invariance of the strong interactions under rotations in isospin space $\mathbb{R}^3_{\text{iso}}$. This is the analogue of invariance under rotations in ordinary Euclidean space $\mathbb{R}^3_{xyz}$. For example, a rotation by π about the 2nd direction in isospin space $U = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ would take $U|p\rangle = |n\rangle$ and $U|n\rangle = -|p\rangle$ in effect reversing the sign of I_3. While the strong interactions are nearly unchanged by this transformation, the proton and neutron have very different EM interactions due to their different electric charges and magnetic moments. Thus, the EM interactions pick out a specific direction in isospin space, the third direction. Now we may define the charge operator, which in the $|p\rangle, |n\rangle$ basis for $\mathbb{C}^2$ is just $Q = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Like other observables acting on the nucleon Hilbert space, it must be a real linear combination of the Pauli matrices and the identity. We find $Q = \frac{1}{2}(\tau_3 + I) = I_3 + \frac{1}{2}$. So Q does not commute with I_1 and I_2. In particular, electric charge is not invariant under rotations in isospin space: EM interactions violate isospin conservation.

• If we use electromagnetic probes to observe particles of definite electric charge, we do not see states like $|p\rangle + |n\rangle$, though they too are elements of the Hilbert space. Turning on EM is like breaking rotation invariance by switching on a magnetic field $\mathbf{B} = B_z \hat{z}$ in the z-direction of xyz-space. The magnetic field breaks the degeneracy in energy (Zeeman splitting) among the $2J_z + 1$ states that differ in their projections of $\mathbf{J}$ on $\hat{z}$.

• The isospin raising and lowering operators $I_\pm = I_1 \pm i I_2$ are defined by analogy with those for angular momentum and allow us to construct representations of the isospin SU(2) algebra. As with ordinary spin, we have $(2I+1)$-dimensional irreducible representations where the Casimir $\mathbf{I}^2 = I_1^2 + I_2^2 + I_3^2$ takes the values $I(I+1)$ for each $I = 0, \frac{1}{2}, 1, 3/2, 2, \ldots$ etc. $I = 0$ is the one dimensional trivial representation where I_α are all represented by the zero matrix. The corresponding group elements are all represented by the identity matrix.

• Looking at masses of strongly interacting particles, one notices that they come in multiplets of particles of nearly equal mass and the same spin and parity, but different electric charge and strangeness. This is due to isospin symmetry, the states of a multiplet differ in their isospin projections I_3, and are related by the action of $I_\pm$. The nucleons as well as u, d form $I = \frac{1}{2}$ doublets. Pions $\pi^\pm(139.6)$ MeV, $\pi^0(135)$ MeV form an $I = 1$ triplet as do their orbital excitations $\rho^\pm, \rho^0$. (K^+, K^0) and $(\bar{K}^0, K^-)$ are isospin doublets. The spin $3/2$ baryons (and anti-baryons) provide several examples of isospin multiplets. The Ω^- is an isospin singlet. Ξ^-, Ξ^0 are an isospin doublet while $\Sigma^-, \Sigma^0, \Sigma^+$ form an $I = 1$ triplet. The Δ's transform in a 4-dimensional $I = 3/2$ representation. The microscopic origin of isospin symmetry lies in the nearly degenerate masses of the up and down quark, which form an $I = \frac{1}{2}$ doublet. The other quarks s, c, t, b and leptons are all isospin singlets. Since the strange and charm quarks have rather different masses, a transformation (like the isospin transformations that relate u and d) that relates them would not be a symmetry of the strong interactions, so it is not useful.

• The empirical formula $Q = I_3 + Y/2$ relates isospin to electric charge where $Y = B + S + C + \tilde{B} + T$ is the hypercharge. B is baryon number, $S, C, \tilde{B}, T$ are the strangeness, charm beauty and topness quantum numbers respectively that count the number of $\bar{s}, c, \bar{b}, t$ (anti-)quarks. This formula can be explained using the quark model.

• E.g., the triplet of pions π_1, π_2, π_3 carry the 3d adjoint representation $\mathbf{3}$ of isospin, where the matrix elements of the isospin operators are the structure constants themselves: $(I_\alpha)_{\beta\gamma} = -i\epsilon_{\alpha\beta\gamma}$. Check that these matrices furnish a representation (the 'adjoint' rep.) of the SU(2) Lie algebra $[I_\alpha, I_\beta] = i\epsilon_{\alpha\beta\gamma} I_\gamma$.

$$I_1 = i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad I_2 = i \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad I_3 = i \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{125}$$

If we denote the components of the pion triplet in this basis as $\pi = (\pi_1, \pi_2, \pi_3)^t$, then the representation is $(I_\alpha \pi)_\beta = (I_\alpha)_{\beta\gamma} \pi_\gamma = -i\epsilon_{\alpha\beta\gamma} \pi_\gamma$. In order to relate these π_α to the physical pions of definite electric charge we must find the eigenstates of $Q = I_3$. I_3 is not diagonal in the $\mathrm{span}(|\pi_1\rangle, |\pi_2\rangle)$ sub-space in this basis. So $|\pi_1\rangle, |\pi_2\rangle$ are not the physical pions. The observed pions are $\pi^\pm = (\pi_1 \pm i\pi_2)/\sqrt{2}$ and $\pi^0 = \pi_3$. Check that $I_3 |\pi^\pm\rangle = \pm |\pi^\pm\rangle$ and $I_3 |\pi^0\rangle = |\pi^0\rangle$.

• As for the nucleon doublet, under an infinitesimal isospin transformation, $\pi \to \pi + \delta\pi$ where $\delta\pi = -i\theta_\alpha I_\alpha \pi$ or in components,

$$(\delta\pi)_\beta = -i\theta_\alpha (I_\alpha)_{\beta\gamma}\pi_\gamma = -i\theta_\alpha(-i\epsilon_{\alpha\beta\gamma})\pi_\gamma = \epsilon_{\beta\alpha\gamma}\theta_\alpha\pi_\gamma. \tag{126}$$

So $\delta\vec{\pi} = \vec{\theta} \times \vec{\pi}$ is the change in the vector $\vec{\pi}$ under an infinitesimal c.c. rotation by $|\vec{\theta}|$ about the $\vec{\theta}$ axis.

• Another interesting isospin multiplet consists of the anti-nucleons $(\bar{n}, \bar{p})$ with $I_3 = (\frac{1}{2}, -\frac{1}{2})$. The vector space spanned by the anti-nucleons carries the conjugate of the doublet representation, denoted $\bar{\mathbf{2}}$ or $\mathbf{2}^*$. For $\mathrm{SU}(2)$ it turns out that $\mathbf{2}$ and $\bar{\mathbf{2}}$ are equivalent representations. In other words, by a suitable choice of basis for the anti-nucleon Hilbert space, the anti-nucleon doublet transforms in the same way as the nucleon doublet. To see this, recall that under a rotation about the second axis, the nucleon doublet transforms as $N \to N' = U(\theta)N$ or $p' = cp - sn$ and $n' = sp + cn$, where $c = \cos\theta/2, s = \sin\theta/2$. Applying charge conjugation $Cp = \bar{p}, Cn = \bar{n}$ we get

$$\bar{n}' = c\bar{n} + s\bar{p}, \quad \text{and} \quad \bar{p}' = -s\bar{n} + c\bar{p}. \tag{127}$$

So when (p, n) transforms via $U(\hat{y}, \theta)$, $(\bar{n}, \bar{p})$ transforms via $U(\hat{y}, -\theta)$. However, if we reverse the sign of $\bar{p}$ and define the anti-nucleon doublet as $\bar{N} = (\bar{n}, -\bar{p})^t$, then $\bar{N}$ transforms in the same way as N:

$$\bar{n}' = c\bar{n} - s(-\bar{p}), \quad \text{and} \quad -\bar{p}' = s\bar{n} + c(-\bar{p}) \tag{128}$$

More generally, check that $\bar{N} = (\bar{n}, -\bar{p})^t$ transforms via $U = \exp(-\frac{1}{2}i\vec{\theta}\cdot\vec{\tau})$, just as $N = (p, n)^t$ does. This shows that the fundamental/defining $\mathbf{2}$ representation of SU(2), and the conjugate representation $\bar{\mathbf{2}}$ are equivalent. This also explains why the isospin-1 triplet of pions is written $\pi^+ = u\bar{d}, \pi^0 = \frac{1}{\sqrt{2}}(d\bar{d} - u\bar{u}), \pi^- = d\bar{u}$ in terms of the quark doublet $(u, d)^t$ and anti-quark doublet $(\bar{d}, -\bar{u})^t$. There is also an isospin singlet pseudoscalar meson which may be written as $\eta = \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d})$ and another isosinglet pseudoscalar meson $\eta' = s\bar{s}$. The pions kaons and η, η' comprise the pseudoscalar meson nonet. All these have zero orbital angular momentum. The corresponding excited states (particles) with orbital angular momentum one are the nonet of vector mesons, ($\rho^\omega, \rho^0, K^*, \phi, \omega$).

• In general, if $D(g)$ are the matrices representing a group G with elements g. Then the matrices $D(g)^*$ also provide a representation of the group. It is called the complex conjugate representation. In the case of the fundamental representation of SU(2), $D(\theta) = e^{-i\theta\cdot\tau/2}$, so $D(\theta)^* = e^{i\theta\cdot\tau^*/2}$. Now check that $-\tau_a^*$ satisfy the same commutation relations as τ_a: $[\tau_a, \tau_b] = 2i\epsilon_{abc}\tau_c$. So while $\vec{\tau}$ furnish the fundamental $\mathbf{2}$ representation of su(2) Lie algebra, $-\vec{\tau}^*$ furnish the $\mathbf{2}^*$ representation. Find a unitary equivalence between the two.

• The rules for addition of angular momenta also apply to isospins. E.g., the total isospin observable for a system with B nucleons is defined as the sum of the isospin operators for each $\mathbf{I} = \mathbf{I}^{(1)} + \ldots + \mathbf{I}^{(1)}$. Isospin multiplets higher than $3/2$ do not appear in the spectrum of hadrons, presumably because mesons made of up and down quarks and anti-quarks must belong to $\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$ and baryons/anti-baryons have three up or down quarks/anti-quarks must belong to $\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} = \frac{1}{2} \oplus \frac{1}{2} \oplus \frac{3}{2}$. Of course, there are mesons with a strange quark or anti-quark and only one up or down quark or anti-quark, such as kaons, they must live in an $I = \frac{1}{2}$ multiplet coming from $\frac{1}{2} \oplus 0 = \frac{1}{2}$. Similarly there are baryons with strangeness (or charm) equal to ± 1 which could belong to $\frac{1}{2} \otimes \frac{1}{2} \otimes 0 = 1 \oplus 0$ (e.g. the $S = -1$ isotriplet $\Sigma^{\pm}, \Sigma^0$) or $S = \pm 2$ which would belong to $\frac{1}{2} \otimes 0 \otimes 0 = \frac{1}{2}$, like the Ξ^-, Ξ^0.

• However, there are nuclei with baryon number $B > 1$ with isospin more than $3/2$. All nuclear states of a given I multiplet are approximately degenerate in energy.

• When electromagnetic and weak interactions can be ignored, the proton and neutron are no longer distinguishable particles but rather like two spin states of an electron. Fermi statistics now implies that the state vector of a multi-nucleon state must be anti-symmetric under exchange of any two nucleons. For example, this helps to constrain the possible states of two-nucleon systems.

6.6 Two nucleon states

• Consider a two nucleon system (in a bound or scattering state). The states of definite total isospin I and I_3 consist of a triplet of $I = 1$ states that we may denote $pp, \frac{1}{\sqrt{2}}(pn + np), nn$ with $I_3 = 1, 0, -1$ and an $I = 0, I_3 = 0$ iso-singlet $\frac{1}{\sqrt{2}}(pn - np)$. The triplet states are symmetric under exchange while the singlet is anti-symmetric. It is generally the case that all members of an irreducible isospin multiplet of a multi-nucleon system behave in the same way under exchange of any pair of nucleons, since $I_{\pm} = \sum_j I_{\pm}^{(j)}$ which transform members of the multiplet into one another are symmetric under exchange. As with ordinary angular momentum, the states of the triplet transform as a vector under rotations in isospin space while the singlet is invariant.

• Now, the near equality of proton and neutron masses and their strong interactions may be interpreted as a symmetry of the strong nuclear hamiltonian H under exchange of neutron and proton. However, this does not require invariance of H under arbitrary rotations in isospin space. Indeed, a term in H such as CI_3^2 has the same value (C/4) for the proton and neutron, but it is only invariant under rotations about the 3rd axis of isospin space. So do we really need the nuclear hamiltonian to be isospin invariant?

• Gottfried and Weisskopf use experimental facts about two nucleon states to argue that nature has chosen invariance under all rotations in isospin space as the realization of the equivalence of n and p. For the 2 nucleon system, the term CI_3^2 has the same value C for $|pp\rangle$ and $|nn\rangle$ but vanishes for $pn, np, pn + np$ and $pn - np$. This means that the states of the $I = 1$ triplet are not degenerate in energy if $C \neq 0$. However, it is found that the energy of a 2 nucleon state depends only on whether the state belongs to an iso-triplet or is an iso-singlet (i.e., on the value of I) and not on the value of I_3 or on whether the state contains two 'like' nucleons or 'unlike' nucleons. Using such data, we infer that the nuclear hamiltonian can depend on I but not on the magnitude of the projection of $\vec{I}$ on any particular axis, i.e., is invariant under rotations in isospin space.

• The difference in energies of triplet and singlet states can be attributed to a term in the hamiltonian such as $H' = \mathbf{I}_1 \cdot \mathbf{I_2}$. Being a scalar product H' is invariant under rotations in isospin space. In fact, $H' = \frac{1}{2}(\mathbf{I}^2 - \mathbf{I}_1^2 - \mathbf{I}_2^2) = \frac{1}{2}(I(I+1) - 3/2)$. So H' takes the values $-3/4$ and $1/4$ for the singlet and triplet and can explain the singlet-triplet energy splitting.

6.7 The Deuteron

• The deuteron is the nucleus of deuterium, the second isotope of Hydrogen. It was discovered and isolated by H Urey (along with Brickwedde and Murphy) in 1931 (Nobel prize in Chemistry, 1934). The initial discovery was through the observation of very weak 'satellite' atomic spectral lines in addition to the usual spectral lines from a sample of hydrogen that also contained deuterium (about 1 part in 10000 in naturally occurring samples). The mass of the nucleus m_N affects the wavelengths of atomic transitions through the reduced mass $m = \frac{m_e m_N}{m_e + m_N}$ that enters the Bohr energy spectrum $-\frac{1}{2}mc^2\alpha^2/n^2$.

• The deuteron is a proton neutron bound state. The deuteron is to nuclear physics what the hydrogen atom is to atomic physics, both are two-body systems to first approximation. The deuteron is a rich source of information on inter-nucleon forces. But it is not as simple as the hydrogen atom since the inter-nucleon force (even in a non-relativistic approximation) unlike the electrostatic Coulomb force is non-central (a.k.a. 'tensor force'), it depends on the orientation of their spins (via $\vec{S}_1 \cdot \vec{r}$ and $\vec{S}_2 \cdot \vec{r}$), as well as on separation r. The deuteron ground state is not quite spherically symmetric (slightly cigar shaped) unlike the spherically symmetric ground state of hydrogen. Nevertheless, like the 1S state of hydrogen, the deuteron has zero electric dipole moment due to cancellations. What is more, nucleons and pions in a nucleus are not elementary particles like electrons, nuclei and photons in an atom. In particular the 1 Fermi size of nucleons is comparable to the size of the deuteron while the sizes of the electron (point-like) and nucleus (fm) are much smaller than the angstrom size of an atom. So despite its apparent simplicity, the deuteron is a somewhat complicated system when viewed in terms of quarks and gluons. Nevertheless, a good deal about the inter-nucleon strong nuclear force can be learned from a study of the deuteron. We confine ourselves to a very brief introduction largely ignoring the non-central nature of the inter-nucleon force.

• There are other two body bound states in classical (e.g. Kepler problem) and quantum physics. An electron-positron bound state is called positronium, its bound state spectrum is given by the Bohr formula for Hydrogen with the reduced mass given by half the electron mass. Positronium is short-lived due to e^+e^- annihilation. Electron-electron bound states (Cooper pairs) occur in BCS superconductors. Here the attractive force is very weak and long ranged, due to phonons (ionic lattice vibrations) in the material, so the electrons in a Cooper pair can be several nanometers apart. Since the attractive force is very weak (milli eV), the temperature must be sufficiently low for thermal fluctuations not to break up the pair, hence T_c is low. Moreover, the electrons in a Cooper pair can either be in an $S = 0$ (anti-symmetric singlet) state or an $S = 1$ (symmetric triplet) state since $\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$. Most of the traditionally studied low-T_c superconducting materials are well modelled by $S = 0$ Cooper pairs. The pair as a whole is a boson and many pairs can Bose condense. However, Fermi statistics applies to the electrons making up the pair. An anti-symmetric $S = 0$ spin state for the pair must go with a symmetric spatial wave function, i.e. an $l = 0$ S-wave. So we say that traditional superconductors involve S-

wave Cooper pairs. There are some superconductors where the spin state of the Cooper pair is a member of the symmetric $S = 1$ triplet, whose spatial wave function must be anti-symmetric. The simplest such possibility is an $l = 1$ P-wave spatial orbital, which is anti-symmetric under exchange ($(-1)^{l=1} = -1$). The latter are called P-wave superconductors.

• The deuteron is the ground state of the two nucleon system, with binding energy 2.2 MeV, which is small enough compared to nucleon rest energies to permit a non-relativistic quantum mechanical treatment. On the other hand, the binding energy is large enough to make beta decay of the neutron into a proton energetically disallowed (more on this when we discuss beta decay). Free neutrons in the early universe decayed to protons; neutrons that are left over today are those that were 'saved' by being bound inside deuterons (and to a lesser extent inside other light nuclei)!

• Since nucleons have spin half and isospin half, the deuteron must have spin either zero or one and isospin either zero or one from $\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$. The deuteron is experimentally found to have a non-zero magnetic dipole moment ($\mu_d = 0.857\mu_B$ where $\mu_B = e\hbar/2m_p c$ is the nuclear 'Bohr' magneton and m_p is the nucleon mass). This is not possible if the deuteron is spin-less, since the magnetic moment would not have a preferred direction to point in. So the deuteron has spin $S = 1$ and must be symmetric under exchange of spins of the two nucleons. What is more, though the deuteron has *zero electric dipole moment*, it has a non-zero *electric quadrupole moment*. A non-zero electric quadrupole moment is disallowed for spin zero and spin half particles, so this too is consistent with deuteron spin being one rather than zero. A further piece of experimental data is that the deuteron has even parity.

• To obey Fermi-Dirac statistics (nucleons have spin half), the wave function must be anti-symmetric under exchange of position and isospin of the two nucleons, since it is symmetric under exchange of spins. Since nucleons have isospin half, the deuteron can have isospin $I = 0$ or 1. In most systems with central potentials (e.g. hydrogen atom), the minimum energy state is one with zero orbital angular momentum, i.e., the symmetric S-wave state. If we ignore the non-central forces between nucleons (of the form $(\vec{S}_1 \cdot \mathbf{r})(\vec{S}_2 \cdot \mathbf{r})/r^2$) and work with the central Yukawa potential between nucleons, then we expect the deuteron ground state to be an $l = 0$ S-wave state which is symmetric under exchange. (When one includes non-central spin dependent forces, the g.s. of the deuteron is found to include a small (few percent) $l = 2$ D-wave ad-mixture. Note that even-l is needed for the even parity $(-1)^l$ of the deuteron.) The Pauli principle then implies the deuteron wavefunction must be anti-symmetric under exchange of isospins, i.e. it must be in an iso-singlet $I = 0$ state.

• Note that a two nucleon bound state with $I = 1$ would be an nn or pp or $pn+np$ isospin state, all of which should have the same energy since isospin rotations are a symmetry of the strong forces (energy should not depend on the projection I_3). However, no pp or nn bound state has been observed. This is interpreted as implying that while pion exchange between nucleons in the $I = 0$ channel provides a sufficiently strong attractive force to produce a deuteron bound state, the force in the $I = 1$ channel is not adequate to produce bound states. However, scattering states with $I = 1$ do exist.

• On the other hand, the attractive nuclear force in the $I = 0$ channel is too short-ranged (~ 2 fm) to support any excited bound states of a proton and neutron with higher orbital angular momentum. So the deuteron is rather weakly bound and lucky to exist. Indeed, the helium nucleus is much more tightly bound with a binding energy of about 28.3 MeV. However, nuclei with baryon number more than two do possess rotational excitations.

• To estimate the spatial dependence of the deuteron wave function, we could work in the approximation where non-central forces are ignored, then we simply have to solve the non-relativistic Schrodinger eigenvalue problem in a Yukawa potential $-ge^{-\mu r}/r$ and find the spherically symmetric g.s. wave function $\psi(r)$. Since it is a short ranged potential, the wave function should vanish exponentially fast for large r. For small r the Yukawa potential is the same as the Coulomb potential so we expect the g.s. wave function to approach a non-zero constant at $r = 0$, as for the hydrogen atom where $\psi_{1S}(r) \propto e^{-r/2a_0}$. Though an analytical gs wave function for the Yukawa potential is not available, excellent approximate solutions are possible, though we do not discuss them here.

• As another application of isospin conservation in the strong interactions, consider the production of deuterons in nucleon-nucleon scattering. There are three strong interaction processes resulting in a deuteron and pion (to conserve charge) in the final state (a) $pp \to d\pi^+$ and (b) $pn \to d\pi^0$ and (c) $nn \to d\pi^-$. The first two happen when protons are fired at a hydrogen or other nuclear target while the third requires a neutron beam. In all cases, the final state has $I = 1$, due to the pion. Since isospin is conserved in strong interactions, the reactions can only proceed via the $I = 1$ channel even if the initial state has both $I = 1$ and $I = 0$ components present. In the pp and nn reactions (a) and (c), the initial state is necessarily one with $I = 1$ and the two are simply related by a rotation in isospin space (i.e. application of I_- twice). So we should expect their cross sections to be equal (as is found experimentally). In collision (b) the tensor product initial state $|p\rangle|n\rangle = (|I = 1, I_3 = 0\rangle + |I = 0, I_3 = 0\rangle)/\sqrt{2}$ could be either in an $I = 0, I_3 = 0$ $(pn - np)/\sqrt{2}$ state or in an $I = 1, I_3 = 0$, $(pn + np)/\sqrt{2}$ state with equal probability. But the reaction proceeds exclusively through the $I = 1$ channel. Thus we expect $\sigma_{pp} : \sigma_{pn} : \sigma_{nn} = 2 : 1 : 2$ as is experimentally observed. An analogous calculation is discussed in more detail in the context of πN scattering in the next section.

• The three proton nucleus Li_3^3 with all protons in S wave states does not exist. Such a state would be symmetric under exchange of positions and isospins. But there are only two possible spin projections for a proton and one cannot accommodate a third proton in the same orbital by Pauli exclusion. Remarkably, analogous three quark states exist. $\Delta^{++} = uuu$ (and similarly $\Omega^- = sss$) is made possible because the state of the three quarks is anti-symmetric in color but symmetric in position, spin and isospin (flavor).

• $\alpha_2^4(nnpp)$ is a 4 nucleon bound state with $J = I = 0$ and a large binding energy of 28.3 MeV. Nuclei with $I = 0$ are particularly tightly bound and stable, they are called magic nuclei.

6.8 Isospin symmetry applied to strong interactions of pions and nucleons

• A classic application of isospin invariance of the strong interactions concerns cross sections for π-nucleon elastic scattering, which played an important role in the discovery and analysis of the $I = 3/2$ multiplet of excited states of the nucleon (Δ resonances) in 1952 by Fermi and collaborators at the 450 MeV Chicago synchrocyclotron. They measured cross sections for scattering of charged pions off a hydrogen target using pion beams of energy up to a few 100 MeV. At these energies, they expected three types of reactions for negative pions: elastic scattering ($\pi^- p \to \pi^- p$), charge exchange ($\pi^- p \to \pi^0 n$) and radiative capture ($\pi^- p \to n\gamma$) but only elastic scattering for π^+, $\pi^+ p \to \pi^+ p$. So they expected the cross sections to satisfy $\sigma_{\pi^- p} \gg \sigma_{\pi^+ p}$. However, they found the opposite! Let us analyze the reactions using isospin invariance.

• Consider the 6 possible strong scattering processes between charged pions and nucleons $\pi N \to \pi N$, some are elastic while others involve charge exchange

$$\begin{array}{llllll} (a) & \pi^+ p \to \pi^+ p & (b) & \pi^- n \to \pi^- n & (c) & \pi^- p \to \pi^- p \\ (d) & \pi^- p \to \pi^0 n & (e) & \pi^+ n \to \pi^+ n & (f) & \pi^+ n \to \pi^0 p \end{array} \qquad (129)$$

An expression such as $\pi^+ p$ for the colliding particles means that π^+ is the projectile and p is the target. We regard this state as living in the tensor product Hilbert space $\mathbb{C}^3 \otimes \mathbb{C}^2$. In fact $|\pi^+\rangle \otimes |p\rangle$ is one of the 6 basis vectors in the uncoupled basis (labelled by $I_3^{(\pi)}, I_3^{(N)}$) for this Hilbert space. Since $I_\pi = 1$ and $I_N = \frac{1}{2}$, a pion nucleon system can have total $I = \frac{1}{2}$ or $I = \frac{3}{2}$. E.g., in (a) and (b) both LHS and RHS must have $I = 3/2$ since $I_3 = 3/2, -3/2$ in these two cases. On the other hand the other 4 combinations $\pi^- p, \pi^0 n, \pi^+ n, \pi^0 p$ are linear combinations of $I = \frac{1}{2}$ and $I = 3/2$ states, as determined by their Clebsch-Gordan coefficients, which relate uncoupled basis states to coupled basis states $|I, I_3\rangle$.

$$\begin{aligned} |\pi^+ p\rangle &= |\frac{3}{2}, \frac{3}{2}\rangle, |\pi^+ n\rangle = \sqrt{\frac{1}{3}}|\frac{3}{2}, \frac{1}{2}\rangle + \sqrt{\frac{2}{3}}|\frac{1}{2}, \frac{1}{2}\rangle, \;\; |\pi^0 p\rangle = \sqrt{\frac{2}{3}}|\frac{3}{2}, \frac{1}{2}\rangle - \sqrt{\frac{1}{3}}|\frac{1}{2}, \frac{1}{2}\rangle, \\ |\pi^0 n\rangle &= \sqrt{\frac{2}{3}}|\frac{3}{2}, -\frac{1}{2}\rangle + \sqrt{\frac{1}{3}}|\frac{1}{2}, -\frac{1}{2}\rangle, \;\; |\pi^- p\rangle = \sqrt{\frac{1}{3}}|\frac{3}{2}, -\frac{1}{2}\rangle - \sqrt{\frac{2}{3}}|\frac{1}{2}, -\frac{1}{2}\rangle, \;\; |\pi^- n\rangle = |\frac{3}{2}, -\frac{3}{2}\rangle. \end{aligned}$$

Since strong interactions are invariant under rotations in isospin space, they can only depend on I, not I_3. So pion nucleon scattering is described by just two isospin amplitudes (or matrix elements) $A_{3/2}$ and $A_{\frac{1}{2}}$. These amplitudes are not fixed by consideration of isospin invariance, but depend on the detailed dynamics (space-time, isospin and spin dependence of hamiltonian). We say that each reaction can proceed through either the $I = 3/2$ or $I = 1/2$ channel. More precisely, the amplitude for a process $i \to f$ is the matrix element $M_{fi} = \langle f|H|i\rangle$ and the cross section σ is proportional to $|M_{fi}|^2$. Here $|i\rangle$ and $|f\rangle$ are initial and final states such as $|\pi^- p\rangle$ and $|\pi^0 n\rangle$ etc., and the proportionality factor K is the same for all reactions (at the same collision energy) ignoring mass differences within an isospin multiplet. The initial and final states are uncoupled basis states, but based on isospin rotation invariance, we know somethings about the matrix elements of the interaction hamiltonian between coupled basis states. So we write the uncoupled basis states as linear combinations of $I = 3/2$ and $I = \frac{1}{2}$ states. Due to isospin conservation the strong interaction hamiltonian H has zero matrix elements between states of different I. So the only two amplitudes that could be non-zero are

$$A_{\frac{1}{2}} = \langle \frac{1}{2}, *|H|\frac{1}{2}, *\rangle \quad \text{and} \quad A_{3/2} = \langle 3/2, **|H|3/2, **\rangle. \qquad (130)$$

where $*, **$ are any values of I_3 (the amplitude cannot depend on them since $[H, I_\pm] = 0$). Now we may express the cross sections σ_a for the 6 reactions in terms of these two amplitudes.

For example, the first two reactions $\pi^+ p \to \pi^+ p$ and $\pi^- n \to \pi^- n$ can only proceed via the $I = 3/2$ channel, so $\sigma_a = \sigma_b = K|A_{3/2}|^2$. For (c) $\pi^- p \to \pi^- p$, (d) $\pi^- p \to \pi^0 n$ (e) $\pi^+ n \to \pi^+ n$ and (f) $\pi^+ n \to \pi^0 p$ we get

$$\sigma_c = K \left| \frac{1}{3} A_{3/2} + \frac{2}{3} A_{\frac{1}{2}} \right|^2, \qquad \sigma_d = K \left| \sqrt{\frac{2}{9}} A_{3/2} - \sqrt{\frac{2}{9}} A_{\frac{1}{2}} \right|^2,$$
$$\sigma_e = K \left| \frac{1}{3} A_{3/2} + \frac{2}{3} A_{\frac{1}{2}} \right|^2, \qquad \sigma_f = K \left| \sqrt{\frac{2}{9}} A_{3/2} - \sqrt{\frac{2}{9}} A_{\frac{1}{2}} \right|^2. \tag{131}$$

In the 1952 Rochester conference, Fermi reported on measurements of cross sections of three of these reactions at the 450 MeV Chicago synchrocyclotron

$$(1 = a) \quad \pi^+ p \to \pi^+ p \quad (2 = d) \quad \pi^- p \to \pi^0 n \quad \text{and} \quad (3 = c) \quad \pi^- p \to \pi^- p. \tag{132}$$

These are selected since it is easier to have a proton target (liquid hydrogen) than a neutron target. They found that the cross sections were in the ratio $\sigma_1 : \sigma_2 : \sigma_3 \approx 9 : 2 : 1$ (the measured cross sections in millibarns are $195 : 45 : 22$). From our calculation (which was first done by Heitler in 1946), we know that

$$\sigma_1 : \sigma_2 : \sigma_3 = |A_{3/2}|^2 : \frac{2}{9} |A_{3/2} - A_{\frac{1}{2}}|^2 : \frac{1}{9} |A_{3/2} + 2A_{\frac{1}{2}}|^2. \tag{133}$$

Fermi concluded that the scattering proceeded primarily through the $I = 3/2$ channel, i.e., that $A_{\frac{1}{2}} \approx 0$.

• The cross section was measured from the attenuation of a pion beam while traversing a liquid hydrogen target. There were several peaks in the cross section, with the first at an invariant mass of 1232 MeV (see fig 3.8 in Perkins 4th ed. p.92) which was interpreted as the Δ resonance. The measured ratio of cross sections at the Δ resonance was

$$\frac{\sigma_1}{\sigma_2 + \sigma_3} = \frac{\sigma(\pi^+ p \to \pi^+ p)}{\sigma(\pi^- p \to \pi^- p \text{ or } \pi^0 n)} = 3. \tag{134}$$

This too is consistent with the assumption that in the neighborhood of the resonance, scattering proceeds through the $I = 3/2$ channel, since if we ignore $A_{\frac{1}{2}}$, then $\sigma_1/(\sigma_2 + \sigma_3) = 3$.

• The πp scattering was analyzed using partial waves. At the low energies considered one expects contributions primarily from $s(l = 0)$ and $p(l = 1)$ waves. If pions and nucleons have spin zero and half, then there are three possibilities for the spectroscopic terms of the πp state: l_j: $s_{\frac{1}{2}}, p_{\frac{1}{2}}, p_{3/2}$ and two possible isospin channels $I = \frac{1}{2}, 3/2$. The phase shifts were determined from the angular distribution[35] and it was found that the phase shift in the $l = 1, J = 3/2, I = 3/2$ channel changed by π around a pion beam kinetic energy of 200 MeV (or total energy 340 MeV) corresponding to the '$3, 3'$ ($I = 3/2, J = 3/2$) resonance Δ with an invariant mass of 1236 MeV.

[35]Phase shift calculations we done with the aid of the MANIAC (Mathematical Analyzer Numerical Integrator And Computer designed by von Neumann and built by N Metropolis) at Los Alamos by N Metropolis and E Fermi.

6.9 Alpha decay

• Alpha decay $N(A+4, Z+2) \to N'(A,Z) + \alpha_2^4$ is the most common type of radioactive decay of heavy unstable nuclei including naturally occurring U-238, Thorium-232 and Radon-222. Decay half lives $t_{\frac{1}{2}}$ vary greatly from billions of years to a few micro seconds even though the range of alpha particle kinetic energies is relatively narrow 5-10 MeV. Moreover the $t_{\frac{1}{2}}$ increases rapidly with decreasing alpha particle kinetic energies. This is encoded in the empirical Geiger-Nuttal rule/law proposed in 1911

$$\log t_{1/2} = \frac{aZ}{\sqrt{E}} + b \quad \text{or} \quad \log W = -\frac{aZ}{\sqrt{E}} + c \tag{135}$$

where $W = \frac{\Gamma}{\hbar} = \frac{1}{\tau} = \frac{\log 2}{t_{1/2}}$ is the decay rate (equal to the exponential decay constant λ) and Γ the energy width of the decay and τ the mean lifetime. a, b are approximately constant and Z is the atomic number of the daughter nucleus.

• Gamow and independently Condon and Gurney proposed a model for alpha decay based on quantum mechanical tunneling, and explained the Geiger-Nuttal law in 1928. First, the process may be treated non-relativistically since the alpha particle k.e. of 5-10 MeV is much less than its rest energy. The alpha particle is held inside the nucleus by the strong nuclear force, which however is short ranged. Outside the nucleus the doubly charged alpha particle sees an electrostatic Coulomb barrier due to the charged daughter nucleus. The nuclear potential is complicated, but it is sufficient for our purposes to model it as a finite attractive square well. Suppose r_0 is the maximum distance between the daughter nucleus and the alpha particle over which the nuclear force operates. Then the potential seen by the alpha particle is taken as $V(r) = -V_0$ for $r \leq r_0$ and $V(r) = Z_\alpha Z_d e^2/4\pi r$ for $r > r_0$. Here $Z_\alpha = 2$ and Z_d is the atomic number of daughter nucleus. The height of the barrier seen by the alpha particle was known to be at least twice as high ($\gtrsim$ 20 MeV) as the typical k.e. E of an emitted alpha particle. However, there is a non-zero probability that it may tunnel through the barrier. Let r_1 be the distance at which the alpha particle emerges after tunneling out. Then $E = \frac{Z_d Z_\alpha e^2}{4\pi r_1}$. The classically forbidden region is $r_0 \leq r \leq r_1$ and the maximum height of the potential is $V(r_0) = Z_d Z_\alpha e^2/4\pi r_0$. Since the tunneling probability is small we use the semi-classical formula obtained via the WKB approximation. A simple way of obtaining this formula is to recall that the semi-classical approximation to the time-independent Schrodinger equation is given by the time-independent Hamilton-Jacobi equation. It is obtained by putting $\psi = e^{iW/\hbar}$ and working to leading order in $\hbar$. One obtains the time-independent HJ equation for Hamilton's characteristic function $W(x)$, $E = H(x, W'(x))$ where p = $W'(x)$. If $H(q,p) = p^2/2m + V(x)$ then $W'(x)^2/2m + V(x) = E$. The solution is $W(x) = \int_{x_0}^x \sqrt{2m(E - V(x))dx}$. Now if we are in a classically forbidden region $E < V$, the WKB amplitude becomes $e^{-\gamma} \equiv e^{-\frac{1}{\hbar}\int_{x_0}^x \sqrt{2m(V-E)}dx}$. The transmission probability for tunneling in 1d across a classically forbidden interval from r_0 to r_1 is therefore

$$e^{-2\gamma} = \exp\left[-\frac{2}{\hbar}\int_{r_0}^{r_1} \sqrt{2m(V(r) - E)}dr\right]. \tag{136}$$

To find the decay probability per unit time or decay rate W, we must multiply this probability with the number of times per unit time that the alpha particle arrives at r_0. If we suppose that the alpha particle is simply bouncing between the walls of the nucleus $2r_0$ apart with a speed determined by $\frac{1}{2}mv^2 = E + V_0$, then it arrived at r_0 $v/2r_0$ times per unit time. Thus

$$W = e^{-2\gamma} \frac{\sqrt{\frac{2}{m}(E+V_0)}}{2r_0}. \tag{137}$$

The integral for γ may be expanded for small r_0/r_1 in a Laurent series in $\sqrt{E}$. Show that

$$2\gamma = \pi\alpha Z_\alpha Z_d \sqrt{\frac{2mc^2}{E}} - 4\sqrt{\alpha Z_\alpha Z_d}\sqrt{\frac{2mc^2 r_0}{\hbar c}} + \mathcal{O}(\sqrt{E}) = \frac{2K_1 Z_d}{\sqrt{E}} - 2K_2\sqrt{Z_d r_0} \tag{138}$$

Here $K_1 = 1.98$ MeV$^{1/2}$ and $K_2 = 1.485$ fm$^{-1/2}$ (see Griffiths, Intro to QM Ch. 8). It is clear that $\log W$ behaves like $1/\sqrt{E}$ to leading order, plus a constant plus a term proportional to $\log E$ coming from $\log(v/2r_0)$. In estimating $\log W$ we may take $V_0 = 0$, its value affects the speed of the alpha particle inside the potential well but only has a sub-leading effect on the energy dependence of $\log W$. A numerical estimate for r_0 may be obtained from the empirical law of approximate constant density of nuclear matter, A/r^3 = constant. The radius of a nucleus of atomic mass A is $r \approx (1.07fm)A^{1/3}$. Use this to estimate the constants in the Geiger-Nuttal law for U-238.

6.10 Brief introduction to nuclear beta decay

• Nuclear beta decay was discovered by H. Becquerel (1896) long before the discovery of the nucleus (1909-13) or the discovery of the neutron (1932). Though discovered by chance in the context of natural radioactivity of uranic salts (at that time, the name 'uranic salts' was applied to salts of several nearby elements in today's periodic table, not just isotopes of uranium), beta decay is the prime example of a weak interaction process and provides a window into the sub-nuclear world of elementary particles. Its theoretical understanding contributed greatly to the theory of weak interactions, parity violation and the electroweak standard model.

• In beta decay, a nucleus with mass number B and atomic number Z decays to a nucleus with same mass number B and atomic number $Z+1$ along with an increase in nuclear charge by $+e$ and the emission of two particles, an electron (β^- particle) and an anti-neutrino. The (anti)neutrino was not initially detected and current constraints suggest that its mass is less than an eV/c^2. The basic reaction is the decay of a neutron inside the nucleus $n \to p^+ + e^- + \bar{\nu}$, so the number of protons and charge increase by one. E.g. Thorium(234,90) beta decays to Protactinium(234,91) with a half life of 24 days. Carbon(14,6) decays to Nitrogen(14,7) $+e^- + \bar{\nu}_e$ with a half life of 5730 years.

• Free neutrons beta decay with a half-life of 10.5 minutes. The mass difference $m_n - m_p - m_e = .782$ MeV manifests as kinetic energy of decay products. But neutron beta decay in a nucleus may or may not happen depending on the nuclear binding energies. Typically, replacing a neutron by a proton in a nucleus costs energy, since the proton is less strongly bound than a neutron, due to electric repulsion by other protons. The binding energy of a neutron in a nucleus is usually of order of a few MeV, and more than that of a proton. This tends to inhibit β^- decay.

Indeed, neutrons in a Helium(4,2) nucleus (α particle) or any other stable nucleus do not decay. But beta decay does take place if the cost of replacing a neutron with a proton in the nucleus is outweighed by the mass difference $m_n - m_p - m_e = .782$.

• More precisely, following Gottfried and Weisskopfs's notation, let $\Phi_j(N, Z)$ denote a nucleus with N neutrons and Z protons in a state labelled by j (j refers to other quantum numbers like angular momentum, as in an atom). Its total mass is denoted $M_j(N, Z)$. The beta decay $\Phi_j(N, Z) \to \Phi_{j'}(N-1, Z+1) + e^- + \bar{\nu}_e$ is energetically allowed if the mass of the initial state exceeds that of the final state

$$M_j(N, Z) - M_{j'}(N-1, Z+1) > m_e + m_\nu \qquad (139)$$

This condition can be written in terms of binding energies. Denote the binding energy of a nucleus with N neutrons and Z protons in state j by $BE_j(N, Z)$. $BE > 0$ is the energy required to dissociate the bound nucleus into its far separated constituent nucleons, so

$$BE_j(N, Z) = Nm_n + Zm_p - M_j(N, Z). \qquad (140)$$

Thus the condition for beta decay to occur is also

$$BE_j(N, Z) - BE_{j'}(N-1, Z+1) < m_n - (m_p + m_e + m_\nu) \approx 0.782\, MeV. \qquad (141)$$

In other words, if the daughter is more tightly bound than the parent (this rarely happens), then beta decay can happen. More interestingly, beta decay is energetically allowed if the daughter is less tightly bound than the parent, but not by so much as to offset the mass difference of .782 MeV.

• Another observed reaction, related by crossing, is the capture of an (inner, typically K shell) atomic electron by a proton in a proton-rich nucleus $p^+ + e^- \to n + \nu_e$, to produce a neutron and neutrino. This is called K-electron capture.

• Free protons have never been observed to decay into neutrons, this reaction is energetically forbidden since $m_p < m_n$. However, inside some nuclei, the charge lowering β^+ decay $p \to n + e^+ + \bar{\nu}_e$ occurs, provided the mass increase is more than compensated by a gain in binding energy. It is related to neutron decay by crossing. E.g., Fluorine(18,9) $\to$ Oxygen(18,8) $+e^+ + \nu_e$. It often happens that replacement of a proton by a neutron lowers the energy of a nucleus, due to the reduced electrostatic repulsion. It could also happen that the neutron can occupy a lower nuclear energy level which was Pauli-excluded from being occupied by the proton that it replaces [p, n are not identical particles if mass differences and EM interactions are included].

6.10.1 Need for the neutrino in beta decay

• Neutrinos were not detected till 1956. Till then, in nuclear beta decay, the only observed outgoing particle was the electron (aside from the recoiling nucleus). Thus it seemed like a 2 body decay $N \to N' + \beta$. We have shown using energy-momentum conservation that in the rest frame of the parent nucleus, the energies of N' and β are fixed by the masses of the three particles. So the beta particle must be mono-energetic. However, experiments[36] showed that the spectrum of emitted beta particle energies was continuous, in apparent violation of energy-momentum conservation. Bohr, in the wake of the recently discovered uncertainty principle suggested that energy conservation may be true only in some statistical sense in quantum physics, so individual beta particles could come out with any energy. But this seemed to be inconsistent with the abrupt end-point in the beta energy spectrum. There was also a problem with angular momentum conservation.

• In three body decay, the daughter particles have a continuous energy spectrum. In 1930 Pauli suggested that there was in fact an undetected light uncharged (no track seen in cloud chambers) particle that was emitted along with the electron in beta decay, it carried away the missing energy. This particle was called the neutrino by Fermi. It is required by angular momentum conservation as well. Consider the beta decay $\Phi_j(N,Z) \to \Phi_{j'}(N-1,Z+1)$. The initial and final nuclei have the same number of nucleons (baryon number $B = N+Z$). Since nucleons have spin half, both nuclei have either even or odd angular momentum in units of $\hbar/2$, according as B is even or odd. It follows that the change in angular momentum is an integer multiple of $\hbar$, which must be carried away by the beta particle and other decay products. In particular, there must be an even number of fermions in the final state, aside form the daughter nucleus. Since the electron is a spin half fermion, the simplest possibility is a three body decay with the neutrino being the other fermion. Measurements of the energy and angular distribution in beta decay show that the neutrino has spin half. Current experimental bounds put $m_\nu < 1$ eV/c^2.

6.10.2 Neutrino vs anti-neutrino, and lepton number

• Neutrinos do interact with matter, though very weakly, with cross sections roughly of order $\sigma_\nu \sim E \times 10^{-11}$ mb where E is the neutrino energy in GeV[37]. For example, we have the following crossed/conjugate versions of neutron beta decay: $\nu_e + n \to p + e^-$ and $\bar\nu_e + p \to n + e^+$. These neutrino/anti-neutrino-nucleon scattering processes are also called inverse beta decay. Since 1956 they have been experimentally observed.

• Since neutrinos are neutral, one wonders whether the neutrino and anti-neutrino are the same particle, as in the case of γ, π^0, Z^0. Experimental evidence against this was obtained by R. Davis. Nuclear reactors produce a large flux of anti-neutrinos/neutrinos from decay of free neutrons. Their interactions with matter in the vicinity was studied experimentally. It was found that only positrons are produced in scattering with nucleons. If $\nu = \bar\nu$ then both the above scattering reactions would take place and we would expect both electrons and positrons to be produced in the neighborhood of reactors. This indicates that the neutrino is not its own

[36]Experiments by Otto Hahn and Lise Meitner in Berlin (1911) (who later collaborated in the discovery of nuclear fission), with further detailed confirmation from Charles Drummond Ellis at the Cavendish in the 1920s. Ellis and Chadwick were together in an internment camp in World War I where they did experiments in a horse stable. They both joined Rutherford's Cavendish lab subsequently, Chadwick studying alpha scattering and Ellis studying beta decay.

[37]This approximate formula for the cross section as a function of neutrino energy is valid for E less than the energy scale of electroweak mixing/unification (80 GeV). Cross sections cannot indefinitely increase with energy, that would violate unitarity (essentially, probabilities cannot exceed one).

anti-particle, which is assumed in the standard model.

• In weak processes, electrons or muons are always accompanied by neutrinos, e.g., beta decay $n \to p + e^- + \bar{\nu}_e$ and muon decay $\mu \to e + \bar{\nu}_e + \nu_\mu$ and pion decay $\pi^- \to \mu^- + \bar{\nu}_\mu$.

• Assuming $\nu \neq \bar{\nu}$ we may assign a quantum number L (lepton number) that is conserved in all particle reactions. We put $L(e^-) = L(\nu) = L(\mu^-) = L(\tau^-) = 1$, $L(e^+) = L(\bar{\nu}) = L(\mu^+) = L(\tau^+) = -1$ for the leptons and anti-leptons and $L = 0$ for the nucleons (and mesons). Lepton number is the analogue of baryon number for leptons. We see that beta decay $n \to p + e + \bar{n}$ conserves lepton number (zero on either side) as do all reactions related to it by crossing. On the other hand, lepton number violating processes such as $\pi^- \to \mu^- \gamma$ have never been observed. The experimental limit on mean lifetime of the L-violating process of neutrinoless double beta decay is $\tau(^{76}Ge \to ^{76} Se + 2e^-) > 10^{26}$ y (see Perkins).

• Refined versions of lepton number. Interestingly, many lepton number conserving reactions that are otherwise not forbidden, e.g. $\mu \to e\gamma$ have not been seen. So conservation of individual electron, muon and tau lepton numbers L_e, L_μ, L_τ was proposed. E.g., $L_e = 1$ for e^-, ν_e, $L_e = -1$ for their anti-particles and $L_e = 0$ for all other elementary particles of the SM. $L_{e,\mu,\tau}$ are conserved in the perturbative SM.

• With the discovery of neutrino flavor oscillations such as $\nu_\mu \leftrightarrow \nu_e$, we have to accept that the individual lepton numbers cannot be exactly conserved. However, no process involving charged leptons that violates L_e, L_μ or L_τ has been seen till now, though it is expected that the L_e and L_μ violating process $\mu \to e\gamma$ will be experimentally detected in careful searches in future.

6.10.3 Fermi's current-current interaction vertex for beta decay and weak interactions

• The typically long half-lives (minutes to years) of nuclei that undergo beta decay attest to the weakness of the force involved, which could not be purely electromagnetic, since neutrinos are uncharged. Alpha decay, could be treated using non-relativistic QM (as Gamow-Condon-Gurney did) since the alpha particle kinetic energies are much less than rest energies. By contrast, beta decay is a relativistic process since neutrinos travel at close to the speed of light and beta particle energies usually exceed their rest mass. So beta decay requires a quantum field theoretic treatment. Fermi proposed a theory of weak interactions in 1934. It was based on an analogy with the tri-linear photon-electron interaction vertex of quantum electrodynamics $qj^\mu A_\mu$ where q is the electron charge. $j^\mu = \bar{e}\gamma^\mu e$ is the electron-positron electromagnetic current. It can annihilate an electron and create an electron or create an electron-positron pair etc. (recall that the field e acting on the vacuum annihilates an electron or creates a positron, while $e^\dagger$ creates an electron or annihilates a positron and that $\bar{e} = e^\dagger\gamma^0$). We say the electromagnetic current j^μ is a neutral current. It cannot change charge since the photon field A_μ to which it couples, is uncharged). By analogy with the e^+e^- electromagnetic current, Fermi introduced 'charge changing' weak currents for pairs of particles participating in the weak interactions, e.g. $\bar{p}\gamma^\mu n$ is the charge raising proton-neutron weak current.

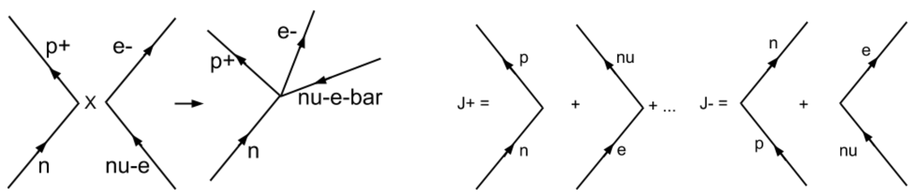
• In 1934, it was not necessary to invoke a force carrier (boson like pion or photon) for weak interactions since they seemed to occur on length scales small compared to nuclear dimensions (now we know the length scale of weak interactions is the Compton wavelength of the W

boson, 10^{-18} m). So weak interactions were modeled as point or 'contact' interactions by Fermi. Since there are four fermions involved in $n \to p + e^- + \bar{\nu}_e$, Fermi's theory involved a 4-fermion current-current vertex given by the interaction Lagrangian

$$\mathcal{L} = \frac{G_F}{\sqrt{2}} \left((\bar{p}\gamma^\mu n)(\bar{e}\gamma_\mu \nu) + (\bar{n}\gamma^\mu p)(\bar{\nu}\gamma_\mu e) \right) = \frac{G_F}{\sqrt{2}} (j^\mu J_\mu^\dagger + J^\mu j_\mu^\dagger) \tag{142}$$

The 2^{nd} term is the hermitian conjugate of the first ($\bar{p} = p^\dagger \gamma^0$ etc). G_F is the Fermi coupling constant.

• Since the four Dirac fields p, n, e, ν have dimensions $M^{3/2}$ (coming from kinetic terms $\int d^4x \bar{p}\gamma^\mu \partial_\mu p + \ldots$ in the action, which must be dimensionless), the Fermi coupling has dimensions $[G_F] = M^{-2}$. Based on measured decay rates ($W = \hbar/\Gamma$, which are proportional to G_F^2) one finds $G_F = 1.16 \times 10^{-5}$ GeV$^{-2} = 1.02 \times 10^{-5}/m_p^2$ [see Perkins]. $j^\mu = \bar{p}\gamma^\mu n$ is called the charge raising proton-neutron weak current since it can annihilate a neutron and create a proton. $J_\mu^\dagger = \bar{e}\gamma_\mu \nu$ is the charge lowering electron-neutrino weak current since it can annihilate a neutrino and create an electron (or annihilate a positron and create an anti-neutrino). So the current-current interaction $j^\mu J_\mu^\dagger$ can correspond to the process $\nu n \to p + e$. Beta decay is related to this reaction by crossing, removing the ν from the lhs and replacing it with a $\bar{\nu}$ on the rhs. The second term in $\mathcal{L}$ corresponds to processes like K-electron capture $p + e \to n + \nu$ and crossed versions thereof.



• There are similar weak currents that we may associate to other pairs of particles that participate in weak interactions, e.g. the $\mu\nu_\mu$ charge lowering weak current is $J_\lambda^\dagger = \bar{\mu}\gamma_\lambda \nu_\mu$. The vertex that governs muon decay $\mu \to e + \nu_\mu + \bar{\nu}_e$ is

$$\mathcal{L} = \frac{G_F}{\sqrt{2}} \left[(\bar{e}\gamma^\lambda \nu_e)\, (\bar{\nu}_\mu \gamma_\lambda \mu) + \text{h.c.} \right]. \tag{143}$$

Remarkably, the same Fermi constant G_F determined by beta decay rates also gives the correct muon decay rate. The charge changing weak interactions are universal in the sense that they involve a single coupling constant. To include all the charge-changing weak interactions (involving leptons and hadrons), we may define a total charge-raising weak vector current

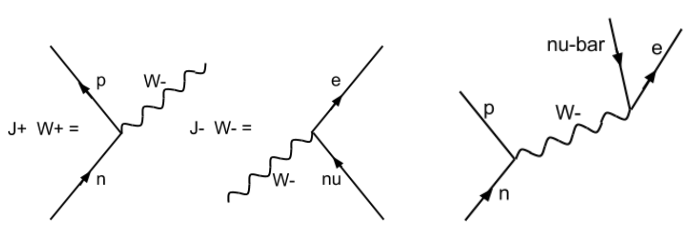
$$J_+^\lambda = \bar{p}\gamma^\lambda n + \bar{\nu}_e \gamma^\lambda e + \bar{\nu}_\mu \gamma^\lambda \mu + \ldots \tag{144}$$

The charge lowering weak current is $J_- = J_+^\dagger$ and the interaction Lagrangian is $\mathcal{L} = \frac{G_F}{\sqrt{2}} \left(\eta_{\mu\nu} J_+^\mu J_-^\nu + \text{h.c.} \right)$. This governs various weak interactions including muon capture by a nucleus $\mu^- p^+ \to n \nu_\mu$.

• For example the decay rate of the muon $W = \Gamma/\hbar = 1/\tau$ (τ is the life-time and Γ the energy width) can be calculated by treating the 4-Fermi interaction to first order in perturbation theory. W is equal to the square of the transition matrix element times phase space factors, so to leading order, $W \propto G_F^2$. Since the electron kinetic energy usually much exeeds its rest mass it is reasonable to treat it as massless. Then the only other dimensional quantity in the rest frame of the muon is the muon mass m_μ (quantities like the energy of the outgoing electron and neutrinos are summed over to arrive at the total decay rate). W being a probability per unit time has inverse mass dimensions, by dimensional analysis we must have $W \propto G_F^2 m_\mu^5$. Evaluation of the matrix element and phase space factors yields $\tau^{-1} = G_F^2 m_\mu^5/192\pi^2$ in the approximation where the electron mass is neglected.

• Subsequent discoveries (experimental and theoretical) necessitated two important modifications to Fermi's theory. Discovery of parity violation along with other experiments showed that weak currents involved in the current-current interaction are not quite vector currents but vector minus axial vector, e.g. $\bar{p}\gamma^\mu n - \bar{p}\gamma^\mu\gamma^5 n$. This was the work of R Marshak and E C G Sudarshan as well as Feynman and Gell-Mann in the late 1950s. The second modification replaced the point-like 4-fermi interaction vertex by a finite range interaction mediated by intermediate vector bosons, the weak gauge bosons $W^\pm$. A theoretical reason for this is that the zero range 4-Fermi interaction leads to a non-renormalizable quantum theory. Attempts to calculate decay rates and cross sections beyond 1st order perturbation theory give infinite answers. Models with coupling constants with negative mass dimension (e.g. $G_{\rm Fermi}$ and $G_{\rm Newton}$) are perturbatively non-renormalizable. On the other hand, the weak interaction is not quite of zero-range. It was fruitful to introduce heavy force carriers called intermediate vector bosons ($W^\pm$) to mediate the short-ranged charge changing weak interactions. Like the tri-linear $qJ_E^\mu A_\mu$ electromagnetic vertex, the weak interaction vertex is $g(J_+^\mu W_\mu^+ + J_-^\mu W_\mu^-)$. g is the dimensionless weak coupling constant like the electromagnetic coupling q. $J_\pm$ are the charge raising and lowering weak currents introduced above. Like the photon field A_μ, $W_\mu^\pm$ are the gauge fields associated to $W^\pm$ gauge bosons. Photon emission and absorption are replaced by emission and absorption of $W^\pm$. W_μ^+ can annihilate a W^+ or create a W^- while W_μ^- can annihilate a W^- or create a W^+.

• Beta decay $n \to p + e + \bar{\nu}_e$ previously thought of as a single vertex 4-fermion interaction is now viewed as two weak vertices $n \to p + W^-$ and $W^- \to e^- + \bar{\nu}_e$ (each proportional to g), separated by the propagator for an exchanged virtual W, as shown in the Feynman diagram.



• However, $W^\pm$ are very heavy, $m_{W^\pm} = 80.4$ GeV. At low momentum transfers relevant to beta decay, muon decay etc, the W propagator $\frac{1}{m_W^2 - q^2} \approx \frac{1}{m_W^2}$. Comparing with the 4-Fermi vertex, $G_F = \frac{\sqrt{2}}{8}\frac{g^2}{M_W^2}$ (the factors are conventional). Viewed from this perspective, the weak interactions are weak since they are mediated by very heavy intermediaries. From the numerical value of G_F, the weak 'fine-structure constant' $g^2/4\pi \approx .03$ is not very different from Sommerfeld's electromagnetic $\alpha = e^2/4\pi \approx 1/137 \approx .007$, indeed the two interactions are mixed in the standard model. To understand these ideas more precisely we will need to develop some field theoretic concepts.

6.11 Gamma decay and absorption

A third type of radioactive decay is gamma decay of a nucleus by photon emission. Rather than discuss gamma decay of nuclei, we will discuss the simpler but analogous topic of gamma decay of excited atoms. Atomic levels and wave functions are known better than nuclear energy levels and wave functions. To do so, we need to understand the quantum theory of the photon field and then the coupling of the photon field to bound atomic electrons. While the photon field must be treated relativistically, it suffices to treat the atomic electrons non-relativistically.

7 Quantum theory of the photon field

• Photons in the form of light are the elementary particles we are most familiar with. So let us begin with a look at the classical and quantum theory of light (radiation).

• The earliest ideas of quantization arose from Planck's attempt to fit the spectrum of black body electromagnetic radiation. The photoelectric effect and Compton scattering reinforced the need to treat light quantum mechanically, light, like electrons and atoms, displayed both wave and particle-like behavior.

• Heisenberg's and Schrödinger's development of point particle non-relativistic quantum mechanics shows that the position and momentum of a particle are subject to quantum fluctuations. This is also motivated by Heisenberg's microscope thought-experiment which suggests that the determination of position with greater accuracy would make the determination of momentum more uncertain. The electromagnetic field is also subject to quantum fluctuations. This can be motivated by Weisskopf's microscope thought-experiment. Here, we try to pin point the position of a charged point particle by measuring the electric field it produces, and fix its momentum by measuring the magnetic field produced by the current of the moving charge. If the electric and magnetic fields could be simultaneously determined, then we might be able to fix the instantaneous position and momentum of the particle, violating the Heisenberg uncertainty principle. To avoid this problem, we expect the electric and magnetic fields to display quantum fluctuations. This gives us a reason to quantize the radiation field. Once quantized, we will be able to identify the photon as a state of the quantized radiation field. However, it is harder to quantize the radiation field than a point particle. This is primarily because a non-relativistic particle or system of non-relativisitc particles has a finite number of degrees of freedom, while the EM field has an infinite number of degrees of freedom (the electric and magnetic fields at each point in space). The quantization of the EM field had to wait till the late 1920s and 1930s work of Dirac, Pauli, Heisenberg, Jordan, Fermi etc., while the quantum theory of a point particle was already formulated by Schrödinger and Heisenberg by 1926-27.

• Another new feature is the possibility for photons to be created or absorbed. The number of atomic electrons in Schrodinger's treatment is fixed. But in radiative transitions the number of photons is not conserved. So the Hilbert space must include a vacuum state of no photons, states with one photon of a definite energy $h\nu$, states with two photons etc as well as creation and annihilation operators that connect these states. These states, which can be labelled by the wave vector and polarizations $\vec{\epsilon}_\lambda$ of the photons $|\mathbf{k}_1\lambda_1;\mathbf{k}_2\lambda_2;\cdots\rangle$ realize the particle-like character of photons seen in the photoelectric and Compton effects. The quantum theory also realizes the wave-like character of light familiar from EM waves. The quantized electric and magnetic field operators do not commute with the Hamiltonian $H = \frac{1}{2}\int(E^2+B^2)d\mathbf{r}$ (recall the Weisskopf microscope). They are in fact linear combinations of photon creation and annihilation operators, and their eigenstates do not contain a definite number of photons or have definite energy. However, as we will see, the matrix element of $\mathbf{E}(\mathbf{r},t)$ (or $\mathbf{B}$) between a one photon state $|\mathbf{k},\lambda\rangle$ and the vacuum takes the form of a plane wave of wave vector $\mathbf{k}$ and frequency $\omega = \hbar|\mathbf{k}|$ with polarization ϵ_λ (or $\hat{k}\times\epsilon_\lambda$)!

• The quantum theory of the EM field is based on Maxwell electrodynamics, to which it must reduce in the classical limit. So we begin by expressing the radiation field of Maxwell theory in a manner that makes quantization straightforward by analogy with particle mechanics.

7.1 Classical radiation from Maxwell's equations in radiation gauge

• Maxwell's equations for the vacuum electric and magnetic fields in rationalized Heaviside-Lorentz units are

$$\nabla\cdot\mathbf{B}=0, \quad \nabla\times E = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}, \quad \nabla\cdot\mathbf{E}=\rho \quad \text{and} \quad \nabla\times\mathbf{B}=\frac{\mathbf{j}}{c}+\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}. \tag{145}$$

where for consistency of the two inhomogeneous equations the charge and current density must satisfy the continuity equation $\frac{\partial\rho}{\partial t}+\nabla\cdot\mathbf{j}=0$. The first two homogeneous Maxwell equations state the absence of magnetic monopoles, and Faraday's law of induction. The second pair of inhomogeneous equations are Gauss' law and Ampere's law with Maxwell's correction term involving the time derivative of the electric field (the displacement current). Gauss was German, Ampere French, Faraday English and Maxwell Scottish. The motion of a charge e in an electromagnetic field is governed by the Lorentz force law

$$\mathbf{F} = e\left[\mathbf{E}+\frac{\mathbf{v}}{c}\times\mathbf{B}\right]. \tag{146}$$

• The first pair of homogeneous Maxwell equations are identically satisfied if the fields are expressed in terms of scalar and vector potentials $(\phi,\mathbf{A})$

$$\mathbf{E} = -\nabla\phi-\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t} \quad \text{and} \quad \mathbf{B}=\nabla\times\mathbf{A}. \tag{147}$$

However, the gauge potentials $(\phi,\mathbf{A})$ are not uniquely determined by the $\mathbf{E}$ and $\mathbf{B}$ fields, more on this momentarily. In terms of the gauge potentials, the Ampere-Maxwell equation becomes (use $\nabla\times(\nabla\times\mathbf{A}) = -\nabla^2\mathbf{A}+\nabla(\nabla\cdot\mathbf{A})$)

$$-\nabla^2\mathbf{A}+\nabla(\nabla\cdot\mathbf{A}) = \frac{\mathbf{j}}{c}-\frac{1}{c}\partial_t\nabla\phi-\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}. \tag{148}$$

So introducing the scalar and vector potentials means that the first pair of homogeneous Maxwell equations have already been solved. The gauge potentials play a very important role in the quantum theory. The hamiltonian for the interaction of a charged particle with the EM field is written in terms of $\mathbf{A}$, rather than in terms of $\mathbf{E}$ or $\mathbf{B}$.

• The inhomogeneous Maxwell equations can be written in a relativistically covariant form by introducing the 4-vectors $A^\mu = (\phi, \mathbf{A})$ and $j^\mu = (c\rho, \mathbf{j})$ and the field strength tensor $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. Then the inhomogeneous Maxwell equations become $\partial_\mu F^{\mu\nu} = \frac{1}{c} j^\nu$ along with the consistency condition $\partial_\mu j^\mu = 0$ which expresses local charge conservation.

• However, $\mathbf{A}$ and ϕ are not uniquely determined by the measurable electric and magnetic fields. Two gauge potentials $(\phi, \mathbf{A})$ and $(\phi', \mathbf{A}')$ which differ by a gauge transformation

$$\mathbf{A}' = \mathbf{A} + \nabla\theta, \quad \phi' = \phi - \frac{1}{c}\frac{\partial\theta}{\partial t}. \tag{149}$$

correspond to the same electromagnetic fields. Gauge transformations form a group $\mathcal{G}$ which acts on the space of gauge potentials $\mathcal{A} = \{(\phi, \mathbf{A})\}$. Each orbit (equivalence class of gauge potentials) corresponds to an electromagnetic field $(\mathbf{E}, \mathbf{B})$ and the space of electromagnetic fields is the quotient $\mathcal{A}/\mathcal{G}$. A choice of orbit representatives is called a gauge choice. It is obtained by imposing condition(s) on the gauge potentials which are satisfied by one set of gauge potentials from each equivalence class.

• A convenient gauge choice is Coulomb gauge $\nabla \cdot \mathbf{A} = 0$. Given a vector potential $\mathbf{A}'$ we find its representative in Coulomb gauge by making the gauge transformation $\mathbf{A} = \mathbf{A}' - \nabla\theta$ with θ chosen to satisfy Poisson's equation $\nabla^2\theta = \nabla \cdot \mathbf{A}'$.

• Gauss' law simplifies in Coulomb gauge: $\nabla \cdot E = -\nabla^2\phi - \frac{\partial \nabla\cdot\mathbf{A}}{\partial t} = 0$ becomes $-\nabla^2\phi = \rho$, whose solution involves the Coulomb potential (this is why $\nabla \cdot A = 0$ is called the Coulomb gauge!) $\phi(\mathbf{r}, t) = \frac{1}{4\pi}\int d^3r' \frac{\rho(\mathbf{r}', t)}{|\mathbf{r}-\mathbf{r}'|}$. In particular, in Coulomb gauge, the scalar potential $\phi(\mathbf{r}, t)$ is not a dynamical quantity, it is entirely fixed by the instantaneous charge density. Now let us specialize to the case where there are no charges present in the interior and boundary of the region of interest, so that $\rho = 0$. Then $\phi = 0$. In the absence of charges, Coulomb gauge is called radiation gauge ($\phi = 0$, $\nabla \cdot \mathbf{A} = 0$), since electromagnetic radiation is most easily described in this gauge. Indeed, $\nabla \cdot \mathbf{A} = 0$ or $\mathbf{k} \cdot \tilde{\mathbf{A}}_k = 0$ in Fourier space means there are only two (transverse) components of the vector potential that are dynamical. These correspond to the two independent polarizations of electromagnetic radiation. In radiation gauge, the Ampere-Maxwell equation becomes

$$\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = \nabla^2\mathbf{A} + \frac{\mathbf{j}}{c}, \qquad \text{(provided } \nabla\cdot\mathbf{A} = 0, \phi = \rho = 0). \tag{150}$$

This is the vector wave equation in the presence of a current source $\mathbf{j}$. One is often interested in EM waves in vacuum, in which case $\mathbf{j} = 0$ and we get the homogeneous vector wave equation.

$$\Box\mathbf{A} \equiv \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} - \nabla^2\mathbf{A} = 0, \qquad \text{(provided } \nabla\cdot\mathbf{A} = 0, \mathbf{j} = 0, \phi = \rho = 0). \tag{151}$$

7.1.1 Fourier decomposition of $\mathbf{A}(\mathbf{r}, t)$, transversality condition and polarization

The wave equation describes EM waves, including traveling plane waves. Since the equation is linear, a super position of plane waves is also a solution. This suggests that we may express the general solution of the wave equation as a superposition of plane waves. This is what Fourier analysis does for us. We first imagine that the EM field is considered in a large cubical box of volume V and write the vector potential as a Fourier series

$$\mathbf{A}(\mathbf{r}, t) = \frac{c}{\sqrt{V}} \sum_{\mathbf{k}} \mathbf{A}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{r}} \tag{152}$$

The Fourier coefficient $\mathbf{A}_{\mathbf{k}}(t)$ is called the Fourier mode corresponding to wave vector $\mathbf{k}$. The pre-factors $c/\sqrt{V}$ are not very important and for later convenience: the formula for the electric field $-\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}$ becomes simpler. The allowed values of $\mathbf{k}$ are determined by the boundary conditions, but are not important to us since we will eventually let $V \to \infty$ so that all $\mathbf{k}$ are allowed. For simplicity, we consider the radiation field in a cubical cavity of volume V with periodic boundary conditions. This allows us to work with Fourier series. We will eventually let the sides of the box go to infinity, and the Fourier series will become Fourier integrals. The distinction is not important for us here.

• The advantage of the Fourier expansion is that the wave equation $\Box \mathbf{A} = 0$ reduces to a system of ODEs, for each mode $\mathbf{k}$

$$\ddot{\mathbf{A}}_{\mathbf{k}}(t) + c^2 \mathbf{k}^2 \mathbf{A}_{\mathbf{k}}(t) = 0 \quad \Rightarrow \quad \ddot{\mathbf{A}}_{\mathbf{k}}(t) = -\omega_k^2 \mathbf{A}_{\mathbf{k}}(t). \tag{153}$$

Thus each mode $A_{\mathbf{k}}$ evolves independently in time like a classical oscillator of angular frequency $\omega_k = c|\mathbf{k}|$. We anticipate that the time dependence of the vector potential may be written as a linear combination of $e^{i\omega_k t}$ and $e^{-i\omega_k t}$. However $\mathbf{A}_{\mathbf{k}}$ is a vector, not a scalar, so it has a direction. Which way does it point? This brings in the concept of polarization.

• The Coulomb gauge condition $\nabla \cdot \mathbf{A} = 0$ becomes

$$\nabla \cdot \mathbf{A} = \frac{ic}{\sqrt{V}} \sum_{\mathbf{k}} \mathbf{k} \cdot \mathbf{A}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{r}} \equiv 0. \quad \Rightarrow \quad \sum_{\mathbf{k}} \mathbf{k} \cdot \mathbf{A}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} = 0 \quad \text{for all} \quad \mathbf{r} \tag{154}$$

The only way for this to happen is for the individual Fourier coefficients to vanish, i.e., $\mathbf{k}\cdot\mathbf{A}_{\mathbf{k}} = 0$ for each $\mathbf{k}$. In other words, the Fourier modes $\mathbf{A}_{\mathbf{k}}$ must each be transversal (orthogonal) to the corresponding wave vectors. Thus the Coulomb gauge condition is also called the transversality condition. We will soon see that $\mathbf{k}$ is the direction of propagation of the corresponding EM wave, so we see that $\mathbf{A}_{\mathbf{k}}$ must be transverse to its wave vector.

• So we may write $\mathbf{A}_{\mathbf{k}} = \sum_\lambda A_{\mathbf{k},\lambda} \epsilon_\lambda$ as a linear combination of two basis polarization vectors $\vec{\epsilon}_1, \vec{\epsilon}_2$, which are perpendicular to $\mathbf{k}$. For convenience we choose them to be mutually orthogonal so that $\epsilon_1, \epsilon_2, \hat{k}$ form an orthonormal system:

$$\hat{\mathbf{k}} \cdot \epsilon_\lambda = 0, \quad \epsilon_\lambda \cdot \epsilon_{\lambda'} = \delta_{\lambda,\lambda'}, \quad \epsilon_1 \times \epsilon_2 = \hat{\mathbf{k}} = \frac{\mathbf{k}}{k} \tag{155}$$

$\epsilon_{1,2}(\mathbf{k})$ of course depend on $\mathbf{k}$, but for brevity, we do not display the $\mathbf{k}$-dependence explicitly.

• A single-mode EM field with $\mathbf{A}_{\mathbf{k},\lambda} \propto \epsilon_\lambda$ would correspond to an EM wave with the electric field $\mathbf{E}_\mathbf{k} \propto -\frac{1}{c}\dot{\mathbf{A}}_\mathbf{k}$ pointing along ("polarized along") the ϵ_λ direction. The corresponding magnetic field is $\mathbf{B}_\mathbf{k} \propto \mathbf{k} \times \mathbf{A}_\mathbf{k}$. We see that for a fixed Fourier mode $\mathbf{k}$, the electric $\mathbf{E}_k \propto \dot{\mathbf{A}}_\mathbf{k}$ and magnetic fields $\mathbf{B}_\mathbf{k} \propto \mathbf{k}\times\mathbf{A}_\mathbf{k}$ are both orthogonal to $\mathbf{k}$, i.e., $\mathbf{k}\cdot\mathbf{E}_\mathbf{k} = 0$ and $\mathbf{k}\cdot\mathbf{B}_\mathbf{k} = 0$. This is the statement that EM waves are transversely polarized. One choice of basis for polarization vectors is[38]

$$\epsilon_1 = \hat{x}, \quad \epsilon_2 = \hat{y} \quad \text{and} \quad \hat{\mathbf{k}} = \hat{z}. \tag{157}$$

Since $\epsilon_1, \epsilon_2, \hat{\mathbf{k}}$ form an orthonormal basis for 3d Euclidean space, we may write the identity matrix as a sum of projections to the subspaces spanned by each

$$I = \epsilon_1\epsilon_1^t + \epsilon_2\epsilon_2^t + \hat{\mathbf{k}}\hat{\mathbf{k}}^t \quad \text{or} \quad \delta_{ij} = (\epsilon_1)_i(\epsilon_1)_j + (\epsilon_2)_i(\epsilon_2)_j + \frac{k_i k_j}{k^2} \tag{158}$$

Thus the transverse projection operator (it appears in the Poisson brackets below) may be expressed as

$$\delta_{ij} - \frac{k_i k_j}{k^2} = \sum_{\lambda=1,2} \epsilon_i^\lambda \epsilon_j^\lambda. \tag{159}$$

7.1.2 Electromagnetic energy, Lagrangian, conjugate momentum and Poisson brackets

• The energy in the electromagnetic field in the radiation gauge becomes

$$H = \frac{1}{2}\int \left(\mathbf{E}^2 + \mathbf{B}^2\right)\, d^3r = \frac{1}{2}\int \left(\frac{1}{c^2}\dot{\mathbf{A}}^2 + (\nabla\times\mathbf{A})^2\right)\, d^3r \equiv \int \mathcal{H}\, d^3r. \tag{160}$$

The instantaneous configuration of the radiation field is specified by the vector potential $\mathbf{A}(\mathbf{r},t)$, subject to $\nabla\cdot\mathbf{A} = 0$. Comparing with the point-particle energy $H = \frac{1}{2}m\dot{q}^2 + V(q) = T + V$, the electric energy is the kinetic energy and the magnetic energy is a potential energy. The corresponding Lagrangian is $T - V$:

$$L = \frac{1}{2}\int \left(\mathbf{E}^2 - \mathbf{B}^2\right)\, d^3\mathbf{r} = \frac{1}{2}\int \left(\frac{1}{c^2}\dot{\mathbf{A}}^2 - (\nabla\times\mathbf{A})^2\right)\, d^3\mathbf{r} \equiv \int \mathcal{L}\, d^3\mathbf{r}. \tag{161}$$

Recall that the momentum conjugate to a coordinate q is $\frac{\partial L}{\partial \dot{q}}$. So the momentum conjugate to A_i is $\pi_i = \frac{\partial \mathcal{L}}{\partial \dot{A}_i} = -\frac{1}{c}E_i$. It is tempting to write Poisson brackets $\{A_i(\mathbf{r},t), -\frac{1}{c}E_j(\mathbf{r},t)\} = \delta_{ij}\delta^3(\mathbf{r}-\mathbf{r}')$. However, this would not be consistent with the radiation gauge condition, which requires that the divergence of the lhs must vanish. In fact the divergence of the lhs in both $\mathbf{r}$ and $\mathbf{r}'$ must vanish since $\nabla\cdot\mathbf{A} = 0$ and $\nabla\cdot\mathbf{E} = 0$ in the absence of charges.

[38]It is also interesting to choose a complex basis of 'right' and 'left' circular polarization vectors. E.g. if $\vec{\epsilon}_1 = \hat{x}, \vec{\epsilon}_2 = \hat{y}$,

$$\epsilon_+ = \frac{1}{\sqrt{2}}(\hat{x} + i\hat{y}) \quad \text{and} \quad \epsilon_- = \frac{1}{\sqrt{2}}(\hat{x} - i\hat{y}), \quad \text{while} \quad \hat{k} = \hat{z}. \tag{156}$$

These are orthonormal in the sense $\epsilon_\lambda^* \cdot \epsilon_{\lambda'} = \delta_{\lambda,\lambda'}$, $\hat{z}\cdot\epsilon_\pm = 0$.

• Poisson bracket relations that respect the transversality constraints are

$$\left\{ \frac{1}{c} E_i(\mathbf{r},t), A_j(\mathbf{r}',t) \right\} = \delta^T_{ij}(\mathbf{r}-\mathbf{r}'), \quad \{A_i(\mathbf{r},t), A_j(\mathbf{r}',t)\} = \{E_i(\mathbf{r},t), E_j(\mathbf{r}',t)\} = 0. \quad (162)$$

Here $\delta^T_{ij}(\mathbf{r}-\mathbf{r}')$ is the transverse projection of the delta function:

$$\delta^T_{ij}(\mathbf{r}-\mathbf{r}') = \int \frac{d^3k}{(2\pi)^3} \left(\delta_{ij} - \frac{k_i k_j}{\mathbf{k}^2} \right) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} = \frac{1}{V} \sum_{\mathbf{k}} \left(\delta_{ij} - \frac{k_i k_j}{\mathbf{k}^2} \right) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}. \quad (163)$$

The transverse delta function is symmetric and divergence-free $\frac{\partial}{\partial \mathbf{r}_i} \delta^T_{ij}(\mathbf{r}-\mathbf{r}') = -\frac{\partial}{\partial \mathbf{r}'_i} \delta^T_{ij}(\mathbf{r}-\mathbf{r}') = 0$.

• These Poisson brackets may seem a bit ad hoc. The justification for any set of Poisson brackets is that they must give the correct equations of motion with the appropriate hamiltonian (and satisfy anti-symmetry and the Jacobi identity). We will verify later that these p.b. imply the vector wave equation for $\mathbf{A}$ (this is easier to check in Fourier space).

• Let us return to the Fourier expansion of the vector potential and write the electromagnetic energy in terms of the modes $\mathbf{A}_{\mathbf{k}}$[39]

$$\mathbf{A}(\mathbf{r},t) = \frac{c}{\sqrt{V}} \sum_{\mathbf{k}} \mathbf{A}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{r}} \quad (165)$$

• The electric and magnetic fields are

$$\mathbf{E} = -\frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \dot{\mathbf{A}}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad \text{and} \quad \mathbf{B} = \frac{ic}{\sqrt{V}} \sum_{\mathbf{k}} (\mathbf{k} \times \mathbf{A}_{\mathbf{k}}) e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (166)$$

• The electric (kinetic) energy is (we use $\int d^3\mathbf{r} \, e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} = V\delta_{\mathbf{k}\mathbf{k}'}$ and $\mathbf{A}_{-\mathbf{k}} = \mathbf{A}^*_{\mathbf{k}}$)

$$\text{K.E.} = \frac{1}{2} \int \mathbf{E}^2 d^3\mathbf{r} = \frac{1}{2V} \sum_{\mathbf{k},\mathbf{k}'} \dot{\mathbf{A}}_{\mathbf{k}} \dot{\mathbf{A}}_{\mathbf{k}'} \int d^3\mathbf{r} e^{i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{r}} = \frac{1}{2} \sum_{\mathbf{k}} \dot{\mathbf{A}}_{\mathbf{k}} \dot{\mathbf{A}}_{-\mathbf{k}} = \frac{1}{2} \sum_{\mathbf{k}} |\dot{\mathbf{A}}_{\mathbf{k}}|^2. \quad (167)$$

While the magnetic (potential) energy is

$$\text{P.E.} = -\frac{1}{2} \frac{c^2}{V} \sum_{\mathbf{k},\mathbf{k}'} (\mathbf{k}\times\mathbf{A}_{\mathbf{k}})\cdot(\mathbf{k}'\times\mathbf{A}_{\mathbf{k}'}) \int e^{i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{r}} d^3\mathbf{r} = \frac{c^2}{2} \sum_{\mathbf{k}} (\mathbf{k}\times\mathbf{A}_{\mathbf{k}})\cdot(\mathbf{k}\times\mathbf{A}_{-\mathbf{k}}) = \frac{c^2}{2} \sum_{\mathbf{k}} |\mathbf{k}\times\mathbf{A}_{\mathbf{k}}|^2. \quad (168)$$

[39]Since $\mathbf{A}$ is real, the Fourier coefficients must satisfy the symmetry $\mathbf{A}^*_{-\mathbf{k}} = \mathbf{A}_{\mathbf{k}}$. Why is this true? This is a general fact about Fourier series. Suppose $f(x) = \sum_{n=-\infty}^{\infty} f_n e^{inx}$. Then $f^*(x) = \sum_n f^*_n e^{-inx}$. But now let us relabel the dummy index of summation as $n' = -n$, then $f^*(x) = \sum_{n'=-\infty}^{\infty} f^*_{-n'} e^{in'x} = \sum_n f^*_{-n} e^{inx}$. But this must equal $f(x)$ for all x, and this is possible only if the Fourier coefficients are all the same, i.e., if $f_n = f^*_{-n}$. To make the reality of $A(\mathbf{r},t)$ manifest, we could also write

$$\mathbf{A}(\mathbf{r},t) = \frac{1}{2} \frac{c}{\sqrt{V}} \sum_{\mathbf{k}} \left(\mathbf{A}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{r}} + \mathbf{A}^*_{\mathbf{k}}(t) e^{-i\mathbf{k}\cdot\mathbf{r}} \right) \quad (164)$$

Thus the electromagnetic energy is

$$H = \frac{1}{2}\sum_{\mathbf{k}} \left(|\dot{\mathbf{A}}_{\mathbf{k}}|^2 + \omega_k^2 |\hat{k} \times \mathbf{A}_{\mathbf{k}}|^2 \right) \quad \text{where} \quad \omega_k = c|\mathbf{k}|. \tag{169}$$

Comparing with the hamiltonian of a particle of mass $m = 1$ in a harmonic oscillator potential, $H_{\text{sho}} = \frac{1}{2}\dot{q}^2 + \frac{1}{2}\omega^2 q^2$ we see that the electromagnetic energy is a sum of energies of a collection of oscillators $\mathbf{A}_{\mathbf{k}}$. This may also be seen from the equation of motion $\Box \mathbf{A} = 0$.

7.1.3 Solution of vector wave equation as superposition of polarized plane waves

• Indeed, the advantage of the Fourier expansion is that the wave equation $\Box \mathbf{A} = 0$ reduces to a system of decoupled ODEs, for each mode $\mathbf{k}$ and each independent polarization λ. Upon dotting with $\vec{\epsilon}_\lambda$,

$$\ddot{\mathbf{A}}_{\mathbf{k}}(t) + c^2\mathbf{k}^2\mathbf{A}_{\mathbf{k}}(t) = 0 \quad \text{becomes} \quad \ddot{A}_{\mathbf{k},\lambda}(t) = -\omega_k^2 A_{\mathbf{k},\lambda}(t) \quad \text{where} \quad \omega_k = c|\mathbf{k}|. \tag{170}$$

Thus each mode $A_{\mathbf{k},\lambda}$ evolves independently in time like a classical oscillator of angular frequency ω_k

$$A_{\mathbf{k},\lambda}(t) = c_{\mathbf{k},\lambda} e^{-i\omega_k t} + c^*_{\mathbf{k},\lambda} e^{i\omega_k t}. \tag{171}$$

The real and imaginary parts of $c_{\mathbf{k},\lambda}$ are the two constants of integration. Thus

$$\mathbf{A}_{\mathbf{k}}(t) = \sum_\lambda \vec{\epsilon}_\lambda \left[c_{\mathbf{k},\lambda} e^{-i\omega_k t} + c^*_{\mathbf{k},\lambda} e^{i\omega_k t} \right]. \tag{172}$$

If we were working with complex polarization vectors $\epsilon_\pm$ we would have

$$\mathbf{A}_{\mathbf{k}}(t) = \sum_\lambda \left[c_{\mathbf{k},\lambda} \vec{\epsilon}_\lambda e^{-i\omega_k t} + c^*_{\mathbf{k},\lambda} \vec{\epsilon}^{\,*}_\lambda e^{i\omega_k t} \right]. \tag{173}$$

For simplicity, let us stick to real polarization vectors. Using these Fourier coefficients we synthesize the vector potential that is the general solution of the vector wave equation incorporating the Coulomb gauge condition[40]

$$\mathbf{A}(\mathbf{r},t) = \frac{c}{\sqrt{V}} \sum_{\mathbf{k},\lambda} \vec{\epsilon}_\lambda \left[c_{\mathbf{k},\lambda} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega_k t)} + c^*_{\mathbf{k},\lambda} e^{-i(\mathbf{k}\cdot\mathbf{r} - \omega_k t)} \right] \tag{175}$$

$\mathbf{A}$ has been expressed as a linear combination of plane waves $\epsilon_\lambda e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}$ traveling in the direction of their respective wave vectors $\mathbf{k}$, and with polarization λ. The corresponding electric field is (here and elsewhere $+$ c.c. denotes addition of the complex conjugate)

$$\mathbf{E}(\mathbf{r},t) = -\frac{1}{c}\dot{\mathbf{A}} = \frac{i}{\sqrt{V}} \sum_{\mathbf{k},\lambda} \vec{\epsilon}_\lambda \omega_k \left(c_{\mathbf{k}\lambda} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} - c.c. \right) \tag{176}$$

[40] • When $V \to \infty$, these Fourier series become integrals $\frac{1}{V}\sum_{\mathbf{k}} \to \int \frac{d^3k}{(2\pi)^3}$ with $\mathbf{k}$ taking all values. For instance,

$$\mathbf{A}(\mathbf{r},t) = c \int \frac{d^3\mathbf{k}}{(2\pi)^3} \sum_\lambda [c(\mathbf{k},\lambda)\epsilon_\lambda e^{(\mathbf{k}\cdot\mathbf{r} - \omega t)} + \text{c.c}] \quad \text{where} \quad c(\mathbf{k},\lambda) = \sqrt{V} c_{\mathbf{k},\lambda}. \tag{174}$$

Here the constant Fourier coefficients $c_{\mathbf{k},\lambda}, c^*_{\mathbf{k},\lambda}$ are determined by initial conditions on A, E. With a slight abuse of notion, it is convenient to define

$$c_{\mathbf{k},\lambda}(t) = c_{\mathbf{k},\lambda} e^{-i\omega_{\mathbf{k}} t}, \quad c^*_{\mathbf{k},\lambda}(t) = c^*_{\mathbf{k},\lambda} e^{i\omega_{\mathbf{k}} t} \quad \text{where} \quad c_{\mathbf{k},\lambda}, c^*_{\mathbf{k},\lambda} \quad \text{are the initial values.} \tag{177}$$

7.1.4 Change of phase space variables from $\mathbf{A}, \mathbf{E}$ to Fourier modes $c_{\mathbf{k},\lambda}, c^*_{\mathbf{k},\lambda}$

It is clear that the Fourier modes $c_{\mathbf{k},\lambda}(t), c^*_{\mathbf{k},\lambda}(t)$ of definite wave number and polarization have a simpler (simple harmonic) time-dependence than the position space E, A fields (which are linear combinations of several modes). Moreover, the hamiltonian does not couple distinct Fourier modes. This motivates a change of phase space variables from A and E to c, c^*. We *define*[41]

$$\begin{aligned}
\mathbf{A}(\mathbf{r},t) &= \frac{c}{\sqrt{V}} \sum_{\mathbf{k},\lambda} \vec{\epsilon}_\lambda \left[c_{\mathbf{k},\lambda}(t) e^{i\mathbf{k}\cdot\mathbf{r}} + c^*_{\mathbf{k},\lambda}(t) e^{-i\mathbf{k}\cdot\mathbf{r}} \right] \\
\mathbf{E}(\mathbf{r},t) &= \frac{i}{\sqrt{V}} \sum_{\mathbf{k},\lambda} \vec{\epsilon}_\lambda \omega_k \left(c_{\mathbf{k}\lambda}(t) e^{i\mathbf{k}\cdot\mathbf{r}} - c^*_{\mathbf{k}\lambda}(t) e^{-i\mathbf{k}\cdot\mathbf{r}} \right)
\end{aligned} \tag{178}$$

These changes of variables are chosen so that the transversality constraints on A and E are automatically satisfied. Note that the electric field is not obtained by differentiating A in time, it is an independent field. On the other hand, the magnetic field is obtained by taking curl of A, it is a dependent field.

• A significant advantage of c, c^* over A, E is that they have simpler p.b. than A and E. Upon quantization c, c^* are related to annihilation and creation operators for photons with definite wave vector and polarization.

• The equal-time Poisson brackets among the components A_i and E_j are satisfied if the modes $c_{\mathbf{k},\lambda}(t), c^*_{\mathbf{k},\lambda}(t)$ satisfy the following equal time p.b.

$$\left\{ c_{\mathbf{k},\lambda}(t), c^*_{\mathbf{k}',\lambda'}(t) \right\} = \frac{1}{2i\omega_k} \delta_{\mathbf{k}\mathbf{k}'} \, \delta_{\lambda\lambda'}, \quad \left\{ c_{\mathbf{k},\lambda}(t), c_{\mathbf{k}',\lambda'}(t) \right\} = \left\{ c^*_{\mathbf{k},\lambda}(t), c^*_{\mathbf{k}',\lambda'}(t) \right\} = 0. \tag{179}$$

Apart from some constant factors, this must remind us of commutators between SHO annihilation and creation operators $[a, a^\dagger] = 1, [a, a] = [a^\dagger, a^\dagger] = 0$. Let us indicate how this is verified in one case. All dynamical variables A, E, c, c^* are evaluated at the same time t which we suppress

$$\begin{aligned}
\left\{ \frac{1}{c} E_i(\mathbf{r}), A_j(\mathbf{r}') \right\} &= \frac{i}{V} \sum_{\mathbf{k},\lambda,\mathbf{k}',\lambda'} \omega_k \left\{ c_{\mathbf{k},\lambda} (\epsilon_\lambda)_i e^{i\mathbf{k}\cdot\mathbf{r}} - \text{c.c.}, \; c_{\mathbf{k}'\lambda'} (\epsilon_{\lambda'})_j e^{i\mathbf{k}'\cdot\mathbf{r}'} + \text{c.c} \right\} \\
&= \frac{i}{iV} \sum_{\mathbf{k},\lambda,\mathbf{k}',\lambda'} \frac{\omega_k}{2\omega_k} \left[\delta_{\mathbf{k}\mathbf{k}'} \delta_{\lambda\lambda'} \epsilon_{\lambda i} \epsilon_{\lambda' j} e^{i(\mathbf{k}\cdot\mathbf{r} - \mathbf{k}'\cdot\mathbf{r}')} + \text{c.c.} \right]
\end{aligned}$$

[41] Of course, we know the time dependence of $c(t), c^*(t)$ from solving the equations of motion. But that explicit time-dependence is not needed now, it simply motivates the following change of phase space dynamical variables. Equal time Poisson brackets between dynamical variables do not depend on their time-dependence nor on what the hamiltonian is.

$$= \frac{1}{2V}\left(\sum_{\mathbf{k}}\left(\delta_{ij} - \frac{k_i k_j}{k^2}\right) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} + \text{c.c.}\right) = \frac{1}{2V} 2V \delta^T_{ij}(\mathbf{r}-\mathbf{r}') = \delta^T_{ij}(\mathbf{r}-\mathbf{r}').$$

We used the completeness/transverse projection formula $\sum_\lambda \epsilon_{\lambda i}\epsilon_{\lambda j} = \delta_{ij} - \hat{k}_i\hat{k}_j$. In the last line, the Fourier series is a real function of $\mathbf{r}-\mathbf{r}'$, since $\delta_{ij} - \hat{k}_i\hat{k}_j$ is an even function of momentum. So the addition of the complex conjugate just doubles it. One may similarly check that the components of the electric field Poisson commute with each other and so too do the components of the vector potential.

7.1.5 Hamiltonian in terms of Fourier modes $c_{\mathbf{k}\lambda}, c^*_{\mathbf{k}\lambda}$

• Let us express the classical hamiltonian in terms of the Fourier modes. We will show below that

$$H = \frac{1}{2}\int (\mathbf{E}^2 + \mathbf{B}^2) d^3\mathbf{r} = 2\sum_{\mathbf{k},\lambda} \omega_k^2 c^*_{\mathbf{k},\lambda} c_{\mathbf{k},\lambda} \tag{180}$$

• To obtain this formula for the hamiltonian, let us work for simplicity with a real o.n. basis for polarizations $\epsilon_1 \times \epsilon_2 = \hat{k}$. The expressions for $\mathbf{A}, \mathbf{E}, \mathbf{B}$ in terms of c, c^* are

$$\begin{aligned}
\mathbf{A}(\mathbf{r},t) &= \frac{c}{\sqrt{V}}\sum_{\mathbf{k},\lambda} \vec{\epsilon}_\lambda \left[c_{\mathbf{k},\lambda}(t) e^{i\mathbf{k}\cdot\mathbf{r}} + c^*_{\mathbf{k},\lambda}(t) e^{-i\mathbf{k}\cdot\mathbf{r}}\right] \\
\mathbf{E}(\mathbf{r},t) &= \frac{i}{\sqrt{V}}\sum_{\mathbf{k},\lambda} \vec{\epsilon}_\lambda \omega_k \left(c_{\mathbf{k}\lambda}(t) e^{i\mathbf{k}\cdot\mathbf{r}} - c^*_{\mathbf{k}\lambda}(t) e^{-i\mathbf{k}\cdot\mathbf{r}}\right) \\
\mathbf{B}(\mathbf{r},t) &= \nabla \times \mathbf{A} = \frac{i}{\sqrt{V}}\sum_{\mathbf{k},\lambda} \omega_{\mathbf{k}} \hat{k} \times \vec{\epsilon}_\lambda \left(c_{\mathbf{k}\lambda}(t) e^{i\mathbf{k}\cdot\mathbf{r}} - c^*_{\mathbf{k}\lambda}(t) e^{-i\mathbf{k}\cdot\mathbf{r}}\right)
\end{aligned} \tag{181}$$

Now we compute the electric energy using orthogonality $\int e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} d^3r = V\delta_{\mathbf{k},\mathbf{k}'}$ and $\epsilon_\lambda \epsilon_{\lambda'} = \delta_{\lambda,\lambda'}$.

$$\begin{aligned}
\frac{1}{2}\int \mathbf{E}^2 d^3\mathbf{r} &= -\frac{1}{2V}\sum_{\mathbf{k},\mathbf{k}',\lambda,\lambda'} \epsilon_\lambda \cdot \epsilon'_\lambda \omega_{\mathbf{k}}\, \omega_{\mathbf{k}'} \int \left[c_{\mathbf{k}\lambda} e^{i\mathbf{k}\cdot\mathbf{r}} - c^*_{\mathbf{k}\lambda} e^{-i\mathbf{k}\cdot\mathbf{r}}\right]\left[c_{\mathbf{k}'\lambda'} e^{i\mathbf{k}'\cdot\mathbf{r}} - c^*_{\mathbf{k}'\lambda'} e^{-i\mathbf{k}'\cdot\mathbf{r}}\right] d^3\mathbf{r} \\
&= -\frac{1}{2}\sum_{\mathbf{k}\lambda} \omega_{\mathbf{k}}^2 \left[c_{\mathbf{k}\lambda} c_{-\mathbf{k}\lambda} + c^*_{\mathbf{k}\lambda} c^*_{-\mathbf{k}\lambda} - 2|c_{\mathbf{k}\lambda}|^2\right].
\end{aligned} \tag{182}$$

The magnetic energy is

$$\frac{1}{2}\int \mathbf{B}^2 d^3\mathbf{r} = -\frac{1}{2V}\sum_{\mathbf{k},\mathbf{k}',\lambda,\lambda'} \omega_{\mathbf{k}}\omega_{\mathbf{k}'} (\hat{k}\times\epsilon_\lambda)\cdot(\hat{k}'\times\epsilon'_\lambda) \int \left[c_{\mathbf{k}\lambda} e^{i\mathbf{k}\cdot\mathbf{r}} - c^*_{\mathbf{k}\lambda} e^{-i\mathbf{k}\cdot\mathbf{r}}\right]\left[c_{\mathbf{k}'\lambda'} e^{i\mathbf{k}'\cdot\mathbf{r}} - c^*_{\mathbf{k}'\lambda'} e^{-i\mathbf{k}'\cdot\mathbf{r}}\right] d^3\mathbf{r}.$$

It is clear that the spatial integrals will produce either $\delta_{\mathbf{k},-\mathbf{k}'}$ or $\delta_{\mathbf{k},\mathbf{k}'}$ in the various terms. Now we use $\hat{k} = \pm\hat{k}'$, orthogonality of wave and polarization vectors, and the scalar and vector triple product identities to simplify

$$(\hat{k}\times\epsilon_\lambda)\cdot(\hat{k}'\times\epsilon'_\lambda) = \hat{k}'\cdot(\epsilon'_\lambda \times(\hat{k}\times\epsilon_\lambda)) = \hat{k}'\cdot[\hat{k}(\epsilon_{\lambda'}\cdot\epsilon_\lambda) - \epsilon_\lambda(\epsilon_{\lambda'}\cdot\hat{k})] = (\hat{k}'\cdot\hat{k})\,\delta_{\lambda\lambda'}. \tag{183}$$

So the magnetic energy becomes

$$\frac{1}{2}\int \mathbf{B}^2 d^3\mathbf{r} = -\frac{1}{2}\sum_{\mathbf{k},\lambda} \omega_{\mathbf{k}}^2 [-c_{\mathbf{k},\lambda}c_{-\mathbf{k},\lambda} - c^*_{\mathbf{k},\lambda}c^*_{-\mathbf{k},\lambda} - 2|c_{\mathbf{k},\lambda}|^2]. \tag{184}$$

We see that the cc and c^*c^* terms cancel between the electric and magnetic energies giving $H = 2\sum_{\mathbf{k},\lambda}\omega_{\mathbf{k}}^2 c^*_{\mathbf{k},\lambda}$. To make the energy look like that of a collection of harmonic oscillators, we define rescaled Fourier modes

$$c_{\mathbf{k},\lambda} = \sqrt{\frac{\hbar}{2\omega_k}}\, a_{\mathbf{k},\lambda}, \quad \text{and} \quad c^*_{\mathbf{k},\lambda} = \sqrt{\frac{\hbar}{2\omega_k}}\, a^*_{\mathbf{k},\lambda}. \tag{185}$$

Then the hamiltonian and p.b. become

$$H = \sum_{\mathbf{k},\lambda} \hbar\omega_k \left(a^*_{\mathbf{k},\lambda}a_{\mathbf{k},\lambda}\right), \qquad \{a_{\mathbf{k},\lambda}, a^*_{\mathbf{k}',\lambda'}\} = \frac{1}{i\hbar}\delta_{\mathbf{k},\mathbf{k}'}\delta_{\lambda,\lambda'}, \quad \{a_{\mathbf{k},\lambda}, a_{\mathbf{k}',\lambda'}\} = \{a^*_{\mathbf{k},\lambda}, a^*_{\mathbf{k}',\lambda'}\} = 0. \tag{186}$$

Note that the introduction of factors of $\hbar$ does not make this a quantum theory, we are simply choosing to measure the energy of each mode in units of $\hbar\omega_k$. We will quantize this hamiltonian dynamical system soon by replacing classical dynamical variables by operators on Hilbert space and p.b. by commutators $\{.,.\} \to \frac{1}{i\hbar}[.,.]$. These re-scalings ensure that the quantum hamiltonian and commutators take a standard form.

• Simple as the hamiltonian and p.b. are, we must still check that they imply the correct time dependence for a and a^*, previously obtained by solving the vector wave equation, i.e., $a_{\mathbf{k},\lambda}(t) = e^{-i\omega_{\mathbf{k}}t}a_{\mathbf{k},\lambda}(0)$. Hamilton's equation for evolution is

$$\dot{a}_{\mathbf{k}\lambda} = \{a_{\mathbf{k},\lambda}, H\} = \left\{a_{\mathbf{k},\lambda}, \sum_{l,\mu}\hbar\omega_l a^*_{l\mu}a_{l\mu}\right\} = \sum_{l\mu}\hbar\omega_l a_{l\mu}\frac{\delta_{\mathbf{k},\mathbf{l}}\delta_{\lambda,\mu}}{i\hbar} = -i\omega_k a_{\mathbf{k}\lambda}. \tag{187}$$

The solution of this equation is $a_{\mathbf{k},\lambda}(t) = a_{\mathbf{k},\lambda}(0)e^{-i\omega_k t}$. Thus we have verified that the Hamiltonian and p.b. we have postulated for the classical radiation field lead to the correct time-evolution. This justifies the 'ad-hoc' introduction of the transverse delta function in the p.b. between $\mathbf{A}$ and $\mathbf{E}$.

• Let us motivate the passage to the quantum theory by recalling how to 'canonically' quantize a harmonic oscillator using creation and annihilation operators.

7.2 Quantization of the harmonic oscillator using creation and annihilation operators

• Newton's equation for a particle of mass m executing simple harmonic motion is $m\ddot{q} = -\omega^2 q$. The energy of such a harmonic oscillator is $E = \frac{1}{2}m\dot{q}^2 + \frac{1}{2}m\omega^2 q^2$. In terms of the momentum $p = m\dot{q}$, the hamiltonian is $H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2$. Position and momentum satisfy the p.b. $\{q,p\} = 1$, $\{q,q\} = \{p,p\} = 0$.

• Since $\hbar\omega$ has dimensions of energy, even classically we may write $H = \hbar\omega(\frac{p^2}{2m\hbar\omega} + \frac{1}{2}\frac{m\omega q^2}{\hbar})$. Defining a constant with unit of inverse length $\beta = \sqrt{\frac{m\omega}{\hbar}}$ we have the dimensionless coordinate and momentum $\xi = \beta q$ and $\mathtt{p} = \frac{p}{\hbar\beta}$ with p.b. $\{\xi, \mathtt{p}\} = \frac{1}{\hbar}$ and $H = \frac{1}{2}\hbar\omega(\xi^2 + \mathtt{p}^2)$. We now define the complex combinations

$$a = \frac{\xi + i\mathtt{p}}{\sqrt{2}} \quad \text{and} \quad a^* = \frac{\xi - i\mathtt{p}}{\sqrt{2}} \quad \text{with} \quad \{a, a^*\} = -\frac{i}{\hbar} \quad \text{and} \quad H = \hbar\omega a^* a = \hbar\omega|a|^2. \tag{188}$$

In the quantum theory, q, p become operators. In the Schrodinger representation $p = -i\hbar\frac{\partial}{\partial q}$ and the p.b. $\{q, p\} = 1$ is replaced by the commutator $[q, p] = i\hbar$ (i.e., multiply the rhs by $i\hbar$). It follows that $[\xi, \mathtt{p}] = i$. In the Schrodinger representation $\mathtt{p} = -i\frac{\partial}{\partial\xi}$, check that this gives the desired commutator $[\xi, \mathtt{p}] = i$.

• Now if $a^\dagger = (\xi - i\mathtt{p})/\sqrt{2}$ denotes the hermitian adjoint of the operator a (quantum version of a^*), then $[a, a^\dagger] = 1$. Moreover,

$$a^\dagger a = \frac{1}{2}\left(\xi^2 + \mathtt{p}^2 + i[\xi, \mathtt{p}]\right) = \frac{1}{2}\left(\xi^2 + \mathtt{p}^2 - 1\right) \tag{189}$$

• The hamiltonian operator

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2 = \frac{1}{2}\hbar\omega(\xi^2 + \mathtt{p}^2) \tag{190}$$

may be written as $H = \hbar\omega(a^\dagger a + \frac{1}{2})$. We have used that fact that $aa^\dagger - a^\dagger a = 1$. $N = a^\dagger a$ is the number operator. We may check using the commutation relations that

$$[N, a] = [a^\dagger a, a] = [a^\dagger, a]a = -a \quad \text{and} \quad [N, a^\dagger] = [a^\dagger a, a^\dagger] = a^\dagger[a, a^\dagger] = a^\dagger. \tag{191}$$

It follows that

$$[H, a] = -\hbar\omega a \quad \text{and} \quad [H, a^\dagger] = \hbar\omega a^\dagger. \tag{192}$$

$a, a^\dagger$ are called the annihilation and creation operators (or lowering and raising operators) because of the way we may interpret these relations. Suppose $|\psi\rangle$ is an energy eigenstate with energy eigenvalue E. Then assuming $a|\psi\rangle$ is not the zero vector, $a|\psi\rangle$ is also an energy eigenstate with a little lower energy $E - \hbar\omega$, since

$$H(a|\psi\rangle) = aH|\psi\rangle - \hbar\omega a|\psi\rangle = (E - \hbar\omega)(a|\psi\rangle) \tag{193}$$

Similarly, $a^\dagger|\psi\rangle$ is also an energy eigenstate with a slightly higher energy $E + \hbar\omega$.

• Now the SHO hamiltonian is a positive operator, in the sense that its diagonal matrix element in any state is positive:

$$\langle\phi|H|\phi\rangle = \frac{1}{2}\hbar\omega + \langle\phi|a^\dagger a|\phi\rangle = \frac{1}{2}\hbar\omega + \langle a\phi|a\phi\rangle = \frac{1}{2}\hbar\omega + |||a\phi\rangle||^2 \geq \frac{1}{2}\hbar\omega. \tag{194}$$

Since eigenvalues are simply expectation values in normalized eigenstates, the energy eigenvalues must all be $\geq \frac{1}{2}\hbar\omega$. Now, if there is one energy eigenstate $|\psi\rangle$ with eigenvalue E, then by repeated application of the lowering operator a, we may produce an energy eigenstate with negative energy, contradicting the positivity of H. To avoid this problem, successive application of a must result in a state $|0\rangle$ (taken to have unit norm) which is annihilated by the lowering operator. This state is the ground state of the hamiltonian $H|0\rangle = \frac{1}{2}\hbar\omega|0\rangle$ with energy $E_0 = \frac{1}{2}\hbar\omega$. $|0\rangle$ is also called the vacuum state. The first excited state is $|1\rangle = a^\dagger|0\rangle$, with an energy $E_1 = \frac{3}{2}\hbar\omega$.

• Note that the ground or vacuum state is not the zero vector. The zero vector $\psi(x) = 0$ is not a state describing a particle, since it has zero probability to be found any where. But the simple harmonic oscillator describes one particle at all times, so every physical state of the SHO must satisfy the normalization condition $\int_{-\infty}^{\infty} |\psi(x)|^2 dx = 1$. One may find the position space wave function of the ground state, i.e., $\langle x|0\rangle = \psi_0(x)$ using the condition $a\psi(x) = \frac{1}{\sqrt{2}}(\xi + \partial_\xi)\psi = 0$. This implies $\psi'/\psi = -\xi$ or $\psi = Ae^{-\xi^2/2} = Ae^{-\beta^2 x^2/2}$. Find the value of constant A to ensure the ground state is normalized to one. Note that though the average value of x in the ground (vacuum) state is zero on account of evenness (parity symmetry) of $\psi_0(x)$, the position does suffer fluctuations, $\langle x^2\rangle_0 \neq 0$. Similarly, one checks that $\langle p\rangle_0 = 0$ but $\langle p^2\rangle_0 \neq 0$.

• We check using the commutation relation $aa^\dagger - a^\dagger a = 1$ that $|1\rangle$ has unit norm

$$\langle 1|1\rangle = \langle 0|aa^\dagger|0\rangle = \langle 0|0\rangle + \langle 0|a^\dagger a|0\rangle = 1. \tag{195}$$

Similarly, the second excited state is $\propto a^\dagger|1\rangle$. The square of its norm is

$$\langle 0|aaa^\dagger a^\dagger|0\rangle = \langle 0|a(a^\dagger a + 1)a^\dagger|0\rangle = \langle 0|aa^\dagger|0\rangle + \langle 0|(a^\dagger a + 1)aa^\dagger|0\rangle = 1 + 1 = 2. \tag{196}$$

So the normalized second excited state is $|2\rangle = \frac{1}{\sqrt{2}}a^\dagger a^\dagger|0\rangle$, with an energy $\frac{5}{2}\hbar\omega$. Proceeding this way[42], one finds that the n^{th} excited state (normalized to one) is $|n\rangle = \frac{1}{\sqrt{n!}}(a^\dagger)^n|0\rangle$ with an energy eigenvalue $E_n = \hbar\omega(n + \frac{1}{2})$.

• In the Schrödinger picture, the states evolve with time, as specified by the Schrodinger equation $i\hbar\frac{\partial|\psi(t)\rangle}{\partial t} = H|\psi(t)\rangle$, while observables like H, x, p are time-independent. Energy levels are stationary, in the sense that they evolve by a phase

$$|n(t)\rangle = e^{-iE_n t/\hbar}|n(0)\rangle. \tag{197}$$

• In the Heisenberg picture, states are time-independent while observables carry the time dependence. By definition, the state in the Heisenberg picture is just the Schrödinger state at $t = 0$. So

$$|\phi(t)\rangle_s = e^{-iHt/\hbar}|\phi(0)\rangle \quad \Rightarrow \quad |\phi(t)\rangle_s = e^{-iHt/\hbar}|\phi\rangle_h. \tag{198}$$

The Heisenberg picture operator A_h corresponding to the Schrödinger picture operator A_s is defined as

$$A_h(t) = e^{iHt/\hbar}A_s e^{-iHt/\hbar} = U^\dagger A_s U, \quad \text{where we denote} \quad U = e^{-iHt/\hbar}. \tag{199}$$

It follows that the hamiltonian is the same in both pictures $H_s = H_h \equiv H$. Further, the states and operators in the two pictures coincide at $t = 0$.

[42]The normalization factors may be obtained by first showing that $a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$ and $a|n\rangle = \sqrt{n}|n-1\rangle$.

• Matrix elements and expectation values (which carry physical significance) may be computed in either picture, resulting in the same values

$$_h\langle\phi|A_h(t)|\psi\rangle_h =_s \langle\phi(t)|UU^\dagger A_s UU^\dagger|\psi(t)\rangle_s =_s \langle\phi(t)|A_s|\psi(t)\rangle_s. \tag{200}$$

Just as the Schrodinger equation governs the evolution of states in the Schrodinger picture, the time evolution of an observable A_h is governed by the Heisenberg equation of motion

$$i\hbar\frac{\partial A_h}{\partial t} = [A_h, H]. \tag{201}$$

Let us derive the Heisenberg equation of motion. Recall $A_h(t) = U^\dagger A_s U$

$$i\hbar\partial_t A_h = i\hbar(\partial_t U^\dagger)A_s U + i\hbar U^\dagger A_s \partial_t U = -HU^\dagger A_s U + U^\dagger A_s U H = -HA_h + A_h H = [A_h, H]. \tag{202}$$

Here we used $U = e^{-iHt/\hbar}$ so that $i\hbar\partial_t U = HU = UH$ since H and U commute. And taking the hermitian adjoint, $-i\hbar\partial_t U^\dagger = HU^\dagger$.

• The Heisenberg equation of motion is the quantum version of Hamilton's classical equations written in p.b. form. Start with $\{A, H\} = \frac{\partial A}{\partial t}$ and replace p.b. by commutators and multiply the rhs by $i\hbar$.

• Let us use the Heisenberg equation of motion[43] to find the time evolution of the SHO creation and annihilation operators in the Heisenberg picture.

$$i\hbar\frac{\partial a_h}{\partial t} = [a_h, H] = ([a, H])_h = \hbar\omega a_h \quad\Rightarrow\quad \frac{\partial a_h}{\partial t} = -i\omega a_h \quad\Rightarrow\quad a_h(t) = e^{-i\omega t}a_h(0) \tag{203}$$

Similarly (or taking the hermitian conjugate), we get $a_h^\dagger(t) = e^{i\omega t}a_h^\dagger(0)$. This is the same time evolution as in a classical harmonic oscillator.

7.3 Quantization of radiation field in radiation gauge

• We will quantize the radiation field by the canonical procedure of replacing Poisson brackets with commutators, as we did for the harmonic oscillator. Indeed, the radiation field can be regarded as an infinite collection of harmonic oscillators, one for each mode labelled by wave vector and polarization $(\mathbf{k}, \lambda)$.

$$H = 2\sum_{\mathbf{k},\lambda} \omega_k^2 c_{\mathbf{k}\lambda}^* c_{\mathbf{k}\lambda} \tag{204}$$

where the vector potential is

$$\mathbf{A}(\mathbf{r}, t) = \frac{c}{\sqrt{V}}\sum_{\mathbf{k}\lambda} \vec{\epsilon}_\lambda \left[c_{\mathbf{k}\lambda}(t)e^{i\mathbf{k}\cdot\mathbf{r}} + c_{\mathbf{k}\lambda}^* e^{-i\mathbf{k}\cdot\mathbf{r}}\right]. \tag{205}$$

[43] We also use $[H, a] = -\hbar\omega a, [H, a^\dagger] = \hbar\omega a^\dagger$ and the relation $[B_h, H] = [U^\dagger BU, H] = U^\dagger[B, H]U = ([B, H])_h$ on account of $[H, U] = 0$ where $U = e^{-iHt/\hbar}$.

Recall the p.b. among the modes of the EM field

$$\left\{c_{\mathbf{k},\lambda}, c^*_{\mathbf{k}',\lambda'}\right\} = \frac{1}{2i\omega_k}\,\delta_{\mathbf{k}\mathbf{k}'}\,\delta_{\lambda\lambda'}, \quad \left\{c_{\mathbf{k},\lambda}, c_{\mathbf{k}',\lambda'}\right\} = \left\{c^*_{\mathbf{k},\lambda}, c^*_{\mathbf{k}',\lambda'}\right\} = 0. \tag{206}$$

To make the Hamiltonian and p.b. look simpler and to follow a notation similar to the one used in quantizing the SHO, we define re-scaled Fourier modes for the radiation field

$$c_{\mathbf{k},\lambda} = \sqrt{\frac{\hbar}{2\omega_k}}\, a_{\mathbf{k},\lambda}, \quad \text{and} \quad c^*_{\mathbf{k},\lambda} = \sqrt{\frac{\hbar}{2\omega_k}}\, a^*_{\mathbf{k},\lambda}. \tag{207}$$

Then the hamiltonian becomes

$$H = \sum_{\mathbf{k}\lambda} \hbar\omega_k a^*_{\mathbf{k}\lambda} a_{\mathbf{k}\lambda}. \tag{208}$$

Then the Fourier decomposition of the vector potential reads

$$\mathbf{A}(\mathbf{r},t) = \frac{c}{\sqrt{V}} \sum_{\mathbf{k},\lambda} \sqrt{\frac{\hbar}{2\omega_k}} \vec{\epsilon}_\lambda \left[a_{\mathbf{k},\lambda}(t) e^{i\mathbf{k}\cdot\mathbf{r}} + a^*_{\mathbf{k}\lambda}(t) e^{-i\mathbf{k}\cdot\mathbf{r}}\right] \tag{209}$$

and the equal-time p.b. among the $a_{\mathbf{k},\lambda}, a^*_{\mathbf{k},\lambda}$ are

$$\left\{a_{\mathbf{k},\lambda}, a^*_{\mathbf{k}',\lambda'}\right\} = \frac{1}{i\hbar}\delta_{\mathbf{k},\mathbf{k}'}\delta_{\lambda,\lambda'}, \quad \text{while} \quad \left\{a_{\mathbf{k},\lambda}, a_{\mathbf{k}',\lambda'}\right\} = \left\{a^*_{\mathbf{k},\lambda}, a^*_{\mathbf{k}',\lambda'}\right\} = 0. \tag{210}$$

Now we canonically quantize this system by analogy with the SHO. The p.b. among the a, a^* are replaced by commutators between a , multiplying the RHS by $i\hbar$. Thus we get the canonical commutation relations

$$[a_{\mathbf{k},\lambda}, a^\dagger_{\mathbf{k}',\lambda'}] = \delta_{\mathbf{k},\mathbf{k}'}\delta_{\lambda,\lambda'}, \quad \text{while} \quad [a_{\mathbf{k},\lambda}, a_{\mathbf{k}',\lambda'}] = [a^\dagger_{\mathbf{k},\lambda}, a^\dagger_{\mathbf{k}',\lambda'}] = 0. \tag{211}$$

The expansion of the vector potential operator now reads

$$\mathbf{A}(\mathbf{r},t) = \frac{c}{\sqrt{V}} \sum_{\mathbf{k},\lambda} \sqrt{\frac{\hbar}{2\omega_k}} \vec{\epsilon}_\lambda \left[a_{\mathbf{k},\lambda}(t) e^{i\mathbf{k}\cdot\mathbf{r}} + a^\dagger_{\mathbf{k}\lambda}(t) e^{-i\mathbf{k}\cdot\mathbf{r}}\right]. \tag{212}$$

We will find its time dependence shortly. There are similar expansions for the electric and magnetic field operators as linear combinations of creation and annihilation operators. The quantum version of the A, E Poisson brackets are the commutators $[E_i, E_j] = [A_i A_j] = 0$ and

$$[E_i(\mathbf{r},t), A_j(\mathbf{r}',t)] = i\hbar c\delta^T_{ij}(\mathbf{r}-\mathbf{r}'). \tag{213}$$

• The expression for the hamiltonian operator is ambiguous in the quantum theory since $a, a^\dagger$ do not commute. Classically we may write several equivalent expressions, $H = \sum \hbar\omega_k a^*_{\mathbf{k}\lambda} a_{\mathbf{k}\lambda} = \sum \hbar\omega_k a_{\mathbf{k}\lambda} a^*_{\mathbf{k}\lambda} = \frac{1}{2}\sum \hbar\omega_k (a^*_{\mathbf{k}\lambda} a_{\mathbf{k}\lambda} + a_{\mathbf{k}\lambda} a^*_{\mathbf{k}\lambda})$. If a, a^* are replaced by $a, a^\dagger$ then one gets hamiltonian operators that differ by an additive constant. If we use the first expression, then

$\hat{H} = \sum_{\mathbf{k},\lambda} \hbar\omega_k a^\dagger_{\mathbf{k},\lambda} a_{\mathbf{k},\lambda}$. But if we use the third (symmetric expression for H, which also corresponds to $\frac{1}{2}(E^2 + B^2)$) then the quantum hamiltonian is

$$\hat{H} = \sum_{\mathbf{k},\lambda} \hbar\omega_k \left(a^\dagger_{\mathbf{k},\lambda} a_{\mathbf{k},\lambda} + \frac{1}{2} \right) \quad \text{since} \quad a_{\mathbf{k},\lambda} a^\dagger_{\mathbf{k},\lambda} = a^\dagger_{\mathbf{k},\lambda} a_{\mathbf{k},\lambda} + 1. \tag{214}$$

The additive constant $\sum_{\mathbf{k},\lambda} \frac{1}{2}\hbar\omega_k$ is called the zero point energy. In the infinite volume limit, it is infinite. However, this is a constant addition to the energy, and can be eliminated by redefining the zero of energy. Henceforth, we define $H = \sum_{\mathbf{k},\lambda} \hbar\omega_{\mathbf{k}} a^\dagger_{\mathbf{k},\lambda} a_{\mathbf{k},\lambda}$. This definition is convenient since it assigns energy zero to the vacuum state.

7.4 Hilbert space of photon states

• To find the spectrum of the hamiltonian, we proceed as we did for the SHO. Indeed, the hamiltonian of the quantized radiation field is a sum of harmonic oscillators, one for each 'mode' labelled by $\mathbf{k}, \lambda$. From the commutation relations, we find as before, that

$$[H, a_{\mathbf{k},\lambda}] = -\hbar\omega_k a_{\mathbf{k},\lambda} \quad \text{and} \quad [H, a^\dagger_{\mathbf{k},\lambda}] = \hbar\omega_k a^\dagger_{\mathbf{k},\lambda}. \tag{215}$$

It follows that $a_{\mathbf{k},\lambda}, a^\dagger_{\mathbf{k},\lambda}$ lower and raise the energy by $\hbar\omega_k$. Thus we have a vacuum state $|0\rangle$ with energy zero, which is annihilated by *all* the lowering operators $a_{\mathbf{k},\lambda}|0\rangle = 0$. (This also means $\langle 0|a^\dagger_{\mathbf{k},\lambda} = 0$ for all $\mathbf{k}, \lambda$.) As before, $N_{\mathbf{k},\lambda} = a^\dagger_{\mathbf{k},\lambda} a_{\mathbf{k},\lambda}$ is a number operator. It has non-negative integers as its eigenvalues, which count the number of photons with wave vector $\mathbf{k}$ and polarization ϵ_λ in an eigenstate. $\mathbf{k}, \lambda$ are together good quantum numbers for photons. $N_{\mathbf{k},\lambda}|0\rangle = 0$. The operator whose eigenvalues are the total number of photons is $\hat{N} = \sum_{\mathbf{k},\lambda} N_{\mathbf{k},\lambda}$. We say that the vacuum state has no photons of any wave vector or polarization. However, the vacuum state is not the zero vector, it has unit norm $\langle 0|0\rangle = 1$. We will see that though the average electric and magnetic fields in the vacuum state are zero, they have non-zero fluctuations in the vacuum state. The free space around us (ignoring the EM fields from cell phone towers etc) is to a reasonable approximation the vacuum state of the photon field. If we measure the electric field, we will get small non-zero values which on average are zero. These small non-zero values are due to quantum fluctuations. This is just like saying that x and p are on average zero in the ground state of the harmonic oscillator. Nevertheless $\langle x^2 \rangle$ and $\langle p^2 \rangle$ are non-zero in the ground state of the SHO. The position x of the particle is the counterpart of the vector potential A, while the particle momentum p is the analog of ($-1/c$ times) the electric field.

• A state with one photon of wave vector $\mathbf{k}$ and polarization λ is $|1_{\mathbf{k},\lambda}\rangle = a^\dagger_{\mathbf{k},\lambda}|0\rangle$. This 1-photon state has energy $\hbar\omega_k$.

• Similarly, a state with two photos is

$$|1_{\mathbf{k},\lambda}, 1_{\mathbf{k}',\lambda'}\rangle = a^\dagger_{\mathbf{k},\lambda} a^\dagger_{\mathbf{k}',\lambda'}|0\rangle. \tag{216}$$

It is an eigenstate of the hamiltonian with energy $\hbar(\omega_k + \omega_{k'})$. Since creation operators commute, it does not matter in what order we write the creation operators, so

$$|1_{\mathbf{k}',\lambda'} 1_{\mathbf{k},\lambda}\rangle = |1_{\mathbf{k},\lambda}, 1_{\mathbf{k}',\lambda'}\rangle. \tag{217}$$

In other words, the state function of a system of two photons is symmetric with respect to exchange of the quantum numbers of the two photons: photons behave as bosons.

• The above two photon state has norm one if the quantum numbers are distinct, i.e., $(\mathbf{k},\lambda) \neq (\mathbf{k}',\lambda')$. If the quantum numbers are the same, then the normalized 2-photon state is

$$|2_{\mathbf{k},\lambda}\rangle = \frac{1}{\sqrt{2!}}(a^\dagger_{\mathbf{k},\lambda})^2|0\rangle. \tag{218}$$

This follows from the commutation relations of creation and annihilation operators just as in the case of the harmonic oscillator.

• More generally a normalised multi-photon state with $n, n', n'', \ldots$ photons with quantum numbers $(\mathbf{k},\lambda), (\mathbf{k}',\lambda'), (\mathbf{k}'',\lambda''), \ldots$ is

$$|n_{\mathbf{k},\lambda}, n'_{\mathbf{k}',\lambda'}, n''_{\mathbf{k}'',\lambda''}, \ldots\rangle = \frac{(a^\dagger_{\mathbf{k},\lambda})^n}{\sqrt{n!}} \frac{(a^\dagger_{\mathbf{k}',\lambda'})^{n'}}{\sqrt{n'!}} \frac{(a^\dagger_{\mathbf{k}'',\lambda''})^{n''}}{\sqrt{n''!}} \ldots |0\rangle. \tag{219}$$

Again, by the commutativity of the creation operators, these multi-photon states are symmetric under exchange of any pair, they describe bosons. These multi-photon states together span the Hilbert space of the quantized radiation field. It is called the bosonic Fock space of photons. The basis we have chosen to describe the Fock space is called the occupation number/photon number basis, since the basis states have definite numbers of photons with specified wave vector and polarization. It is in this way that the quantum theory accommodates the particle-like nature of photons discovered by Planck, Einstein et. al. On the other hand, a linear combination such as $|1_{\mathbf{k},\lambda}\rangle - |2_{\mathbf{k}',\lambda'}\rangle$ is also a valid state of the radiation field, but it does not have a definite number of photons, a measurement of the number of photons may result either in the answer one or two. Such states play a role in the ability of the quantum theory to accommodate the wave-like character of light, as we will see.

7.5 Fluctuations of $\mathbf{E}$ and $\mathbf{B}$ fields, EM Waves from matrix elements

• In the quantum theory the transverse photon vector potential field in radiation gauge is the hermitian operator

$$\mathbf{A}(\mathbf{r},t) = \frac{c}{\sqrt{V}} \sum_{\mathbf{k},\lambda} \vec{\epsilon}_\lambda \sqrt{\frac{\hbar}{2\omega_k}} \left[a_{\mathbf{k},\lambda}(t) e^{i\mathbf{k}\cdot\mathbf{r}} + a^\dagger_{\mathbf{k}\lambda}(t) e^{-i\mathbf{k}\cdot\mathbf{r}} \right]. \tag{220}$$

The time dependence of $\mathbf{A}$ as well as the electric and magnetic field operators is determined by that of the creation and annihilation operators. To find their time-dependence, we use the Heisenberg equation of motion which is the quantised version of Hamilton's equation $\dot{a} = \{a, H\}$ obtained by the replacement $\{\cdot,\cdot\} \to [\cdot,\cdot]/i\hbar$

$$i\hbar \frac{da_{\mathbf{k},\lambda}}{dt} = [a_{\mathbf{k},\lambda}, H] = \hbar\omega_k a_{\mathbf{k},\lambda} \quad \Rightarrow \quad a_{\mathbf{k},\lambda}(t) = e^{-i\omega_k t} a_{\mathbf{k},\lambda}(0) \tag{221}$$

Similarly, $a^\dagger_{\mathbf{k},\lambda}(t) = e^{i\omega_k t} a^\dagger_{\mathbf{k},\lambda}(0)$. We may regard $a_{\mathbf{k},\lambda}(t)$ as the annihilation operator in the Heisenberg picture while $a_{\mathbf{k},\lambda}(0)$ is the annihilation operator in the Schrodinger picture. We will often omit the argument of a, and hope it is clear from the context. Thus our Fourier mode expansion of the quantized vector potential is

$$\mathbf{A}(\mathbf{r},t) = \frac{c}{\sqrt{V}} \sum_{\mathbf{k},\lambda} \vec{\epsilon}_\lambda \sqrt{\frac{\hbar}{2\omega_k}} \left[a_{\mathbf{k},\lambda}(0) e^{i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)} + a^\dagger_{\mathbf{k}\lambda}(0) e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)} \right]. \tag{222}$$

We notice that the time dependence is the same as in classical radiation theory (this is generally true when the hamiltonian is quadratic in fields and commutators are canonical). It follows that $\mathbf{E}$ and $\mathbf{B}$ are the hermitian field operators

$$\begin{aligned} \mathbf{E}(\mathbf{r},t) &= -\frac{\dot{\mathbf{A}}}{c} = i \sum_{\mathbf{k},\lambda} \sqrt{\frac{\hbar\omega_k}{2V}} \vec{\epsilon}_\lambda \left(a_{\mathbf{k}\lambda} e^{i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)} - a^\dagger_{\mathbf{k}\lambda} e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)} \right) \\ \mathbf{B}(\mathbf{r},t) &= \nabla \times \mathbf{A} = i \sum_{\mathbf{k},\lambda} \sqrt{\frac{\hbar\omega_k}{2V}} (\hat{k} \times \vec{\epsilon}_\lambda) \left(a_{\mathbf{k}\lambda} e^{i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)} - a^\dagger_{\mathbf{k}\lambda} e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)} \right) \end{aligned} \tag{223}$$

Being linear combinations of creation and annihilation operators, the electric and magnetic fields do not commute either with the number operator or hamiltonian (nor with each other). Their eigenstates do not have definite energy or number of photons in general. On the other hand, states of definite energy (like the vacuum) are not eigenstates of $\mathbf{E}$ or $\mathbf{B}$. For instance, acting on the vacuum the electric field operator produces a linear combination of one photon states with all possible wave vectors and polarizations.

$$\mathbf{E}(\mathbf{r},t)|0\rangle = -i \sum_{\mathbf{k},\lambda} \sqrt{\frac{\hbar\omega_k}{2V}} \vec{\epsilon}_\lambda e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)} |1_{\mathbf{k},\lambda}\rangle \tag{224}$$

It follows that the matrix element of the electric (or magnetic field) between the vacuum and 1 photon state $1_{\mathbf{k},\lambda}$ is a transversely polarized plane EM wave

$$\langle 1_{\mathbf{k},\lambda}|\mathbf{E}(\mathbf{r},t)|0\rangle = -i\sqrt{\frac{\hbar\omega}{2V}} \epsilon_\lambda e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega t)} \quad \text{and} \quad \langle 1_{\mathbf{k},\lambda}|\mathbf{B}(\mathbf{r},t)|0\rangle = -i\sqrt{\frac{\hbar\omega}{2V}} (\epsilon_\lambda \times \hat{k}) e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega t)}. \tag{225}$$

Since plane waves satisfy Maxwell's equations, we see that these matrix elements of the electric and magnetic fields in the quantum theory satisfy the same wave equations (Maxwell's equations) as the classical electric and magnetic fields. The wave nature of light follows from Maxwell's equations. So this is one of the ways in which the quantum theory of the photon field accommodates the wave nature of light while also manifesting the particle-like nature of photons. From the viewpoint of the quantum theory, we may regard Maxwell's equations as determining these matrix elements of the fields. This is a useful point of view, since it also applies to the Klein-Gordon and Dirac equations.

• Historically, the Dirac and KG equations were introduced as relativistic quantum equations to describe a single electron or single pion. This interpretation was inconsistent in situations with significant relativistic effects: due to the possibility for particle creation and annihilation, particle number is not conserved and it does not make sense to look for a theory of a definite number of particles. The apparent successes of the '1 particle' Dirac equation (like the prediction of the magnetic moment of the electron or fine structure of the hydrogen spectrum) are all in the regime where relativistic effects are very small. In the current view, the Dirac and KG equations are not one particle wave equations like the non-relativistic Schrodinger equation, but rather classical wave-field equations, on the same footing as Maxwell's equations. The appearance of factors of $\hbar$ in the Dirac and KG equations does not make them 'quantum', but is due to a conventional choice of units for momenta and energies. Strictly, these classical relativistic field equations do not admit particle interpretation at all. However, when quantised via the process of 'field' quantization (somewhat misleadingly also known as 'second' quantization) that we have just carried out for the EM field, we arrive at the quantised Dirac and KG fields. States in the Hilbert space of these quantum fields now admit physical interpretation in terms of particles. Remarkably, the matrix elements of these quantised fields (Dirac, KG, Maxwell) between the vacuum and 1 particle states, satisfy the classical wave-field equations that one started with.

• By contrast, the 1 particle (or n-particle) non-relativisitic Schrodinger wave equation is already quantised. Unlike KG, Dirac or Maxwell, it is not to be regarded as a classical field equation awaiting quantization. It already deals with operators and states in Hilbert space.

• As the Weisskopf microscope thought experiment suggested, the electromagnetic field displays quantum fluctuations. To see this, consider the simplest of states, the vacuum $|0\rangle$ with no photons. In this state, $\langle 0|\mathbf{A}|0\rangle = 0$ since the annihilation operators will kill the ket-vacuum while the creation operators kill the bra-vacuum. Since $\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}$ and $\mathbf{B} = \nabla \times \mathbf{A}$, the electric and magnetic fields also have zero expectation values in the vacuum state. However, just as $\langle 0|x^2|0\rangle > 0, \langle 0|p^2|0\rangle > 0$ in the g.s of the SHO, one checks that $\langle 0|\mathbf{B}^2|0\rangle > 0$ and $\langle 0|\mathbf{E}^2|0\rangle > 0$ since $\mathbf{E}^2, \mathbf{B}^2$ include terms of the form $aa^\dagger$ which have non-zero vacuum expectation values. It follows that there are vacuum fluctuations in the electromagnetic fields, even when their mean values are zero. E.g., free space around us (ignoring the EM fields from cell phone towers etc) is to a reasonable approximation the vacuum state of the photon field. If we measure the electric field, we will get small non-zero values which on average are zero. These small non-zero values are due to quantum fluctuations.

• Question: How is $\langle 0|\mathbf{E}^2|0\rangle > 0$ and $\langle 0|\mathbf{B}^2|0\rangle > 0$ consistent with $\langle 0|H|0\rangle = 0$? Ans: our hamiltonian $H = \sum \hbar\omega a^\dagger a$ differs from $\int \frac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2) = \sum \frac{1}{2}\hbar\omega(aa^\dagger + a^\dagger a)$ by an additive constant ('zero point energy' $\sum \frac{1}{2}\hbar\omega_{\mathbf{k}}$). The non-zero vacuum fluctuations in $\mathbf{E}$ and $\mathbf{B}$, in a sense, add up to give this 'zero point energy'.

• Heuristically we may say that in the vacuum, though there are no real photons, there can be virtual photons that pop in and out of existence here and there, for short periods of time. These virtual photons are a way to visualize, for example, the evaluation of the expectation value $\langle 0|aa^\dagger|0\rangle$. Virtual photons are not directly detected, they are not present in the initial or final state. But virtual photons have real effects such as (1) The measurable vacuum fluctuations in the electric and magnetic fields. (2) The Casimir force between metal plates in vacuum. (3) The spontaneous gamma decay of atoms, nuclei and hadrons from excited states. While stimulated emission of photons from excited atoms is understandable, it was found that excited atoms can spontaneously emit photons. In a sense, virtual photons 'stimulate' the atom to 'spontaneously' decay.

7.6 Polarization, helicity, spin and angular momentum of the photon

• We have seen that a one photon state can have two linearly independent polarizations, say the linear polarizations $\vec{\epsilon}_1 = \hat{x}, \vec{\epsilon}_2 = \hat{y}$ for $\mathbf{k} = k\hat{z}$ or the basis of right and left circular polarizations $\epsilon_\pm = \frac{1}{\sqrt{2}}(\hat{x} \pm i\hat{y})$. The circular polarization vectors have the advantage of transforming into multiples of themselves under an infinitesimal rotation by angle θ about the z axis:

$$\delta\hat{x} = -\theta\hat{y}, \quad \delta\hat{y} = \theta\hat{x} \quad \Rightarrow \quad \delta\epsilon_+ = i\theta\epsilon_+ \quad \text{and} \quad \delta\epsilon_- = -i\theta\epsilon_-. \tag{226}$$

This is precisely how the $|m = 1\rangle$ and $|m = -1\rangle$ angular momentum states of the $j = 1$ angular momentum multiplet transform under infinitesimal rotations about the z-axis

$$e^{i\theta J_z/\hbar}|m\rangle \approx (1 + i\theta J_z/\hbar)|m\rangle = (1 + i\theta m)|m\rangle \quad \Rightarrow \quad \delta|m\rangle = i\theta m|m\rangle \tag{227}$$

So the right and left circular polarized 1 photon states transform under rotations about the propagation direction $\hat{k}$, like states having angular momentum projection $m = \pm 1$ along $\hat{k}$. In other words, a general one photon state must be a linear combination of helicity $h = +1$ and helicity $h = -1$ eigenstates ($h = \vec{J} \cdot \hat{p}$). We say that the photon (when in a state of definite helicity) has helicity ± 1. A longitudinal polarization vector pointing along the direction of propagation would be unchanged under rotations about $\mathbf{k}$ and transform like the $|m = 0\rangle$ state. So a longitudinally polarized photon would have zero helicity, it is forbidden for photons and classical EM waves. On the other hand, phonons or massive gauge vector bosons like the $W^\pm$ and Z^0 can be longitudinally polarized.

• Going by the definition of the spin of a particle as the maximum value of its helicity, the photon has spin one, though it can only be in a linear combination of $|s, h\rangle = |1, 1\rangle$ and $|1, -1\rangle$ states.

• What is more, the spin of the photon (or any other massless particle) is not related to the SO(3) rotation group (or its double cover SU(2)), but rather, to the group E(2) of Euclidean motions (translations and rotations) in a two dimensional plane (not ordinary position space). The 'spin group' is the subgroup of the Lorentz group that leaves the momentum vector of the relevant particle invariant, it is also called Wigner's little group or the isotropy subgroup or stabiliser. For a massive particle the momentum vector is time like, e.g. $(m, \vec{0})$ and it is clear that rotations in Euclidean space leave this momentum vector invariant. An isomorphic copy of this group leaves any other time-like momentum 4-vector invariant. For a massless particle, the momentum vector is light like, e.g., $(E, 0, 0, E)$ and one may show that its stabiliser is isomorphic to E(2). The representations of the little group determine the allowed values of helicity. This is responsible for the absence of the zero helicity photon. More generally, a spin s massless particle has only two helicities $\pm s$, nothing in between, unlike massive particles.

• So far, we have only discussed the intrinsic angular momentum (spin) of the photon. A photon could be in a linear combination of states with definite linear momenta $c_1|\mathbf{k}, \lambda\rangle + c_2|\mathbf{k}', \lambda\rangle$. In this manner, it is possible to build a 1 photon state of definite orbital angular momentum l. The total angular momentum includes both the orbital and spin contributions. However, a photon cannot be in a state of total angular momentum $j = 0$, if it were, then the projection of the angular momentum on any axis would be zero, and in particular, the helicity would be zero. So there is no $j = 0$ monopole photon field.

• On the other hand, there are 1 photon states with higher total angular momentum $j = 1, 2 \cdots$ and definite parity, they are called multipole fields. These correspond to electric and magnetic dipole ($j = 1$), electric and magnetic quadrupole ($j = 2$), octupole ..., 2^j-pole fields of classical EM. The electric multipole fields are symmetric linear combinations of helicity $h = \pm 1$ (and angular momentum j) while magnetic multipole fields are anti-symmetric linear combinations.

8 Interaction of atomic electrons with radiation

8.1 Overview of atomic structure & transitions

• Let us begin with a brief overview of atomic structure and transitions. Though weak interactions do play a tiny role (especially in parity-violating effects), atomic structure is determined to an excellent approximation simply by applying non-relativistic quantum mechanics to the Coulomb interaction between electrons and nuclei, while imposing Pauli's exclusion principle. The mass of the nucleus m_N is too large compared to m_e to affect atomic wave functions and energies much. To a good approximation, they depend only on the charge (via the fine structure constant $\alpha = e^2/4\pi\hbar c$) and mass of the electron m_e and the nuclear charge Z. There are two length scales associated with atomic electrons, their reduced Compton wavelength $\hbar/mc = 1/m \approx 4 \times 10^{-14}$m and the atomic size (Bohr radius $a_0 = 1/\alpha m$). The typical atomic size, binding energy and speed of electrons in atoms can be estimated using the Bohr model of the H-atom, or variationally by using Heisenberg's uncertainty principle. Heuristically (replacing $\langle 1/r \rangle$ by $1/\langle r \rangle$ etc.), the expectation value of energy is $E \approx \frac{(\Delta p)^2}{2m} - \frac{\alpha}{\Delta x}$. Using $\Delta p \Delta x \sim 1$ we get $E \approx \frac{1}{2m(\Delta x)^2} - \frac{\alpha}{\Delta x}$. Minimizing in Δx we get $a_0 \approx (\Delta x)_{\rm min} = \frac{1}{\alpha m}$. The resulting energy is $E_{\rm min} \sim -\frac{1}{2} m\alpha^2 \equiv -1{\rm Ry} = -13.6$ eV. Putting $|E| \approx \frac{1}{2} mv^2$ we find that the speed of an electron in an atom is roughly $v = \alpha$ in units of c. We notice that (1) the size of the atom is much larger than the Compton wavelength of its constituents $a_0 = 1/\alpha m_e \gg 1/m_e \gg 1/m_N$; (2) the electron moves non-relativistically and (3) the binding energy $\frac{1}{2} m\alpha^2$ is much less than the rest energy of the constituents (Ry $\ll m_e c^2 \ll m_N c^2$). These are general features of a non-relativistic bound state. To a good approximation, a nucleus is a non-relativistic bound state of nucleons, the solar system is a non-relativistic bound state of sun and planets. A hadron is very different: it is a relativistic bound state of quarks and gluons.

• The solution of the Schrödinger equation for the hydrogen spectrum suggests that an atom would remain forever in an excited stationary state. In the presence of an external EM field, atoms can be stimulated to make transitions between stationary states. The rate of stimulated emission or absorption is proportional to the intensity of light (energy density of the stimulating EM radiation). However, spectroscopists have known for long that atoms in excited states spontaneously decay in about a nanosecond through emission of light, even in the absence of any stimulation. How is this to be explained theoretically? Einstein showed that to understand thermodynamic equilibrium between atoms and radiation (whose spectral distribution is governed by Planck's blackbody law) in a cavity, the rate for spontaneous decay from excited states must be non-zero. Remarkably, Einstein's 1917 argument preceded the formulation of quantum mechanics and the development of time dependent perturbation theory (by Dirac)! However, Einstein's argument does not explain how there can be spontaneous decay even in the absence of external EM fields. The explanation for this is provided by the quantum theory of radiation. Even in the vacuum state where the mean electric and magnetic fields are zero, there are vacuum fluctuations which, in a sense, 'induce' spontaneous emission[44]!

• The interaction of atomic electrons with radiation leads to decay of atoms in excited states to lower energy levels via spontaneous emission of photons. As for any non-relativistic bound state, the reduced wavelength $\lambda = 1/\Delta E$ of photons emitted is large compared to the atomic size a_0. To see this, note that the energy difference $\Delta E \sim \frac{1}{2}m\alpha^2 \sim 1/2ma_0^2$. So $\lambda/a_0 \sim ma_0 = 1/\alpha = 1/v \gg 1$. The radiation field of the decaying atom may be expressed as a multipole expansion in powers of the small parameter $a_0/\lambda \sim \alpha$. We will see that the leading term corresponds to electric dipole radiation (E1), which is followed by electric quadrupole E2 and magnetic dipole (M1) terms.

• Not all transitions between atomic levels are allowed, there are selection rules based on parity and angular momentum conservation. In one photon emission, the angular momentum j_i of the atom in its initial state must equal the combined angular momentum of the photon (j_γ) and final state atom (j_f). By the rules for combining angular momenta we must have $|j_f - j_\gamma| \leq j_i \leq |j_f + j_\gamma|$. Dipole radiation has $J_\gamma = 1$, so dipole transitions must satisfy the selection rule $\Delta j = 0, \pm 1$. Now a 1 photon state (dipole or not) cannot have zero angular momentum, $j_\gamma \neq 0$. It follows that j_i, j_f cannot both be zero, there are no 1-photon transitions between a pair of zero angular momentum states. Parity conservation implies $\Pi_i = \Pi_f \Pi_\gamma$. We will see that $\Pi_\gamma = -1$ for E1 ($\Pi_\gamma = 1$ for M1), so it follows that parity must be reversed in E1 transitions.

8.2 Coupling of atomic electrons to the EM field

• To systematically study the interaction of electrons in an atom with the radiation field, we begin with the atomic hamiltonian H_0 and treat the interaction with the EM field as a perturbation H_1. For simplicity, let us consider a one electron atom such as hydrogen, ignoring the spin of

[44] Note that spontaneous absorption is almost never seen to occur, an atom in its ground state in vacuum is rarely found to spontaneously get excited. A statistical mechanics argument for this may be offered, using the principle of equal a priori probabilities: in equilibrium, all states of a system with the same energy are equally likely. Consider an atom in the presence of electromagnetic radiation present in the vacuum. Suppose the energy difference between the ground and first excited state of the atom is ΔE. There is only one way in which this quantum of energy can be possessed by the atom: by being in the first excited state. On the other hand, this energy can be kept in the radiation field in very many ways, essentially, since the electromagnetic field has very many degrees of freedom, the electric and magnetic fields at each point of space. Since a priori all these possibilities are equally probable, it is infinitely more likely for the quantum of energy to be stored in the electromagnetic field than in the atom. This explains why atoms are typically found in their ground states and are not seen to spontaneously absorb radiation from the vacuum and get excited.

the electron. In rationalized Heaviside-Lorentz units,

$$H_0 = \frac{\mathbf{p}^2}{2m} - \frac{e^2}{4\pi r} \tag{228}$$

We know that the eigenstates of H_0 are $|nlm\rangle$ with energies

$$E_{nlm} = -\frac{me^4}{2(4\pi)^2 n^2 \hbar^2} = -\frac{e^2}{2(4\pi)n^2 a_0} = -\frac{mc^2\alpha^2}{2n^2} \quad \text{where} \quad a_0 = \frac{(4\pi)\hbar^2}{e^2 m} \quad \text{and} \quad \alpha = \frac{e^2}{4\pi\hbar c} \tag{229}$$

are the Bohr radius and fine structure constant. Here the principal quantum number n takes the values $1, 2, \ldots$ corresponding to the K,L,M shells. The angular momentum/azimuthal quantum number $l = 0, 1, \ldots n-1$ corresponds to the s, p, d, f orbitals and the magnetic quantum number $m_l = -l, -l+1, \ldots, l-1, l$ corresponds to the various possible projections of the angular momentum on the z-axis. Each level has a degeneracy of n^2. If we included the spin of the electron, then the energies are not altered but the degeneracies are doubled to $2n^2$ on account of the two possible spin projections on the z-axis $m_s = \pm\frac{1}{2}$. If we have a hydrogenic atom with Z protons in the nucleus e^2 is replaced by Ze^2.

• According to the Schrodinger equation, all the eigenstates $|n, l, m\rangle$ are stable if the hydrogen atom is considered in isolation. However, when we consider the hydrogen atom coupled to the electromagnetic field, we find that all except the ground state $|100\rangle$ are unstable to decay by emission of one or more photons, as is experimentally observed. The decay may be either stimulated by external EM radiation or 'spontaneous'. Even in the vacuum state, the electromagnetic field displays quantum fluctuations and these quantum fluctuations can cause spontaneous emission.

• The interaction of a charged particle with an EM field is given by the Lorentz force law. The interaction of an electron (charge e) with an EM field (given by the vector potential $\mathbf{A}$) may be derived from a Hamiltonian. It is obtained by replacing the electron momentum $\mathbf{p}$ by $\mathbf{p} - e\mathbf{A}/c$. This is called the minimal coupling or Lorentz prescription (it can be proved by checking that the resulting equation of motion for $\mathbf{r}(t)$ is Newton's equation with the Lorentz force). Thus

$$\frac{\mathbf{p}^2}{2m} \to \frac{1}{2m}\left(\mathbf{p} - \frac{e\mathbf{A}}{c}\right)^2 = \frac{\mathbf{p}^2}{2m} - \frac{e}{2mc}\left(\mathbf{p}\cdot\mathbf{A} + \mathbf{A}\cdot\mathbf{p}\right) + \frac{e^2}{2mc^2}\mathbf{A}^2. \tag{230}$$

• In the quantum theory, $\mathbf{p} = -i\hbar\nabla$. So $\mathbf{p}$ and $\mathbf{A}(\mathbf{r})$ do not commute in general, $[p_i, A_j] = -i\hbar\partial_i A_j$. In particular,

$$\mathbf{p}\cdot\mathbf{A} - \mathbf{A}\cdot\mathbf{p} = \sum_i [p_i, A_i] = -i\hbar\nabla\cdot\mathbf{A}. \tag{231}$$

However, in Coulomb gauge, $\nabla\cdot\mathbf{A} = 0$, so $\mathbf{p}\cdot\mathbf{A} = \mathbf{A}\cdot\mathbf{p}$. Thus the hamiltonian becomes

$$H = \left(\frac{p^2}{2m} - \frac{e^2}{4\pi r}\right) - \frac{e}{mc}\mathbf{A}\cdot\mathbf{p} + \frac{e^2}{2mc^2}\mathbf{A}^2 = H_0 - \frac{e}{mc}\mathbf{A}\cdot\mathbf{p} + \frac{e^2}{2mc^2}\mathbf{A}^2 \tag{232}$$

where H_0 is the standard hydrogen hamiltonian including kinetic energy and Coulomb potential. The first interaction term linear in $\mathbf{A}$ is usually the dominant one, responsible for decay by single photon emission in perturbation theory (e.g. electric dipole radiation). The quadratic term in A is usually a small correction, it is called the dielectric term. It is responsible for simultaneous two photon emission processes. We will ignore it. It becomes more important when the effect due to the first term vanishes (as in the decay of the 2S state of hydrogen).

• In addition to this 'minimal' electromagnetic coupling of the electron, there is Pauli's magnetic moment interaction between the spin of the electron and the magnetic field

$$\frac{e\hbar}{2mc}\sigma\cdot\mathbf{B} \quad \text{where} \quad \mathbf{B}=\nabla\times\mathbf{A}. \tag{233}$$

Though an ad hoc addition to the hamiltonian in the non-relativistic treatment of the atom, Pauli's magnetic moment interaction also arises via the minimal coupling Lorentz prescription in Dirac's relativistic theory of the electron. The magnetic moment interaction is usually smaller than the electric dipole interaction and we will ignore it. It is responsible for magnetic dipole radiation from atoms.

8.3 Golden rule for radiative emission rate in first order perturbation theory

We would like to study a radiative transition of an atom from an initial (excited) state $|i\rangle$ to a final state $|f\rangle$ while emitting a single photon with wave vector and polarization $\mathbf{k},\lambda$. There is the analogous absorption process as well. The hamiltonian that governs these processes in the leading approximation mentioned above is

$$H=\left(\frac{p^2}{2m}-\frac{e^2}{4\pi r}\right)-\frac{e}{mc}\mathbf{A}\cdot\mathbf{p}=H_0+H_1(t). \tag{234}$$

$\mathbf{A}(\mathbf{r},t)$ is dependent on time. So we have a time dependent perturbation to the atomic hamiltonian, which we wish to treat to first order. For emission, our initial state at $t=0$, $|i\rangle\otimes|0\rangle$ consists of an atom in state $|i\rangle$ and the radiation field in its vacuum state $|0\rangle$. After a time T, we ask for the amplitude to make a transition to the state $|f\rangle\otimes|1_{\mathbf{k},\lambda}\rangle$ consisting of the atom in state $|f\rangle$ and the radiation field in the given 1-photon state. From 1st order perturbation theory, assuming $i\neq f$, the transition probability is the absolute square of the

$$\text{Amplitude}(f,1_{\mathbf{k}\lambda}\leftarrow i;T)=-\frac{i}{\hbar}\int_0^T\langle f,1_{\mathbf{k},\lambda}|H_1(t)|i\rangle\, e^{-i(E_i-E_f)t/\hbar}\,dt+\cdots \tag{235}$$

Recalling that

$$\mathbf{A}(\mathbf{r},t)=\frac{c}{\sqrt{V}}\sum_{\mathbf{k}',\lambda'}\sqrt{\frac{\hbar}{2\omega_{k'}}}\vec{\epsilon}_{\lambda'}\left[a_{\mathbf{k}',\lambda'}e^{i(\mathbf{k}'\cdot\mathbf{r}-\omega_{k'}t)}+a^\dagger_{\mathbf{k}'\lambda'}e^{-i(\mathbf{k}'\cdot\mathbf{r}-\omega_{k'}t)}\right], \tag{236}$$

we see that only the term involving the creation operator $a^\dagger_{\mathbf{k}',\lambda'}e^{i\omega_{k'}t}$ for $\mathbf{k}'=\mathbf{k}$ and $\lambda'=\lambda$ can have a non-vanishing matrix element between the vacuum initial state and the final 1-photon state $|1_{\mathbf{k}\lambda}\rangle$. On the other hand, for absorption of a photon with wave vector $\mathbf{k}$ and polarization λ, only the annihilation operator term $a_{\mathbf{k},\lambda}e^{-i\omega_k t}$ can contribute.

• Thus, without loss of generality, we may factor $H_1(t) = \tilde{H}_1 e^{\pm i\omega_k t}$ where $\tilde{H}_1$ is time-independent. The $+$ sign is for emission and $-$ sign for absorption. Thus the amplitude for emission is

$$\text{Ampl}_{f,1_{\mathbf{k}\lambda}\leftarrow i}(T) = \frac{1}{i\hbar}\langle f, 1_{\mathbf{k}\lambda}|\tilde{H}_1|i\rangle \int_0^T e^{-i(E_i-E_f-\hbar\omega_k)t/\hbar} = \frac{1}{\hbar}\langle f, 1_{\mathbf{k}\lambda}|\tilde{H}_1|i\rangle \frac{\left[e^{-i(E_i-E_f-\hbar\omega_k)T/\hbar}-1\right]}{(E_i-E_f-\hbar\omega_k)/\hbar} \tag{237}$$

The transition probability is its absolute square

$$\text{Prob}_{f,1_{\mathbf{k}\lambda}\leftarrow i}(T) = \frac{|\langle f|\tilde{H}_1|i\rangle|^2}{\hbar^2}\frac{4\sin^2(\Omega T/2)}{\Omega^2}. \tag{238}$$

where $\Omega = (E_i - E_f - \hbar\omega_k)/\hbar$ and $|e^{i\theta} - 1|^2 = 4\sin^2(\theta/2)$. Plot $\sin^2(\Omega T/2)/\Omega^2$ as a function of Ω for various times T and notice that this function is increasingly concentrated around $\Omega = 0$ as T grows. For long times T, the transition probability is significant only when $\Omega = (E_i - E_f - \hbar\omega_k)/\hbar \approx 0$. Recalling the representation of the Dirac δ function,

$$\frac{2}{\pi}\lim_{T\to\infty}\frac{\sin^2\frac{1}{2}\Omega T}{\Omega^2 T} = \delta(\Omega). \tag{239}$$

we see that for long times, the transition probability is proportional to the time:

$$\text{Prob}_{f,1_{\mathbf{k}\lambda}\leftarrow i}(T) \to \frac{2\pi}{\hbar}|\langle f|\tilde{H}_1|i\rangle|^2\, T\, \delta(E_i - E_f - \hbar\omega_k) \tag{240}$$

We used $\hbar^{-1}\delta(\Omega) = \delta(\hbar\Omega)$. Dividing by T, the transition probability per unit time (or transition rate) approaches a constant for long times

$$\text{Rate}_{f,1_{\mathbf{k}\lambda}\leftarrow i} \to \frac{2\pi}{\hbar}|\langle f|\tilde{H}_1|i\rangle|^2\, \delta(E_i - E_f - \hbar\omega_k) \tag{241}$$

The same formula holds for absorption with the change $-\hbar\omega_k \to +\hbar\omega_k$.

• In the case of emission, when the volume of our box $V \to \infty$, there is a continuous energy spectrum of possible final state photons which could have wave vectors pointing in various directions. So it is interesting to find the transition rate for photons emitted into an elemental solid angle $d\Omega$ around the direction (θ, ϕ) and having an energy lying between $\hbar\omega$ and $\hbar(\omega + d\omega)$. This rate is given by the product of the above rate by the number of photon states in this range. We will eventually sum/integrate over all the possible states (energies, directions and polarizations) of the emitted photon to find the total decay rate, but we go in steps.

• Now the energy of a photon is $E = \hbar\omega_k = \hbar c|\mathbf{k}|$. So photon states in a given energy range lie in a spherical shell in $\mathbf{k}$-space. In general, we associate one quantum state to a phase region of volume $d^3\mathbf{r}d^3\mathbf{p}/h^3$. So the number of photon states in a volume V (with fixed polarization - we will sum over polarizations later) with wave vectors in the range $[\mathbf{k}, \mathbf{k} + d\mathbf{k}]$ is ($\mathbf{p} = \hbar\mathbf{k}$)

$$dn = \frac{V\, d^3\mathbf{p}}{(2\pi)^3\hbar^3} = \frac{V\, d^3\mathbf{k}}{(2\pi)^3} = \frac{1}{(2\pi)^3}V\, k^2\, dk\, d\Omega \tag{242}$$

upon transforming to spherical polar coordinates. For photons emitted into the solid angle $d\Omega$, let us denote the number of photon states with energy in the interval $[E, E + dE]$ by $\rho(E,\Omega) dE d\Omega$. Then

$$\rho(E,\Omega)\, dE\, d\Omega = dn = \frac{1}{(2\pi)^3} V\, k^2\, dk\, d\Omega. \tag{243}$$

Thus, the density of states ($E = \hbar\omega = \hbar ck$)

$$\rho(E,\Omega)\, dE\, d\Omega = \frac{1}{(2\pi)^3} V\, k^2\, d\Omega\, \frac{dk}{dE}\, dE = \frac{V k^2 d\Omega}{(2\pi)^3 \hbar c} dE = \frac{V \omega^2 d\Omega}{(2\pi)^3 \hbar c^3} dE. \tag{244}$$

Multiplying by the previously obtained rate and integrating over photon energies (which is fixed by the energy conserving δ-function), we obtain Fermi's Golden Rule for the emission rate of a photon with polarisation λ and wave vector $\mathbf{k}$ pointing in the solid angle $d\Omega$ around the direction defined by θ, ϕ:

$$w(\Omega) d\Omega = d\Omega \frac{2\pi}{\hbar} \int |\langle f|\tilde{H}_1|i\rangle|^2\, \delta(E_i - E_f - \hbar\omega_k) \rho(E,\Omega) dE = \frac{2\pi}{\hbar} |\langle f|\tilde{H}_1|i\rangle|^2 \rho(E_{\mathbf{k}},\Omega)\, d\Omega \tag{245}$$

where $E_{\mathbf{k}} = \hbar\omega_{\mathbf{k}} = E_i - E_f$. The letter w is a commonly used symbol to denote the rate of a process in physics. This formula for w is called Fermi's golden rule.

• Now let us apply this to the case of photon emission. The relevant interaction hamiltonian is the coefficient of the creation operator $a^\dagger_{\mathbf{k},\lambda} e^{i\omega_{\mathbf{k}} t}$ in $-(e/mc)\mathbf{A}\cdot\mathbf{p}$:

$$\tilde{H}_1 = -\frac{e}{mc} \frac{c}{\sqrt{V}} \sqrt{\frac{\hbar}{2\omega_k}} e^{-i\mathbf{k}\cdot\mathbf{r}}\, \vec{\epsilon}_\lambda \cdot \mathbf{p}. \tag{246}$$

Note that $\langle 1_{\mathbf{k}\lambda}|a^\dagger_{\mathbf{k}\lambda}|0\rangle = 1$, so we do not indicate the photon creation operator or photon state any more. Thus the rate for photon emission into $d\Omega$ is

$$w(\Omega) d\Omega = \frac{2\pi}{\hbar} \frac{e^2\hbar}{2m^2\omega_k V} |\langle f|e^{-i\mathbf{k}\cdot\mathbf{r}} \vec{\epsilon}_\lambda \cdot \mathbf{p}|i\rangle|^2 \frac{V\omega_k^2}{(2\pi)^3} \frac{d\Omega}{\hbar c^3} = \frac{e^2\omega_k}{8\pi^2 m^2 \hbar c^3} |\langle f|e^{-i\mathbf{k}\cdot\mathbf{r}} \vec{\epsilon}_\lambda \cdot \mathbf{p}|i\rangle|^2\, d\Omega \tag{247}$$

Notice that the factor of V in the density of states cancels the $1/V$ from the square of the matrix element leaving a finite limit as $V \to \infty$.

8.4 Electric dipole approximation

To determine the emission rate, we must evaluate the matrix element

$$\langle f|e^{-i\mathbf{k}\cdot\mathbf{r}}\, \vec{\epsilon}_\lambda \cdot \mathbf{p}|i\rangle \tag{248}$$

between the initial and final atomic states $|i\rangle, |f\rangle$. Here $\mathbf{k}, \vec{\epsilon}_\lambda$ are the photon wave vector and polarization while $\mathbf{r}, \mathbf{p}$ are the position and momentum operators of the electron. Computing this matrix element between atomic energy eigenstates is in general quite difficult since it involves the exponential of the position operator. To make progress we would like to expand the exponential in a series.

$$e^{-i\mathbf{k}\cdot\mathbf{r}} = 1 - i\mathbf{k}\cdot\mathbf{r} - (\mathbf{k}\cdot\mathbf{r})^2 + \cdots . \tag{249}$$

This is a reasonable approximation if we are considering EM radiation in the visible/UV/IR region of the spectrum. The wave number of $k = \frac{2\pi}{\lambda}$ of visible light corresponds to wave lengths of several thousands of angstroms while atomic wave functions are spread over lengths of the order of an angstrom. So the order of magnitude of

$$\langle \mathbf{k}\cdot\mathbf{r}\rangle \;\text{ is }\; \frac{\text{size of atom}}{\text{wave length of light}} \sim 10^{-3}. \tag{250}$$

The electric dipole approximation E1 consists in approximating $e^{-i\mathbf{k}\cdot\mathbf{r}}$ by 1. Retaining the next term $e^{-i\mathbf{k}\cdot\mathbf{r}} \approx 1 - i\mathbf{k}\cdot\mathbf{r}$ is called the electric quadrupole approximation E2. We expect the quadrupole term to be a thousand times smaller than the dipole term.

• To compute the transition rate in the long wavelength E1 approximation, we need to find the matrix element

$$\langle f|\vec{\epsilon}_\lambda\cdot\mathbf{p}|i\rangle = \vec{\epsilon}_\lambda\cdot\langle f|\mathbf{p}|i\rangle. \tag{251}$$

The matrix elements of position between hydrogen energy levels are somewhat easier to compute (by direct integration), than the momentum matrix elements. We may relate them using a trick: the commutator of the hydrogen hamiltonian with position is proportional to momentum:

$$[\mathbf{r}, H_0] = [\mathbf{r}, \frac{p^2}{2m}] = i\frac{\hbar}{m}\mathbf{p} \tag{252}$$

Bearing in mind that $|i\rangle$ and $|f\rangle$ are eigenstates of H_0 with energies $E_{i,f}$ we get

$$i\frac{\hbar}{m}\langle f|\mathbf{p}|i\rangle = \langle f|[\mathbf{r}, H_0]|i\rangle = (E_i - E_f)\langle f|\mathbf{r}|i\rangle = \hbar\omega_k\langle f|\mathbf{r}|i\rangle. \tag{253}$$

Thus

$$\langle f|\vec{\epsilon}_\lambda\cdot\mathbf{p}|i\rangle = -im\omega\vec{\epsilon}_\lambda\cdot\langle f|\mathbf{r}|i\rangle \quad \text{and} \quad |\langle f|\vec{\epsilon}_\lambda\cdot\mathbf{p}|i\rangle|^2 = m^2\omega^2|\vec{\epsilon}_\lambda\cdot\langle f|\mathbf{r}|i\rangle|^2 \tag{254}$$

So the transition rate in the E1 approximation becomes

$$w(\Omega)d\Omega = \frac{e^2\omega^3}{8\pi^2\hbar c^3}|\vec{\epsilon}_\lambda\cdot\langle f|\mathbf{r}|i\rangle|^2\, d\Omega_k. \tag{255}$$

8.5 Selection rules for E1 transitions

• Selection rules state that the matrix element for electric dipole transitions $\langle f|\mathbf{r}|i\rangle$ vanish for certain initial and final states. So there can be no E1 transitions for certain quantum numbers of the initial and final states.

• The parity selection rule states that an E1 transition is forbidden if the initial and final atomic levels $|i\rangle, |f\rangle$ have the same parity. To see this first recall that parity acts as

$$\mathbf{P}\psi(\mathbf{r}) = \psi(-\mathbf{r}) \quad \text{and} \quad \mathbf{P}^2 = I \quad \Rightarrow \quad \mathbf{P}^{-1} = \mathbf{P}. \tag{256}$$

It follows that parity anti-commutes with position, for

$$(\mathbf{Pr} + \mathbf{rP})\psi(\mathbf{r}) = -\mathbf{r}\psi(-\mathbf{r}) + \mathbf{r}\psi(-\mathbf{r}) = 0. \tag{257}$$

Therefore $\mathbf{P}^{-1}\mathbf{r}\mathbf{P} = \mathbf{P}\mathbf{r}\mathbf{P} = -\mathbf{r}$. Using this, the dipole matrix element satisfies

$$\langle f|\mathbf{r}|i\rangle = -\langle f|\mathbf{PrP}|i\rangle = -P_f P_i \langle f|\mathbf{r}|i\rangle \quad \Rightarrow \quad (1 + P_f P_i)\langle f|\mathbf{r}|i\rangle = 0. \tag{258}$$

Hence, either the matrix element vanishes or the product of parities $P_f P_i$ is -1. So the parity must change in an E1 transition.

• Recall that the parity of the hydrogen level $|nlm\rangle$ is $(-1)^l$. So the E1 transition $|nlm\rangle \to |n'l'm'\rangle$ is forbidden if $l + l'$ is odd. In particular, $l' - l = 0$ is forbidden. So for instance there cannot be E1 transitions between two S-wave states or two P-wave states or two D-wave states etc. In fact, combining with the angular momentum selection rule (see below), this implies $\Delta l = \pm 1$.

• Angular momentum selection rule. As we argue heuristically below, in E1 transitions the emitted photon carries angular momentum $j_\gamma = 1$ in units of $\hbar$. (See QM2 lecture notes for a more detailed treatment or see a book on quantum mechanics) Suppose j_i, j_f are the total angular momentum quantum numbers of the initial and final electronic states. Then the selection rule states that an E1 transition is forbidden if

$$\Delta j = j_f - j_i \neq 0, \pm 1. \tag{259}$$

Moreover the transition from $j_i = 0$ to $j_f = 0$ is also forbidden. Heuristically, in the matrix element $\langle f|\mathbf{r}|i\rangle$, the operator $\mathbf{r}$ behaves as if it has angular momentum one[45]. Then we are adding the angular momentum j_i of the initial state to this, $j_i \otimes 1$ and we know that if $j_i \neq 0, \frac{1}{2}$, the resulting system behaves as if it has angular momentum $j_f = j_i - 1$ or j_i or $j_i + 1$. So j_i and j_f must differ by 0 or ± 1. If $j_i = \frac{1}{2}$ then j_f must be $\frac{1}{2}$ or $3/2$. If $j_i = 0$ then $j_f = 1$, so $0 \to 0$ E1 transition is forbidden.

• In particular, the 'metastable' 2S state of hydrogen is stable to radiative decay in the E1 approximation. In fact, it is stable to decay via all 1 photon electric and magnetic multipole transitions, it decays via 2 photon emission. This accounts for its unusually long mean lifetime of 0.12 seconds.

8.6 Polarization and direction averaged E1 emission rate

So far we have found that the E1 transition rate $i \to f$ accompanied by the emission of a photon into solid angle $d\Omega$ with polarization λ is

$$w(\Omega)d\Omega = \frac{e^2\omega^3}{8\pi^2\hbar c^3}|\vec{\epsilon}_\lambda \cdot \langle f|\mathbf{r}|i\rangle|^2 \, d\Omega. \tag{260}$$

[45]To motivate this, notice that $\mathbf{r} = (x, y, z) = r(\cos\theta, \sin\theta\cos\phi, \sin\theta\sin\phi)$ so $\frac{1}{r}(z, x \pm iy) = (\cos\theta, \sin\theta e^{\pm i\phi})$ which we notice are proportional to the $l = 1$ spherical harmonics $Y_{10}, Y_{1,\pm 1}$.

From here on, we are interested in the rate of emission, irrespective of the direction or polarization of the outgoing photon. To do this averaging, it is convenient to write the square of the dot product in terms of the angle Θ_λ between the unit polarization vector $\vec{\epsilon}_\lambda$ and the dipole vector matrix element $\langle f|\mathbf{r}|i\rangle \equiv \mathbf{r}_{fi}$

$$|\mathbf{r}_{fi}\cdot\epsilon_\lambda| = |\mathbf{r}_{fi}|\ \cos\Theta_\lambda. \tag{261}$$

Then

$$w(\Omega)\,d\Omega = \frac{e^2\omega^3}{8\pi^2\hbar c^3}|\mathbf{r}_{fi}|^2\ \cos^2\Theta_\lambda\,d\Omega \tag{262}$$

To do the sum over polarizations λ, let us work in a real right-handed orthonormal basis $\epsilon_1, \epsilon_2, \hat{\mathbf{k}}$ with $\epsilon_1\times\epsilon_2 = \hat{\mathbf{k}}$ and define spherical polar coordinates θ, ϕ: $\mathbf{r}_{fi}$ makes an angle θ with $\mathbf{k}$ and the projection of $\mathbf{r}_{fi}$ onto the $\epsilon_1 - \epsilon_2$ plane makes an angle ϕ with ϵ_1. Then

$$\cos\Theta_1 = \sin\theta\cos\phi \quad \text{and} \quad \cos\Theta_2 = \sin\theta\sin\phi. \tag{263}$$

Then the sum over polarizations just gives a $\sin^2\theta$ factor

$$\sum_{\lambda=1}^{2} w(\Omega)d\Omega = \frac{e^2\omega^3}{8\pi^2\hbar c^3}|\mathbf{r}_{fi}|^2\ (\cos^2\Theta_1+\cos^2\Theta_2)d\Omega\ = \frac{e^2\omega^3}{8\pi^2\hbar c^3}|\mathbf{r}_{fi}|^2\ \sin^2\theta\,d\Omega \tag{264}$$

• There is another way of summing over polarizations to show that

$$\sum_\lambda |\vec{\epsilon}_\lambda\cdot\langle f|\mathbf{r}|i\rangle|^2 = |\mathbf{r}_{fi}|^2\sin^2\theta \tag{265}$$

where θ is the angle between the dipole matrix element and $\mathbf{k}$. To do this we abbreviate $\langle f|\mathbf{r}|i\rangle \equiv \mathbf{r} = (r_1, r_2, r_3)$, write in components and use the completeness of the orthonormal system $\epsilon_1, \epsilon_2, \hat{k}$:

$$\begin{aligned}\sum_\lambda |\vec{\epsilon}_\lambda\cdot\langle f|\mathbf{r}|i\rangle|^2 &= \sum_\lambda r_i\epsilon_{\lambda i}r_j^*\epsilon_{\lambda j} = \sum_{i,j=1}^{3} r_i r_j^*\sum_\lambda \epsilon_{\lambda i}\epsilon_{\lambda j}\\ &= \sum r_i r_j^*(\delta_{ij}-\hat{k}_i\hat{k}_j) = |r|^2 - |\mathbf{r}\cdot\hat{\mathbf{k}}|^2 = r^2\sin^2\theta.\end{aligned} \tag{266}$$

• Before performing the average over directions we remark that to obtain the energy radiated per unit time into solid angle $d\Omega$ we must multiply the above polarization averaged rate by the energy per photon $\hbar\omega$

$$\text{Power}_{d\Omega} = \frac{e^2\omega^4}{8\pi^2 c^3}|\mathbf{r}_{fi}|^2\ \sin^2\theta\,d\Omega \tag{267}$$

The power is proportional to the fourth power of ω and to the square of the matrix element of the electric dipole moment of the electron $e\mathbf{r}$. Moreover, it has a $\sin^2\theta$ angular dependence where θ is the angle between the dipole moment matrix element and the direction of propagation $\hat{k}$. No energy is radiated along the direction of the dipole moment ('it is darkest underneath the candle'!). This is reminiscent of the formula for the intensity of energy radiated by an oscillating electric dipole[46] $\mathbf{p}(t) = 2p_0 \cos\omega t \; \hat{z}$ in classical E & M, which may be described by the flux of the time-averaged Poynting vector $\langle \mathbf{S} \rangle$ across the element $r^2 d\Omega \hat{r}$. In HL units (see Griffiths and put $\epsilon_0 = 1$)

$$\langle \mathbf{S} \rangle \cdot r^2 d\Omega \hat{r} = \frac{\omega^4}{8\pi^2 c^3} p_0^2 \frac{\sin^2\theta}{r^2} \hat{r} \cdot r^2 d\Omega \hat{r} = \frac{\omega^4}{8\pi^2 c^3} p_0^2 \sin^2\theta d\Omega. \tag{268}$$

Here $2p_0 = 2q_0 s$ is the maximum value of the dipole moment and s is the separation between the oscillating charges $\pm 2q(t) = \pm 2q_0 \cos\omega t$.

• The ω^4 dependence is also expected classically from the Larmor formula which states that the power radiated is proportional to the square of acceleration [acceleration $\sim \omega^2$ from the formula for $\mathbf{p}(t)$].

• Returning to the rate of E1 transitions $w_{d\Omega}$, we follow the polarization sum by the integral over directions

$$\int_0^\pi \int_0^{2\pi} \sin^2\theta \sin\theta \, d\theta \, d\phi = 2\pi \int_{-1}^1 \sin^2\theta \, d(\cos\theta) = 2\pi \times 2 \int_0^1 (1 - t^2) \, dt = \frac{8\pi}{3}. \tag{269}$$

So the polarization and direction averaged emission rate is

$$w = \sum_\lambda \int w_{d\Omega} = \frac{e^2 \omega^3}{3\pi\hbar c^3} |\mathbf{r}_{fi}|^2 = \left(\frac{e^2}{4\pi\hbar c} \right) \frac{4\omega^3}{3c^2} |\mathbf{r}_{fi}|^2 = \frac{4\alpha\omega^3}{3c^2} |\mathbf{r}_{fi}|^2 \tag{270}$$

It has dimensions of inverse time. Here $\alpha = \frac{e^2}{4\pi\hbar c}$ is the fine structure constant. We have computed this rate to first order in perturbation theory in the electric dipole approximation. Multiplying by the photon energy $\hbar\omega$ we also get the power radiated in all directions

$$\text{Power} = \frac{e^2 \omega^4}{3\pi c^3} |\mathbf{r}_{fi}|^2. \tag{271}$$

This too is reminiscent of the classical formula for the power radiated by the above oscillating electric dipole. In HL units (see Griffiths and mind the factor of two in the definition of dipole moment)

$$\langle \text{Power} \rangle_{\text{classical}} = \int \langle \mathbf{S} \rangle \cdot \hat{r} \, r^2 d\Omega = \frac{\omega^4 p_0^2}{3\pi c^3}. \tag{272}$$

This increase in power with the fourth power of the frequency is used to explain the blueness of sunlight scattered by the atmosphere.

• Sakurai points out that Heisenberg obtained the formula for the rate w prior to the development of the quantum theory of radiation, by a use of the correspondence principle.

8.7 Life-time of 2p state of hydrogen: Lyman α transition

• The mean lifetime τ of state $|i\rangle$ is defined as the reciprocal of the sum of transition rates to all possible final states $|f\rangle$ allowed by the selection rules and energy conservation

$$\frac{1}{\tau_i} = \sum_f w_{f \leftarrow i}. \tag{273}$$

[46] The factors of 2 are because $e^{i\omega t} + e^{-i\omega t} = 2\cos\omega t$ and we used exponentials in the quantum theory.

The sum over final states $\sum_f w_{f\leftarrow i}$ is called the total decay rate of the state i, the summands being the partial decay rates. $\Gamma_{f\leftarrow i} = \hbar w_{f\leftarrow i}$ is called the partial energy width, $\Gamma_{tot} = \sum_f \Gamma_{f\leftarrow i}$ is called the total energy width of the unstable state i. The lifetime $\tau = \hbar/\Gamma_{tot}$. Note that we do not speak of partial life-times.

• Now consider the first excited states of hydrogen, they are of course degenerate, including the 2s and 2p levels. The only lower level they can decay to is 1s (Note that 2p $\to$ 2s has zero rate since $\omega = 0$). The 2s $\to$ 1s electric dipole transition is forbidden since they have the same parity. The 2p to 1s E1 transition is allowed by the selection rules. Spectroscopists call it the Lyman α transition. The Lyman α transition has been a useful tool in cosmology. The Lyman alpha 'forest': absorption lines in light from very far away quasars due to excitation of hydrogen in inter-stellar gas. One sees not one line but several lines, indeed a forest of lines, because of the shifting of spectral lines due to relative motions of the various clouds of intervening gas. But all the lines are believed to correspond to the same Lyman α atomic transition.

• τ_{2p} in the E1 approximation is given by

$$\tau_{2p}^{-1} = w = \frac{4\alpha\omega^3}{3c^2}|\langle 1s0|\mathbf{r}|2pm_l\rangle|^2 \tag{274}$$

The rate is the same for all the values of $m_l = 0, \pm 1$. Let us compute it for $m_l = 0$. Recall that the hydrogen wave functions are given by $\psi_{nlm}(r,\theta,\phi) = R_{nl}(r)Y_{lm}(\theta,\phi)$. We have

$$R_{1s} = \frac{2}{\sqrt{a^3}}e^{-r/a}, \quad Y_{00} = \frac{1}{\sqrt{4\pi}}, \quad R_{2p} = \frac{1}{\sqrt{24a^3}}\frac{r}{a}e^{-r/2a}, \quad Y_{10} = \sqrt{\frac{3}{4\pi}}\cos\theta \tag{275}$$

Note that the R_{2p} decays twice as slowly as R_{1s}, excited states are more spread out. Also $R_{2p}(r)$ has one more node than R_{1s} as we expect of the first excited state. The pre-factors are fixed by normalisation $\int |\psi|^2 r^2 dr d\Omega = 1$, $\int |Y_{lm}|^2 d\Omega = 1$. So

$$\langle 1s0|\mathbf{r}|2p0\rangle = \int dr d\theta d\phi\, r^2 \sin\theta \,\frac{1}{\sqrt{4\pi}}\sqrt{\frac{3}{4\pi}}\cos\theta\,\frac{2}{\sqrt{a^3}}e^{-r/a}\,\frac{1}{\sqrt{24a^3}}\frac{r}{a}e^{-r/2a}\,(x,y,z) \tag{276}$$

Now $\mathbf{r} = (x,y,z) = r(\cos\theta, \sin\theta\cos\phi, \sin\theta\sin\phi)$. We see that only z component contributes to the matrix element, the x, y components vanish since they are proportional to the integrals $\int_0^{2\pi}\cos\phi d\phi$ and $\int_0^{2\pi}\sin\phi\, d\phi$. So we only need

$$\langle 1s0|z|2p0\rangle = \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta\,\frac{1}{\sqrt{4\pi}}\cdot\cos\theta\cdot\sqrt{\frac{3}{4\pi}}\cos\theta\int_0^{\infty} dr\, r^2\frac{2}{\sqrt{a^3}}e^{-r/a}\cdot r\cdot\frac{1}{\sqrt{24a^3}}\frac{r}{a}e^{-r/2a}. \tag{277}$$

where $a = \frac{\hbar}{mc\alpha} = \frac{4\pi\hbar^2}{me^2} = .53$ Angstroms is the Bohr radius. Doing the integrals we get a matrix element that is of order of the Bohr radius

$$\langle 1s0|z|2p0\rangle = \frac{256}{243\sqrt{2}}a = 0.74a = 3.9\times 10^{-9}\text{cm} \quad \Rightarrow \quad |\mathbf{r}_{fi}|^2 = 0.55\,a^2 = 1.52\times 10^{-17}\,\text{cm}^2. \tag{278}$$

The energy of the emitted photon is

$$\hbar\omega = E_{2p} - E_{1s} = -\frac{mc^2\alpha^2}{2}\left(\frac{1}{2^2} - \frac{1}{1^2}\right) = \frac{3mc^2\alpha^2}{8} = -\frac{e^2}{8\pi a}\left(\frac{1}{2^2} - \frac{1}{1^2}\right) = \frac{3e^2}{32\pi a} = 13.6 \times \frac{3}{4} = 10.2\,\text{eV}. \tag{279}$$

Since $\hbar = 6.52\times10^{-16}$ eV.s, $\omega = 1.56\times10^{16}$/s corresponding to the wave length $\lambda = 2\pi c/\omega = 1216$ Angstroms of the Lyman α line. With $\alpha = 1/137$ and $c = 3 \times 10^{10}$ cm/s we get the numerical values of the 2p-1s Lyman α transition rate and mean lifetime of the 2p level

$$w = \frac{4\alpha\omega^3}{3c^2}|\mathbf{r}_{fi}|^2 = 0.6 \times 10^9\ s^{-1} \quad \Rightarrow \quad \tau_{2p} \approx 1.6\,\text{ns} \tag{280}$$

• The energy widths of excited levels that decay via E1 transitions are of order $\Gamma = \hbar/\tau \sim \frac{\alpha\omega^3}{c^2}|\mathbf{r}_{fi}|^2$. Now the dipole matrix element is of order the Bohr radius $|\mathbf{r}_{fi}|^2 \sim a_0^2$ while $\hbar\omega$ is of the order of a Rydberg $\text{Ry} = \frac{1}{2}mc^2\alpha^2$. Let us write the width in natural units, where $a_0 = \frac{1}{\alpha m}$. We get $\Gamma \sim \text{Ry}\,\alpha^3 = m\alpha^5$. This gives a simple way of estimating its numerical value, $\tau \sim \frac{1}{m\alpha}\frac{1}{\alpha^4}$. We can restore factors of $c, \hbar$ by not ing that $c\tau$ is a length, so it must be a multiple of the Compton wave length of the electron, so $c\tau = \frac{\hbar}{mc\alpha}\frac{1}{\alpha^4}$. The first factor is the Bohr radius a_0. Putting in $a_0 = .53$ Angstroms, $\alpha = 1/137$ and $c = 3 \times 10^8$ m/s we get $\tau \sim 10^{-9}$ s.

• A similar calculation can be performed to find the rates for E1 transitions between other hydrogen levels that are not forbidden by the selection rules. For small values of n, the dipole matrix element $|\mathbf{r}_{fi}|$ is of order the Bohr radius and $\omega \approx 10^{16}$ Hz, resulting in lifetimes on the order of nanoseconds or tens of nanoseconds. Some of these allowed decays are 2p-1s, 3s-2p, 3p-1s, 3p-2s, 3d-2p e.t.c.

• When a decay is forbidden in the electric dipole approximation we go to the next approximation: the electric quadrupole approximation E2 coming from the second term in $e^{-i\mathbf{k}\cdot\mathbf{r}} = 1 - i\mathbf{k}\cdot\mathbf{r} + \ldots$. One must also consider the magnetic dipole approximation M1 due to Pauli's coupling of the magnetic dipole moment of electrons to the radiation field. The lifetimes of excited states that decay to leading order via E2 and M1 transitions are about a million times as long as those that decay via E1

$$\tau_{\text{E2,M1}} \sim \left(\frac{\lambda}{a_0}\right)^2 \tau_{\text{E1}} \sim 10^{-3}\ \text{s}. \tag{281}$$

• The 2s level cannot decay in the E1 approximation due to the parity selection rule, the only lower level 1s has the same parity. In fact is is forbidden to decay even via E2 or M1. It eventually decays via 2 photon emission. This is in fact the superposition of two amplitudes, one coming from treating the interaction hamiltonian $-\frac{e}{mc}\mathbf{A}\cdot\mathbf{p}$ to second order in perturbation theory and by treating the dielectric term $(\frac{e^2}{2mc^2}\mathbf{A}^2)$ in first order perturbation theory. The 2s level consequently has a long lifetime $\tau_{2s} \approx .12$ s.

9 Spin zero, half and one relativistic fields and their interactions

9.1 Need for quantum fields for matter particles

• We are chiefly concerned with phenomena where sub-nuclear particles have energies much more than their rest masses, so that they travel at relativistic speeds. Moreover, these phenomena involve length scales comparable to or less than particle de Broglie wavelengths, necessitating a quantum mechanical treatment. The theory of quantum fields is a formalism that (among other things) is able to incorporate both special relativistic and quantum mechanical principles.

• Classical Maxwell theory deals with propagating disturbances in electric and magnetic fields, they involve infinitely many degrees of freedom. The concept of a propagating field was needed to accommodate the finite speed of propagation, it replaced notions of instantaneous action at a distance. With the discovery that light photons also displayed quantum mechanical and particle-like behavior, it became necessary to quantize the EM field and deal with field operators for the photon. The quantum field theory of photons was developed from 1928-29 onwards.

• On the other hand, particles like electrons in atoms had already been given a successful quantum mechanical treatment involving finitely many degrees of freedom (described by the coordinates x_i). Coordinates and momenta are hermitian operators satisfying Heisenberg's commutation relations $[\hat{x}^i(t), \hat{p}^j(t)] = i\hbar\delta^{ij}$ and time is a parameter. A familiar consequence of $[\hat{x}, \hat{p}_x] = i\hbar$ is that if a particle can be found at all possible locations $-\infty < x < \infty$, then $\hat{p}_x = -i\hbar\partial_x$ also has a continuous spectrum $-\infty < p_x < \infty$.

• However, this framework needed to be generalized to incorporate special relativity. It may not be obvious why a relativistic generalization would require the concept of an electron field. Indeed, as Weisskopf and Gottfried point out, one could try to develop a relativistic quantum mechanics of one electron (or a definite number of electrons) with finitely many degrees of freedom, by defining 4-vectors of quantum mechanical operators $\hat{x}^\mu = (\hat{t}, \hat{x}, \hat{y}, \hat{z})$ and $\hat{p}^\mu = (\hat{H}/c, \hat{p}_x, \hat{p}_y, \hat{p}_z)$ (perhaps parametrized by proper-time). Generalizing the Heisenberg commutation relations in a Lorentz covariant manner leads to $[\hat{x}^\mu, \hat{p}^\nu] = -i\hbar\eta^{\mu\nu}$. This formula takes the same form in all inertial frames $x^{'\mu} = \Lambda^\mu_\nu x_\nu, p^{'\nu} = \Lambda^\nu_\rho p^\rho$ since $\eta^{\mu\nu}$ does not change under Lorentz transformations. It also reduces to the Heisenberg relations for the spatial components. In particular, we must have $[\hat{x}, \hat{H}] = 0$ or $\frac{d}{d\tau}\hat{x} = 0$. It appears that the particle would have to have zero proper velocity in any state. This seems to be too stringent a dynamical restriction to accommodate the observed phenomena. Moreover, if we assume that the time operator has a continuous spectrum (as observed), then the relation $[\hat{t}, \hat{H}/c] = -i\hbar$ would imply that the hamiltonian must have a continuous spectrum, including arbitrarily negative values, making the stability of matter problematic. These predictions are not borne out by the behavior of relativistic electrons. On the other hand, what is observed is that at relativistic speeds, the number of electrons is not conserved due to pair production and annihilation, necessitating a framework with an indefinite number of degrees of freedom.

• This attempt at elevating time to an operator so that Lorentz transformations could transform position and time operators into each other does not succeed. Another option is to demote x, y, z to parameters just like time and introduce instead quantum fields $\phi_i(x, y, z, t)$ that depend on space and time parameters and transform among each other under the Poincare group. Indeed, this is what the analogy with the electromagnetic field would suggest since E, B are functions of x, y, z, t and transform into each other under Lorentz transformations. The field concept also brings in an infinite number of degrees of freedom as required by experiment. As history has shown, the quantum field concept has succeeded in particle physics far beyond the expectations of the initiators of the subject (Dirac, Pauli, Heisenberg et. al.). Thus we embrace the field concept both for force carriers like the photon and matter particles like the electron.

9.2 Lagrangian, equations of motion and Noether's theorem

• The classical dynamics of fields is governed by relativistic wave equations such as the Maxwell, Klein-Gordon, Dirac/Pauli/Weyl and Yang-Mills equations which we will study shortly. These follow from various Lagrangian densities (the word density is often dropped) $\mathcal{L}$ via Hamilton's principle of extremal action $\delta \int d^4x \mathcal{L} = 0$. For example, if ϕ_a are a collection of fields and $\mathcal{L} = \mathcal{L}(\phi_a, \partial_\mu \phi_a)$, then the Euler-Lagrange field equation for each component a is

$$\frac{\partial \mathcal{L}}{\partial \phi_a} = \partial_\mu \frac{\partial \mathcal{L}}{\partial \partial_\mu \phi_a} \quad \text{or} \quad \frac{\partial \mathcal{L}}{\partial \phi_a} = \partial_\mu \pi_a^\mu \quad \text{where} \quad \pi_a^\mu = \frac{\partial \mathcal{L}}{\partial \partial_\mu \phi_a} \tag{282}$$

$\phi_a(x,t)$ specify the instantaneous configuration of the system. $\pi_a^0 = \frac{\partial \mathcal{L}}{\partial \dot{\phi}_a}$ is called the momentum conjugate to ϕ_a and is conserved if ϕ_a does not appear in $\mathcal{L}$. The energy/hamiltonian density $\mathcal{H} = \pi_a^0 \dot{\phi}_a - \mathcal{L}$ is time-independent if the Lagrangian is not explicitly dependent on time. The canonical Poisson brackets are $\{\phi_a(\mathbf{x},t), \pi_b^0(\mathbf{y},t)\} = \delta_{ab}\delta^3(\mathbf{x}-\mathbf{y})$.

• Noether's theorem: suppose $\mathcal{L}$ is invariant under an infinitesimal symmetry transformation $\phi_a \to \phi_a + \delta\phi_a$, then

$$0 = \delta \mathcal{L} = \frac{\partial \mathcal{L}}{\partial \phi_a} \delta\phi_a + \frac{\partial \mathcal{L}}{\partial \partial_\mu \phi_a} \delta \partial_\mu \phi_a = (\partial_\mu \pi_a^\mu)\delta\phi_a + \pi_a^\mu(\partial_\mu \delta\phi_a) = \partial_\mu(\pi_a^\mu \delta\phi_a) \tag{283}$$

by use of the equations of motion. Thus the current $j^\mu = \pi_a^\mu \, \delta\phi_a$ is locally conserved $\partial_\mu j^\mu = 0$ or $\frac{\partial j^0}{\partial t} + \nabla \cdot \vec{j} = 0$ where $j^\mu = (j^0, \vec{j})$. It follows that the 'Noether charge' in a region V, $Q \equiv \int_V j^0 \, d^3x$ is independent of time, $\dot{Q} = \int_V \dot{j}^0 d^3x = -\int_V \nabla \cdot \vec{j} d^3x = -\int_{\partial V} \vec{j} \cdot \hat{n} dA = 0$ provided the fields are such that the flux of $\vec{j}$ across the boundary of V is zero. Note that the charge $Q = \int \pi_b^0(y)\delta\phi_b(y) d^3y$ is a linear combination of the conjugate momenta π_a^0. It follows that $\{\phi_a(x), Q\} = \delta\phi_a(x)$. So the conserved Noether charge generates the infinitesimal symmetry transformation via the Poisson bracket. This is as expected. We know from classical mechanics that the p.b. of any quantity with an observable is equal to the change in the quantity under the infinitesimal canonical transformation generated by the observable.

9.3 Klein Gordon scalar field for spin zero particles

• The simplest classical relativistic wave equation, the KG equation, describes free propagation of classical massive scalar fields. Its quantization (upon including some interactions) describes spin zero (scalar) particles such as the Higgs boson and π mesons. It is obtained from the relativistic energy-momentum dispersion relation $E^2 = \mathbf{p}^2c^2 + m^2c^4$ via the 'correspondence rule' $E = i\hbar\frac{\partial}{\partial t}$, $\mathbf{p} = -i\hbar\nabla$.

$$-\hbar^2\frac{\partial^2\phi}{\partial t^2} = -\hbar^2c^2\nabla^2\phi + m^2c^4\phi \quad \text{or} \quad (\Box + \frac{m^2c^2}{\hbar^2})\phi(\mathbf{r}, t) = 0 \tag{284}$$

where $\Box = (1/c^2)\partial_t^2 - \nabla^2$ is the d'Alembert or wave operator. $\hbar/mc$ is the Compton wavelength corresponding to the mass m. The equation is relativistically invariant if ϕ transforms as a scalar under the Poincare group $\phi'(x') = \phi(x)$. Let $\eta_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$ be the Minkowski metric. The contravariant and covariant components of the position, gradient and momentum 4-vectors are

$$\begin{aligned} x^\mu &= (x^0, \vec{x}) = (ct, \vec{x}), \quad x_\mu = \eta_{\mu\nu}x^\nu = (ct, -\vec{x}), \quad \partial^\mu = \frac{\partial}{\partial x_\mu} = \left(\frac{1}{c}\frac{\partial}{\partial t}, -\nabla\right), \quad \partial_\mu = \left(\frac{1}{c}\frac{\partial}{\partial t}, \nabla\right), \\ p^\mu &= (E/c, \vec{p}) = i\hbar\left(\frac{1}{c}\partial_t, -\nabla\right) = i\hbar\partial^\mu \quad \text{and} \quad p_\mu = \eta_{\mu\nu}p^\nu = (E/c, -\vec{p}) = i\hbar\partial_\mu. \end{aligned} \tag{285}$$

Then $p^2 = p^\mu p_\mu = -\hbar^2\partial^\mu\partial_\mu = -\hbar^2\Box$ and the KG equation is $(p^2 - m^2c^2)\phi = 0$ or $(\partial^\mu\partial_\mu + m^2c^2/\hbar^2)\phi = 0$. Despite the appearance of factors of $\hbar$, this is not to be regarded as a quantum mechanical wave equation. The quantity $\lambda = \hbar/mc$ has dimensions of length and may simply be regarded as a parameter of a classical field model with equation of motion $(\Box + \frac{1}{\lambda^2})\phi = 0$. Such an equation could be used to model propagation of scalar classical EM waves (i.e. ignoring polarization) through a plasma, where λ is a Debye length. Interpretation in terms of particles must await quantization as in the case of Maxwell theory.

• The KG equation may be studied either for a real or complex scalar field. The complex case is more relevant to the standard model of particle physics since the Higgs scalar field is a (pair of) complex scalars.

• The KG equation is the **Euler-Lagrange equation** following from the **action** $S = \int \mathcal{L}\, d^4x$ where

$$\mathcal{L} = (\partial_\mu\phi)^*(\partial^\mu\phi) - m^2|\phi|^2 = |\partial_t\phi|^2 - |\nabla\phi|^2 - m^2|\phi|^2. \tag{286}$$

We regard ϕ, ϕ^* as a pair of independent fields. Up to a 4-divergence, which does not contribute to the action with decaying boundary conditions, this is equal to $\mathcal{L} = \phi^*(-\partial_\mu\partial^\mu - m^2)\phi$. Extremization with respect to ϕ^* gives the KG equation for a complex scalar ϕ (i.e the EL Lagrange equation reduces to $\frac{\partial\mathcal{L}}{\partial\phi^*} = 0$ since $\mathcal{L}$ written this way, does not involve derivatives of ϕ^*). Similarly, the EL equation from extermination in ϕ gives the complex conjugate KG equation $(\Box + m^2)\phi^* = 0$.

• From $\mathcal{L}$ we get $\pi_\phi^\mu = \partial^\mu\phi^*$ and $\pi_{\phi^*}^\mu = \partial^\mu\phi$. The momenta conjugate to ϕ, ϕ^* are $\pi_\phi^0 = \dot{\phi}^*$ and $\pi_{\phi^*}^0 = \dot{\phi}$. The corresponding conserved energy density defined as $\mathcal{H} = \pi_\phi^0\dot{\phi} + \pi_{\phi^*}^0\dot{\phi}^* - \mathcal{L}$ reduces to $\mathcal{H} = |\partial_t\phi|^2 + |\nabla\phi|^2 + m^2|\phi|^2$. We interpret $\mathcal{T} = |\dot{\phi}|^2$ as the kinetic energy density and $\mathcal{V} = |\nabla\phi|^2 + m^2|\phi|^2$ as potential energy density. So $\mathcal{L} = \mathcal{T} - \mathcal{V}$ while $\mathcal{H} = \mathcal{T} + \mathcal{V}$. The KG equation (without a mass term and for a real scalar in one spatial dimension) is also used to model small oscillations of a stretched string whose height is $\phi(x, t)$. In that case, integration by parts yields $\int(\partial_x\phi)^2 dx = -\int \phi\partial_x^2\phi dx$ with suitable boundary conditions. The latter may be interpreted as a bending energy of the string, it grows with the curvature (second derivative) of the string. It is a potential energy in the sense that it is due to the tensional forces exerted on a piece of string by the neighbouring pieces of string, which tend to straighten out the string.

• One may recall from a course on relativistic QM that a complex KG field admits a local conservation law $\partial_t \rho(x,t) + \nabla \cdot j(x,t) = 0$ for the density and current

$$\rho(x,t) = -\frac{ie}{c^2}\left(\phi^* \dot{\phi} - \phi \dot{\phi}^*\right) \quad \text{and} \quad \vec{j}(x,t) = ie\left(\phi^* \nabla \phi - \phi \nabla \phi^*\right). \tag{287}$$

Note that $\rho, \vec{j}$ are identically zero for a real scalar field. e is a constant with the physical meaning of electric charge, as we shall see. In Lorentz invariant form, this is the conservation of the 4-vector current. Check using KG, that

$$j^\mu = \left(c\rho, \vec{j}\right) = -ie(\phi^* \partial^\mu \phi - (\partial^\mu \phi)^* \phi) \quad \Rightarrow \quad \partial_\mu j^\mu = 0. \tag{288}$$

This is Noether's conserved current for the global $\mathrm{U}(1)$ symmetry $\phi \to e^{ie\theta}\phi$, $\phi^* \to e^{-ie\theta}\phi$ corresponding to the infinitesimal transformations $\delta\phi = ie\phi$, $\delta\phi^* = -ie\phi^*$. One finds the conserved current $-ie\theta(\phi^* \partial^\mu \phi - (\partial^\mu \phi)^* \phi)$ which is just j^μ upon omitting the small parameter θ.

• In the non-relativistic limit $\phi(r,t) = e^{-imc^2 t/\hbar}\psi(r,t)$, where ψ is slowly varying in time. The KG equation reduces to the non-relativistic Schrödinger equation for ψ and the components of $-\hbar j^\mu/2me$ reduce to $(cP, \vec{J})$ where $P = |\psi|^2$ and $\vec{J} = \frac{\hbar}{2mi}(\psi^* \nabla \phi - \nabla \phi^* \psi)$ are the conserved Schrodinger probability density and current. However, the conserved KG density ρ is not always positive and we cannot interpret ϕ as a probability amplitude. Pauli and Weisskopf reinterpreted the KG equation as a scalar field equation and the current j^μ as the electromagnetic current of charged spinless particles, such as pions. (Note: j^μ needs to be modified to be conserved in the presence of an EM field, see the next section). When quantized, the creation and annihilation operators of the scalar field create/destroy scalar particles. The 'charge' e should be the eigenvalue of the conserved charge operator $Q = \int j^0 d^3x$ in a one-scalar particle state.

• A **self-interacting scalar field** (such as the Higgs field) is obtained by including a real potential $V(\phi^*, \phi)$ in the Lagrangian $\mathcal{L} = |\partial\phi|^2 - m^2|\phi|^2 - V(\phi^*, \phi)$ leading to the eqn of motion $(\partial^2 + m^2)\phi + \frac{\partial V}{\partial \phi^*} = 0$ (some times the $m^2|\phi|^2$ mass term is included in V). To retain the above $\mathrm{U}(1)$ global symmetry, V must only be a function of the combination $\phi^*\phi$. Of particular interest is the quartic 'Mexican hat' potential $V = (\lambda/4)(|\phi|^2 - v^2)^2$ which is minimal when $|\phi| = v$, the so-called vacuum expectation value of the scalar field. The field equation for a massive *real* scalar field $(\partial^2 + m^2)\phi + V'(\phi) = 0$ follows from $\mathcal{L} = \frac{1}{2}\partial^\mu\phi\partial_\mu\phi - \frac{1}{2}m^2\phi^2 - V(\phi)$. Why the factor of two difference between complex and real scalar fields?

9.4 Complex scalar coupled to EM field

Though the above complex scalar field possesses a global $\mathrm{U}(1)$ symmetry, the lagrangian $(\partial_\mu\phi)^*(\partial^\mu\phi) - V(\phi^*\phi)$ isn't invariant under local $\mathrm{U}(1)$ phase transformations $\phi' = e^{ie\theta(x)}\phi(x)$, where $\theta(x)$ is a function on Minkowski space. This is because $\partial_\mu\phi$ does not transform in the same way as ϕ does. There is an extra additive term (inhomogeneous transformation) in the partial derivative of ϕ':

$$\partial_\mu \phi' = e^{ie\theta}\left[\partial_\mu \phi + (ie\partial_\mu \theta)\phi\right]. \tag{289}$$

The reason to seek invariance under local phase changes is that it naturally leads to coupling to the electromagnetic field. Indeed. this extra term can be gotten rid of by introducing an additional field A_μ which transforms to $A'_\mu = A_\mu + \partial_\mu \theta$. In other words, the Lagrangian may be made local $U(1)$ invariant by 'minimally coupling' the scalar field to an electromagnetic vector gauge potential $A^\mu = (\varphi, \mathbf{A})$ (note that φ is the electric potential, distinct from the scalar field ϕ. ϕ is a scalar while φ transforms as the time component of a 4-vector). This is done by replacing partial derivatives by covariant derivatives in the Lagrangian of the scalar field

$$D_\mu \phi = (\partial_\mu - ieA_\mu)\phi \quad \text{and} \quad \mathcal{L} = |D_\mu \phi|^2 - V(\phi^* \phi). \tag{290}$$

Under a local $U(1)$ gauge transformation, the covariant derivative of ϕ transforms in the same way as the field itself. Check that

$$A'_\mu = A_\mu + \partial_\mu \theta \quad \text{and} \quad \phi' = e^{ie\theta}\phi \quad \Rightarrow \quad (D_\mu \phi)'(x) = e^{ie\theta(x)} D_\mu \phi(x) \tag{291}$$

so that $\mathcal{L}$ is gauge invariant. The number e is the charge of the scalar field, and we have described the interaction of a charged scalar field (necessarily complex) with an external electromagnetic field. It is instructive to write out the Lagrangian

$$\mathcal{L} = (\partial_\mu + ieA_\mu)\phi^*(\partial^\mu - ieA^\mu)\phi - V(\phi^*, \phi) = |\partial_\mu \phi|^2 + ieA_\mu(\phi^* \partial^\mu \phi - \partial^\mu \phi^* \phi) + e^2 A_\mu A^\mu |\phi|^2 - V(\phi^* \phi). \tag{292}$$

The tri-linear 'charge interaction' terms $ieA_\mu(\phi^* \partial^\mu \phi - \partial^\mu \phi^* \phi)$ are derivative interactions. A scalar that does not have such a derivative interaction with the gauge potential is said to be uncharged (neutral). The Higgs is a neutral scalar particle. $\pi^\pm$ are charged scalar particles.

• The resulting equation of motion for the scalar is

$$\left(\partial^2 - 2ieA \cdot \partial - ie(\partial \cdot A) - e^2 A^2\right)\phi + \frac{\partial V}{\partial \phi^*} = 0 \quad \text{or} \quad D_\mu D^\mu \phi + \frac{\partial V}{\partial \phi^*} = 0 \tag{293}$$

and its complex conjugate. The second form is manifestly gauge covariant, both terms transform by multiplication by $e^{ie\theta(x)}$.

• The gauge-invariant generalization of the conserved current is

$$j^\mu = -ie\left(\phi^* D^\mu \phi - (D^\mu \phi)^* \phi\right) = -ie(\phi^* \partial^\mu \phi - (\partial^\mu \phi)^* \phi) - 2e^2 A^\mu |\phi|^2 \tag{294}$$

Check using the eom that $\partial_\mu j^\mu = 0$. j^μ is called the electromagnetic current of the charged scalar field. We will see that it enters as a source in Maxwell's equations[47], $\partial_\mu F^{\mu\nu} = j^\nu$ so that its conservation is crucial to the consistency of Maxwell's equations.

9.5 Maxwell field equations for massless spin one particles

• The photon field A_μ coupled to the complex scalar in the last section was an external field, it did not have any dynamics of its own. The dynamics of the EM field is governed by Maxwell's

[47] Technical aside: Note that the coupling of the scalar to the gauge field in the Lagrangian cannot simply be written as $\mathcal{L} = |\partial\phi|^2 - \frac{1}{4}F^2 - j^\mu A_\mu$ due a factor of 2 in j^μ. Since j_μ depends on A_μ this does not lead to the correct equation of motion.

equations. In rationalized Heaviside-Lorentz units they are

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times E = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \cdot \mathbf{E} = \rho \quad \text{and} \quad \nabla \times \mathbf{B} = \frac{\mathbf{j}}{c} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}. \tag{295}$$

where for consistency of the two inhomogeneous equations the electric charge and current density must satisfy the continuity equation $\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0$. For example, $\rho, \vec{j}$ may be the electromagnetic current due to a charged scalar field. The first two homogeneous Maxwell equations state the absence of magnetic monopoles, and Faraday's law of induction. The second pair of inhomogeneous equations are Gauss' law and Ampere's law with Maxwell's correction term involving the time derivative of the electric field (the displacement current).

• The first pair of homogeneous Maxwell equations are identically satisfied if the fields are expressed in terms of scalar and vector potentials $(\varphi, \mathbf{A})$

$$\mathbf{E} = -\nabla \varphi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} \quad \text{and} \quad \mathbf{B} = \nabla \times \mathbf{A}. \tag{296}$$

• The inhomogeneous Maxwell equations can be written in a relativistically covariant form by introducing the real 4-vectors $A^\mu = (\varphi, \mathbf{A})$ and $j^\mu = (c\rho, \mathbf{j})$ and the field strength tensor $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. Then the inhomogeneous Maxwell equations become $\partial_\mu F^{\mu\nu} = \frac{1}{c} j^\nu$ along with the consistency condition $\partial_\mu j^\mu = 0$ which expresses local charge conservation.

• The electric and magnetic fields $\mathbf{B} = \nabla \times \mathbf{A}$ and $\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \nabla \varphi$ with cartesian components $E^i, B^i = \epsilon^{ijk}\partial_j A^k$ are then the components of the field strength $F^{0i} = \partial^0 A^i - \partial^i A^0 = -E^i$ and

$$F^{ij} = \partial^i A^j - \partial^j A^i = -\partial_i A^j + \partial_j A^i = -\epsilon^{ijk} B^k \quad \Rightarrow \quad F^{\mu\nu} = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{pmatrix}. \tag{297}$$

The covariant components are given by $F_{\mu\nu} = \eta_{\mu\rho}\eta_{\nu\sigma}F^{\rho\sigma}$. So $F_{0i} = -F^{0i}$ and $F_{ij} = F^{ij}$.

• However, $\mathbf{A}$ and ϕ are not uniquely determined by the measurable electric and magnetic fields. Two gauge potentials $A^\mu = (\phi, \mathbf{A})$ and $A'^\mu = (\phi', \mathbf{A}')$ which differ by a gauge transformation

$$\mathbf{A}' = \mathbf{A} + \nabla \theta, \quad \varphi' = \varphi - \frac{1}{c}\frac{\partial \theta}{\partial t} \quad \text{or} \quad A'^\mu = A^\mu + \partial^\mu \theta \tag{298}$$

correspond to the same electromagnetic fields and therefore leave Maxwell's equations invariant.

• The gauge transformations may be composed by addition $\theta_1(x) + \theta_2(x)$. So we have a group of gauge transformations $g(x) = e^{ie\theta(x)}$ with $\theta(x)$ living in the Lie algebra. The group elements $g(x)$ appeared in the gauge transformation of the scalar field $\phi' = e^{ie\theta(x)}\phi$. The gauge transformation of the photon field may also be written as $A'_\mu = gA_\mu g^{-1} - \frac{1}{ie} g\partial_\mu g^{-1}$. At any one space-time point, $g = e^{ie\theta}$ is just a phase living in the abelian group $U(1)$ and θ is a real number living the Lie algebra of the circle group $U(1)$. The gauge group (also known as structure group) of Maxwell theory is $U(1)$. A related concept is the group of gauge transformations, it is defined as the set of functions $g(x)$ from Minkowski space to the group $U(1)$, which tend to the identity at infinity ($\theta(x) \to 0$ as $x \to \infty$). By replacing the real $\theta(x)$ by a hermitian $N \times N$ matrix $\theta^a_b(x)$ at each x we will get a U(N) non-abelian (or Yang-Mills) gauge theory for the hermitian $N \times N$ matrix-valued gauge field $[A_\mu(x)]^a_b$. In this case, the single complex scalar field is replaced with an N-component $(\phi_1, \cdots, \phi_N)^t$ complex scalar which transforms as $\phi \to e^{ie\theta(x)}\phi(x)$ under gauge transformations. For $N = 2$ we get the complex doublet of Higgs scalar fields that appears in the the standard model.

• If we restrict to traceless hermitian matrices we get an SU(N) gauge theory. The gauge group of the standard model is U$(2)\times$ SU(3) corresponding to the electroweak and strong interactions. U(2) is a 4d group with a 4d Lie algebra generated by the identity and Pauli matrices. Certain linear combinations of these 4 generators correspond to the photon, $W^\pm$ and Z fields. SU(3) is the color group with generators corresponding to 8 gluons.

• If we replace the indices a, b in the internal space of the Lie algebra of the gauge group with space-time indices, then the gauge potential $[A_\mu]^a_b(x)$ becomes $\Gamma^\rho_{\mu\sigma}$ which are the Christoffel connection coefficients of general relativity/Riemannian geometry. So the gauge potential and its inhomogeneous transformation law are the analogues of the Christoffel connection and its inhomogeneous transformation law. Moreover the group of gauge transformation is now replaced with the space of functions $g(x)$ from Minkowski space to Minkowski space, these are the general coordinate transformations encountered in GR. The group of gauge transformations is the analogue of the group of space-time diffeomorphisms. Moreover, we will see that the generalization of the EM field strength to a non-abelian gauge theory is the matrix field $[F_{\mu\nu}]^a_b = \partial_\mu [A_\nu]^a_b - \partial_\nu [A_\mu]^a_b - i[A_\mu, A_\nu]^a_b$. Again, replacing internal indices with space-time indices we see the emergence of the Riemann curvature tensor $R^\rho_{\mu\nu\sigma}$. So the electric and magnetic fields (and their non-abelian generalizations) are to be regarded as the curvature of the gauge connection in the same way as R is the curvature of the space-time Christoffel connection.

• Maxwell's equations describe propagation of electromagnetic waves at the speed of light: each component of $\mathbf{E}$ and $\mathbf{B}$ satisfy the d'Alembert wave equation or the massless KG equation. When quantized, they describe massless particles (photons). Photons have spin one: in radiation gauge ($\phi = 0, \nabla \cdot \mathbf{A} = 0$) the dynamical variable is the vector potential $\mathbf{A}$, and we have seen that components of a 3d vector transform like a spin one multiplet. However, unlike a massive spin one particle, the transversality condition $\nabla \cdot \mathbf{A} = 0$ ensures that the photon has only two helicities $h = \pm 1$ (or spin projections).

• Maxwell's equations $\partial^\mu F_{\mu\nu} = j_\nu$ are the Euler-Lagrange equations following from the Lorentz and gauge-invariant action (gauge invariance under $A'_\mu = A_\mu + \partial_\mu \theta$ requires current conservation $\partial_\mu j^\mu = 0$)

$$S = \int \mathcal{L}\, d^4x, \quad \text{where} \quad \mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - j^\mu A_\mu = \frac{1}{2}(\mathbf{E}^2 - \mathbf{B}^2) - j^\mu A_\mu. \tag{299}$$

Here we assumed that j^μ does not itself depend on A (there are interesting exceptions, see below). The sign of the Lagrangian does not affect the equations of motion, but is fixed by the convention that it be the difference of EM kinetic ($\mathbf{E^2}$) and potential ($\mathbf{B}^2$) energies. Find the conjugate field momenta, calculate the Legendre transform and show that the electromagnetic energy density $\mathcal{H} = \frac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2)$

• The Lagrangian for a charged complex scalar coupled to the EM field is ($D_\mu\phi = (\partial_\mu - ieA_\mu)\phi$)

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^*(D^\mu\phi) = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + |\partial\phi|^2 + ieA_\mu(\phi^*\partial^\mu\phi - (\partial^\mu\phi^*)\phi) + e^2A^2|\phi|^2 \quad (300)$$

Check that the resulting equations of motion for the gauge and scalar fields are

$$\partial_\mu F^{\mu\nu} = j^\nu \quad \text{and} \quad D^\mu D_\mu\phi = 0 \quad \text{where} \quad j^\mu = -ie(\phi^*\partial^\mu\phi - \partial^\mu\phi^*\phi) - 2e^2A^2|\phi|^2. \quad (301)$$

• If ϕ is subject to a self-interaction $V(\phi^*\phi)$ such that $|\phi|$ is non-vanishing in the ground state, then the $e^2|\phi|^2A^2$ term is like a mass term M^2A^2 for the gauge field. This idea is exploited in the Higgs mechanism to give a mass $e|\phi|$ to the gauge field.

9.5.1 Magnetic monopole and charge quantization

• While Maxwell's equations allow for electric charges (monopoles) as sources for electric fields $\nabla \cdot \mathbf{E} = \rho$, they do not permit magnetic charges $\nabla \cdot \mathbf{B} = 0$. Magnetic dipoles exist (bar magnets, magnetic moments of particles, steady current in a loop) but no magnetic monopoles have so far been found, despite experimental searches. Remarkably, Dirac found that the existence of a magnetic monopole could explain the observed quantization of electric charge in multiples of a basic unit. If magnetic monopoles exist, they may be very heavy and rare particles. Grand unified theories predict the existence of certain types of magnetic monopoles.

• By analogy with a point electric charge e at the origin producing an electric field $\mathbf{E} = e\frac{\hat{r}}{r^2}$, a magnetic monopole of strength g produces a magnetic field $\mathbf{B} = g\frac{\hat{r}}{r^2}$. Since we are concerned with the motion of charges, we use Gaussian units where 4π's appear in Maxwell's equations rather than in Coulomb's law, and also set $c = 1$. Dirac's discovery may be explained using Saha's 1936 analysis of the motion of an electric charge in the field of a magnetic monopole. The Newton-Lorentz equation for the rate of change of momentum $\mathbf{p} = m\mathbf{v}$ is $\frac{d\mathbf{p}}{dt} = eg\mathbf{v} \times \frac{\hat{r}}{r^2}$. Saha evaluated the rate of change of angular momentum of the charge and got a non-zero answer despite the apparent spherical symmetry of the field around the monopole

$$\frac{d\mathbf{L}}{dt} = \mathbf{r} \times \dot{\mathbf{p}} + \mathbf{v} \times m\mathbf{v} = eg\,\mathbf{r} \times (\mathbf{v} \times \frac{\mathbf{r}}{r^3}) = eg\frac{1}{r^3}\mathbf{r} \times (\mathbf{v} \times \mathbf{r}). \quad (302)$$

It turns out that to find a conserved angular momentum one must include the contribution of the electromagnetic field. We may find this conserved total angular momentum without evaluating the field angular momentum ($\frac{1}{4\pi}\int \mathbf{r} \times (\vec{E} \times \vec{B})\, d^3r$) by noticing that the RHS is in fact a total time derivative $\frac{1}{r^3}\mathbf{r} \times (\mathbf{v} \times \mathbf{r}) = \frac{d\hat{r}}{dt}$! Indeed,

$$\frac{1}{r^3}\mathbf{r} \times (\mathbf{v} \times \mathbf{r}) = \frac{1}{r^3}[r^2\mathbf{v} - (\mathbf{r} \cdot \mathbf{v})\mathbf{r}] = \frac{\mathbf{v}}{r} - \frac{\dot{r}}{r^2}\mathbf{r} = \frac{d}{dt}\left(\frac{\mathbf{r}}{r}\right). \quad (303)$$

In the penultimate equality we used $r^2 = \mathbf{r} \cdot \mathbf{r}$ to write $\dot{r} = \mathbf{r} \cdot \mathbf{v}$. Thus we have a conserved total angular momentum $\mathbf{J} = \mathbf{r} \times \mathbf{p} - eg\hat{r} = \mathbf{L} - eg\frac{\mathbf{r}}{r}$.

• Now the component of total angular momentum along the line joining the monopole to the charge is $\mathbf{J} \cdot \hat{r} = -eg$. In quantum mechanics, the component of angular momentum in any direction is quantized to be an integer or half odd integer multiple of $\hbar$. Thus $eg = n\hbar$ where $n \in \{0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \pm 2, \ldots\}$. So if the least non-zero strength of a magnetic monopole is g, then all electric charges must be integer multiples of $\hbar/2g$.

9.6 Pauli-Weyl equation for massless spin half particles

• Spin half particles have two spin projections so it is natural to look for a wave equation for two component spinors. The Schrodinger hamiltonian for a free massive spin half particle (like the electron) $H = \frac{\mathbf{p}^2}{2m} \otimes I_{2\times 2}$ is proportional to the identity in spin space. Notice that it can be written as $H = \frac{1}{2m}(\vec{\sigma}\cdot\mathbf{p})^2$ where $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is the vector of Pauli matrices, (the vector spin observable is $\mathbf{S} = \frac{1}{2}\hbar\vec{\sigma}$). This follows from the identity $\sigma_i\sigma_j = \delta_{ij} I + i\epsilon_{ijk}\sigma_k$. Note that

$$\sigma\cdot\mathbf{p} = \begin{pmatrix} p_3 & p_1 - ip_2 \\ p_1 + ip_2 & -p_3 \end{pmatrix} \quad \text{and} \quad \det(\sigma\cdot\mathbf{p}) = -\mathbf{p}^2. \tag{304}$$

Since $\mathbf{S}$ and $\mathbf{p}$ transform as vectors under rotation, the Schrodinger-Pauli equation $i\hbar\partial_t\psi = (\sigma\cdot\mathbf{p})^2\psi$ is rotation invariant. However, it is not Lorentz covariant: it is second order in space derivatives but first order in time derivatives.

• We seek a wave equation for a two component spinor that is first order in space and time derivatives, which nevertheless has the property that each component satisfies the relativistic scalar wave equation (KG equation). Notice that the first order operator $\vec{\sigma}\cdot\mathbf{p}$ is a square-root of the operator $\mathbf{p^2}\otimes\mathbf{I}$. Pauli found a relativistic generalization by defining $\sigma^\mu = (I, \vec{\sigma}), p_\mu = (E/c, -\mathbf{p})$ and considering the Minkowski inner product

$$\sigma\cdot p = \sigma^\mu p_\mu = p_0 I - \vec{\sigma}\cdot\mathbf{p} = \begin{pmatrix} p_0 - p_3 & -(p_1 - ip_2) \\ -(p_1 + ip_2) & p_0 + p_3 \end{pmatrix} = i\hbar\sigma\cdot\partial = i\hbar\begin{pmatrix} \partial_t - \partial_z & \partial_x - i\partial_y \\ \partial_x + i\partial_y & \partial_t + \partial_z \end{pmatrix}. \tag{305}$$

$\sigma\cdot p$ is a hermitian operator. The Pauli wave equation for a massless spin half particle is

$$(\sigma\cdot\partial)\psi = 0 \quad \text{or} \quad (\partial_t + \vec{\sigma}\cdot\nabla)\psi = 0 \quad \text{or} \quad p_0\psi = (\vec{\sigma}\cdot\mathbf{p})\psi. \tag{306}$$

In components, if

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \quad \text{then} \quad \partial_t\psi_1 + \partial_z\psi_1 + \partial_x\psi_2 - i\partial_y\psi_2 = 0 \quad \text{and} \quad \partial_t\psi_2 - \partial_z\psi_2 + \partial_x\psi_1 + i\partial_y\psi_1 = 0. \tag{307}$$

So the Pauli equation couples the two components. Nevertheless each component satisfies the massless KG equation. Indeed $(E^2 - \mathbf{p}^2)I\psi = (E^2 - (\vec{\sigma}\cdot\mathbf{p})^2\psi) = (E + \sigma\cdot\mathbf{p})(E - \sigma\cdot\mathbf{p})\psi = 0$, the last equality following from the Pauli equation. Thus $\hbar^2\Box\psi = 0$.

• In the 1930s, Pauli proposed his equation $(\partial_t + \vec{\sigma}\cdot\nabla)\psi = 0$ for neutrinos, which were believed to be massless. However, he found that the equation is not invariant under parity $E \to E, \mathbf{p} \to -\mathbf{p}, \vec{\sigma} \to \vec{\sigma}$ (spin is an axial-vector like angular momentum). Under parity the Pauli equation $(E - \vec{\sigma}\cdot\mathbf{p})\psi = 0$ turns into $(E + \vec{\sigma}\cdot\mathbf{p})\psi = 0$. Till 1957, parity was mistakenly believed to be a symmetry of the weak interactions, so Pauli rejected his equation for the neutrino.

• Note that a positive energy solution of the Pauli equation $(\vec{\sigma}\cdot\mathbf{p})\psi = E\psi$ must have positive helicity. On the other hand, positive energy solutions of the parity reflected Pauli equation $(\vec{\sigma}\cdot\mathbf{p})\psi = -E\psi$ have negative helicity. The two Pauli equations are on a common footing. When quantized, they describe parity-violating massless spin-half particles of positive or negative helicity. Together we write them as

$$(E - \vec{\sigma}\cdot\mathbf{p})\psi_+ = 0 \quad \text{and} \quad (E + \vec{\sigma}\cdot\mathbf{p})\psi_- = 0. \tag{308}$$

For $E > 0$, ψ_+, ψ_- are positive and negative helicity Pauli spinors. Sometimes, they are also called right and left-handed Pauli spinors. Weyl also considered the same pair of equations so Pauli spinors are also called Weyl spinors.

9.7 Dirac field equation for massive spin half particles

• Dirac discovered a parity-invariant relativistic wave equation for massive spin half particles. This equation is relevant for the propagation of charged leptons (Dirac had the electron in mind), quarks, nucleons and possibly neutrinos. Rather than follow his reasoning (which is similar to the one used in arriving at the Pauli equation above; see Dirac's book or our notes on relativistic QM or QM3), we will arrive at Dirac's equation from Pauli's equation. Since a massive particle can have either helicity, we expect a Dirac spinor to be made from two Pauli spinors of opposite helicity. We can get a parity invariant equation by combining Pauli equations for two Pauli spinors ψ_-, ψ_+ of opposite helicity, which are exchanged under parity $(\Pi\psi_\pm = \psi_\mp)$

$$(E - \vec{\sigma}\cdot\mathbf{p})\psi_+ = 0 \quad \text{and} \quad (E + \vec{\sigma}\cdot\mathbf{p})\psi_- = 0. \tag{309}$$

To describe a massive spin half particle, we include mass terms that couple the two Pauli spinors, since a massive particle can have either helicity. Putting $E = i\hbar\partial_0$ and $\mathbf{p} = -i\hbar\nabla$, the simplest possibility we can concoct is the Dirac equations

$$\begin{aligned} i(\partial_0 + \vec{\sigma}\cdot\nabla)\psi_+ - \frac{mc}{\hbar}\psi_- = 0 \qquad &\& \qquad i(\partial_0 - \vec{\sigma}\cdot\nabla)\psi_- - \frac{mc}{\hbar}\psi_+ = 0 \\ \text{or} \quad (E - \vec{\sigma}\cdot\mathbf{p})\psi_+ = m\psi_- \quad &\text{and} \quad (E + \vec{\sigma}\cdot\mathbf{p})\psi_- = m\psi_+ \end{aligned} \tag{310}$$

$\hbar/mc$ is the Compton wavelength corresponding to the mass m. We put $\hbar = c = 1$ in the second line above. The sign of the mass terms can be fixed by studying the physical consequences, the energy of the particle should be a little more than mc^2 at low speeds, for instance. These can be combined as an equation for a 4-component spinor

$$\begin{pmatrix} 0 & i(\partial_0 + \vec{\sigma}\cdot\nabla) \\ i(\partial_0 - \vec{\sigma}\cdot\nabla) & 0 \end{pmatrix} \begin{pmatrix} \psi_- \\ \psi_+ \end{pmatrix} = m \begin{pmatrix} \psi_- \\ \psi_+ \end{pmatrix}. \tag{311}$$

If $m = 0$ the two Pauli spinors evolve independently. Each of the components of ψ_+ and ψ_- satisfies the massive KG equation. For example,

$$(E^2 - \mathbf{p}^2)\psi_+ = (E + \vec{\sigma}\cdot\mathbf{p})(E - \vec{\sigma}\cdot\mathbf{p})\psi_+ = (E + \vec{\sigma}\cdot\mathbf{p})m\psi_- = m^2\psi_+. \tag{312}$$

These equations may be compressed by introducing the 4-component Dirac spinor and 4×4 Dirac matrices

$$\psi = \begin{pmatrix} \psi_- \\ \psi_+ \end{pmatrix}, \quad \gamma^0 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \quad \text{and} \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}. \tag{313}$$

Then the Dirac equation takes the familiar form $(i\gamma^\mu \partial_\mu - m)\psi = 0$. We are free to choose a different basis for Dirac spinors and matrices. The above basis is called the Weyl or chiral basis since for $m = 0$ the upper and lower components of the Dirac spinor are Pauli spinors with definite (negative and positive) helicity and also because the chirality matrix γ_5 (see below) is diagonal. The chiral representation is convenient for studying parity, chirality, helicity and the ultra-relativistic $m \to 0$ limit.

• For example, under parity $x^\mu \to x'^\mu = (x^0, -\mathbf{x})$. We had noted that if $\psi_\pm$ are exchanged under parity, then the two Pauli equations go into eachother. This exchange is implemented by γ^0. Check that $\psi'(x') = \gamma^0 \psi(x')$ satisfies the parity transformed equation $i(\gamma^\mu \partial'_\mu - m)\psi'(x') = 0$. We are also free to multiply by a phase in defining the parity transformed Dirac spinor.

• The condition that the components of the Dirac spinor each satisfy the 2nd order KG equation imposes the anti-commutator conditions $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ which define the Dirac or Clifford algebra. Distinct Dirac matrices anti-commute while $(\gamma^0)^2 = -(\gamma^i)^2 = I$. It follows that γ^μ are traceless (e.g. take the trace of $\gamma^0 \gamma^\mu \gamma^0 = -\gamma^\mu$). Hermiticity of the Dirac hamiltonian ($H = \alpha \cdot \mathbf{p} + \beta m$, look up Dirac's approach) implies that γ^0 is hermitian while γ^i are anti-hermitian. Verify that these relations are satisfied in the chiral representation. The γ matrices are unique up to changes of basis $\gamma' = S^{-1}\gamma S$ where S is an invertible 4×4 complex matrix. For example $(\gamma^\mu)^*$ and $-(\gamma^\mu)^*$ also satisfy the Dirac algebra and so each set must be related to the γ^μ by a change of basis.

• Dirac's original basis for his matrices is different from the chiral basis. In his basis γ^0 is diagonal, it facilitates passage to the non-relativistic limit.

$$\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad \text{in Dirac's basis.} \tag{314}$$

What is γ_5 in Dirac's basis?

• The chirality matrix is defined as $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. Use the anti-commutation relations to show that γ_5 anti-commutes with γ^μ, $\gamma_5^2 = I$ and that it is traceless. So the eigenvalues of γ_5 are ± 1 with multiplicity two each. In the chiral basis $\gamma_5 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}$ is diagonal. Eigenspinors of γ_5 with eigenvalue $+1$ are called right-handed, while eigenspinors with chirality -1 are called left-handed. In the chiral basis $\psi_R = \begin{pmatrix} 0 \\ \psi_+ \end{pmatrix}$ and $\psi_L = \begin{pmatrix} \psi_- \\ 0 \end{pmatrix}$ are the right- and left-handed spinors

$$\gamma_5 \psi_R = \psi_R \quad \text{or} \quad \frac{1}{2}(I + \gamma_5)\psi_R = \psi_R \quad \text{and} \quad \gamma_5 \psi_L = -\psi_L \quad \text{or} \quad \frac{1}{2}(I - \gamma_5)\psi_L = \psi_L. \tag{315}$$

$P_{R,L} = \frac{1}{2}(I \pm \gamma_5)$ are projections to right and left handed spinors. Notice that chirality is defined independent of the value of mass m. If $m = 0$, then ψ_+ and ψ_- satisfy Pauli equations

$(E-\vec{\sigma}\cdot\mathbf{p})\psi_+ = 0$ and $(E+\vec{\sigma}\cdot\mathbf{p})\psi_- = 0$ and are seen to have positive and negative helicities for $E > 0$. So for massless Dirac particles, helicity and handedness are equivalent notions: RH (+1 chirality) is the same as positive helicity and LH (-1 chirality) is the same as negative helicity.

• Adjoint of the Dirac equation: Just as we have the KG equation for ϕ and its complex conjugate for ϕ^* (for a complex scalar field), we have the Dirac equation for ψ and the adjoint equation for $\psi^\dagger$ adjoint. The adjoint equation is more conveniently written in terms of $\bar{\psi} = \psi^\dagger\gamma^0$, which is called the Pauli adjoint spinor. It is obtained by taking the complex conjugate transpose of the Dirac equation and then right multiplying by the invertible matrix γ^0 and using hermiticity/anti-hermiticity of γ^0/γ^i and the anti-commutation relations

$$-i(\partial_\mu\psi^\dagger)(\gamma^\mu)^\dagger = m\psi^\dagger \quad\Rightarrow\quad -i(\partial_0\psi^\dagger\gamma^0 - \partial_i\psi^\dagger\gamma^i) = m\psi^\dagger \quad\Rightarrow\quad -i(\partial_0\bar{\psi}\gamma^0 + \partial_i\bar{\psi}\gamma^i) = m\bar{\psi}. \tag{316}$$

• The Dirac equation $(i\gamma\cdot\partial - m)\psi = 0$ (and its adjoint $-i\bar{\psi}\gamma\cdot\partial - m\bar{\psi} = 0$) (the derivative is understood to act to the left) follow from the **Lagrangian** $\mathcal{L} = \bar{\psi}(i\gamma\cdot\partial - m)\psi$ upon varying with respect to the Pauli adjoint $\bar{\psi} = \psi^\dagger\gamma^0$ and ψ respectively.

$$i\gamma\cdot\partial\psi = m\psi \quad \text{and} \quad -i\partial_\mu\bar{\psi}\gamma^\mu = m\bar{\psi} \quad \text{or} \quad \bar{\psi}(-i\gamma\cdot\partial - m) = 0. \tag{317}$$

• **The Dirac equation admits a conserved 'vector' current** $j^\mu = \bar{\psi}\gamma^\mu\psi$. This follows from the Dirac equations for ψ and $\bar{\psi}$:

$$\partial_\mu j^\mu = \partial_\mu\bar{\psi}\gamma^\mu\psi + \bar{\psi}\gamma\cdot\partial\psi = im\bar{\psi}\psi + \bar{\psi}(-im)\psi = 0. \tag{318}$$

As a consequence the total charge $Q = \int j^0\, d^3x$ is independent of time. This is the conservation law that follows from Noether's theorem applied to the global 'vector' $U(1)$ symmetry $\psi \to \psi' = e^{ie\theta}\psi$ of the Dirac Lagrangian.

• Examples of conserved charges arising from such global $\mathrm{U}(1)$ symmetries are quark number (one third of baryon number) and lepton number, corresponding to global phase changes in the quark and lepton fields, which are Dirac spinors in the standard model. Neutrinos were modeled as massless Dirac fields (in effect satisfying the Pauli equations aside from interactions), but now we know that they too are massive. It is still not established whether massive neutrinos are Dirac fields, they could be Majorana fields.

• Under a Lorentz transformation $x' = \Lambda x$ (where $\Lambda = e^{-(i/2)\omega_{\mu\nu}J^{\mu\nu}}$ and $J_{\mu\nu}$ generate boosts and rotations) a Dirac spinor transforms to $\psi'(x') = S(\Lambda)\psi(x)$ where $S(\Lambda) = e^{-(i/4)\omega_{\mu\nu}\sigma^{\mu\nu}}$ and $\sigma^{\mu\nu} = \frac{i}{2}\{\gamma^\mu,\gamma^\nu\}$.

• It can be shown that $\bar{\psi}\psi$ transforms as a scalar field under Lorentz transformations while $\bar{\psi}\gamma_5\psi$ is a pseudo scalar scalar field. A scalar field ϕ transforms as $\phi'(x') = \phi(x)$. To see the behavior under parity, recall that parity is implemented on Dirac spinors via γ^0: $\psi(x) \to \psi'(x') = \gamma^0\psi(x')$. Thus under parity $\bar{\psi}\psi \to \bar{\psi}\gamma^0\gamma^0\psi = \bar{\psi}\psi$ while $\bar{\psi}\gamma^5\psi \to \bar{\psi}\gamma^0\gamma^5\gamma^0\psi = -\bar{\psi}\gamma^5\psi$. $\bar{\psi}\gamma^\mu\psi$ is a polar vector, called the vector current, which follows from the $\psi \to e^{i\theta}\psi$ symmetry of the Lagrangian. A polar vector is one that transforms as $j'_\mu = \Lambda_\mu^\nu j_\nu$ under Lorentz transformations, while under parity $\Pi(j^0,\vec{j}) = (j^0,-\vec{j})$. Check that $\bar{\psi}\gamma^\mu\gamma_5\psi$ is an axial vector (under parity its spatial components retain their signs while the time component reverses its sign). It is called the axial vector current (to be discussed below). If we define $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$, then $\bar{\psi}\sigma^{\mu\nu}\psi$ transforms as a rank two antisymmetric tensor field.

• The Dirac $\mathcal{L}$ may be expressed in terms of left and right-chiral projections

$$\mathcal{L} = \bar{\psi}_L i\partial\!\!\!/ \psi_L + \bar{\psi}_R i\partial\!\!\!/ \psi_R - m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \tag{319}$$

Notice that the mass terms couple the different chiralities. The weak interactions treat the LH and RH quark and lepton fields differently, they transform as LH doublets and RH singlets under weak isospin. It follows that mass terms for quarks and leptons violate gauge invariance. So in the SM, all quarks and leptons are massless. They get their mass through their interactions with a scalar field via the Higgs mechanism. The Yukawa interaction of Dirac spinors and scalars also couples LH and RH spinors but in a gauge invariant manner.

• When $m = 0$, the Lagrangian posses an additional $\mathrm{U}(1)$ 'chiral' symmetry under which the RH and LH spinors transform via complex conjugate phases $\psi_{R,L} \to e^{\pm i\phi}\psi_{R,L}$ or in short $\psi \to e^{i\gamma_5\phi}\psi$. The axial vector current $j_5^\mu = \bar{\psi}\gamma^\mu\gamma_5\psi$ is conserved as a consequence. Of course, when $m = 0$ we are free to change the phases of $\psi_{R,L}$ by unrelated angles. The vector and axial-vector $\mathrm{U}(1)$ symmetries are a convenient way of formulating these two symmetries, since $\mathrm{U}(1)_V$ survives as a symmetry even when $m \neq 0$.

9.8 Interactions of Dirac fields with Dirac, EM and scalar fields

• Terms which involve a product of more than two fields may be introduced into the Lagrangian density to describe interactions ('interaction vertices' where three or more lines meet in a Feynman diagram). A minimal requirement is that the interaction term must be Lorentz invariant, which is ensured if Lorentz indices are contracted. Additional requirements (like gauge invariant, behavior under parity, renormalizability etc.) may be imposed on possible interaction terms, depending on the system we wish to model.

9.8.1 Four Fermi interaction

• Fermi's theory of beta decays involved a point interaction among four Dirac spinor fields (neutron, proton, electron and neutrino). It was a vector current-current interaction of the form $\mathcal{L}_{int} = G_F(\bar{p}\gamma^\mu n)(\bar{e}\gamma_\mu\nu)+$ h.c. Each parenthesis contains a vector current. The 4-Fermi interaction preserves the parity invariance of the Dirac Lagrangians of the four particles.

• Following the discovery of parity violation in beta decay in 1956, the Fermi theory had to be modified. Since the interaction among the four particles was very short-ranged, it was modeled as an interaction between a pair of 'currents', each of which is quadratic in Dirac spinors. Lorentz covariance permits five bilinears constructed from Dirac spinors: scalar, pseudoscalar, polar vector and axial vector and tensor. Marshak and Sudarshan (with contributions from Feynman and Gell-Mann etc.) found that the experimental data on weak decays is consistent with vector minus axial vector (V-A) weak currents. Since vector and axial vector currents transform differently under parity, their difference does not have a definite parity. This allows for a parity violating interaction. For example, the $e\nu_e$ current is $j^\mu_{e\nu_e} = \bar{e}\gamma^\mu(1-\gamma_5)\nu_e$ and similarly for the $\mu\nu_\mu$, pn and each pair of quarks in a generation. The interaction Lagrangian is a contraction of two such currents, e.g. $G_F \eta_{\mu\nu} j^\mu_{e\nu_e} j^\nu_{\mu\nu_\mu} + \text{h.c.}$ is relevant to muon decay. Since $P_L = \frac{1}{2}(1-\gamma_5)$ projects to LH spinors and $\bar{e}\gamma^\mu(1-\gamma_5)\nu = \bar{e}(1+\gamma_5)\gamma^\mu\nu = ((1-\gamma_5)e)^\dagger\gamma^0\gamma^\mu\nu$, we see that the weak currents only involve the LH fields. The charge changing weak interactions couple only to the LH quark and lepton fields. The RH components of these particles simply do not participate in the charge-changing weak interactions. This is incorporated in the SM by making the LH fields transform as doublets under weak isospin while the RH fields are singlets do not transform at all (so they do not feel the charge changing weak force at all). E.g. the RH electron e_R does not participate in the charge changing weak interactions. An extreme case of this are the RH neutrinos, not only do they not participate in the charge changing weak interactions, they do not even find a place in the standard model. They either do not exist or, more likely, are very heavy and have not been detected so far.

9.8.2 Dirac field coupled to photons

• As in the case of complex scalars, we may couple the Dirac field to the EM field by gauging the above $U(1)$ global symmetry (minimal coupling): $\partial_\mu\psi \to D_\mu\psi = (\partial_\mu - ieA_\mu)\psi$. The covariant derivative of ψ transforms in the same way as $\psi \to e^{ie\theta}\psi$ provided $A_\mu \to A_\mu + \partial_\mu\theta(x)$. The Dirac equation in the presence of an external EM field is

$$(i\gamma\cdot D - m)\psi = 0 \quad \text{or} \quad (i\gamma^\mu(\partial_\mu - ieA_\mu) - m)\psi = 0 \tag{320}$$

It follows from the Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma\cdot D - m)\psi = \bar{\psi}(i\gamma\cdot\partial - m)\psi + ej^\mu A_\mu \tag{321}$$

$j^\mu = \bar{\psi}\gamma^\mu\psi$ is the electromagnetic current of the Dirac field. e is called the gauge coupling, the strength of the interaction between the Dirac particles and the gauge field. e is also (up to a sign!) the electric charge of the particle annihilated by ψ or the anti-particle created when ψ acts on the vacuum. To find out, one must compute the eigenvalue of the conserved charge operator $Q = -e\int j^0 d^3x = -e\int \psi^\dagger\psi d^3x$ when acting on the 1 anti-particle state created when ψ acts on the vacuum. [Check whether the signs are chosen so that we may interpret ψ to annihilate electrons of charge $-e$ (and create positrons of charge e) by checking whether $Q = -e$ is the eigenvalue of the conserved charge operator $Q = -e\int j^0 d^3x = -e\int \psi^\dagger\psi d^3x$ in the one electron state.]

• If we include the dynamics of the gauge field we arrive at the Lagrangian of quantum electrodynamics

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma\cdot D - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \bar{\psi}(i\gamma\cdot\partial - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + ej^\mu A_\mu. \tag{322}$$

The Maxwell equations in the presence of the charged Dirac field are $\partial_\mu F^{\mu\nu} + e\bar{\psi}\gamma^\nu\psi = 0$. $\mathcal{L}_{\text{QED}}$ is Lorentz and gauge-invariant by construction. Check that it is also invariant under parity. The mass dimensions of A, ψ are $M^1, M^{3/2}$. It follows that e is dimensionless. A dimensionless coupling leads to a perturbatively renormalizable quantum theory.

9.8.3 Coupling of Dirac field to scalars

• Coupling of Dirac field to scalar particles is called Yukawa coupling since Yukawa considered a related model for nucleon-pion interactions. The Dirac equation modified in the presence of a real scalar,

$$i\gamma\cdot\partial\psi = m\psi + g\phi\psi, \tag{323}$$

follows from the Lagrangian

$$\mathcal{L} = \bar\psi(i\gamma\cdot\partial - m)\psi - g\bar\psi\psi\phi. \tag{324}$$

$\bar\psi\gamma\cdot\partial\psi, \bar\psi\psi, \phi$ are all Lorentz scalars, so the Lagrangian is Lorentz invariant. It is also invariant under $\mathrm{U}(1)_V$ phase changes of ψ. The phase of ϕ cannot be changed since ϕ is a real scalar. The Yukawa coupling constant g is dimensionless since $[\phi] = M, [\psi] = M^{3/2}$, leading to a perturbatively renormalizable model. g measures the strength of the scalar-fermion coupling (tri-linear vertex). Notice that if ϕ is a constant, then the Yukawa coupling term behaves like a mass term for the Dirac field. This idea is used in the Higgs mechanism to give masses to the quarks and charged leptons.

• If we include the Klein-Gordon dynamics of the scalar, the Lagrangian becomes

$$\mathcal{L} = \bar\psi(i\gamma\cdot\partial - m)\psi - g\phi\bar\psi\psi + \frac{1}{2}|\partial\phi|^2 - \frac{1}{2}\mu^2|\phi|^2. \tag{325}$$

The Dirac field appears as a scalar source term $\bar\psi\psi$ in the KG field equation for ϕ

$$(\Box + \mu^2)\phi = -g\bar\psi\psi. \tag{326}$$

The Yukawa coupling $-g\phi\bar\psi\psi$ is a (true) scalar coupling since $\bar\psi\psi$ transforms as a true scalar under parity. One can similarly consider coupling of Dirac fields to pseudoscalars.

9.8.4 Pseudo-scalar coupling of Pion triplet to nucleon doublet

• Attempts to model the inter-nucleon interaction using pions to match nucleon-nucleon scattering data and deuteron properties lead to a pseudo-scalar meson coupling to nucleons. Let us write a Lagrangian for pion-nucleon strong interactions, it must be invariant under SU(2)-isospin rotations and of course Poincare invariant. Pions are not true scalars but pseudo-scalars, they couple with opposite sign to the right and left handed components of nucleons (which are spin half Dirac particles). To begin with, we ignore isospin and treat the nucleon field as a single Dirac spinor and the pion field as a single real scalar field. Since $\gamma_5 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}$ in the chiral basis, this opposite sign coupling to RH and LH nucleons is modeled by the pseudo-scalar coupling

$$i\gamma\cdot\partial = m\psi - g\phi\gamma_5\psi \quad \text{and} \quad (\Box + \mu^2)\phi = -g\bar\psi\gamma_5\psi. \tag{327}$$

These equations follow from the Dirac-Yukawa Lagrangian

$$\mathcal{L} = \bar\psi(i\gamma\cdot\partial - m)\psi - g(\bar\psi\gamma_5\psi)\phi + \frac{1}{2}(|\partial\phi|^2 - \mu^2|\phi|^2). \tag{328}$$

Now let us incorporate isospin. The nucleon field ψ_i has two components $i = 1, 2$, each of which is a Dirac spinor. $\psi_1^\dagger$ creates a proton while acting on the vacuum while $\psi_2^\dagger$ creates a neutron. The pion field is an isospin one real triplet $\vec{\phi} = (\phi_1, \phi_2, \phi_3)$. The physical pions of definite charge are $\pi^0 = \phi_3, \pi^\pm = \frac{1}{\sqrt{2}}(\phi_1 \pm i\phi_2)$. The pseudo-scalar coupling of pions to nucleons is given by

$$\mathcal{L} = \bar{\psi}_i(\delta_{ij} i\gamma^\mu \partial_\mu - M_N \delta_{ij})\psi_j + \frac{1}{2}\left(\partial^\mu \vec{\phi} \cdot \partial_\mu \vec{\phi} - \mu^2 \vec{\phi} \cdot \vec{\phi}\right) - g(\bar{\psi}_i \gamma_5 \vec{\tau}_{ij} \psi_j) \cdot \vec{\phi}. \tag{329}$$

M_N is the nucleon mass, the proton and neutron must have a common mass for isospin to be a symmetry. We have used vector notation for the isospin triplet and indices $1 \leq i, j \leq 2$ for the isospin doublet. The contraction of indices and dot products make it clear that this Lagrangian is a scalar under rotations in isospin space, thus implementing the isospin symmetry. Note that the combination of two $I = \frac{1}{2}$ nucleon doublets in $N^\dagger \vec{\tau} N$ transforms as an $I = 1$ triplet which can then be dotted with the iso-triplet $\vec{\phi}$ to get a scalar under isospin. This should not come as a surprise $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$: $N^\dagger \vec{\tau} N$ is the iso-triplet while $N^\dagger N$ is the iso-singlet.

• Note that in natural units ψ and ϕ have dimensions of $M^{3/2}$ and M^1 while g has mass dimension zero; the model is perturbatively renormalizable. However $g^2/\hbar c$ is not small (it can be as large as ≈ 40 based on the strength of the inter-nucleon force). So perturbation expansions in g cannot be trusted in general for pion nucleon interactions.

9.9 Charge conjugation for the Dirac field

• Under charge conjugation, the Dirac equation $(i\gamma \cdot (\partial - ieA) - m)\psi = 0$ for a spinor ψ annihilating a particle of electric charge e turns into the Dirac equation for the charge conjugate spinor ψ_c (to be found below), which describes particles of opposite electric charge $-e$ but same mass (anti-particles), $(i\gamma \cdot (\partial + ieA) - m)\psi_c = 0$. Based on this observation, Weyl and Oppenheimer pointed out that the anti-electron could not be the proton, as originally proposed by Dirac.

• To find ψ_c we take the complex conjugate of the Dirac equation $(-i\gamma^* \cdot (\partial + ieA) - m)\psi^* = 0$ and try to make it look like the Dirac equation with $e \to -e$. Since $-(\gamma^\mu)^*$ satisfy the same Dirac algebra as γ^μ, they are related by a similarity transformation $-\gamma^{\mu *} = S^{-1}\gamma^\mu S$ with S conventionally written as $S = C\gamma^0$ where C is called the charge conjugation matrix. Then left multiplying the complex conjugate equation by S we see that $\psi_c \equiv C\gamma^0\psi^*$ satisfies the Dirac equation with $e \to -e$:

$$(i\gamma^\mu C\gamma^0(\partial_\mu + ieA_\mu) - C\gamma^0 m)\psi^* = 0 \quad \Rightarrow \quad (i\gamma \cdot (\partial + ieA) - m)\psi_c = 0. \tag{330}$$

To find C we must solve the equation $S\gamma^{\mu *} = -\gamma^\mu S$. S is of course defined only up to a phase. The explicit matrix $C = S\gamma^0$ depends on the basis chosen. C can be found in the same way in both the Dirac and Weyl bases as in both these bases, $\gamma^0, \gamma^1, \gamma^3$ are real while γ^2 is purely imaginary. So in any such basis, the condition on S is that it commute with γ^2 but anti-commute with $\gamma^{0,1,3}$. A choice that does the job is $S = \gamma^2$ (up to a phase). Thus in the

Dirac and Weyl bases $\psi_c = \gamma^2\psi^*$ and $C = \gamma^2\gamma^0$. Explicitly,

$$C_{\text{Dirac}} = \begin{pmatrix} 0 & -\sigma_2 \\ -\sigma_2 & 0 \end{pmatrix} \quad \text{and} \quad C_{\text{Weyl}} = \begin{pmatrix} \sigma_2 & 0 \\ 0 & \sigma_2 \end{pmatrix}. \tag{331}$$

We may now check that charge conjugation takes LH spinors to RH spinors and vice versa. For example, in the Chiral basis a LH spinor is of the form $\psi_L = \begin{pmatrix} \psi_- \\ 0 \end{pmatrix}$ and its charge conjugate spinor $\psi_{Lc} = \gamma^2\psi_L^* = \begin{pmatrix} 0 \\ -\sigma_2\psi_-^* \end{pmatrix}$ is right-handed. Upon quantization we may say that the anti-particle of, say a left-handed neutrino ν_L is a right-handed anti-neutrino and anti-particles corresponding to the LH weak isospin doublet $\begin{pmatrix} \nu \\ e \end{pmatrix}_L$ is the RH weak isospin doublet $\begin{pmatrix} \bar{\nu} \\ \bar{e} \end{pmatrix}_R$.

9.9.1 Remarks on anti-particles

• Dirac introduced the anti-particle concept in 1929-31 using holes in the filled negative energy sea [look at a field theory or relativistic QM text for this]. The name was given by de Broglie in 1934. Anti-particles are just like particles, they can possess position, energy, momentum (and therefore fall in a gravitational field) and charge (and so leave tracks in detectors) etc.

• For each particle discovered so far, there is a corresponding anti-particle with the same mass and spin. Relative to a particle, the anti-particle's quantum numbers like electric charge, electric and magnetic moment, strangeness, baryon and lepton number etc. have opposite sign, colors are also reversed red to anti-red, so to speak or more precisely anti-quarks transform under the conjugate $\mathbf{3}^*$ of the fundamental representation $\mathbf{3}$ of color SU(3).

• If a particle has no attributes other than energy, momentum, spin and angular momentum, then its anti-particle is defined to be itself. If it has any other attributes like charge, then its anti-particle is distinct and has the opposite charge.

• The antiparticle of the electron e^- is the positron denoted e^+ or $\bar{e}$, it was the first anti-particle to be discovered, in 1932 in cloud chamber photographs of C D Anderson and Blackett and Occhialini. It is a matter of convention that we call the electron the particle and e^+ the anti-particle. $\pi^\pm$ are anti-particles of each other as are $\mu^\pm$. In these cases particles and anti-particles were discovered at about the same time and they are all unstable, and have roughly equal abundances, so there is not much meaning to calling μ^- matter and μ^+ anti-matter. Anti-quarks are distinct from quarks and most of matter is made of u, d quarks. Some electrically neutral particles are their own anti-particles, e.g. the photon, neutral π^0 meson, η^0 meson and the Z^0 boson. However, not every neutral particle is its own anti-particle, for example the anti-neutron is distinct from the neutron as is evident from the valence quark content $n = udd, \bar{n} = \bar{u}\bar{d}\bar{d}$. Moreover neutron and anti-neutron have opposite magnetic moments and most spectacularly, a neutron and anti-neutron can annihilate producing energy, while this does not happen when a pair of neutrons are brought nearby (as in a nucleus or in a neutron star).

• Particle-antiparticle annihilation produces a state with energy and possibly momentum, but no charge, baryon or lepton number etc. Before they annihilate, particle and anti-particle can form short-lived bound states (e.g. π^0 which is made of $u\bar{u}$ and $d\bar{d}$.). Particle-anti particle collisions at high energy are a way of creating new particles. E.g., this was done with e^+e^- at SLAC (Stanford) and LEP (CERN) and with $p\bar{p}$ at CERN and Fermilab.

9.10 Yang-Mills field for spin 1 particles

Recall that minimal coupling of the EM field to a Dirac field (or complex scalar) could be obtained by gauging a global U(1) symmetry $\psi' = e^{ie\theta}\psi$ of the matter field Lagrangian, say $\mathcal{L} = \bar{\psi}(i\gamma\cdot\partial - m)\psi$. Invariance under local gauge transformations $\psi' = g(x)\psi$ where $g = e^{ie\theta(x)} \in U(1)$ and θ is a function on space-time, is ensured by replacing partial derivatives by covariant derivatives $D_\mu = \partial_\mu - ieA_\mu$ with the associated gauge transformation $A'_\mu = A_\mu + \partial_\mu\theta$ which leaves the field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ invariant[48]. The equations of electrodynamics $\partial_\mu F^{\mu\nu} = -ej^\nu$ then follow from the Lagrangian $\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(i\gamma\cdot D - m)\psi$ where $j^\mu = \bar{\psi}\gamma^\mu\psi$. We will now generalize from the abelian gauge group U(1), to gauge transformations $g(x) = e^{ie\theta(x)}$ living in a non-abelian gauge group G, such as SU(N). Here $\theta(x)$ is an $N\times N$ hermitian matrix field. The construction works for any compact semi-simple G such as SU$(2)\times$ SU(3) as well as a group that has U(1) factors. The resulting non-abelian gauge theory was first proposed by R Shaw and independently by C N Yang and R L Mills in 1954 for the isospin group $G =$ SU(2). As it turns out, isospin is only an approximate global symmetry (neutron and proton are not exactly degenerate in mass) and is not the one that nature has 'gauged' (i.e., made a local symmetry). But the resulting framework of non-abelian gauge theory *does work* in nature when applied instead to weak isospin SU(2) and color SU(3) (as well as the abelian gauge groups: $U(1)$ weak-hypercharge and U(1) of electric charge).

• It is helpful to simplify notation by absorbing the coupling e into $\theta(x)$ and $A_\mu(x)$ so that $g = e^{i\theta(x)}$, $\psi' = g\psi$, $A'_\mu = gA_\mu g^{-1} + ig\partial_\mu g^{-1}$, $D_\mu\psi = (\partial_\mu - iA_\mu)\psi$ and $\mathcal{L} = -\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma\cdot D - m)\psi$. Though we do not do so here, one may work with the purely imaginary field $\mathcal{A}_\mu(x) = -iA_\mu$ which lives in the Lie algebra of $U(1)$.

• Suppose there are N matter fields (say Dirac spinors) that transform under the fundamental representation of $SU(N)$ ($\psi' = g\psi$ where g is an $N\times N$ special unitary matrix) leaving the Dirac Lagrangian invariant. This may be made a local symmetry $\psi' = g(x)\psi(x)$ by defining the **covariant derivative** $D_\mu\psi = (\partial_\mu - iA_\mu)\psi$. The gauge potential $A_\mu(x)$ is a traceless hermitian $N\times N$ matrix-valued field. Let t_a be a basis (with $a = 1,\ldots \dim SU(N) = N^2-1$) of traceless hermitian $N\times N$ generators for the Lie algebra, satisfying the commutation relations $[t_a, t_b] = iC_{ab}^c t_c$ with real structure constants, and conventionally normalized according to $\operatorname{tr}\, t_a t_b = \frac{1}{2}\delta_{ab}$. Then we may expand $A_\mu = A_\mu^a t_a$, where $A_\mu^a(x)$ are real. The t_a span a linear space of dimension $n = N^2 - 1$ which carries the adjoint representation of the Lie algebra. So for a gauge group of dimension n, there are n independent gauge fields or gauge bosons. The gauge field A_μ at a fixed location x^μ is said to live in the adjoint representation. The adjoint representation of a Lie algebra or group always has dimension equal to the dimension of the group or algebra.

• The covariant derivative transforms in the same way as ψ, i.e., $(D_\mu\psi)' = g(D_\mu\psi)$ provided

[48]Notice that we may express the transformation of the gauge potential as $A'_\mu = gA_\mu g^{-1} + \frac{i}{e}g\partial_\mu g^{-1}$.

the gauge field transforms as $A'_\mu = gA_\mu g^{-1} + ig\partial_\mu g^{-1}$. This ensures that the matter field Lagrangian $\mathcal{L} = \bar\psi(i\gamma\cdot D - m)\psi$ is gauge invariant as the following calculation shows.

$$\begin{aligned}
\mathcal{L} &\to \bar\psi g^{-1}(i\gamma\cdot(\partial - igAA^{-1} + g\partial g^{-1}))g\psi \\
&= \bar\psi g^{-1}(i\gamma\cdot\partial g + ig\gamma\cdot\partial + \gamma\cdot gAg^{-1}g + i\gamma\cdot g(\partial g^{-1})g)\psi \\
&= \bar\psi\left(ig^{-1}\gamma\cdot\partial g + i\gamma\cdot\partial + \gamma\cdot A + i\gamma\cdot(\partial g^{-1})g\right)\psi = \mathcal{L}.
\end{aligned} \tag{332}$$

In the last equality we used $g^{-1}\partial g + (\partial g^{-1})g = 0$ which follows from $\partial(g^{-1}g) = \partial(I) = 0$. Thus, the Dirac Lagrangian minimally coupled to A_μ is gauge invariant if ψ, A transform in the above manner.

• To find the gauge field dynamics we must find a non-abelian generalization of the field strength tensor since $\partial_\mu A_\nu - \partial_\nu A_\mu$ is not gauge invariant nor homogeneous under gauge transformations. Check this for an infinitesimal gauge transformation

$$g(x) \approx I + i\theta(x), \quad \delta\psi = i\theta\psi, \quad \delta A_\mu = \partial_\mu\theta - i[A_\mu,\theta] \equiv D_\mu\theta. \tag{333}$$

Here we have defined the **covariant derivative** of the matrix field θ by the last equality. The formula for the covariant derivative of a field depends on the representation of the gauge group to which the field belongs, it is different for the matter fields that live in the fundamental rep and for fields like θ or A that live in the adjoint rep. This is similar to the different formulae for the covariant derivatives of various tensor fields in Riemannian geometry: $\nabla_\mu f = \partial_\mu f$ for scalar functions, $\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma V$ for components of a vector field, $\nabla_\mu g_{\rho\sigma} = \partial_\mu g_{\rho\sigma} + \Gamma\Gamma g - \Gamma\Gamma g$ for the components of a covariant second rank tensor field, etc. Here Γ^a_{bc} are the Christoffel connection coefficients. Fill in the indices in the formulae for covariant derivatives!

• To find the correct generalization of the electromagnetic field strength, notice that in the abelian case, we may express the field strength $\partial_\mu A_\nu - \partial_\nu A_\mu$ as the commutator of covariant derivatives, check that $F_{\mu\nu}\psi = i[D_\mu, D_\nu]\psi$. The field strength measures the departure from 'flatness' of the connection A_μ, and just as in Riemannian geometry, the curvature is given by the commutator of covariant derivatives. In the non-abelian case, the hermitian matrix field

$$F_{\mu\nu} = i[D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu], \quad F_{\mu\nu} = F^a_{\mu\nu} t_a. \tag{334}$$

transforms homogeneously under a gauge transformation. This follows from the behavior under gauge transformation $(D_\mu\psi) \to gD_\mu\psi, D_\mu D_\nu\psi \to gD_\mu D_\nu\psi$, so that $F_{\mu\nu}\psi \to gF_{\mu\nu}\psi = gF_{\mu\nu}g^{-1}g\psi$. So we infer that $F_{\mu\nu} \to gF_{\mu\nu}g^{-1}$ transforms homogeneously by conjugation. Under an infinitesimal gauge transformation $g \approx I + i\theta(x)$, $\delta F_{\mu\nu} = -i[F_{\mu\nu}, \theta]$. Unlike in the abelian case, the field strength $F_{\mu\nu}$ is *not* gauge invariant. But we can easily find a gauge invariant quantities, they are quantities constructed from $F_{\mu\nu}$ that are invariant under conjugation, for instance by taking the trace of a polynomial in F. The simplest possibility is $\mathrm{tr}\, F_{\mu\nu}$, however this is zero in an SU(N) gauge theory since A_μ is traceless and the trace of a commutator vanishes. A gauge-invariant and Lorentz invariant Lagrangian that is quadratic in derivatives and reduces to $\frac{1}{2}(E^2 - B^2)$ in the abelian case is the Yang-Mills Lagrangian

$$\mathcal{L} = -\frac{1}{2e^2}\,\mathrm{tr}\, F_{\mu\nu}F^{\mu\nu} = -\frac{1}{4e^2}F^a_{\mu\nu}F^{a\mu\nu} \quad \text{since} \quad \mathrm{tr}\, t_a t_b = \frac{1}{2}\delta_{ab}. \tag{335}$$

The pre-factor of $1/e^2$ would be absent if we did not absorb e into the gauge potentials. e is called the gauge coupling, it is a measure of the strength with which the gauge fields A couple to themselves and to the matter fields ψ. Thus the Lagrangian of an SU(N) non-abelian Yang-Mills gauge theory coupled to Dirac spinors in the fundamental representation of SU(N) is

$$\mathcal{L} = -\frac{1}{2e^2} \text{ tr } F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma^\mu(\partial_\mu - iA_\mu) - m)\psi \tag{336}$$

An example is Quantum Chromodynamics, an SU(3) Yang-Mills theory coupled to quarks, which transform in the fundamental representation of the color group SU(3).

• The classical Yang-Mills field equations that follow are a generalization of Maxwell's equations

$$D_\mu F^{\mu\nu} = \partial_\mu F^{\mu\nu} - ie[A_\mu, F^{\mu\nu}] = -e(\bar{\psi}\gamma^\nu t_a \psi)t_a. \tag{337}$$

Unlike Maxwell's equations, but like Einstein's equations, these are non-linear equations even in the absence of quark matter sources on the RHS. Indeed $D_\mu D^{\mu\nu}$ is cubically non-linear in the gauge fields A. The non-abelian gauge fields A_μ^a (e.g. $a = 1, \cdots 8$ gluons in QCD) interact with themselves, unlike photons in an abelian gauge theory. The interaction vertices are obtained by expanding the Lagrangian and identifying terms that involve products of more than two fields. Omitting some indices and displaying the coupling constant e explicitly,

$$\mathcal{L} = -\frac{1}{2e^2} \text{ tr } (e\partial A - e\partial A - ie^2[A, A])(e\partial A - e\partial A - ie^2[A, A]) + i\bar{\psi}(i\gamma \cdot (\partial - ieA)) \tag{338}$$

The quadratic terms in $\mathcal{L} \sim \partial A \partial A, \bar{\psi}(i\gamma \cdot \partial - m)\psi$ describe free propagation of the *massless* gauge fields A_μ^a (say 8 gluons in the case of SU(3)) as well as the massive Dirac fields ψ (say quarks). We see that there are 3-gluon 'cubic/trilinear' vertices of strength e coming from $\partial A[A, A]$ terms (these are derivative interactions) as well as 4-gluon vertices of strength e^2 coming from the $[A, A][A, A]$ terms. The possibility of self-interactions of gauge fields is due to the non-abelian nature of the gauge group, for an abelian group like U(1) of EM, the commutator terms vanish. Photons do not self-interact, while gluons, weak gauge bosons (and gravitons) do self-interact. In addition, we have the cubic vertex $e\bar{\psi}A\psi$ where a gluon can pair produce a quark and anti-quark.

• The local gauge symmetries that have so far been identified in nature are SU$(3)_C$ for color, U$(1)_Q$ for electric charge, SU$(2)_W$ for weak isospin, U$(1)_Y$ for weak hypercharge and $\text{Diff}(M)$ space-time diffeomorphisms (general coordinate transformations in general relativity). It turns out that U$(1)_Q$ is a combination of U$(1)_Y$ and the third component of SU$(2)_W$. More on this later. It is often said that the electroweak gauge symmetry $SU(2)_W \times U(1)_Y$ is spontaneously broken to U$(1)_Q$. What this means is that the ground state of the theory does not display the full gauge symmetry, only the U$(1)_Q$ of electric charge. Nevertheless, the theory as a whole still possesses the full gauge symmetry.

• A key difference between the abelian gauge theory QED and the non-abelian gauge theory QCD is that QED (as we know it) does not confine electrons and positrons and photons inside bound states like positronium. Positronium can be split (actually, it decays spontaneously) liberating electrons and positrons as well as photons. On the other hand, QCD in the phase that describes hadrons (as we know them) confines quarks and gluons inside hadrons. Hadrons may be unstable (like positronium), but the decay or induced breakup of hadrons produces more hadrons or leptons and photons, but not quarks and gluons.

• Not every non-abelian gauge theory is confining in the phase of interest. The $SU(2) \times U(1)$ electroweak theory does not confine the W, Z, γ, leptons or quarks. In part, this is because the gauge symmetry is spontaneously broken in the ground state.

• Even QCD may have a phase where quarks and gluons are not confined, a 'deconfinement' transition is hypothesized at sufficiently high temperatures.

10 Hadrons, quark model and the strong interactions

10.1 Nucleons are not elementary; Pion exchange does not quite explain strong interactions

• The Dirac equation for a charged spin half particle minimally coupled to the EM field gives the correct value for the magnetic moment of the electron. The magnetic dipole energy is $H = -\vec{\mu} \cdot \vec{B}$ where $\vec{\mu} = g\mu_B \frac{\vec{S}}{\hbar}$. The electron Bohr magneton is $\mu_B = \frac{e\hbar}{2m_e c}$ and $g = 2$ according to the Dirac equation. Small deviations from $g = 2$ could be explained by including quantum fluctuations in the EM field (QED - Schwinger 1948).

• If the proton and neutron were also spin half elementary Dirac particles, then we expect that the neutron should have zero magnetic moment since it is neutral while the proton should have a magnetic moment obtained by replacing the electron Bohr magneton with the nucleon Bohr magneton (replace $m_e \to m_p$ and reverse the sign of charge) $\mu_N = e\hbar/2m_p c$. But the measured value (O Stern, 1933, Nobel prize 1943, also measurements by I I Rabi 1934 onwards leading to development of NMR) $\mu_p = 2.79\mu_N$ is significantly different. The neutron, being uncharged should have no magnetic moment according to the Dirac theory, but $\mu_n = -1.9\mu_B$. So nucleons cannot be elementary spin half particles in the same way as electrons.

• Electron proton elastic scattering (carried out especially by R Hofstadter at SLAC) showed that the charge of the proton is spread over a charge radius of about a fermi, it is not point-like. Electric 'form factors' which are, loosely, Fourier transforms of charge distributions, were measured, and a charge radius could be extracted. Nucleons were certainly non point-like, unlike electrons which are point-like down to 10^{-18} m.

• The discovery by E Fermi, H Anderson et. al. (Chicago, 1954) that the proton has an excited state (the Delta resonance) suggests that it is not elementary. In fact there is a whole tower of nucleon resonances with increasing angular momenta and masses. By analogy with excited atoms, it appeared that excited hadrons might result for instance, from additional orbital motion of the constituents of the lowest lying hadrons.

• What is more, through the 1940s, 50s and 60s hundreds of strongly interacting particles were discovered (1432 hadrons had been identified by 1967!). Most of them were short lived resonances that decayed in 10^{-23} s while others were more stable and decayed weakly or electromagnetically. In 1968, R Hagedorn found exponential growth by plotting the density of hadronic states as a function of mass! [see plot in Huang] It was hard to believe that all the hadrons were elementary. Moreover, if the density of hadronic states $\rho(E) \propto e^{E/E_0}$ grows exponentially fast, then the statistical mechanical partition function $Z(T) = \sum_E \rho(E) e^{-E/kT}$ would cease to exist for $kT_h > E_0$. So a statistical mechanical treatment of a system of hadrons would not make sense at sufficiently high temperatures. One possibility is that the statistical mechanical treatment does hold, but that there is a phase transition around T_h, above which the density states grows more slowly. This is the hypothetical Hagedorn transition. If hadrons have more elementary constituents, then it is possible that they may organize themselves in a different way at higher temperatures.

• Though pion exchange gives a first approximation to inter-nucleon forces relevant to low energy scattering, at higher energies and at short distances, pion exchange is not adequate. Among other things, heavier (spin 1) vector mesons like the $\rho^{\pm}, \rho^0$ could also mediate the strong force and indeed they lead to corrections to the inter-nucleon potential, including a short range repulsion (due to their vector nature, like photons between like charges) that is measured. In fact, the force between like charges changes sign as the spin of the exchanged boson is increased by one. Spin zero scalar pion exchange leads to an attractive force, spin one photon exchange is repulsive, and spin two graviton exchange is attractive (all between like charges). In fact, if one wishes to explain strong interactions using meson exchange, there is no reason to exclude other mesons from being exchanged. It was clear that Yukawa's simple picture was not the whole story, and attempts to fix it became rather more complicated.

• These and other features of hadrons motivated the search for composite models for hadrons and other possible dynamical principles to explain the forces between hadrons. These investigations eventually lead to quarks as constituents of hadrons, with gluons as mediators of the strong force, summarized in QCD in the 1970s. This development took many decades, we discuss some steps along the way.

10.2 Excitation energy, sizes and strong decay widths of hadrons

• Let us begin by recalling some of the characteristic orders of magnitude associated with strong interactions of hadrons. The excitation energies of baryons and mesons are typically 100s of MeV. E.g. $\Delta(1232)$ is the lowest lying baryonic excitation ($J = 3/2$) above the nucleons $p(938), n(939)$ which have $J = \frac{1}{2}$. Similarly, the $\rho(776)$ mesons (J=1) are excited states of $J = 0$ pions $\pi(140)$. So the excitation energies are comparable to the masses of the ground states. So the quark and gluon motions responsible for increase in angular momentum that give rise to hadronic excitations are relativistic in such hadrons[49] Now, the radius R of a hadron is about a fermi or a fraction thereof. If this is taken as the uncertainty in position of quarks $\Delta x = R$, then by the uncertainty principle we would estimate $\Delta p \sim \hbar/R \approx 200$ MeV$/c$, which is of the order of the excitation energy $\Delta E = 300$ MeV for $N - \Delta$. Furthermore, the mean life-times of hadrons that decay via the strong interactions is $\tau \approx 10^{-23} s$. τ is simply the time it takes light to traverse a hadron. Check that $\tau \approx R/c$ where $R \approx 1/200 - 1/300$ MeV^{-1} or equivalently that the energy widths of these resonances are $\Gamma = \hbar/\tau \approx 300$ MeV. In

[49]There are also mesons built of heavy quark-anti-quark pairs such as $\psi = c\bar{c}$ and $\Upsilon = b\bar{b}$, where the excitation energies are much less than the rest energy of the ground state, in these cases the quarks can be treated non-relativistically.

summary, excitation energies, widths and characteristic (inverse) hadron radii are all of order a few hundred MeV.

10.3 Fermi-Yang composite model

• Shortly after the experimental discovery of (charged) pions in 1947 (Powell et. al.), Fermi and Yang (1948-49) suggested that the three pions be regarded as bound states of nucleons and anti-nucleons, this time in the isotriplet spin zero ($I = 1, J = 0$) combination as opposed to the isosinglet spin-triplet combination that produces the deuteron. They found that the charges and isospin quantum numbers of the pions follow from those of the constituents if we define three spin zero states

$$(I_3 = 1)\ \ \pi^+ = \bar{n}p, \quad (I_3 = -1)\ \ \pi^- = \bar{p}n, \quad (I_3 = 0)\ \ \pi^0 = \bar{n}n - \bar{p}p. \qquad (339)$$

Here (n, p) form an isospin half doublet while $(\bar{n}, -\bar{p})$ form another isospin doublet. The negative sign is due to the Condon-Shortley phase convention for charge conjugation. (p, n) transforms in the $\mathbf{2}$ representation and $(\bar{n}, -\bar{p})$ also transforms in the same way, for SU(2) the fundamental representation $\mathbf{2}$ and its conjugate $\bar{\mathbf{2}}$ are equivalent.

• The odd parity of pions cold also be accommodated, since nucleons have even parity and an anti-fermion has opposite parity compared to a fermion. Thus the Fermi-Yang proposal explains the pseudo-scalar nature of pions. However, the dynamical details of this idea did not work since the inter-nucleon force in the iso-triplet channel does not lead to bound states. It would also have to be a very peculiar bound state since the masses of the constituent nucleons (938-939 MeV) are so much greater than the pion masses (135-140 MeV) and very large binding energies would be needed to make the pions so light. However, the Fermi-Yang model did get the quantum numbers right. In a sense, some group-theoretic aspects were correct though n and p aren't the correct constituents of pions. With the benefit of hindsight, the Fermi-Yang idea is correct if we replace the nucleon iso-doublet (n, p) with the fractionally charged quark iso-doublet (u, d) subject to gluon interactions. However, this was not realized till the early 1970s. The next step was to include strange particles in the scheme of Fermi and Yang.

10.4 Discovery of strange particles, GNN relation

• V particles (which are among the strange particles) were discovered in cosmic ray events seen in cloud chamber tracks that looked like a V: e.g., some charged $V^\pm$ particles decayed into a charged + neutral particle (e.g. $K^+ \to \mu^+\nu_\mu$, $\Sigma^- \to n\pi^-$) leading to a bend in the track while some neutral V^0 particles decayed to two charged particles (e.g. $\Lambda^0 \to p\pi^-$) in a V shaped cloud chamber photograph. Some $V^\pm$ decayed to three charged particles (e.g. $K^- \to \pi^-\pi^-\pi^+$). As in these examples, the decay end products were pions, muons, nucleons etc. i.e., not strange particles (and the decays were typically two body decays leaving no scope for neutral strange daughter particles that did not leave a track). The V particles were very long-lived (e.g. 10^{-10}s, for $\Sigma^+ \to p^+\pi^0$) compared to the typical time-scale $\tau \sim 10^{-23}$s of the strong interactions. V particle events among cosmic ray events were rare: they weren't seen till the late 1940s. But interestingly, there were a lot of photographs with two V particles. This could not be a coincidence: the probability of two V's should be vanishingly small if they were independently produced. Pais and Nambu conjectured that they must be produced in the same reaction in association with each other. V particles were strange since they were copiously produced with cross sections typical of strong interactions in pion-nucleon or nucleon-nucleon scattering but decayed rather slowly on time scales more typical of weak interactions.

• Gell-Mann, Nishijima and Nakano proposed that strange particles carry an additive strangeness quantum number that is conserved in strong (and electromagnetic) interactions but not in weak interactions. Around 1953 they noticed that the centre of charge of a strange isospin multiplet is displaced relative to a non-strange isospin multiplet. E.g. Lambda is displaced by half a unit to left compared to center of charge of nucleon doublet when the isospin multiplets are drawn one below the other, with charges aligned:

$$\begin{pmatrix} S\backslash Q & -1 & 0 & 1 & I \\ 0 & & n^0 & p^+ & \frac{1}{2} \\ -1 & & \Lambda^0 & & 0 \\ -1 & \Sigma^- & \Sigma^0 & \Sigma^+ & 1 \\ -2 & \Xi^- & \Xi^0 & & \frac{1}{2} \end{pmatrix} \qquad (340)$$

Strangeness could be defined as twice this displacement. We also see that electric charge increases with I_3. Along with the constant shift given by strangeness, this lead to GNN formula $Q = I_3 + (B+S)/2$. $Y = B + S$ was called hypercharge.

• With the benefit of hindsight, strangeness counts the number of anti-strange quarks minus number of strange quarks. Strong interactions do not change quark flavor, so strange and anti-strange quarks must be produced in pairs in gluonic interactions, explaining the associated production of two V particles in the same strong reaction. The lightest strange particles do not decay via the strong interactions since there are no lighter strange particles with the appropriate quantum numbers. This explains the longevity of strange particles which typically decay weakly in strangeness changing charged current weak interactions $s \to uW^-, \bar{s} \to \bar{u}W^+$.

• More examples of strangeness conserving strong interactions: when a secondary K^- beam of several GeV strikes a proton, several strong interactions are seen, all of which conserve strangeness and also serve as a way of creating other strange particles

$$(a)\;\; K^-p \to K^-p\pi^+\pi^-\pi^0 \quad (b)\;\; K^-p \to \Sigma^-\pi^+ \quad (c)\;\; K^-p \to \Lambda^0\pi^0 \quad (d)\;\; K^-p \to K^-pp\bar{p}. \qquad (341)$$

10.5 Sakata model for hadrons with p, n, Λ as building blocks

In 1956 S Sakata tried to extend the Fermi-Yang composite model to include strange hadrons. Setting aside the problems with the Fermi-Yang model, he postulated that the three baryons p, n, Λ^0 make up other mesons and baryons. n, p were assigned $I_3 = \mp\frac{1}{2}$ and strangeness zero, while Λ^0 was assigned $I_3 = 0, S = -1$. Then the charges, isospins and strangeness of pions and kaons could be accounted for if they were composed of a baryon and anti-baryon from the basic triplet p, n, Λ

$$\pi^+ = \bar{n}p, \;\; \pi^- = \bar{p}n, \;\; \pi^0 = \bar{n}n - \bar{p}p, \;\; K^+ = \bar{\Lambda}p, \;\; K^0 = \bar{\Lambda}n, \;\; K^- = \bar{p}\Lambda, \;\; \bar{K}^0 = \bar{n}\Lambda. \quad (342)$$

What is more, Sakata could accommodate the spin half strange baryons Σ, Ξ as well. Being spin half fermions, baryons had to be composed of an odd number of baryons from Sakata's basic triplet (or a baryon and a meson, etc.)

$$\Sigma^+ = \Lambda\pi^+ = \Lambda\bar{n}p, \;\; \Sigma^0 = \Lambda\pi^0, \;\; \Sigma^- = \Lambda\pi^- = \Lambda\bar{p}n, \;\; \Xi^0 = \Lambda\bar{K}^0 = \Lambda\Lambda\bar{n}, \;\; \Xi^- = \Lambda K^- = \Lambda\Lambda\bar{p}. \quad (343)$$

A further consequence of Sakata's composite model was a simple explanation of the GNN relation $Q = I_3 + \frac{1}{2}(B + S)$. It follows by assuming that I_3, B, S are additive quantum numbers in forming the composite state. This explanation for the GNN relation survives in the quark model.

• The Sakata model acquired even more predictive power when it was postulated that the $su(2)$ isospin symmetry between n, p could be extended to an approximate $su(3)$ symmetry under which p, n, Λ transform in the fundamental triplet representation. Then the composite states would be expected to appear in irreducible multiplets of $su(3)$. For example, there are 3×3 states composed of a pair of baryons from Sakata's basic trio $nn, pp, \Lambda\Lambda, \cdots$. These 9 states may be divided into two irreducible multiplets, 6 symmetric states $pn + np, nn, p\Lambda + \Lambda p$ etc., and three anti-symmetric states $pn - np, n\Lambda - \Lambda n, p\Lambda - \Lambda p$. We will see that $\mathbf{3} \otimes \mathbf{3} = \mathbf{6} + \mathbf{3}^*$. Of more interest is the decomposition of states of a baryon and an anti-baryon from Sakata's trio. We will see soon that $\mathbf{3} \otimes \mathbf{3}^* = \mathbf{8} \oplus \mathbf{1}$. This appears to be in line with the observation of an octet of pseudoscalar mesons $\pi, K, \eta = \bar{p}p + \bar{n}n - 2\bar{\Lambda}\Lambda$ and the singlet $\eta' = \bar{p}p + \bar{n}n + \bar{\Lambda}\Lambda$. Moreover, the mass difference between the kaons and pions as well as between η' and the octet can be attributed to the Λ being heavier than the nucleons.

• However, Sakata's model ran into trouble with the baryons. Like mesons, the lowest lying baryons also appear in an octet of nearly degenerate states (N, Σ, Ξ, Λ), while Sakata's model placed three of these (p, n, Λ) in a triplet and treats them differently from the remaining 5. Moreover, there is no 5d representation of $su(3)$ to accommodate the $\Sigma^\pm, \Sigma^0, \Xi^-, \Xi^0$.

• Though the Sakata model had to be abandoned, several key features of the model survive in the quark model of Gell-Mann and Zweig, including the manner in which the GNN relation is explained. In retrospect, the idea of three basic constituents is correct (at least for hadrons with zero charm, beauty and topness), as is the idea of an $su(3)$ symmetry among them. Rather than try to construct all hadrons from a few basic hadrons, the idea that worked was to focus on the group theory without trying to identify constituents from among existing particles. Eventually, nature threw up new more elementary constituents – quarks from which all hadrons are composed.

10.6 SU(3) flavor symmetry

• The mass spectrum of hadrons displays multiplets of nearly degenerate (a few percent) particles with the same spin and parity (e.g. nucleons, pions, Δ baryons). These multiplets could be interpreted as carrying irreducible representations of the $su(2)$ isospin symmetry Lie algebra. The states within a multiplet are labelled by their I_3 values. I_3 can be taken to generate the Cartan subalgebra of $su(2)$. A Cartan subalgebra is a maximal abelian subalgebra (i.e. all its elements commute), all its generators can be simultaneously diagonalized, so their values for each of the states of an irreducible representation can be specified simultaneously. A Cartan subalgebra is not unique. I_1, I_2, I_3, I_1+3I_3 etc each span a one dimensional Cartan sub-algebra of $su(2)$. Check that adding any linearly independent element to one of these results is a non-abelian subalgebra. A basis for a Cartan subalgebra consists of the so-called Cartan generators. BY a suitable choice of basis, the Cartan generators may all be simultaneously diagonalized in any representation. The dimension of a Cartan sub-algebra is called the *rank* of the Lie algebra. $su(2)$ has rank one, $su(3)$ rank 2, $su(n)$ rank $n-1$. This is easy to understand if we recall that $su(n)$ consists of traceless hermitian matrices. So by a suitable choice of basis, the Cartan subalgebra consists of traceless diagonal hermitian matrices, and this is an $n-1$ dimensional space. Similarly, $u(n)$ has rank n.

• Strange particles like (K^+, K^0) and $(\bar{K}^0, K^-)$ also form isospin multiplets. Moreover, there appeared to be larger multiplets consisting of several isospin multiplets, all with the same J^P but differing in strangeness and masses (~ 150 MeV separating isospin multiplets). Examples include the pseudoscalar and vector meson octets, the $\frac{1}{2}^+$ nucleon octet and the $\frac{3}{2}^+$ baryon decuplet. The degeneracy among masses is broken much more strongly ($15-20\%$) than in the case of isospin multiplets (5%). The states are shown below on $I_3 - Y$ plots. So states in the octets and decuplets are labelled by I_3 and Y, which means that both I_3 and Y must be diagonal, suggesting that they span a two-dimensional Cartan subalgebra of an enlarged rank two symmetry algebra containing isospin su(2) as well as strangeness/hypercharge. The simplest candidate is su(3) and it worked. The above multiplets could be regarded as carrying irreducible representations of $su(3)$. We will see that while su(2) has a 1-dimensional Cartan sub-algebra spanned by I_3, $su(3)$ has a 2-dimensional Cartan sub-algebra spanned by I_3 and Y.

• **Baryon octet:** Lowest mass baryons with $J^P = \frac{1}{2}^+$ Strangeness increases vertically while I_3 increases to the right. Electric charge is constant along the NW-SE diagonal. Check the GNN relation.

$I_3=-1$	$I_3=-\frac{1}{2}$	$I_3=0$	$I_3=\frac{1}{2}$	$I_3=1$	S	I	$\bar{m}$ (MeV)
	n		p		0	$\frac{1}{2}$	939
	udd		uud				
Σ^-		$\Sigma^0(uds)$		Σ^+	-1	1	1193
dds		$\Lambda^0(uds)$		uus			1116
	Ξ^-		Ξ^0		-2	$\frac{1}{2}$	
	dss		uss				1318

• **Anti-baryon octet** with opposite I_3 and S quantum numbers and identical masses.

• The spin half baryon and anti-baryon octets consist of stable baryons. They do not decay through the strong interactions, and are long lived on the $10^{-23}s$ time-scale of the strong interactions. With the exception of the proton, they decay weakly or electromagnetically. E.g. the beta decay $n \to pe\bar{\nu}_e$, strangeness changing weak decays $\Lambda^0 \to p\pi^-$, $\Sigma^- \to n\pi^-$, $\Sigma^+ \to p\pi^0$

and the EM decay $\Sigma^0 \to \Lambda^0\gamma$.

• $J^P = \frac{1}{2}^-$ singlet baryon Λ with mass 1405 MeV has the wrong parity to be included with the baryon octet.

• **Baryon decuplet:** $J^P = \frac{3}{2}^+$ resonances decay via the strong interactions (e.g. $\Delta \to \pi N$ with appropriate charges, $\Sigma^{0*} \to \Lambda^0\pi^0$ etc.).

$I_3 = -\frac{3}{2}$	$I_3 = -1$	$I_3 = -\frac{1}{2}$	$I_3 = 0$	$I_3 = \frac{1}{2}$	$I_3 = 1$	$I_3 = \frac{3}{2}$	S	$\bar{m}$(MeV)	I
$\Delta^-(ddd)$		$\Delta^0(uud)$		$\Delta^+(duu)$		$\Delta^{++}(uuu)$	0	1232	$\frac{3}{2}$
	$\Sigma^{-*}(dds)$		$\Sigma^{0*}(dus)$		$\Sigma^{+*}(uus)$		-1	1384	1
		$\Xi^{-*}(dss)$		$\Xi^{0*}(uss)$			-2	1533	$\frac{1}{2}$
			$\Omega^-(sss)$				-3	1672	0

• **Pseudo-scalar meson octet:** Lowest lying mesons have spin parity $J^P = 0^-$ and form the pseudo-scalar meson octet. They are stable with respect to strong decays, are long lived and decay weakly or electromagnetically (e.g. the weak decay $\pi^+ \to \mu^+\nu_\mu$, EM decay $\pi^0 \to 2\gamma$, the strangeness changing weak decays $K^+ \to \mu^+\nu_\mu$, $K^0 \to \pi^+\pi^-$). In addition we have the $SU(3)$ flavor singlet pseudo scalar meson $\eta'(u\bar{u} + d\bar{d} + s\bar{s})$ at 958 MeV.

$I_3 = -1$	$I_3 = -\frac{1}{2}$	$I_3 = 0$	$I_3 = \frac{1}{2}$	$I_3 = 1$	S	I	m (MeV)
	K^0		K^+		1	$\frac{1}{2}$	498, 494
	$d\bar{s}$		$u\bar{s}$				
π^-		$\pi^0(d\bar{d} - u\bar{u})$		π^+	0	1	140, 135, 140
$d\bar{u}$		$\eta^0(d\bar{d} + u\bar{u} - 2s\bar{s})$		$u\bar{d}$	0	0	549
	K^-		$\bar{K}^0$				
	$s\bar{u}$		$s\bar{d}$		-1	$\frac{1}{2}$	494, 498

• **Vector meson octet:** $J^P = 1^-$ vector meson resonances. There is also the nearby 1^- singlet $\phi^0(1019)$.

$I_3 = -1$	$I_3 = -\frac{1}{2}$	$I_3 = 0$	$I_3 = \frac{1}{2}$	$I_3 = 1$	S	I	$\bar{m}$ (MeV)
	$K^{*0}(d\bar{s})$		$K^{*+}(u\bar{s})$		1	$\frac{1}{2}$	892
$\rho^-(d\bar{u})$		$\rho^0(d\bar{d} - u\bar{u})$		$\rho^+(u\bar{d})$	0	1	776
		ω^0			0	0	783
	$K^{*-}(s\bar{u})$		$\bar{K}^{*0}(s\bar{d})$		-1	$\frac{1}{2}$	892

• Some features emerge from the observed multiplets of common J^P. Mesons appear in nonets consisting of an octet and a singlet with same J^P and nearby masses. Baryons fall into decuplets, octets and singlets. There does not appear to be a $\frac{1}{2}^+$ baryon of mass comparable to that of the nucleon octet to form a nonet. Gell-Mann and Neeman postulated an $su(3)$ flavor symmetry of hadrons, extending isospin symmetry based on the observation that $su(3)$ has 1,8, and 10 dimensional irreps. The appearance of octets motivated Gell-Mann to call the scheme the eight-fold way. Using $su(3)$ flavor symmetry, the mass and decay properties of the Ω^- particle were predicted. It was discovered soon after, in 1964, with the expected properties lending support to global $su(3)$ flavour symmetry. However, $su(3)$ also has irreps of dimension 3, 6, 15, 24 etc., though there do not appear to be hadrons that transform in any of these representations, this needed an explanation.

• In 1964 Gell-Mann and Zweig independently suggested that the fundamental 3d representation of $su(3)$ is carried by 3 hypothetical quarks, up down and strange of baryon number $1/3$ and that the anti-quarks $\bar{u}, \bar{d}, \bar{s}$ transform in the conjugate $\bar{3}$ irrep. Moreover, mesons were supposed to be made from a quark and an anti-quark while baryons and anti-baryons were made of three quarks and three anti-quarks respectively. Recall that higher dimensional irreps of su(2) can be obtained by decomposing tensor products of the fundamental spin half representation into irreps. Similarly, the possible meson and baryon multiplets could be found by decomposing $3 \otimes \bar{3}$ and $3 \otimes 3 \otimes 3$ into $su(3)$ irreps. We will find that $3 \otimes \bar{3} = 1 \oplus 8$ and $3 \otimes 3 \otimes 3 = 1 \oplus 8 \oplus 8 \oplus 10$, which agree with the observed multiplets of mesons and baryons. This gave a simple explanation for the appearance of singlets and octets of mesons and singlets, octets and decuplets of baryons.

• However, it remained a mystery why particles transforming in other $su(3)$ flavour multiplets (most notably the 3 and $\bar{3}$ quarks and anti-quarks) were not observed among the hadrons. This was eventually resolved by postulating an additional $su(3)$ color quantum number (distinct from the approximate $su(3)$ flavour quantum numbers) for quarks. Quarks were postulated to transform in the fundamental representation of both su(3) flavour and su(3) color, so q_a^α denotes a quark of flavor $a = 1, 2, 3 = u, d, s$ and color $\alpha = 1, 2, 3 = red, green, blue$. So q_1^1 is a red up quark. Color $su(3)$ was postulated to be an exact local gauge symmetry for the quark fields, along with 8 associated su(3) gauge bosons called gluons which transform in the adjoint representation of su(3). We now believe that the dynamics of this color gauge theory QCD conspires to confine quarks and anti-quarks within hadrons, and ensures that the only detected hadrons are color singlets. States that transform non-trivially under color are presumably infinitely massive in isolation, and are expelled from the spectrum of hadrons. Color confinement has not been established in QCD, but is believed to be the reason why many of the other $su(3)$ flavor representations are not realized in the spectrum of hadrons, they correspond to particles that transform non-trivially under color su(3). More on this later.

10.6.1 Remark on Cartan generators vs Casimir operators

• Note that the generators of the Cartan sub-algebra are distinct from Casimir operators. In general, Casimir operators are not elements of the Lie algebra, they are not required to be linear combinations of the basis generators. Casimir operators are (possibly non-linear) quantities that commute with all elements of the Lie algebra, constructed by taking linear combinations of products of elements of the Lie algebra. E.g. $L^2 = L_1^2 + L_2^2 + L_3^2$ is a quadratic Casimir of the angular momentum Lie algebra $[L_i, L_j] = i\hbar\epsilon_{ijk}L_k$, since it commutes with the three generators L_i. The Poincare Lie algebra has two independent Casimirs, the square of momentum $P^\mu P_\mu$ and the square of the Pauli-Lubanski spin vector $W^\mu W_\mu$ where $W^\mu = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}P_\nu M_{\rho\sigma}$. The Lorentz group also has two independent Casimirs J_+^2 and J_-^2 where $\vec{J}_\pm = \vec{L} \pm i\vec{K}$ and $\vec{L}$ and $\vec{K}$ are the angular momentum and boost generators.

• Occasionally an element of a Lie algebra may commute with all elements, then it is a Casimir. Such a Casimir is said to lie in the center of the Lie algebra. The center of a Lie algebra is the sub-algebra of elements that commute with all other elements.

• Casimir operators are a multiple of the identity in an irreducible representation. If a Casimir operator is not a multiple of the identity in a representation, then the representation is reducible: e.g. the representation of orbital angular momentum $L_i = -i\hbar\epsilon_{ijk}r_j\partial_k$ on the space of functions on the sphere $L^2(S^2)$ is reducible, write out the L^2 matrix in the Y_{lm} representation and see that it admits a block decomposition into $(2l+1)$ dimensional blocks where it is the multiple of the identity $\hbar^2 l(l+1)$ for $l = 0, 1, 2, \ldots$. Different irreps may be labeled by the values Casimirs take in the irreps. For example representations of SU(2) are labeled by s which is related to the eigenvalue $\hbar^2 s(s+1)$ of S^2. Representations of the Poincare group are labelled by mass and spin, which are related to the eigenvalues of P^2 and W^2. Representations of the Lorentz group are labeled by a pair of spins s_+, s_- where the eigenvalues of $J_\pm^2$ are $\hbar^2 s_\pm(s_\pm + 1)$.

• For a given irreducible representation the basis vectors of the vector space that carries the representation can be chosen to to be eigenvectors of Cartan generators, since Cartan generators are simultaneously diagonalizable. The basis states can be labeled by the values of the Cartan generators. For example L_3 spans a Cartan subalgebra of su(2) and so its eigenvalues $\hbar m$ can be used to label the (2l+1) linearly independent states of the representation space.

10.6.2 $su(3)$ Lie algebra, Gell Mann matrices, quark basis

• $\mathrm{SU}(3)$ consists of 3×3 unitary matrices of determinant one. The Lie algebra denoted $\underline{\mathrm{SU}(3)}$ or $su(3)$ consists of the 8-dimensional space of traceless anti-hermitian matrices. It is conventional to pull out an i and use the hermitian Gell-Mann matrices λ_a as generators for $su(3)$, in terms of which, the group elements are $U = \exp(-\frac{i}{2}\theta_a\lambda_a)$.

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = 2I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix},$$
$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \sqrt{3}Y = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{344}$$

This is the defining representation of su(3) expressed in the so-called quark basis $x_i = (u, d, s)$.

• Note that flavor SU(3) is only an approximate global symmetry of the strong interactions, we have seen that mass degeneracies within octets and decuplets is violated at the 20% level. In particular, it is a symmetry of the QCD Lagrangian only in the approximation where the u,d, and s quarks have the same mass:

$$\mathcal{L} = \sum_{i=1}^{3} \bar{\psi}_i(i\gamma \cdot D - m_a)\psi_a - \frac{1}{4e^2}\,\mathrm{tr}\,F^2. \tag{345}$$

Here $i = 1, 2, 3$ label the three quark flavors. Under an SU(3) flavor transformation, $\psi_i \to U_{ij}\psi_j$ and $\bar{\psi}_i \to (U^\dagger)_{ij}\bar{\psi}_j$. So under SU(3) the mass term transforms to $\bar{\psi}U^\dagger \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & m_s \end{pmatrix} U\psi$. Only if the mass matrix is a multiple of the identity, can we use $U^\dagger U = I$ to cancel the $U's$ and get an invariant $\mathcal{L}$.

• In addition to the approximate global flavor $su(3)$ we have an exact global $U(1)$ symmetry corresponding to baryon number conservation, which acts in the same way on all the three quarks u, d, s, so $x \to e^{-i\theta BI}x$. $B = 1/3$ for the quarks. $B = -1/3$ for the anti-quarks.

• The third component of isospin $I_3 = \lambda_3/2$ is diagonal with the values $\pm\frac{1}{2}, 0$ for u, d, s. λ_8 is also diagonal and is chosen so that $Y = \lambda_8/\sqrt{3}$ is the hypercharge. $Y = B + S = (\frac{1}{3}, \frac{1}{3}, -\frac{2}{3})$ for u, d, s. λ_3, λ_8 span a maximal abelian subalgebra (a Cartan subalgebra). Check that the charge matrix $Q = I_3 + Y/2$ is diagonal in this basis with entries $2/3, -1/3, -1/3$ corresponding to the charges of u, d, s quarks.

• The factors in $\lambda_3 = 2I_3$ and $\lambda_8 = \sqrt{3}Y$ ensure that the λ_a are normalized according to $\mathrm{tr}\ \frac{1}{2}\lambda_a \frac{1}{2}\lambda_b = \frac{1}{2}\delta_{ab}$. The numerical factors also ensure that the 'weight diagrams' in the $\frac{1}{2}\lambda_3$-$\frac{1}{2}\lambda_8$ plane for the baryon and meson octets are regular hexagons.

• Under an infinitesimal SU(3) transformation, the quark triplet vector x transforms to $x' = Ux \approx (I - \frac{i}{2}\theta_a\lambda_a)x$.

• $su(2)$ may be embedded in $su(3)$ in infinitely many ways. Three of these are $su(2)$ transformations among (1) (u, d) (I-spin, isospin), (2) (d, s) (U-spin) and (3) (u, s) (V-spin). Indeed $I_{1,2,3} = \frac{1}{2}\lambda_{1,2,3}$ give an embedding of I-spin (Pauli matrices) in the 1-2 (ud) subspace. $U_1 = \frac{1}{2}\lambda_6, U_2 = \frac{1}{2}\lambda_7$ and $U_3 = \frac{1}{2}(\frac{3}{2}Y - I_3) = \mathrm{diag}(0, 1, -1)$, give another embedding of U-spin $su(2)$ in the 2-3 (ds) subspace. $V_1 = \frac{1}{2}\lambda_4, V_2 = \frac{1}{2}\lambda_5$ and $V_3 = \frac{1}{2}(\frac{3}{2}Y + I_3) = \mathrm{diag}(1, 0, -1)$ give an embedding of V-spin $su(2)$, in the 1-3 (us) subspace.

• A general $su(3)$ element is conventionally written $u = \frac{1}{2}\sum_a \theta_a\lambda_a$. Just as $I_a = \frac{1}{2}\sigma_a$ we sometimes write $F_a = \frac{1}{2}\lambda_a$.

• Up to anti-symmetry, there are 9 nonzero $su(3)$ structure constants $[\frac{1}{2}\lambda_a, \frac{1}{2}\lambda_b] = if_{abc}\frac{1}{2}\lambda_c$ in this basis:

$$f_{123} = 1, \quad f_{147} = \frac{1}{2}, \quad f_{156} = -\frac{1}{2}, \quad f_{246} = f_{257} = f_{345} = \frac{1}{2}, \quad f_{367} = -\frac{1}{2}, \quad f_{458} = f_{678} = \frac{\sqrt{3}}{2}. \tag{346}$$

• A more convenient basis for studying representations is the Cartan-Weyl basis consisting of Cartan generators I_3, Y and a bunch of raising and lowering operators defined by analogy with $su(2)$:

$$I_\pm = F_1 \pm iF_2, \quad U_\pm = F_6 \pm iF_7, \quad V_\pm = F_4 \pm iF_5. \tag{347}$$

Due to the complex linear combinations, $I_3, Y, I_\pm, U_\pm, V_\pm$ are a basis for the complexification of the $su(3)$ Lie algebra rather than the real $su(3)$ Lie algebra[50]. Real linear combinations of Gell Mann matrices give us the real $su(3)$ Lie algebra. Though the complex combinations do not live in the original Lie algebra, they are very useful to understand its representations, as we know from angular momentum theory.

• Check the commutators $[I_+, I_-] = 2I_3$ and similarly for $U_\pm, V_\pm$. Check that $I_\pm$, $V_\pm$ and $U_\mp$ raise and lower the eigenvalue of I_3 by one, half and half respectively:

$$[I_3, I_\pm] = \pm I_\pm, \quad [I_3, U_\pm] = \mp\frac{1}{2}U_\pm, \quad [I_3, V_\pm] = \pm\frac{1}{2}V_\pm. \tag{348}$$

On the other hand $U_\pm$ and $V_\pm$ raise and lower the eigenvalue of Y by one while $I_\pm$ do not

[50] Show that the complexification of $su(2)$ Lie algebra is the Lie algebra of traceless complex matrices. The complexification of the SU(2) group is the group $SL_2(\mathbb{C})$ of complex matrices with unit determinant.

change it

$$[Y, I_\pm] = 0, \quad [Y, U_\pm] = \pm U_\pm, \quad [Y, V_\pm] = \pm V_\pm. \tag{349}$$

• To get some idea of how the raising and lowering operators act on states of a representation, consider the 8d representation $\mathbf{8}$ on the v.s. $\mathbb{C}^8$ spanned by the 8 baryons of the $\frac{1}{2}^+$ octet N, Σ, Λ, Ξ. This turns out to be the adjoint representation, so the generators of su(3) are represented by the 8×8 matrices with entries $(t_a)_{bc} = -if_{abc}$. Mark the baryons on an $I_3 - Y$ plot. Then we see that $I_\pm$ take Σ^0 to $\Sigma^\pm$ and annihilate the isosinglet Λ^0. On the other hand, V_+ takes Σ^- to n, raising I_3 by half and Y by one. U_- takes Σ^- to Ξ^-, lowering Y by one and raising I_3 by half. V_+ takes Λ to p etc.

• Like the square of total isospin $I^2 = \sum_a I_a^2$ which is a Casimir of $su(2)$, check that $su(3)$ possesses the Casimir operator $F^2 = \sum_a F_a^2 = \frac{1}{2}\{I_+, I_-\} + I_3^2 + \frac{1}{2}\{U_+, U_-\} + \frac{1}{2}\{V_+, V_-\} + F_8^2$, i.e., $[F^2, F_a] = 0$.

• The trace provides an inner product on the $su(3)$ Lie algebra $\langle u, v \rangle = \text{ tr } uv$. λ_a are normalized to $\text{tr } \lambda_a \lambda_b = 2\delta_{ab}$ so that $\text{tr } u^\dagger u = \frac{1}{4}\theta_a^* \theta_b \text{ tr } \lambda_a \lambda_b = \frac{1}{2}\sum_a |\theta_a^2|$.

10.6.3 Representations of $su(N), su(3)$ on spaces of tensors

• The Gell-Mann matrices are a convenient set of generators for su(3) in its defining fundamental $\mathbf{3}$ representation on the vector space $\mathbb{C}^3$. The basis vectors of $\mathbb{C}^3$ are $u = (1,0,0)^t, d = (0,1,0)^t, s = (0,0,1)^t$. The u, d, s quark vector $x_i \in \mathbb{C}$ transforms in the irrep $\mathbf{3}$ ($x'_{i'} = U_{i'i} x_i$ where $U_{ij} \approx \delta_{ij} - \frac{1}{2} i\theta_a (\lambda_a)_{ij}$) while the anti-quarks x^i transform in $\mathbf{3}^*$ ($x'^{j'} = (U^*)^{j'j} x^j$. So λ_a are replaced with $-\lambda_a^*$ in the anti-fundamental representation if we retain the same convention:

$$x'^i \approx \left(\delta^{ij} - \frac{1}{2} i (-\lambda_a^*)^{ij} \theta_a \right) x^j. \tag{350}$$

In particular I_3 and Y being real matrices in the quark basis, simply reverse signs for the anti-quarks. Since in addition $B = -1/3$ for the anti-quarks, it follows that I_3, Y, S and Q all reverse sign for the anti-quarks.

• By taking tensor products of several copies of $\mathbf{3}$, we get *reducible* representations of $su(3)$ on higher rank tensors denoted $x_{i_1 \cdots i_n}$ which transform as $x_{i_1} \cdots x_{i_n}$:

$$x'_{i'_1 \cdots i'_n} = U_{i'_1 i_1} U_{i'_2 i_2} \cdots U_{i'_n i_n} x_{i_1 \cdots i_n} \tag{351}$$

The space of such tensors in general admits invariant subspaces allowing us to decompose these tensor product representations as direct sums of irreducible representations. It is conventional to denote irreps by their dimension in bold face with additional symbols to distinguish between inequivalent representations of the same dimension. E.g. $\mathbf{3} \otimes \mathbf{3} = \mathbf{6} \oplus \mathbf{3}^*$ and $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{10}$.

• There is a connection between the representations of $su(N)$ (in fact $gl(N)$) and those of the permutation group S_n, called Schur-Weyl duality. Tensors can be decomposed as a sum of those belonging to definite symmetry classes under permutation of indices (symmetric & anti-symmetric in a manner to be specified). It turns out that tensors of a given symmetry class span an invariant irreducible subspace and thus carry an irrep of $su(3)$. The fundamental representation is carried by vectors x_i. As a more interesting example, consider totally anti-symmetric 3rd rank tensors x_{ijk}. Being totally antisymmetric, $x_{ijk} = \lambda \epsilon_{ijk}$ is a multiple of the Levi-Civita symbol. Moreover under $su(3)$, it transforms to $x'_{i'j'k'} = \lambda U_{i'i} U_{j'j} U_{k'k} \epsilon_{ijk} = \lambda \det U \epsilon_{ijk} = x_{ijk}$ since $\det U = 1$. Thus totally anti-symmetric 3rd rank tensors are invariant under SU(3) and thus carry the trivial representation $\mathbf{1}$. All the finite dimensional irreps of $su(3)$ may be obtained in this way by decomposing tensor products of the fundamental representation.

• To get some idea of how permutations relate to representations of su(N) or gl(N), we note that the permutation group acts on tensors by permuting indices (e.g. $P_{12} x_{ijk} = x_{jik}$). su(N) also acts on tensors as given above. These two actions commute, in other words we may permute indices and then make a change of basis or change basis and then permute indices, the resulting tensor is the same. For example, consider

$$P_{12} U_{ki} U_{lj} x_{ij} = U_{li} U_{kj} x_{ij} \quad \text{while} \quad U_{ki} U_{lj} P_{12} x_{ij} = U_{ki} U_{lj} x_{ji}. \tag{352}$$

If we relabel $i \leftrightarrow j$ then we see that the two expressions are the same. Now the 'eigentensors' of P_{12} are tensors that are either symmetric or anti-symmetric in the first pair of indices. So an eigentensor of permutations remains an eigentensor under permutations after the action of SU(N). So tensors with definite behavior under permutation of indices retain that behavior after being transformed by SU(N). This makes plausible why tensors of a definite symmetry class can be an invariant subspace for the action of SU(N), in fact they carry irreducible representations.

• Spaces of tensors also carry representations of the more familiar rotation group SO(3). We recall from the study of orbital angular momentum in QM that the irreps of SO(3) are labeled by $l = 0, 1, 2, 3, \ldots$ and are of dimension $2l + 1$. The angular momentum l representation is in fact carried by the space of real symmetric tranceless tensors of rank l. Indeed, the space of all Cartesian tensors on $\mathbb{R}^3$ of rank l is 3^l, since each index can take three possible values, and the different tensors components are independent. Now the symmetric tensors among these span a space of dimension $(l + 2)(l + 1)/2$. To count these we may enumerate symmetric tensors by saying how many of the n indices are 1's, how many are 2's and how many are 3's. To find this number, we imagine placing all the 1's first followed by the 2's and then then 3's, like in $111|2|333$. Assuming there is at least one two, this is equivalent to the problem of inserting two vertical dividers in $l + 1$ possible places, there are $\binom{l+1}{2}$ ways of doing this. To this we must add the cases where there are no 2's. This corresponds to placing both the vertical dividers in the same place, there are $l + 1$ ways of doing this. Thus the space of symmetric rank l real tensors has dimension $\binom{l+1}{2} + \binom{l+1}{1} = \binom{l+2}{2} = (l + 1)(l + 2)/2$. Now we restrict further to traceless symmetric rank l tensors, by imposing the conditions $x_{iii_3 \cdots i_l} = 0$. There are as many conditions as there are symmetric rank $l - 2$ tensors, i.e., $\binom{l}{2}$ conditions. Thus the space of rank l symmetric traceless tensors has dimension $(l+2)(l+1)/2 - l(l-1)/2 = 2l+1$. Notice that this is the dimension of the angular momentum l irrep of SO(3).

10.6.4 Young tableaux and symmetry classes of tensors

• Young diagrams (tableaux) give a convenient way of labelling tensors of a given symmetry class and thus give a way of labelling irreps of $su(3)$, and more generally, of $su(N)$.

• The table with no boxes is a tensor with no indices, i.e. a scalar, which carries the trivial 1d representation of $su(N)$.

• For example, a symmetric tensor $s_{ijk\cdots n}$ is denoted by a row of as many boxes as there are indices $\begin{array}{|c|c|c|c|c|}\hline i & j & k & \cdot\cdot & n \\ \hline\end{array}$. A single box corresponds to a vector x_i. A totally anti-symmetric tensor $a_{ij\cdots n}$ is denoted by a column of as many boxes $\begin{array}{|c|}\hline i \\ \hline j \\ \hline \cdot \\ \hline n \\ \hline\end{array}$. More general Young tables stand for tensors with mixed symmetry, symmetric in some indices and anti-symmetric in others.

• We will not derive all the above statements about representations of $su(3)$ here but will learn to work with Young Tableaux through examples. Our presentation is based on Kerson Huang's book Quarks Leptons and Gauge Fields. We first define Young Tableaux for $su(N)$ and then specialise to $su(3)$.

• Tensors of a definite symmetry class are obtained from $x_{i_1\cdots i_n}$ through the following process. We fix a partition $n = n_1 + n_2 + \cdots + n_k$ of n with summands arranged in decreasing order $(n_1 \geq n_2 \geq \cdots \geq n_k > 0)$. In particular $k \leq N$.

(1) We pick n_1 indices from $i_1, \cdots, i_n$ and symmetrize among them. This operation of symmetrization is displayed via a row of n_1 boxes, with one index per box. E.g. if $n_1 = 2$ it is written as $\begin{array}{|c|c|}\hline & \\ \hline\end{array}$ or with indices filled in as $\begin{array}{|c|c|}\hline i & j \\ \hline\end{array} = \begin{array}{|c|c|}\hline j & i \\ \hline\end{array}$. If $n = n_1 = 2$ then the corresponding tensor is just $\frac{1}{2}(x_{ij} + x_{ji})$.

(2) Then we pick n_2 of the remaining indices, symmetrize among them and display them in a row of n_2 boxes. The process is repeated for the remaining indices. The k rows of $n_1, n_2, \cdots n_k$ boxes are stacked one below the other to form a Young table. E.g. if $n = 5, n_1 = 2, n_2 = 2, n_3 = 1$, then we get $\begin{array}{|c|c|}\hline & \\ \hline & \\ \hline \\ \cline{1-1}\end{array}$. The lengths of rows is non-increasing as we go down the diagram.

(3) Finally, we anti-symmetrize among the indices in each column of the table (by anti-symmetrization we mean an average over all permutations weighted by the sign of the permutation).

• This procedure results in a tensor of a definite symmetry class, which is symbolically represented by the corresponding Young diagram. Young diagrams with n boxes are in 1-1 correspondence with partitions of n. Thus there are as many symmetry classes of rank-n tensors as there are partitions of n. As mentioned before, tensors of a given symmetry class carry an irrep of $su(N)$. Note that there is no restriction on the tensor rank n in comparison to the rank $N - 1$ of $su(N)$.

• A Young diagram with boxes left empty represents a symmetry class of tensors. When the boxes are filled with indices, then it represents a specific tensor of that symmetry class, constructed from $x_{i_1\cdots i_n}$.

• E.g. $\begin{array}{|c|c|}\hline i & j \\ \hline\end{array} = \frac{1}{2}(x_{ij} + x_{ji})$ are symmetric rank two tensors.

• E.g. $\begin{array}{|c|}\hline i \\ \hline j \\ \hline\end{array} = \frac{1}{2}(x_{ij} - x_{ji})$ are anti-symmetric rank two tensors.

• E.g. $\begin{array}{|c|}\hline i \\ \hline j \\ \hline k \\ \hline\end{array} = \frac{1}{6}(x_{ijk} - x_{jik} - x_{ikj} - x_{kji} + x_{jki} + x_{kij})$ is an anti-symmetric rank three tensor. We sum over permutations weighted by the sign of the permutation and divide by the number of permutations, i.e., the order of the permutation group S_3 in this case. In the case of su(3), this tensor is a multiple of the Levi-Civita tensor ϵ_{ijk}. For su(2) this tensor is simply zero since each index can take only two values.

• E.g. $\begin{array}{|c|c|}\hline i & j \\ \hline k \\ \cline{1-1}\end{array} = \frac{1}{2}\left[\frac{1}{2}(x_{ijk}+x_{jik}) - \frac{1}{2}(x_{kji}+x_{jki})\right]$. We first symmetrized in i,j and then antisymmetrize the resulting tensor in k,i. The resulting tensor is anti-symmetric under $k \leftrightarrow i$ but is not symmetric under $i \leftrightarrow j$.

• Remarks: Indices in any given column of a table are anti-symmetric under permutation among themselves, e.g., $\begin{array}{|c|c|}\hline i & j \\ \hline k \\ \cline{1-1}\end{array} = -\begin{array}{|c|c|}\hline k & j \\ \hline i \\ \cline{1-1}\end{array}$. Indices in a row that do not contain any box underneath are symmetric among themselves. E.g. $\begin{array}{|c|c|c|}\hline i & j & k \\ \hline l \\ \cline{1-1}\end{array} = \begin{array}{|c|c|c|}\hline i & k & j \\ \hline l \\ \cline{1-1}\end{array}$. But if an index in a row has been anti-symmetrized with other indices below, then it does not have any particular symmetry property under permutation with other indices in the same row, so $\begin{array}{|c|c|}\hline i & j \\ \hline k \\ \cline{1-1}\end{array}$ and $\begin{array}{|c|c|}\hline j & i \\ \hline k \\ \cline{1-1}\end{array}$ are independent tensors.

10.6.5 Young tableau and irreducible representations of $su(3)$

• Let us now specialize to $su(3)$ where tensor indices can only take three values $1, 2, 3$. Since indices in a column of a Young table are anti-symmetrized, columns cannot have length more than three. What is more, a column with 3 boxes, which stands for a totally anti-symmetric tensor

$$\begin{array}{|c|}\hline i \\ \hline j \\ \hline k \\ \hline\end{array} = \frac{1}{|S_3|}\sum_{\sigma \in S_3} (-1)^{\text{ sgn } \sigma} x_{\sigma(i)\sigma(j)\sigma(k)} = \frac{1}{6}\left(x_{ijk}+x_{kij}+x_{jki}-x_{ikj}-x_{jik}-x_{kji}\right) \qquad (353)$$

in fact carries the trivial representation of $su(3)$, i.e., is unchanged under the group action, as we argued above. A tensor of rank zero (no indices, zero boxes) also carries the trivial 1d representation denoted $\mathbf{1}$. We write $\begin{array}{|c|}\hline \; \\ \hline \; \\ \hline \; \\ \hline\end{array} = \mathbf{1}$. In fact any column of three boxes can be omitted from a Young table, it corresponds to a scalar. So an $su(3)$ Young diagram for a non-trivial irrep can have at most two rows, with $m+n$ and m boxes respectively. The corresponding irrep is denoted (m,n). $(0,0)$ denotes the trivial representation. Thus, the general Young table for $su(3)$ is of the form

$$\begin{array}{|c|c|c|c|c|c|c|c|}\hline k_1 & k_2 & \cdot\cdot & k_m & i_1 & i_2 & \cdot\cdot & i_n \\ \hline l_1 & l_2 & \cdot\cdot & l_m \\ \cline{1-4}\end{array} = x_{i_1 \cdots i_n; k_1 \cdots k_m; l_1 \cdots l_m} \qquad (354)$$

Since it is anti-symmetric in each pair (k_λ, l_λ), it is convenient to contract with $\epsilon^{j_\lambda k_\lambda l_\lambda}$ and replace each (k_λ, l_λ) pair by a single upper index j_λ. Thus, the Young table for the (m,n) irrep corresponds to the tensor

$$X^{j_1 \cdots j_m}_{i_1 \cdots i_n} = \epsilon^{j_1 k_1 l_1} \cdots \epsilon^{j_m k_m l_m} x_{i_1 \cdots i_n; k_1 \cdots k_m; l_1 \cdots l_m}. \qquad (355)$$

The resulting X^J_I tensor is symmetric in the upper j-type indices and separately symmetric in the lower i-type indices. Furthermore, the tensor is traceless in the sense that contracting an upper with a lower index gives zero: $\sum_j X^{ij_2 \cdots j_m}_{ii_2 \cdots i_n} = 0$. To see why, we go back to the definition

$$\sum_j X^{i_1 j_2 \cdots j_m}_{i_1 i_2 \cdots i_n} = \epsilon^{i_1 k_1 l_1} \epsilon^{j_2 k_2 l_2} \cdots \epsilon^{j_n k_n l_n} x_{i_1 i_2 \cdots i_m; k_1 k_2 \cdots k_n; l_1 l_2 \cdots l_n} \qquad (356)$$

It will help to look at the simplest case $m = n = 1$, the general case is similar:

$$X_i^i = \epsilon^{ikl} x_{i;k;l} = \frac{1}{4}\epsilon^{ikl}[(x_{ikl} + x_{kil}) - (x_{ilk} + x_{lik})]. \tag{357}$$

We have written out the symmetrization in the indices of the first row of the table followed by the anti-symmetrization in the indices of the first column. Now ϵ^{ikl} is anti-symmetric in both $i \leftrightarrow k$ and $i \leftrightarrow l$. The first two terms in parantheses are symmetric under $i \leftrightarrow k$ and the next two are symmetric under $i \leftrightarrow l$. Thus $X_i^i = 0$.

• By assumption, the lower indices on x_I^J transform in the fundamental representation of $su(3)$. Remarkably, the upper j-'type' indices on x_I^J obtained by contracting $x_{I;K;L}$ with ϵ transform in the conjugate of the fundamental representation! For instance, this means the anti-symmetric second rank tensors carry the $\mathbf{3}^*$ representation. To see this it suffices to consider one upper index $X^j = \epsilon^{jkl} x_{kl}$. We know that this transforms to

$$X'^j = \epsilon^{jkl} U_{kk'} U_{ll'} x_{k'l'} \tag{358}$$

Our claim is that this is the same as

$$X'^j = (U^*)^{j\tilde{j}} X^{\tilde{j}} = (U^*)^{j\tilde{j}} \epsilon^{\tilde{j}k'l'} x_{k'l'}. \tag{359}$$

To show that these two are the same, it suffices (since $x_{k'l'}$ is arbitrary) to show that

$$\epsilon^{jkl} U_{kk'} U_{ll'} \stackrel{?}{=} (U^*)^{j\tilde{j}} \epsilon^{\tilde{j}k'l'} \tag{360}$$

Now multiply by $U_{jj'}$ on either side and sum over j. This produces an equivalent condition, since U is an invertible matrix. So it suffices to show that

$$\begin{aligned} \epsilon^{jkl} U_{jj'} U_{kk'} U_{ll'} &\stackrel{?}{=} (U^*)^{j\tilde{j}} U_{jj'} \epsilon^{\tilde{j}k'l'} \\ \Leftrightarrow \quad \det U \, \epsilon_{j'k'l'} &\stackrel{?}{=} (U^\dagger)^{\tilde{j}j} U_{jj'} \epsilon^{\tilde{j}k'l'} \\ \Leftrightarrow \quad \epsilon_{j'k'l'} &\stackrel{?}{=} \delta_{\tilde{j}j'} \epsilon^{\tilde{j}k'l'} \end{aligned} \tag{361}$$

which is an identity! We use the fact that $U^\dagger = U^{*t}$ and that $U^\dagger U = I$. Thus we have shown that an anti-symmetric pair of lower (quark, fundamental indices) when contracted with ϵ and converted to an upper index, transforms in the conjugate of the fundamental representation. So in the irreducible tensor X_I^J, the J are anti-quark indices and I are quark indices. From here on we will always raise anti-symmetric pairs of lower indices and rename X_I^J as x_I^J.

10.6.6 Dimension of irrep (m, n)

• The linear space of (m, n)-tensors $x_{i_i \cdots i_n}^{j_1 \cdots j_m}$ (denoted by the Young table with $m + n$ boxes in the first row and m boxes in the second row) carry the (m, n) irrep of $su(3)$. These components aren't all independent since they are symmetric in the j's and i's separately and also traceless. The number of independent tensors $D(m, n)$ is the dimension of the irrep, which we calculate here.

• E.g. Consider first symmetric rank n tensors, corresponding to the table $\boxed{1}\boxed{2}\boxed{3}\boxed{\cdot}\boxed{\cdot}\boxed{n}$. We have already argued (or see below) that the space of symmetric rank n tensors has dimension $\binom{n+2}{2}$. Thus we must have $D(0,n) = \frac{1}{2}(n+2)(n+1)$.

• More generally, the number of symmetric strings $j_1 \cdots j_m$ is the number of ways of choosing σ_1 1's σ_2 2's and σ_3 3's with $\sigma_1 + \sigma_2 + \sigma_3 = m$ and $\sigma_i \geq 0$. This number is $P_m = \binom{m+1}{2} + \binom{m+1}{1} = \frac{1}{2}(m+1)(m+2)$[51]. Thus the number of ways of choosing J and I is $P_m P_n$. But these aren't all independent due to the trace condition, which imposes $P_{m-1}P_{n-1}$ constraints. Thus

$$D(m,n) = P_m P_n - P_{m-1}P_{n-1} = \frac{1}{2}(m+1)(n+1)(m+n+2). \tag{362}$$

The i and j indices on the tensors x_I^J in the (m,n) representation transform respectively in the fundamental and anti-fundamental representations

$$x'^{j_1'\cdots j_m'}_{i_1'\cdots i_n'} = U^*_{j_1'j_1}U^*_{j_2'j_2}\cdots U^*_{j_m'j_m}\, U_{i_1'i_1}U_{i_2'i_2}\cdots U_{i_n'i_n} x^{j_1\cdots j_m}_{i_1\cdots i_n}. \tag{363}$$

From this it is clear that $(n,m) \equiv (m,n)^*$ is the conjugate representation to (m,n) transforming via the complex conjugate matrices. (n,n) is self-conjugate and has dimension $(n+1)^3$.

Weights by hooks formula for $D(m,n)$:

• There is another way of writing the formula for the dimension $D(m,n)$ of the (m,n) irrep of $su(3)$, which we state here. It expresses $D(m,n)$ as the quotient of the product of 'weights' and the product of 'hooks'. The weights are numbers that we put in the boxes of the Young table, starting with $N=3$ in the LH top corner box and increasing by one to the right. In the second row, the weights start with $N-1=2$ and increase by one in successive boxes to the right. E.g. the weights for the $\mathbf{8}$ irrep are $\begin{smallmatrix}\boxed{3}\boxed{4}\\ \boxed{2}\phantom{\boxed{4}}\end{smallmatrix}$. In general, for a Young table with $m+n$ boxes in the top row and m boxes in the second row, the weights in the first row are $3, 4, \cdots, (m+n+2)$ and the weights in the second row are $2, 3, \cdots, m+1$.

3	4	··	··	··	··	··	$m+n+2$
2	3	··	$m+1$				

(364)

So the product of weights is

$$\prod \text{weights} = \frac{1}{2}(m+n+2)!(m+1)! \tag{365}$$

The hook lengths are numbers placed in each box of the Young diagram. The hook length of a box is the number of boxes to its right plus the number of boxes below it plus one. E.g., the hooks for the $\mathbf{8}$ irrep are $\begin{smallmatrix}\boxed{3}\boxed{1}\\ \boxed{1}\phantom{\boxed{1}}\end{smallmatrix}$. In general the hooks in the lower row beginning from the left

[51] Imagine arranging $\sigma_1 \geq 0$ 1's followed by a $|$ followed by $\sigma_2 \geq 0$ 2's followed by a $|$ followed by $\sigma_3 \geq 0$ 3's where $\sigma_1 + \sigma_2 + \sigma_3 = m$. E.g. $11|2|333$, $|22|3$, $11|33$, $1|2|$ etc. The number of such arrangements with $\sigma_2 > 0$ is the number of ways of placing two $|$'s in $m+1$ possible slots, i.e., $\binom{m+1}{2}$. To this we must ass the possibilities with $\sigma_2 = 0$. For the latter we need only insert one $|$ in one of $m+1$ locations, which may be done in $\binom{m+1}{1}$ ways. Adding we get $P_m = \binom{m+1}{2} + \binom{m+1}{1}$.

are $m, m-1, \cdots, 1$. The hooks in the upper row beginning from the left are $m+n+1, m+n, \cdots, n+2; n, n-1, \cdots, 1$.

$$\begin{array}{|c|c|c|c|c|c|c|c|}\hline m+n+1 & \cdot & \cdot & n+2 & n & \cdot & 2 & 1 \\ \hline m & m-1 & \cdot & 2 & 1 \\ \cline{1-5}\end{array} \tag{366}$$

Thus the product of hook lengths is

$$\prod \text{hooks} = \frac{m!n!(m+n+1)!}{(n+1)!} = \frac{m!(m+n+1)!}{(n+1)}. \tag{367}$$

It is now easy to check that the ratio of product of weights to product of hooks is equal to the previously computed dimension $D(m, n)$

$$\frac{\prod \text{weights}}{\prod \text{hooks}} = \frac{1}{2}(m+n+2)(m+1)(n+1) = D(m, n). \tag{368}$$

10.6.7 Examples of low-dimensional irreducible reps of $su(3)$

• We summarise some results about $su(3)$ irreps here. The (m, n) irrep (integers $m, n \geq 0$) has a Young table with first row with $m+n$ boxes and second row with m boxes and corresponds to a tensor $x_{i_1 \cdots i_n}^{j_1 \cdots j_m}$ with m upper (anti-quark) indices and n lower (quark) indices transforming via $U = e^{-\frac{1}{2} i \theta_a \lambda_a}$ and $U^* = e^{\frac{1}{2} i \theta_a \lambda_a^*}$ respectively. Dimension of the (m, n) irrep is $D(m, n) = \frac{1}{2}(n+1)(m+1)(n+m+2)$. (m, n) is inequivalent to (m', n') if $n \neq n'$ or $m \neq m'$. Value of the Casimir in the (m, n) irrep is $F^2 = \frac{1}{3}(m^2 + mn + n^2) + (m+n)$. Particles transforming in the irrep (m, n) can have baryon number $B = (n-m)/3$ since quarks and anti-quarks have baryon number $\pm 1/3$. We will see that they may also have certain other baryon numbers $B = (n-m)/3 + j$ for some integer integer j.

• Examples: $(0, 0)$ is the trivial 1d representation $\mathbf{1}$ carried by the space of tensors with no indices i.e, scalar multiples of 1. It may be represented by the Young tableau with no boxes. As mentioned before, totally anti-symmetric 3rd rank tensors $\lambda \epsilon_{ijk}$ also carry the trivial representation and are labelled by $\begin{array}{|c|}\hline \\ \hline \\ \hline \\ \hline\end{array}$.

• $(0, 1) = \square$ is the defining or fundamental 3d (quark) representation $\mathbf{3}$ carried by x_i. $(1, 0) = \mathbf{3}^* = \begin{array}{|c|}\hline \\ \hline \\ \hline\end{array}$ is the anti-fundamental representation carried by x^i (anti-quarks).

• $(1, 1) = \mathbf{8} = \mathbf{8}^* = \begin{array}{|c|c|}\hline & \\ \hline \\ \cline{1-1}\end{array}$ is the self-conjugate octet carried by a traceless tensor x_i^j with one quark and one anti-quark index (the pseudo scalar pion octet or vector meson ρ octet). The octet is also equivalent to the adjoint representation of $su(3)$, where the matrix elements are the same as the structure constants. How is this related to the baryon (nucleon) octet?

• Recall that x_i^j was obtained by contracting the original tensor $\begin{array}{|c|c|}\hline & \\ \hline \\ \cline{1-1}\end{array} = x_{i;k;l}$ by ϵ^{jkl}, $x_i^j = \epsilon^{jkl} x_{i;k;l}$. In general, the totally anti-symmetric ϵ_{ijk} can be used to lower anti-quark indices and turn them into a pair of anti-symmetric quark indices $A_{ij} = \epsilon_{ijk} x^k$ with A_{ij} transforming as $x_i x_j$. Thus we may view the upper anti-quark index in the x_i^j as an anti-symmetric pair of quark indices. By this process we may identify the pion octet with the baryon octet, they behave in the same way under $su(3)$. However, the process of replacing an anti-quark index with an anti-symmetric pair of quark indices increases the baryon number of the states by one.

• Similarly, one may replace a quark index by an anti-symmetric pair of anti-quark indices by raising using ϵ^{ikl}. In this manner one obtains the tensors carrying the anti-nucleon octet from those of the pion octet $x_i^j \epsilon^{ikl}$. As far as $su(3)$ is concerned, the pseudo scalar mesons, vector mesons, nucleons and anti-nucleons all transform in the octet $\mathbf{8}$.

• The possibility to replace a quark index with a pair of anti-quark indices and an anti-quark index with a pair of quark indices means that that (m, n) could correspond to particle multiplets with various baryon numbers, all congruent to $(n - m)/3$ modulo 1.

• $(0, 3) = \mathbf{10}$ denoted by □□□ consists of symmetric tensors with 3 quark indices x_{ijk} while $(3, 0) = \mathbf{10}^* =$ □□□
□□□ corresponds to symmetric tensors with 3 anti-quark indices x^{ijk}. The Δ baryon decuplet and the $\bar{\Delta}$ anti-baryon decuplet transform in these inequivalent irreps. More examples of $su(3)$ irreps are given in the table.

dim and irrep	(m, n)	Tableau $\frac{m+n}{m}$ boxes	Tensor $x^{j_1 \cdots j_m}_{i_1 \cdots i_n}$, particles	F^2	$B\,(\mathrm{mod}\,1)$
1=trivial	(0,0)	□ □ □	1, singlet	0	0
3=fundamental	(0,1)	□	x_i, colored quarks	4/3	$\frac{1}{3}$
$\bar{3}$ =anti-fund	(1,0)	□ □	x^i, colored anti-quarks	4/3	$-\frac{1}{3}$
6	(0,2)	□□	x_{ij}, colored diquark	10/3	$\frac{2}{3}$
$\bar{6}$	(2,0)	□□ □□	x^{ij}, colored diantiquark	10/3	$-\frac{2}{3}$
$8 = \bar{8}$= adjoint	(1,1)	□□ □	$x^i_j, x^i_i = 0$, pion octet	3	0
10	(0,3)	□□□	x_{ijk}, $\frac{3}{2}^+$ baryon decuplet	6	1
$\bar{10}$	(3,0)	□□□ □□□	x^{ijk} , $\frac{3}{2}^+$ $B = -1$ decuplet	6	-1
15	(1,2)	□□□ □	$x^i_{jk}, x^i_{ik} = 0$	16/3	$\frac{1}{3}$
$\bar{15}$	(2,1)	□□□ □□	$x^{ij}_k, x^{ij}_i = 0$	16/3	$-\frac{1}{3}$
$15'$	(0,4)	□□□□	x_{ijkl}	28/3	$\frac{4}{3}$
$\bar{15}'$	(4,0)	□□□□ □□□□	x^{ijkl}	28/3	$-\frac{4}{3}$
24	(1,3)	□□□□ □	$x^i_{jkl}, x^i_{ikl} = 0$	25/3	$\frac{2}{3}$
$\bar{24}$	(3,1)	□□□□ □□□	$x^{ijk}_l, x^{ijk}_i = 0$	25/3	$-\frac{2}{3}$
$27 = \bar{27}$	(2,2)	□□□□ □□	$x^{ij}_{kl}, x^{ij}_{il} = 0$	8	0

• We notice that the $su(3)$ irreps with non-integer baryon number are not realized in the spectrum of hadrons. This is not explained by flavor $su(3)$ symmetry but by the hypothesis that color is confined within hadrons.

10.6.8 Weight diagrams for $su(3)$ irreps

• The generators of the $su(3)$ Lie algebra may be divided into Cartan generators λ_3, λ_8 which span a maximal abelian subalgebra and the raising and lowering operators $I_\pm, U_\pm, V_\pm$.

• The Cartan generators $\frac{1}{2}\lambda_3, \frac{\sqrt{3}}{2}\lambda_8$ are diagonal both in the fundamental representation $\mathbf{3}$ and also in any irrep (m, n). Thus each state of the irrep (m, n) is characterized by a pair of eigenvalues (say I_3, Y) of the Cartan generators. A weight diagram displays the $D(m, n)$ states of an irreducible $su(3)$ multiplet as points on the $I_3 - Y$ plane. The weight diagram for $\mathbf{3} = \square$ (carried by x_i) is an isoceles triangle with vertex (s) 'pointing downwards', the

3 vertices are $u(1/2, 1/3), d(-1/2, 1/3), s(0, -2/3)$. The triangle is not equilateral since I_3 and Y (unlike λ_3 and λ_8) do not have the same length with respect to the Cartan metric. The weight diagram for $\mathbf{3}^*$ carried by x^i is an isoceles triangle with vertex 'pointing upwards'. To obtain weight diagrams for other representations, it is convenient to define a 'diamond' lattice on the I_3-Y plane whose sites are integer linear combinations of the three lattice vectors $\mathbf{x}_1 = (1/2, 1/3), \mathbf{x}_2 = (-1/2, 1/3), \mathbf{x}_3 = (0, -2/3)$ which are the position vectors of u, d, s in the weight diagram of $\mathbf{3}$. Then the weight diagram of the (m, n) irrep carried by $x_{i_1 \cdots i_n}^{j_1 \cdots j_m}$ consists of $D(m, n)$ lattice sites, one each for the $D(m, n)$ linearly independent tensor components of the representation. Each linearly independent tensor component $x_{i_1 \cdots i_n}^{j_1 \cdots j_m}$ is assigned the lattice vector $\mathbf{x}_{i_1} + \cdots + \mathbf{x}_{i_n} - (\mathbf{x}_{j_1} + \cdots + \mathbf{x}_{j_m})$. Several tensor components may occupy the same site in a weight diagram. E.g. the origin of the hexagonal weight diagram of the octet $\mathbf{8}$ is doubly occupied. Moreover, the weight diagram is purely an $su(3)$ property, it does not know about baryon number or spin. So the nucleon, anti-nucleon and pseudoscalar meson and vector meson octets have the same weight diagrams.

• E.g. Weight diagram of $\mathbf{8}$ corresponding to traceless tensors x_i^j. There are 7 distinct lattice sites for the 8 independent states, forming a hexagon with center doubly occupied. x_1^1, x_2^2, x_3^3 all correspond to the zero weight vector $\mathbf{x}_1 - \mathbf{x}_1 = \mathbf{x}_2 - \mathbf{x}_2 = \mathbf{x}_3 - \mathbf{x}_3 = \vec{0}$. By $x_i^i = 0$ only two of these are independent and it is conventional to choose the combinations $\frac{1}{\sqrt{2}}(x_1^1 - x_2^2)$ and $\frac{1}{\sqrt{2}}(x_1^1 + x_2^2)$ since these transform under isospin as an isosinglet and in an isotriplet with x_2^1 and x_1^2. The lattice sites for the remaining 6 tensor components are

$$x_2^1 \to \mathbf{x}_2 - \mathbf{x}_1, \quad x_1^2 \to \mathbf{x}_1 - \mathbf{x}_2, \quad x_3^2 \to \mathbf{x}_3 - \mathbf{x}_2, \quad x_2^3 \to \mathbf{x}_2 - \mathbf{x}_3, \quad x_1^3 \to \mathbf{x}_1 - \mathbf{x}_3, \quad x_3^1 \to \mathbf{x}_3 - \mathbf{x}_1, \tag{369}$$

Draw the weight diagram for $\mathbf{8}$ and also for $\mathbf{10}$ and $\mathbf{10}^*$.

• Recall from the $su(3)$ commutation relations that $I_\pm, V_\pm, U_\mp$ raise and lower the weight I_3 by one, half and half respectively. $I_\pm$ leave Y unchanged while $U_\pm, V_\pm$ raise and lower Y by one. Thus the raising and lowering operators annihilate states with maximal or minimal values of I_3 and Y. Such highest or lowest weight states (also known as 'maximally stretched states') lie on the periphery of the weight diagram. In the case of $\mathbf{8}$, each of the 6 states on the periphery of the hexagon are annihilated by one or more of the raising or lowering operators.

10.6.9 Decomposition of tensor products of $su(3)$ representations

• Just as we have rules for decomposing tensor products of $su(2)$ representations as direct sums of irreps, e.g., $j \otimes j' = |j + j'| \oplus |j + j' - 1| \oplus \cdots \oplus |j - j'|$ there are rules for decomposing tensor products of $su(3)$ irreps. These rules are stated in terms of the Young tableaux (see Georgi, Lie algebras in particle physics or other group theory books). Suppose we take the tensor product of two representations A, B with Young tableaux A and B. $A \otimes B$ is the direct sum of several irreps. Of course the product of the dimensions of A and B must equal the sum of the dimensions of the irreps appearing in the decomposition. Moreover, the irreps appearing in the decomposition $A \otimes B$ are the same as those appearing in $B \otimes A$. However, it is easier to find the decomposition when the second factor has a simpler Young table compared to the first, using the following rules. The simplest cases of $A \otimes B$ are when $B = \square$ is the fundamental.

Then the rule is simply to affix a box in all possible ways to A to arrive at a sum of legal Young tables. Let us look at some examples.

• $\mathbf{3}\otimes\mathbf{3} = \yng(1)\otimes\yng(1) = \yng(2)\oplus\yng(1,1)$. Thus $\mathbf{3}\otimes\mathbf{3} = \mathbf{6}\oplus\mathbf{3}^*$. The rule here is that we affix the box from the second factor in all possible ways onto the Young tableau of the first representation to arrive at a legal Young table. In terms of tensors, this decomposes $x_i x_j$ as the sum of its symmetric and anti-symmetric part and identifies the anti-symmetric part A_{ij} with the anti-fundamental representation $\mathbf{3}^*$ by contracting with ϵ^{kij}. It follows that we do not 'need' the anti-fundamental representation, it appears as a summand in the decomposition of tensor products of $\mathbf{3}$. In fact, all irreps arise as summands in the decomposition of tensor products of copies of the fundamental $\mathbf{3}$. $\mathbf{6}$ does not appear in the spectrum of hadrons, it has fractional baryon number.

• $\mathbf{3}^*\otimes\mathbf{3} = \yng(1,1)\otimes\yng(1) = \yng(2,1)\oplus\yng(1,1,1). = \mathbf{8}\oplus\mathbf{1}$. The dimensions are 9 on either side. In other words, by combining a quark and an anti-quark we get an octet and a singlet. This is realised in the pion octet along with nearby singlet η' and the vector meson octet with nearby singlet ϕ.

• $\mathbf{6}\otimes\mathbf{3} = \yng(2)\otimes\yng(1) = \yng(3)\oplus\yng(2,1) = \mathbf{10}\oplus\mathbf{8}$. Soon we will use this to 'make' baryons from a product of three copies of $\mathbf{3}$ using the distributivity of tensor product over direct sum. The decomposition $\mathbf{6}\otimes\mathbf{3} = \mathbf{10}\oplus\mathbf{8}$ corresponds to the decomposition of the product $x_{ij}y_k$ of a symmetric tensor $(\mathbf{6})$ and a vector $(\mathbf{3})$ as the sum of a symmetric 3rd rank tensor $(\mathbf{10})$ and a tensor carrying $\mathbf{8}$, via the identity

$$x_{ij}y_k = \frac{1}{3}\left(x_{ij}y_k + x_{ik}y_j + x_{kj}y_i\right) + \frac{1}{3}\left[(x_{ij}y_k - x_{kj}y_i) + (x_{ij}y_k - x_{ki}y_j)\right] \quad (370)$$

The first term is clearly a symmetric 3rd rank tensor (note, x_{ij} is already symmetric), say u_{ijk} which we know carries $\mathbf{10}$. The second parenthesis contains the rest, we write it as a sum of two terms related by $i \leftrightarrow j$ symmetrization. To see that $x_{ij}y_k - x_{kj}y_i$ transforms as $\mathbf{8}$ we define a $(1,1)$ tensor by contracting with ϵ, $z^p_j = \epsilon^{ikp}(x_{ij}y_k - x_{kj}y_i)$. z^p_j is traceless $z^p_p = 0$, so it transforms in the $\mathbf{8}$ representation.

• $\mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3} = (\mathbf{6}\oplus\mathbf{3}^*)\otimes\mathbf{3} = \mathbf{6}\otimes\mathbf{3}\oplus\mathbf{3}^*\otimes\mathbf{3} = \mathbf{10}\oplus\mathbf{8}\oplus\mathbf{8}\oplus\mathbf{1} = \yng(3)\oplus\yng(2,1)\oplus\yng(2,1)\oplus\yng(1,1,1)$. Note that the dimensions are 27 on either side and that $\mathbf{8}$ appears with multiplicity 2 in the decomposition. Thus a system of three light quarks can transform as a decuplet, octet or singlet under $su(3)$. Examples include the $J^P = \frac{3}{2}^+$ Δ decuplet, the $J^P = \frac{1}{2}^+$ nucleon octet and the Λ^0 singlet.

• $\mathbf{8}\otimes\mathbf{3} = \yng(2,1)\otimes\yng(1) = \yng(3,1)\oplus\yng(2,2)\oplus\yng(2,1,1) = \mathbf{15}\oplus\mathbf{6}^*\oplus\mathbf{3}$. The column of three boxes corresponds to the trivial representation and was eliminated from the last diagram. Note that the dimensions are 24 on either side.

• Under complex conjugation $A\otimes B = C\oplus D\oplus\ldots$ becomes $A^*\otimes B^* = C^*\oplus D^*\oplus\ldots$. In particular we must have $\mathbf{3}^*\otimes\mathbf{3}^* = \mathbf{6}^*\oplus\mathbf{3} = \yng(2,2)\oplus\yng(1)$ etc.

• To decompose a tensor product of irreps $A\otimes B$ when B has more than one box we need to preserve the anti-symmetry in any pair of indices that appear in a common column of A or B, provided the indices appear in the final tableau. We state without proof (see Georgi) the rules for decomposing $A\otimes B$ by adding to A, the boxes of B. We place the label a in each box of the first row of B and the label b in the second row of B. Step 1: Take the boxes labelled a

and stick them onto A in all possible ways (to the right or below existing boxes) to form several legal tableaux while ensuring two as do not appear in the same column. Step 2: Take the b boxes and add them to the tables from step 1 in all possible ways (to the right or below existing boxes) to form Young tables subject to one condition. Reading along the rows from right to left and from top row downwards, the number of a's must be $\geq$ the number of b's at any stage.

• Apply these rules to obtain the Clebsch-Gordan decompositions (1) $\mathbf{3} \otimes \mathbf{3}^* = \mathbf{8} \oplus \mathbf{1}$ (note $\mathbf{8}$, $\mathbf{1}$ are both self-conjugate.) (2) $\mathbf{3}^* \otimes \mathbf{3}^* = \mathbf{6}^* \oplus \mathbf{3}$ and (3) $\mathbf{8} \otimes \mathbf{8} = \mathbf{27} \oplus \mathbf{10} \oplus \mathbf{10}^* \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1}$ (both reps are 64 dimensional).

10.6.10 Mass formula of Gell-Mann and Okubo

• Isospin $su(2)$ (spanned by $\lambda_1, \lambda)2, \lambda_3$) is a very good symmetry of the strong interactions (masses in a multiplet are degenerate to a percent or better), and strangeness or hypercharge (λ_8) is conserved in all strong interactions. But the rest of flavor $su(3)$ is not quite as good a symmetry: masses within the baryon octet vary from 939 to 1318 MeV. Gell-Mann suggested a division of strong interactions into those governed by hamiltonian H_0, that preserve $su(3)$ (H_0 commutes with all λ_a) and those governed by H_1, that commutes only with isospin $\lambda_1, \lambda_2, \lambda_3$ and hypercharge λ_8 generators of $su(3)$. H_1 was to be treated as a perturbation to H_0. The idea was that H_0 is responsible for the main common mass M_0 of the members of an $su(3)$ multiplet, while H_1 is responsible for the mass splittings between different isospin multiplets within the larger su(3) multiplet. Based purely on group theoretic arguments (i.e. that H_1 transforms as λ_8 (recall λ_8 commutes with $\lambda_{1,2,3}$ and λ_8) and use of the Wigner-Eckart theorem, but without detailed knowledge of H_0 or H_1), Gell-Mann and Okubo independently obtained a formula for the masses of members of a multiplet. For the nucleon octet, the masses of various isospin multiplets are (see Georgi)

$$M_N = M_0 - \frac{2X}{\sqrt{12}} + \frac{Y}{\sqrt{12}}, \qquad M_\Sigma = M_0 + \frac{X}{\sqrt{12}} + \frac{Y}{\sqrt{12}},$$
$$M_\Lambda = M_0 - \frac{X}{\sqrt{12}} - \frac{Y}{\sqrt{12}}, \qquad M_\Xi = M_0 + \frac{X}{12} - \frac{2Y}{\sqrt{12}}. \tag{371}$$

These are four predictions in terms of three unknown parameters. The constants M_0, X, Y are not determined by symmetry arguments, they depend on the detailed dynamics of strong interactions. But M_0, X and Y can be eliminated to arrive at the parameter-free Gell-Mann Okubo mass formula

$$2(M_N + M_\Xi) = 3M_\Lambda + M_\Sigma. \tag{372}$$

Putting in the measured values $M_N = 940$, $M_\Sigma = 1190$, $M_\Xi = 1320$ the mass formula predicts $M_\Lambda = 1110$ which compares favorably with the experimental value $M_\Lambda = 1115$ MeV.

• For the baryon decuplet the Gell-Mann-Okubo mass formula predicts equal spacing between isospin multiplets that differ by one unit of strangeness: $M_{\Sigma^*} - M_\Delta = M_{\Xi^*} - M_{\Sigma^*} = M_{\Omega^-} - M_{\Xi^*}$. The measured masses $M_\Delta = 1230, M_{\Sigma^*} = 1385$, $M_{\Xi^*} = 1530, M_{\Omega^-} = 1672$ are nearly equally spaced: $\Delta M = 155, 145$ and 142. In the quark model this may be understood as due to the addition of one strange quark in place of a u or d quark each time S decreases by one. This also suggests that the strange quark has a mass of about $100 - 140$ MeV.

• In fact, the Ω^- had not been discovered when Gell-Mann and Okubo proposed their mass formula. Based on the structure of the $su(3)$ $\mathbf{10}$ representation and mass formula, Gell-Mann and Ne'eman (independently) in 1962 predicted the existence of the $S = -3$ baryon Ω^- at a mass of about 1680 MeV, and also predicted that the Ω^- would be long lived unlike the other baryon resonances. This is because it is not heavy enough to decay in a strangeness conserving strong process to two or more lighter hadrons with total baryon number one (e.g. an $S = -2$ Ξ^* baryon and an $S = -1$ kaon). Thus they predicted that it would decay weakly in a $\Delta S = 1$ transition to $\Xi^-\pi^0, \Xi^0\pi^-$ or $\Lambda^0 K^-$. The $\Omega^-(1672)$ was discovered at Brookhaven by Barnes et. al., in 1964 near the predicted mass and underwent weak decay as predicted. This was spectacular confirmation of the approximate $su(3)$ flavor symmetry of the strong interactions.

10.7 Gell-Mann and Zweig quark model

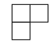
• In the discussion of $su(3)$ flavor symmetry, u, d and$s$ quarks were simply names for three independent states of the fundamental representation $\mathbf{3} = \square$. They did not have a dynamical significance as interacting particles with mass, spin etc. Gell-Mann and Zweig (1964) suggested that quarks may in fact be constituents of hadrons. The difficulty was that quarks had not (and still have not) been detected as isolated fractionally charged particles. The forces among quarks would have to be such that quarks are confined within hadrons. There was no precedent for this in physics nor a mechanism to achieve this in the 1960s. Today, we have good empirical reasons to believe that quarks do exist, and that they are confined within hadrons due to the strong force mediated by gluons and described by QCD. More on QCD later.

• The quark model postulates that hadrons are bound states of particles called quarks and their anti-particles. We will discuss their spin, mass etc later. Baryons are assumed to be made of three quarks and mesons of a quark and anti-quark. For this to work, quarks (anti-quarks) must have $B = \pm\frac{1}{3}$. Baryon number is the conserved qty corresponding to global $U(1)$ phase changes of quark fields. B is additive for a composite system of quarks and anti-quarks. $U(1)_B$ commutes with flavor SU(3). The u, d, s quarks are assigned quantum numbers Q, I_3 and Y and $S = Y - B$ (and their negatives for anti-quarks) as discussed in the context of $su(3)$ flavor symmetry. Each of these is diagonal in the fundamental representation with diagonal elements given below

$$I_3 \to (\frac{1}{2}, -\frac{1}{2}, 0), \quad Y \to (\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}), \quad S \to (0, 0, -1), \quad Q = I_3 + \frac{Y}{2} \to (\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}). \tag{373}$$

• The $su(3)$ irrep (m, n) carried by the space of tensors $x_{i_1 \cdots i_n}^{j_1 \cdots j_m}$ is regarded as a multiplet of $D(m, n)$ particles composed of quarks. I_3 and Y (and therefore Q and S) for a composite system of quarks and anti-quarks is additive. This also follows from from $su(3)$ representation theory: if a generator λ is diagonal in the fundamental rep $\mathbf{3}$, (e.g. the Cartan generators λ_3, λ_8) then it is diagonal in any irrep. Moreover, if the eigenvalues of λ are c_1, c_2, c_3 in $\mathbf{3}$, then its eigenvalues in the irrep (m, n) carried by $x_{i_1 \cdots i_n}^{j_1 \cdots j_m}$ are $c_{i_1} + \cdots + c_{i_n} - (c_{j_1} + \cdots + c_{j_m})$.

• By requiring all the particles of an irreducible multiplet (m, n) to have a given fixed baryon number, their quark composition can be fixed. To fix the quark composition, recall that each lower index i can be regarded either as a quark constituent or an anti-symmetric pair of anti-quark constituents, by contraction with ϵ^{jki}. These two possibilities differ in baryon number $B = 1/3$ or $-2/3$. Specifically, x_1 can be a u quark or an anti-symmetrized $(23) = (\bar{d}\bar{s})$ pair. Anti-symmetrization refers to anti-symmetrization in flavor quantum numbers $(23) = \frac{1}{2}\left(\bar{d}(1)\bar{s}(2) - \bar{s}(2)\bar{d}(1)\right)$. Similarly, $x_2 = d$ or $(31) = (\bar{s}\bar{u})$ and $x_3 = s$ or $(12) = (\bar{u}\bar{d})$. Similarly, an upper index j can be an anti-quark or an anti-symmetric pair of quarks: $x^1 = \bar{u}$ or $(23) = (ds)$, $x^2 = \bar{d}$ or (su) and $x^3 = \bar{s}$ or (ud).

• E.g. The traceless tensors x_i^j carrying the adjoint rep $\mathbf{8} =$  refer to an octet of $q\bar{q}$ bound states (mesons) if $B = 0$ and to an octet of 3 quark bound states (baryons) if $B = 1$. The anti-symmetrization in two of the quarks enters the baryon wave function, as we will discuss later. The tensors x_i with baryon number $1/3$ are just the triplet of quarks. The symmetric tensors x_{ijk} with $B = 1$ are the decuplet of baryons etc.

• Since no hadron with non-integer B has been detected, we suppose there is some deeper dynamical reason (color confinement in QCD) which explains why multiplets with fractional B are not observed. The allowed multiplets are those whose Young tableaux contain a number of boxes that is divisible by three. Recall that the irrep (m, n) has a Young table with $2m + n$ boxes and corresponds to the tensor $x_{i_1 \cdots i_n}^{j_1 \cdots j_m}$. The tensor has n quark and m anti-quark indices which corresponds to a multiplet with $B = (m - n)/3$ modulo one. For B to be an integer, $n - m$ must be a multiple of 3, or equivalently $n - m + 3m = n + 2m$ must be a multiple of three. Examples of such multiplets are $\mathbf{1}, \mathbf{8}, \mathbf{10}, \mathbf{10}^*, \mathbf{27}$ etc.

• Let us consider the possible quark composition of $\mathbf{8}$, i.e. traceless $(1,1)$ tensors x_i^j. This could describe a multiplet of mesons (i.e., $B = 0$) since there is one quark and one anti-quark index. The 8 states in the weight diagram of $\mathbf{8}$ then correspond to mesons with the following quark content

$$
\begin{array}{llll}
x_2^1 = \bar{u}d = \pi^-, & x_1^2 = \bar{d}u = \pi^+, & x_3^2 = \bar{d}s = \bar{K}^0, & x_2^3 = \bar{s}d = K^0, \\
x_1^3 = \bar{s}u = K^+, & x_3^1 = \bar{u}s = K^-, & x_1^1 + x_2^2 = \bar{u}u + \bar{d}d = \pi^0, & x_1^1 - x_2^2 = \bar{u}u - \bar{d}d = \eta
\end{array} \tag{374}
$$

These mesons have been identified with the members of the pseudoscalar pion octet, the same quark composition also applies to the vector meson octet ρ, K^*, ω. Note that there is some mixing between the iso-singlet $\bar{u}u - \bar{d}d$ which is a member of $\mathbf{8}$ and $\bar{u}u + \bar{d}d + \bar{s}s$, which is a member of $\mathbf{1}$ ($su(3)$ singlet and $su(2)$ singlet) to form the physical states η, η'.

• $\mathbf{8}$ could also describe a multiplet of baryons ($B = 1$) if we convert the upper anti-quark index to an anti-symmetric pair of lower quark indices $\epsilon_{jkl} x_i^j$. It could also describe an anti-baryon multiplet by raising the lower quark index $\epsilon^{ikl} x_i^j$. To obtain the quark content of the baryons we simply replace each anti-quark label, such as $\bar{u}$ by the corresponding anti-symmetrized pair of quark indices $\bar{u} \to (ds)$ where $(ds) = d(1)s(2) - d(2)s(1)$ where $1, 2$ identify the two quarks and $u(1), d(1), s(1)$ are a basis for the su(3) flavour states of the first quark and similarly $u(2), d(2), s(2)$ are a basis for the flavor states of the second quark. Thus the quark compositions of the 8 states in the weight diagram are

$$
\begin{array}{lll}
x_2^1 = (ds)d = \Sigma^-, & x_1^2 = (su)u = \Sigma^+, & x_3^2 = (su)s = \Xi^0, \\
x_2^3 = (ud)d = n, & x_1^3 = (ud)u = p, & x_3^1 = (ds)s = \Xi^-, \\
 & x_1^1 + x_2^2 = (ds)u + (su)d = \Sigma^0, & x_1^1 - x_2^2 = (ds)u - (su)d = \Lambda^0
\end{array} \tag{375}
$$

We have identified these 3-quark baryons with members of the nucleon octet. Mark these on a weight lattice.

• Similarly, find the anti-quark content of the 8 states of the anti-baryon octet and indicate them on a weight diagram.

• $\mathbf{10}$ carried by the symmetric 3rd rank tensors x_{ijk} can have baryon number one and correspond to the baryon decuplet (the lowest lying decuplet is the Δ decuplet). The quark compositions of the states in the weight diagram are easily identified (Y increases upwards and I_3 to the right)

$$\begin{array}{cccc} x_{222} = ddd = \Delta^-, \quad x_{221} = ddu = \Delta^0, & & x_{211} = duu = \Delta^+, \quad x_{111} = uuu = \Delta^{++} \\ x_{322} = sdd = \Sigma^{-*}, \quad x_{321} = & sdu & = \Sigma^{0*}, \quad x_{311} = sdu = \Sigma^{+*}, \\ x_{332} = ssd = \Xi^{-*}, & & x_{331} = ssu = \Xi^{0*}, \\ x_{333} = & sss & = \Omega^-. \end{array} \tag{376}$$

Similarly, one may arrive at the quark composition of the anti-baryon decuplet which carries the inequivalent conjugate representation $\mathbf{10}^*$.

• However, $\mathbf{10}$ could also correspond to a multiplet with $B = 0$, by converting a quark index to an anti-symmetric pair of anti-quark indices $\epsilon^{lmi} x_{ijk}$. Such a decuplet would consist of mesons composed of two quarks and two anti-quarks ($qq\bar{q}\bar{q}$). Such mesons are not common at low masses and are called *exotic*, they are probably very unstable, though they are allowed by color confinement. There are several experimentally detected hadronic resonances that may be exotic, such as X(3872),Y(3940), Y(4140),Zc(3900), Z(4430) with masses in MeV indicated. In fact, 10 could even correspond to exotic anti-baryons $\epsilon^{npj}\epsilon^{lmi} x_{ijk}$ with the quark content $\bar{q}\bar{q}\bar{q}\bar{q}q$. Similarly, there are exotic $B = 2$ states that cannot be separated into a pair of color singlet $B = 1$ states ('exotic' deuteron) and so on for higher baryon number. These exotic hadrons are not the same as conventional nuclei, they are probably more massive and more unstable and do not seem to play a role in nuclear physics except possibly at the very high densities in the early universe.

• The quark model also provides an explanation for why $su(2)$ isospin is not quite an exact symmetry of the strong interactions: the up and down quarks are not equally massive ($3 - 7$ MeV). Flavor $su(3)$ symmetry is even less exact since the strange quark has a mass of order 100 MeV - 150 MeV. Moreover, the Gell-Mann-Okubo formula for the constancy of the mass differences between isospin multiplets within the Δ decuplet may simply be interpreted as due to the addition of a strange quark as strangeness decreases by one.

10.8 Flavor and spin: $su(4)$ and $su(6)$

• So far we have only discussed the quark flavor composition of hadron multiplets with given baryon number. The relevant symmetry groups are $U(1)$ for baryon number, $su(2)$ for isospin or its enlarged version flavor $su(3)$. Now we would like to include the other degrees of freedom of quarks, such as spin and location. As a first step, consider only flavor and spin, so that the space-time dynamics is still ignored. Given the success of $su(2)$ and $su(3)$ there were attempts to find enlarged (approximate) symmetry groups whose representations were realized on multiplets of hadrons. The simplest such possibility is $su(4)$ which comes from combining $su(2)$ isospin with $su(2)$ spin in a non-relativistic approximation. Recall from atomic physics that spin-orbit coupling is a relativistic effect that may be derived from the Dirac equation. In the non-relativistic approximation distinct spin projections have the same energy. $su(4)$ symmetry in its defining fundamental representation is acts on the 4d vector space spanned by $u\uparrow, u\downarrow, d\uparrow, d\downarrow$ with arrows indicating spin projections of the two quark flavors. It is supposed that special unitary transformations in the 4d space spanned by these, are an approximate symmetry of the strong interactions. A treatment where spin is separated from orbital angular momentum is typically justified only in non-relativistic qm, so $su(4)$ can only be a very approximate symmetry, especially for light quarks. Similarly, Gursey, Radicati and Sakita suggested an approximate $su(6)$ symmetry by including the strange quark.

• If $su(4)$ were a symmetry, then by analogy with $su(3)$, the mesons and baryons would come in multiplets that appear in the decomposition of $\mathbf{4}\otimes\mathbf{4}^*$ and $\mathbf{4}\otimes\mathbf{4}\otimes\mathbf{4}$ into irreps. Interestingly, this seems to be roughly true. Use the rules for decomposing tensor products and the weights by hooks rules to show that

$$\mathbf{4}\times\mathbf{4}\times\mathbf{4} = \square\otimes\square\otimes\square = \mathbf{20}\oplus\mathbf{20'}\oplus\mathbf{20'}\oplus\mathbf{4}^* = \ydiagram{3}\oplus\ydiagram{2,1}\oplus\ydiagram{2,1}\oplus\ydiagram{1,1,1} \qquad (377)$$

$\mathbf{4}\otimes\mathbf{4}\otimes\mathbf{4}$ includes a $\mathbf{20}$ in its decomposition, corresponding to *symmetric* 3rd rank tensors. These 20 states can be identified with the 4 spin-isospin states of the nucleon doublet and the 16 spin-isospin states of the spin $3/2$ Δ quartet. These 20 hadronic states $n\uparrow, n\downarrow, p\uparrow, p\downarrow$, $\Delta^-(3/2), \Delta^-(1/2), \Delta^-(-1/2), \Delta^-(-3/2), \cdots$ carry the **20** representation of SU(4).

• Similarly, for $su(6)$, we have the decompositions $\mathbf{6}^*\otimes\mathbf{6} = \mathbf{1}+\mathbf{35}$.

$$\ydiagram{1,1,1,1,1}\otimes\square = \ydiagram{1,1,1,1,1,1}\oplus\ydiagram{2,1,1,1,1} = \mathbf{1}\oplus\mathbf{35}. \qquad (378)$$

The states of $\mathbf{35}$ with $B = 0$ can be identified with the 8 flavor states of the pseudoscalar pion octet + 24 spin-flavor states of the spin-1 vector meson (ρ) octet and the 3 spin states of the vector meson su(3) singlet ϕ. The spin-0 su(3) singlet η' can be identified with the $su(6)$ singlet in the decomposition $\mathbf{35}\oplus\mathbf{1}$.

• Moreover, there is a $\mathbf{56}$ carried by symmetric 3rd rank tensors with $B = 1$ in the decomposition (use the weights by hook lengths formula to get the dimensions of irreps)

$$\mathbf{6}\times\mathbf{6}\times\mathbf{6} = \mathbf{56}\oplus\mathbf{70}\oplus\mathbf{70}\oplus\mathbf{20} = \ydiagram{3}\oplus\ydiagram{2,1}\oplus\ydiagram{2,1}\oplus\ydiagram{1,1,1} \qquad (379)$$

These 56 baryon states can be identified with the union of the 16 spin-flavor states of the spin-half N-octet and the 40 spin-flavor states of the Δ spin-$3/2$ decuplet.

• While isospin and flavor su(3) are internal symmetries, spin is a space-time symmetry (spin SU(2) arises in the study of the representations of the Pioncare group). However, treating spin in a manner distinct from orbital angular momentum is only justified in a non-relativistic approximation. Given the above successes, physicists tried to combine space-time and internal symmetries in non-trivial ways (i.e. other than as a cartesian product) into larger groups that could be the symmetries of a relativistic quantum system. However, no non-trivial way was found. In fact, Coleman and Mandula argued that if a Lie algebra combined both space-time and internal symmetry generators in a non-trivial way, a field theory that possessed this symmetry would have a trivial S-matrix in $3+1$ dimensions, i.e., the fields would not interact and particles would not scatter. This Coleman-Mandula theorem is based on some assumptions, which can be relaxed to get non-trivial interacting field theories that are invariant under symmetry groups that combine space-time and internal symmetries in non-trivial ways. One such possibility is provided by dropping the condition that the infinitesimal symmetries form a Lie algebra. If some of the generators satisfy commutation relations and some satisfy anti-commutation relations, one can get a $\mathbb{Z}_2$ graded Lie algebra also known as a Lie super-algebra, using which one may circumvent the Coleman Mandula no-go theorem. This idea has been realized in supersymmetric (SUSY) field theories. However, nature seems not to have chosen to unite space-time and internal symmetries via SUSY in a relativistic theory of hadrons. Instead, nature seems to have selected a non-trivial generalization of electromagnetism, i.e., non-abelian gauge theory [based on a new and unexpected internal color symmetry] for the strong interactions of hadrons.

10.9 Need for colored quarks

• Now we try to extend the quark model by including translational degrees of freedom so that quarks are described by wave functions that depend on position, spin projection and flavour. We ignore the mass differences between u, d, s quarks and treat quarks as identical particles, the different quarks are simply different flavor states (i.e. with different I_3 or Y). Being spin half particles the total wave function must be anti-symmetric under exchange of space, spin and flavour of all quarks. The simplest possibility, by analogy with multi-electron atoms or nuclear shell models is to suppose that quarks move in a mean field produced by the other quarks and occupy single particle orbitals that are suitably anti-symmetrized to satisfy Fermi statistics. Now the su(4) and su(6) models say that the lowest lying baryons nucleon octet and delta decuplet transform in the symmetric rank three tensor representations. So the wave functions of the quarks in these baryons must be symmetric under exchange of spin and flavor. Particularly simple examples are the $J_z = 3/2$ states of Δ^{++}, Δ^- and Ω^- which have 3 up, 3 down and 3 strange quarks respectively, all with the same flavor and spin-projection. To satisfy Pauli exclusion, the wave function must be anti-symmetric under exchange of positions of quarks in the baryon. In particular the quarks cannot all be in the same lowest S wave orbital which would be expected to minimize energy. However, this is not supported by experimental facts, to say nothing of the theoretical difficulty in constructing a rotationally-invariant model of forces between quarks where the ground state is *not* spherically symmetric. For instance, pions and nucleons have roughly the same size (charge radii). In a meson, the quark and anti-quark can both be in the same S-wave orbital since they are not identical particles, this would minimize energy. Since nucleons have the same charge radii as pions, the data favors all quarks to occupy the same spatial S orbital. Moreover, if the quarks were in different spatial orbitals, at least one of them would have a node, which would lead to a zero in the electric form factor (a measure of the charge distribution), which is not seen in the data. Something is not right. If we trust the $su(4)$ and $su(6)$ symmetries then we cannot seem to be able to satisfy Pauli exclusion and match experiment. It appeared that quarks behave like bosons to the extent that at most three quarks could occupy the same state, despite having spin half. To resolve this puzzle, it was suggested that quarks may satisfy a new kind of 'parastatistics' different from Bose and Fermi or that the $su(4)$ and $su(6)$ arguments were not to be trusted since they relied on a non-relativistic treatment. The resolution of this puzzle that has subsequently been confirmed by experiment is that quarks possess an additional 'color' degree of freedom (having nothing to do with color of light) and transform in the fundamental representation of $su(3)$ color. So each flavor of quark comes in three possible colors having the same mass and charge. SU(3) rotates among the three color states. $su(3)$ color is believed to be an exact symmetry, unlike the approximate $su(3)$ flavor symmetry. The three quarks in a baryon of the N octet or Δ decuplet are anti-symmetric under interchange of color, allowing all three of them to occupy the same spatial orbital and still satisfy Fermi-Dirac statistics.

• There are several other reasons for believing that quarks come in precisely $N_c = 3$ colors, both theoretical and experimental.

• The current theory of strong interactions QCD is a non-abelian gauge theory with color playing the role that electric charge plays in the abelian gauge theory QED. There are good reasons (numerical and from simpler models) to believe that quantum chromodynamics confines color so that all hadrons must be color neutral (i.e., color singlets, transforming in the trivial representation of $su(3)$ color). If this were true, it would also explain why observed hadrons have integer baryon number. This is because the trivial representation of color has the Young tables with no boxes or a column of three or two columns of three each. The number of boxes is equal to one third the baryon number, since each box can be regarded as coming from one quark, which has baryon number $1/3$. Color confinement would also explain why quarks are confined within hadrons and have not been isolated.

• Recall that the neutral pion predominantly decays to 2 photons. To leading order, the decay amplitude is encoded in a 'triangle diagram' with a virtual quark loop. One must sum over all possible quarks that can appear in the loop, and square the resulting amplitude to arrive at the decay rate. One finds that the measured decay rate is about 9 times what is obtained if quarks came in a single color. This is evidence for $N_c^2 = 9$, or $N_c = 3$ colors.

• Ratio R of cross-sections for e^+e^- annihilation at a collider at CM energy of $\sqrt{s}$:

$$R = \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)} \tag{380}$$

R has been measured over a wide range of CM energies, it is plotted for instance in Huang Fig 2.8 (p.39) and also Halzen and Martin Fig 11.3 (p.229) up to about $\sqrt{s} = 40$ GeV. To leading order the basic scattering process is electromagnetic with e^+e^- annihilating to form a virtual photon which then produces a quark anti-quark pair or a charged lepton-anti-lepton pair. The leading order QED cross section at CM energy s is $\sigma(e^+e^- \to \mu^+\mu^-) = \frac{4\pi\alpha^2}{3s}$. If we had a $q\bar{q}$ pair produced in place of $\mu^+\mu^-$, then we would have to multiply by the square of the charge of the quark $e_q^2 = (2/3)^2$ or $(1/3)^2$ in units of the electron charge. What is more, if each quark came in N_c colors, then we would have to multiply the cross section by N_c to account for the

N_c possible color-anti-color pairs in the final state, i.e.,

$$\sigma(e^+e^- \to q\bar{q}) = N_c e_q^2 \sigma(e^+e^- \to \mu^+\mu^-). \tag{381}$$

Thus in this leading order QED 'hard scattering' approximation [we ignore the corrections to the cross section due to the process of hadronization of the $q\bar{q}$ pair], $R = N_c \sum_q e_q^2$. Now, as s is increased, the thresholds for production of more massive quarks like s, c, b are crossed. So one expects R to increase like a step function at these thresholds. R is sensitive to the number of colors and $N_c = 3$ is in reasonable agreement with the data. Find the predicted values of R between the various thresholds.

10.10 Quark model wavefunctions for baryons and mesons

• BARYONS: As before, we work in the non-relativistic independent quark approximation (so that we may use single particle orbitals and factorize the position dependence from the spin and flavor), imposing Fermi statistics for 3 quarks in a baryon. Here we briefly describe the structure of the wave functions, in their dependence on color, flavor, spin and position. Such a discussion is justified in a non-relativistic approximation where the number of quarks in a system is independent of time[52].

• We argued that the spin-flavor wave function of a baryon in the N and Δ multiplets must be symmetric (as it belongs to $\mathbf{56}$ of $su(6)$). Moreover since these are the lowest lying baryons, all three quarks occupy the same single particle spatial $l = 0$ S-wave orbital $\psi(r_1)\psi(r_2)\psi(r_3)$. The dependence on position can be determined when one specifies a model for the force between quarks: we do not pursue that here (in principle this is determined by QCD - the potential between quarks is roughly Coulombic at short distances and linear at long distances). So the angular momentum of these baryons arises entirely from combining the quark spins. The color part of the wave function is completely antisymmetric, i.e., the color singlet representation $\mathbf{1}$ that appears in the decomposition $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{10}$ (here $\mathbf{3}$ is fundamental rep of color $su(3)$, not flavor). Explicitly, if r_i, g_i, b_i refer to the three basis vectors $(1,0,0), (0,1,0), (0,0,1)$ for color space $\mathbb{C}^3$ of the i^{th} quark, then the color wave function is

$$\frac{1}{\sqrt{6}} \sum_{\sigma \in S_3} (-1)^{\,\text{sgn}\,\sigma} r_{\sigma(1)} g_{\sigma(2)} b_{\sigma(3)} = \frac{1}{\sqrt{6}} \left[r_1 g_2 b_3 + b_1 r_2 g_3 + g_1 b_2 r_3 - g_1 r_2 b_3 - r_1 b_2 g_3 - b_1 g_2 r_3 \right] \tag{382}$$

To understand the notation for the color wave function, observe that it is a vector in the color Hilbert space $\mathbb{C}^3 \otimes \mathbb{C}^3 \otimes \mathbb{C}^3$ of a three quark system. For instance, it says that the amplitude for finding the first quark to be green, second to be blue and third to be red is $1/\sqrt{6}$ and that the amplitude for the first two quarks to be red and the third quark to be green is zero etc.

• It remains to specify the spin-flavor wave functions for the baryons. It must be symmetric under exchange of any pair of quarks. The $J = 3/2$ Δ decuplet is a bit easier than the N-octet

[52] In a fully relativistic treatment (say, based on QCD), it would not ordinarily be consistent to speak of a wave function of a system of three quarks, since the number of quarks is not a conserved quantity. A fully relativistic treatment has not yet been developed, except in some simplified versions of QCD.

since if we consider the $m = 3/2$ states, then all the spins must point upwards and we only need to symmetrize in flavors (the $m = 1/2, -1/2, -3/2$ states can be obtained by applying the lowering operator J_-). We have already determined the flavor composition of quarks in the **10** weight diagram. So for the $m = 3/2$ states of Δ^{++} we must have $\Delta^{++} = u_\uparrow u_\uparrow u_\uparrow$. Δ^+ has 2u's and a d, so we must symmetrize to obtain $\Delta^+ = \frac{1}{\sqrt{3}}\{u_\uparrow u_\uparrow d_\uparrow + d_\uparrow u_\uparrow u_\uparrow + u_\uparrow d_\uparrow u_\uparrow\} \equiv (u_\uparrow u_\uparrow d_\uparrow)_s$ where $(\cdots)_s$ denotes symmetrization. Similarly we have

$$\begin{aligned}
\Delta^0 &= (u_\uparrow d_\uparrow d_\uparrow)_s, \quad \Delta^- = d_\uparrow d_\uparrow d_\uparrow, \quad \Sigma^{*+} = (u_\uparrow u_\uparrow s_\uparrow)_s, \quad \Sigma^{*-} = (d_\uparrow d_\uparrow s_\uparrow)_s \\
\Sigma^{*0} &= (u_\uparrow d_\uparrow s_\uparrow)_s = \frac{1}{\sqrt{6}}(u_\uparrow d_\uparrow s_\uparrow + s_\uparrow u_\uparrow d_\uparrow + d_\uparrow s_\uparrow u_\uparrow + d_\uparrow u_\uparrow s_\uparrow + u_\uparrow s_\uparrow d_\uparrow + s_\uparrow d_\uparrow u_\uparrow) \\
\Xi^{*0} &= (u_\uparrow s_\uparrow s_\uparrow)_s, \quad \Xi^{*-} = (d_\uparrow s_\uparrow s_\uparrow)_s, \quad \Omega^- = s_\uparrow s_\uparrow s_\uparrow. \qquad (383)
\end{aligned}$$

To obtain the spin-flavour wave functions for the $m = \frac{1}{2}$ states we simply apply $J_- = J_{1-} + J_{2-} + J_{3-}$ and normalize, which would convert $u_\uparrow d_\uparrow u_\uparrow \to \frac{1}{\sqrt{3}}(u_\downarrow d_\uparrow u_\uparrow + u_\uparrow d_\downarrow u_\uparrow + u_\uparrow d_\uparrow u_\downarrow)$ for instance. The $m = -\frac{1}{2}$ and $m = -3/2$ wave functions are obtained by simply reversing the spin projections of all the spins in the $m = \frac{1}{2}$ and $m = 3/2$ wave functions.

• Next we address the spin-flavor wave functions of the $J = \frac{1}{2}$ baryons of N-octet, again they must be symmetric under exchange of any pair of quarks.

• To illustrate, let us focus on the proton, say in the $m = \frac{1}{2}$ up spin state. To get an $m = -\frac{1}{2}$ proton we may reverse all spin projections. And to get the corresponding neutron states, we may exchange u's with d's. To construct the proton state function we imagine combining a u and a d into an isospin zero state and then adding another u quark to get the desired isospin half of the proton. In more detail, let us first put one of the u, d pairs in the isospin zero state $\frac{1}{\sqrt{2}}(ud - du)$ state. For the wave function to be symmetric under exchange of this pair, they must be in an anti-symmetric spin zero ($S = S_z = 0$) state $\frac{1}{\sqrt{2}}(\uparrow\downarrow - \downarrow\uparrow)$. So the state of this ud pair is

$$\frac{1}{2}(ud - du)(\uparrow\downarrow - \downarrow\uparrow) = u_\uparrow d_\downarrow - u_\downarrow d_\uparrow - d_\uparrow u_\downarrow + d_\downarrow u_\uparrow. \qquad (384)$$

The RHS should explain the notation of the LHS. This state has $I = J = 0$. This is good since we can now combine this with a $u_\uparrow$ to get an $I = I_3 = \frac{1}{2}$, $S = S_3 = \frac{1}{2}$ state representing the proton with $S_z = \frac{1}{2}$.

• Now we include an up quark in the up spin state to this so that $I_3 = \frac{1}{2}$ and $J_3 = \frac{1}{2}$:

$$\frac{1}{2}(udu - duu)(\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow) \qquad (385)$$

However, this state is not symmetric under exchanges involving the third quark; it needs to be symmetrized (in both spin and flavor) by adding the result of interchanging quarks 2 & 3 as well as interchanging quarks 1 & 3. We get, upto a normalization constant,

$$p_\uparrow \propto (udu - duu)(\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow) + (uud - duu)(\uparrow\uparrow\downarrow - \downarrow\uparrow\uparrow) + (udu - uud)(\uparrow\downarrow\uparrow - \uparrow\uparrow\downarrow). \quad (386)$$

These 12 terms may be combined to get

$$p_\uparrow = \frac{1}{\sqrt{18}}\left[2u_\uparrow d_\downarrow u_\uparrow + 2d_\downarrow u_\uparrow u_\uparrow + 2u_\uparrow u_\uparrow d_\downarrow - u_\downarrow d_\uparrow u_\uparrow - d_\uparrow u_\downarrow u_\uparrow - u_\downarrow u_\uparrow d_\uparrow - d_\uparrow u_\uparrow u_\downarrow - u_\uparrow d_\uparrow u_\downarrow - u_\uparrow u_\downarrow d_\uparrow\right]. \qquad (387)$$

Reason for the normalization factor: the 9 terms in square-brackets above are mutually orthogonal. So the norm square is just $2^2 + 2^2 + 2^2 + 6 \times 1^2 = 18$. Verify that this state is symmetric under exchange of any pair of the three quarks. Similarly write down the neutron $n_\uparrow$ state. The spin-flavor wave functions of some of the other baryons in the N-octet can be obtained by suitable replacements: $\Sigma^+(uus)$ by replacing $d \to s$ in p, $\Sigma^-(dds)$ by $u \to s$ in n, $\Xi^0(uss)$ by $d \to s$ in n and $\Xi^-(dss)$ by $u \to s$ in p. Λ^0 and Σ^0 need to be worked out, like we did for the proton [it should also be possible to get the Σ^0 state vector by applying I_+ to the one for Σ^-].

• MESONS: For a meson composed of a quark and an anti-quark, the Pauli principle places no restriction, since they are distinguishable particles. But the meson must be in a color singlet state, i.e., in the trivial rep $\mathbf{1}$ in the decomposition $\mathbf{3} \otimes \mathbf{3}^* = \mathbf{1} \oplus \mathbf{8}$ of color $su(3)$. Explicitly, the color wave function of a meson is

$$\frac{1}{\sqrt{3}} \left[r_q \bar{r}_{\bar{q}} + g_q \bar{g}_{\bar{q}} + b_q \bar{b}_{\bar{q}} \right] \tag{388}$$

Here $\bar{r}_{\bar{q}}$ is the first basis vector for the dual color space $\mathbb{C}^{3*}$ of the anti-quark. The flavors of the quark and anti-quark in the pion octet have been discussed in the context of the weight diagram for $\mathbf{8}$.

10.11 Magnetic moments from quark model

• Dirac's equation for a point-like spin half particle of mass m and charge e coupled to an electromagnetic field predicts that the magnetic moment of the particle is $\vec{\mu} = \frac{e\hbar}{2mc} \frac{\vec{S}}{\hbar}$ where $\vec{S}$ is the spin (the magnetic dipole energy is $-\vec{\mu} \cdot \vec{B}$). The coefficient of $\vec{S}/\hbar$ is often loosely called the magnetic moment. For a nucleon $m = m_N$ so one defines the nucleon Bohr magneton as $\mu_N = \frac{e\hbar}{2m_N c}$. So Dirac's theory predicts that for the proton $\mu_p = \mu_N$ and $\mu_n = 0$ as the neutron is uncharged. However, the measured values are

$$\mu_p = 2.79\mu_N \quad \text{and} \quad \mu_n = -1.91\mu_N. \tag{389}$$

These values are further indication that the proton and neutron are not elementary spin half particles. The negative value of μ_n says that the neutron's magnetic moment is oriented opposite to its spin. The nucleon magnetic moments are anomalously large. μ_p was first measured by O Stern in 1933 (Stern is also famous for the Stern-Gerlach experiment). He got the Nobel prize in 1943 for the discovery of the anomalously large proton magnetic moment. I I Rabi improved greatly on measurements of the magnetic moments of the proton and deuteron by developing the NMR technique, for which he got the 1944 Nobel prize. μ_n was first estimated from knowledge of μ_p and μ_d treating it as a spin one bound state. The anomalous nucleon magnetic moments remained mysterious till the static constituent quark model offered an explanation.

• Note that the electric dipole moments of the neutron and proton are zero to current experimental precision. Theoretically, CP symmetry of the strong interactions prevents strong contributions to the neutron EDM. CP violation in the weak interactions implies a very small non-zero neutron EDM, but it has not been experimentally measured.

• In the constituent quark model for nucleons, the proton and neutron are modelled as composed of three valence quarks, each of mass $m_N/3$. These are not the 'current' quarks whose masses are only a few MeV. The constituent quarks may be loosely regarded as current quarks dressed by the gluon field, which contributes most of the nucleon mass. The up and down quark charges are $q_u = 2e/3$ and $q_d = -e/3$ and treating them as elementary spin half Dirac particles, their magnetic moments are

$$\mu_u = \frac{(2e/3)\hbar}{2(m_N/3)c} = \frac{e\hbar}{m_N c} = 2\mu_N \quad \text{and} \quad m_d = \frac{(-e/3)\hbar}{2(m_N/3)c} = -\frac{e\hbar}{2m_N c} = -\mu_N. \qquad (390)$$

Now we need to use the rules of angular momentum addition to determine the magnetic moments of the proton (uud) and neutron (udd).

• Simpler than the proton and neutron is the Δ. The simplest example is the spin $3/2$ Δ^{++} which has three $u_\uparrow$ quarks all in the same spatial orbital. S_z for Δ^{++} can be $3/2(\uparrow\uparrow\uparrow), 1/2, -1/2, -3/2$, all these states are symmetric in spin. $\mu^z_{\Delta^{++}} = \mu_{\Delta^{++}} \frac{S_z}{\hbar}$ where the magnetic moment $\mu_{\Delta^{++}}$ is independent of S_z and must simply be the sum of the magnetic moments of the three quarks each with the same spin projection, so $\mu_{\Delta^{++}} = 3\mu_u = 6\mu_N$.

• Next we consider the Δ^+ in an $S_z = 3/2$ state. We have seen that its spin-flavor wave function is $\frac{1}{\sqrt{3}}(u_\uparrow u_\uparrow d_\uparrow + d_\uparrow u_\uparrow u_\uparrow + u_\uparrow d_\uparrow u_\uparrow)$ with all quarks in the same spatial orbital. To get this we may imagine first combining the spins of the two u-quarks, which must necessarily be in a symmetric (spin one, $S_z = 1$) state since a pair of u quarks is symmetric in flavor: the magnetic moment of this pair is just $2\mu_u$. To this we add a $d_\uparrow$ to make a state with $S_z = 3/2$. Since the three quarks all have spin projection $S_z = \frac{1}{2}$, we may simply add their magnetic moments to get $\mu_{\Delta^+} = 2\mu_u + \mu_d = 3\mu_N$. In terms of Clebsch-Gordon coefficients for addition of spin 1 and spin half $|3/3, 3/2\rangle = |1,1\rangle \otimes |\frac{1}{2}, \frac{1}{2}\rangle$. So an $S_z = 3/2$ Δ^+ is with probability one in a state where the up quarks have $S = S_z = 1$ and the down quark has $S = S_z = \frac{1}{2}$.

• However, the magnetic moments of the Δ baryons are not easy to measure since they are very short-lived. The proton and neutron are more accessible.

• So consider the proton in an $S_z = \frac{1}{2}$ state. We first combine the spins of the two up quarks which must be in a symmetric $s_1 = 1$ state as the two up quarks are symmetric in flavor. Then we must add the $s_2 = \frac{1}{2}$ of the down quark to get a spin half state belonging to the $|s = \frac{1}{2}, s_z = \frac{1}{2}\rangle$ multiplet. The CG coefficients expressing this coupled basis state as a linear combination of the product states are familiar from $\pi - N$ scattering (there we needed the opposite decomposition!)

$$|\frac{1}{2}, \frac{1}{2}\rangle = \sqrt{\frac{2}{3}}|1\rangle|\downarrow\rangle - \frac{1}{\sqrt{3}}|0\rangle|\uparrow\rangle. \qquad (391)$$

So we may say that the proton is with probability $2/3$ in a state with the two up quarks in a state with $S_z = 1$ and down quark with $S_z = -1/2$ and with probability $1/3$ in a state with two up quarks with $S_z = 0$ and the down quark with $S_z = \frac{1}{2}$.

• Thus the magnetic moment of the proton predicted by the constituent quark model is

$$\mu_p = (2/3)(2\mu_u - \mu_d) + (1/3)\mu_d = (2/3)(4\mu_N + \mu_N) + \frac{1}{3}(-\mu_N) = 3\mu_N. \qquad (392)$$

This is not too far from the measured value of $2.79\mu_N$.

One may also arrive at this result from our formula for the spin-flavor wave function of the proton

$$p_\uparrow = \frac{1}{\sqrt{18}}\left[2u_\uparrow d_\downarrow u_\uparrow + 2d_\downarrow u_\uparrow u_\uparrow + 2u_\uparrow u_\uparrow d_\downarrow - u_\downarrow d_\uparrow u_\uparrow - d_\uparrow u_\downarrow u_\uparrow - u_\downarrow u_\uparrow d_\uparrow - d_\uparrow u_\uparrow u_\downarrow - u_\uparrow d_\uparrow u_\downarrow - u_\uparrow u_\downarrow d_\uparrow\right]. \tag{393}$$

The first three terms correspond to a pair of up quarks in the $S_z = 1$ state and down quark in $S_z = -\frac{1}{2}$ state, while the remaining terms correspond to a pair of up quarks in the $S_z = 0$ state and d quark in an $S_z = \frac{1}{2}$ state. From the coefficients we see that the probability for two u quarks in $S_z = 1$ is $(2^2 + 2^2 + 2^2)/18 = 2/3$ while the probability for two u quarks with $S_z = 0$ is $(6 \times 1^2)/18 = 1/3$.

• The quark model prediction for the neutron magnetic moment is obtained by exchanging $u \leftrightarrow d$ in μ_p, thus

$$\mu_n = (2/3)(2\mu_d - \mu_u) + (1/3)\mu_u = -2\mu_N. \tag{394}$$

Again, this is in reasonable agreement with the measured $\mu_n = -1.91\mu_N$. Thus, for the purposes for magnetic moments, the three constituent quarks in a baryon behave roughly as if they are elementary spin half Dirac particles, each with a mass equal to a third of the baryon mass and with charges as given above.

10.12 QCD

Quantum Chromodynamics (QCD) is the current microscopic theory of strong interactions. It is an SU$(3)_C$ non-abelian gauge theory of 8 gluon fields coupled to Dirac fermions (quarks) in the fundamental representation of color SU(3). There are $N_f = 6$ flavors of quarks (u,d,c,s,t,b) with masses going from a few MeV to 175 GeV. Thus the Lagrangian of QCD is

$$\mathcal{L} = -\frac{1}{2g^2}\,\mathrm{tr}\,F^{\mu\nu}F_{\mu\nu} + \sum_{i=1}^{N_f}\bar{\psi}_i[i\gamma\cdot D - m_i]\psi_i. \tag{395}$$

Aside from the six quark masses, which are input parameters, the gauge coupling constant g is classically a dimensionless input parameter. In the quantum theory, it turns out that the procedure of regularization and renormalization introduces a new dimensional parameter (usually the energy scale Λ_{QCD}) into the theory. Suitably defined, $\Lambda_{\rm QCD} \approx 200$ MeV from fits to experimental data. The coupling constant g becomes a running coupling $g(\mu)$, whose dependence on energy scale is predicted by the renormalization group equation. Given these 7 input parameters, QCD should predict all strong interaction cross sections, decay rates, masses of hadrons and nuclei (as pure numbers times Λ_{QCD}), 'wave functions' (structure functions) of hadrons and nuclei etc. Though much progress has been made, at present our tools to solve this theory are very limited. Weak coupling perturbation theory is one approach, but is reliable primarily in high energy scattering processes (e.g. some aspects of deep inelastic lepton hadron scattering) where the quark-gluon coupling g is small due to asymptotic freedom. Perturbation theory is not adequate to study the formation of bound states (hadrons) from quarks and gluons. For instance, we cannot reliably calculate the mass of the pion or proton from first principles (even numerically), nor show the confinement of color. Numerical simulation of the path integral in a discretized version of QCD (lattice QCD) provides a computational approach which has seen much progress in the last 40 years, though there are many challenges that remain. Another promising approach is an expansion of appropriate physical quantities in inverse powers of the number of colors $\sum_0^\infty N_c^{-n} a_n$. Though $N_c = 3$ in nature, the theory simplifies somewhat when $N_c = \infty$ (only planar Feynman diagrams survive), but is expected to retain many of its essential physical features. It is hoped that an understanding of large-N_c QCD (i.e., the leading term in a large-N_c expansion) will provide a first step towards QCD. This is an area of current and future research.

10.13 Spontaneous global symmetry breaking

• Even without a solution of QCD, there are specific strong interaction phenomena that can be treated within approximate 'effective models', by isolating the relevant degrees of freedom and identifying interactions among these degrees of freedom that are consistent with the symmetries. One enduring example is the sigma model for pions and nucleons. It predates QCD but remains interesting and has found application in many areas of physics.

• Pions are particularly light ($m_\pi = 135 - 140$ MeV) compared to nucleons as well as all other hadrons. For some purposes they may even be treated as massless. Nambu used the idea of spontaneous global symmetry breaking (SSB) to explain why the pions are so light. Nambu's model (developed with Jona Lasinio), is more complicated than the sigma model that we discuss and use to illustrate the idea of SSB.

• When a Lagrangian field theory possesses a continuous global symmetry, then there can be a continuous family of vacua related by the symmetry. If initial conditions are such that the system is in one of the possible ground states, then the symmetry is not manifest in the ground state. The symmetry is said to be spontaneously broken if the g.s is not invariant under the symmetry transformation. However, the system can be moved from one vacuum to a neighboring one by an infinitesimal symmetry transformation costing an arbitrarily small amount of energy. In other words, there can be excitations with arbitrarily low energy or arbitrarily long wave length. These zero modes lead to massless particles in the spectrum of the QFT, they are called Nambu-Goldstone bosons. There are as many NG bosons as there are independent directions in which the 'potential' is 'flat'. We may regard pions as the three Nambu-Goldstone bosons of a spontaneously broken $O(4)$ symmetry. Spin waves are NG bosons of spontaneously broken rotation invariance in a system of spins on a lattice, which all point in a common direction in a ferromagnetic ground state.

• SSB is a subtle phenomenon. It is present in classical point particle mechanics, 'goes away' due to tunneling in quantum systems with finitely many degrees of freedom, but returns in systems with infinitely many degrees of freedom.

• Consider a classical non-relativistic particle subject to the potential $V(\mathbf{r})$. In its ground state, it is at rest at a minimum $\mathbf{r}_0$ of the potential: $\partial_i V(\mathbf{r}_0) = 0$. Taylor expanding the potential,

$$V(\mathbf{r}) = V(\mathbf{r}_0) + \frac{1}{2}\partial_i\partial_j V(\mathbf{r}_0)(r_i - r_{0i})(r_j - r_{0j}) + \cdots \quad (396)$$

The hessian of the potential $\partial_i \partial_j V(\mathbf{r}_0)$ is a real symmetric matrix with non-negative eigenvalues, since $\mathbf{r}_0$ is a minimum of V. It controls the behavior of V in the neighborhood of $\mathbf{r}_0$. For example, consider the Mexican hat potential for a particle moving on the plane $V(x, y) = (\lambda/4)(x^2 + y^2 - v^2)^2$. The ground states (vacua) lie on the circle $x^2 + y^2 = v^2$. We say the vacuum manifold is a circle. V is invariant under the group O(2) comprising rotations in the x-y plane. But none of the vacua is invariant. Instead, rotations take one vacuum to another vacuum. In this case, the Hessian at any of the minima $x^2 + y^2 = v^2$ is

$$\partial^2 V = 2\lambda \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix}. \tag{397}$$

The eigenvalues are 0 and $2\lambda v^2$. The zero eigenvalue corresponds to the eigenvector $\begin{pmatrix} -y \\ x \end{pmatrix}$ which points in the $\hat{\theta}$ direction tangential to the vacuum manifold. This 'zero-mode' corresponds to infinitesimal rotations that take one vacuum to a neighboring vacuum. A ball can 'roll' with arbitrarily low energy from one vacuum to another one. This is the simplest version of a Nambu-Goldstone mode. The eigenvector corresponding to the non-zero eigenvalue $2\lambda v^2$ points in the radial direction $\vec{r} = (x, y)$ at any of the minima. It corresponds to a radial oscillatory mode about any ground state. Taylor expansion of the potential about any minimum leads to a 1d SHO in the radial direction $V \approx \frac{1}{2}(2\lambda v^2)(r - v)^2$. Thus the frequency of the corresponding radial oscillation is $\omega = \sqrt{2\lambda} v/m$. In field theory, the radial mode would correspond to a massive particle (such as the sigma meson below or the Higgs particle).

• In general, whenever a continuous symmetry is broken by the ground state, there are zero modes that point in directions (in configuration space) along which the potential is flat to quadratic order. This is not the case for discrete symmetries with degenerate vacua. For example, the reflection symmetric double well potential in one dimension $V(x) = (\lambda/4)(x^2 - v^2)^2$ has two ground states $x = \pm v$. Each of the ground states spontaneously breaks the $x \to -x$ symmetry of the potential. But V'' has one non-zero eigenvalue at each ground state, there are no zero modes. Discrete global symmetries can break spontaneously but do not lead to NG modes.

• For example, if we had a particle moving in 3d subject to the $O(3)$ (rotations and reflections of 3d space) invariant potential $V(\mathbf{r}) = \lambda(\mathbf{r}^2 - v^2)^2$. Every point (x, y, z) on the sphere of radius v (e.g. $(x = 0, y = 0, z = v)$) is a vacuum, and elements of the symmetry group O(3) can rotate one vacuum into another. However, the symmetry is not completely broken. If $\mathbf{r}_0$ is a ground state, then rotations about the vector $\mathbf{r}_0$ (and more generally rotations and reflections in the plane perpendicular to $\mathbf{r}_0$) leave the ground state invariant. We say that O(3) spontaneously breaks to O(2). In this case, the hessian of V at any vacuum has two zero eigenvalues corresponding to two zero modes that point in flat directions of V along the surface of the sphere. The hessian also has one positive eigenvalue $\propto \lambda v^2$ which is proportional to the square of the frequency of small oscillations in the radial direction.

• In quantum mechanics, even if the potential has degenerate minima, the ground state is unique since the particle can tunnel between minima and lower the energy. The ground state is not concentrated around any one minimum of the potential, but is a linear combination of states concentrated around the classical minima. If the potential has a symmetry, then the linear combination with lowest energy eigenvalue is invariant under the symmetry [symmetric combination of wave packets concentrated around the two minima of a double well potential]. Spontaneous symmetry breaking does not occur in quantum systems with finitely many degrees of freedom. However, if the number of degrees of freedom is infinite, tunneling is suppressed and SSB can occur, the system can get 'stuck' in a ground state that is *not* invariant under the symmetry.

10.14 Pions and the Gell-Mann Levy linear sigma model

• Pions can be regarded as Nambu-Goldstone bosons of a spontaneously broken global symmetry of an 'effective' field model proposed by M. Gell-Mann and M. Levy. We have seen that a spontaneously broken $O(2)$ symmetry has one zero mode and one massive excitation while $O(3)$ leads to two zero modes and one massive mode. Since there are three very light pions, it is natural to consider an $O(4)$ invariant Lagrangian in 3+1 space time dimensions for a four component real scalar field $\phi = (\phi_1, \phi_2, \phi_3, \phi_4) = (\vec{\phi}, \phi_4)$:

$$T = \frac{1}{2}\sum_{i=1}^{4} |\partial \phi_i|^2, \quad V = (\lambda/4)\,(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 - F_\pi^2)^2. \tag{398}$$

λ is a dimensionless coupling constant and $F_\pi \approx 150$ MeV is a constant with dimensions of mass, called the pion decay constant for historical reasons [it also controls the rate of weak decay of charged pions in a different model]. The vacua are points on a 3-sphere of radius F_π in R^4, i.e., $\sum_i \phi_i^2 = F_\pi^2$. Suppose initial conditions are such that in the vacuum state, ϕ points along the 4th direction $\phi_4 = F_\pi$ and $\phi_{1,2,3} = 0$ (in the quantum theory these are the vacuum expectation values of the four fields). This vacuum breaks the O(4) symmetry. O(4) transformations in general take this vacuum to other vacua. However, this vacuum is clearly invariant under the subgroup O(3) that rotates the first three components $\vec{\phi}$ among each other. We say that O(4) is broken to O(3) by the non-zero vev of ϕ. Moreover, at any ground state (point on 3-sphere of radius F_π), there are three independent directions tangent to the three sphere, along which the potential does not change. In other words, the hessian evaluated at a vacuum has three zero eigenvalues.

$$\partial_i V = 2\lambda 2\phi_i \delta_{ij}\left(\sum_{k=1}^{4}\phi_k^2 - F_\pi^2\right) \;\Rightarrow\; \partial_i\partial_j V = 4\lambda\delta_{ij}\left(\sum_{k=1}^{4}\phi_k\phi_k - F_\pi^2\right) + \sum_{k=1}^{4} 4\lambda\phi_j 2\phi_k\delta_{ki} = 8\lambda\phi_i\phi_j = 8\lambda F_\pi^2$$

The corresponding zero modes are three massless particles which we identify with the pions. The only non-zero eigenvalue of the hessian is $8\lambda F_\pi^2$. It is the square of the mass of a 'radial' excitation called the σ particle $m_\sigma = 2\sqrt{2\lambda}F_\pi$. The sigma particle has been identified in the hadronic spectrum, but is very short-lived and decays to two pions. Since pions are pseudoscalars, the sigma is a true scalar. The model is named after the sigma particle as well as its originators Gell-Mann and Levy. It is called a linear sigma model since the fields ϕ take values in a linear space $\mathbb{R}^4$: they are maps from 3+1 dimensional Minkowski space to $\mathbb{R}^4$.

• There is a related model called the non-linear sigma model. In the NLSM, the potential V is absent (in a sense $\lambda \to \infty$), but the fields satisfy the constraint $\sum_{k=1}^{4} \phi_i^2 = F_\pi^2$. So fields in the NLSM take values on a 3-sphere (the so-called target space), which is not a linear space, hence its name. Remarkably in addition to low lying excitations which we may identify with the pi mesons, the NLSM also has more massive topologically non-trivial excitations which can be identified with baryons (nucleons). This was discovered by T H R Skyrme. A specific realization is the Skyrme model, where baryons arise as topological solitons of the pion field. This 'effective' model is quite successful in predicting properties of pions, nucleons as well as nuclei.

• More general non-linear sigma models are studied, where fields take values in a 'target space' that is a coset space of a group. For example, string theories are often defined as $1+1$ dimensional non-linear sigma models, with the target space playing the role of a curved 'background' space-time. Of particular interest is the type IIB string sigma model where the target space is the product of a sphere and an anti-deSitter space, $\text{AdS}_5 \times \text{S}^5$. The 1+1 dimensional space time is reinterpreted as the string world sheet while the target space is interpreted as space-time.

• The linear sigma model can be Yukawa-coupled to nucleons, which are Dirac spinors. The pions must couple to the pseudoscalar bilinear in Dirac fields while the sigma particle must couple to the scalar bilinear. Let us rename $\phi_4 = \sigma$, then the Lagrangian is

$$\mathcal{L} = \bar{\psi} i \gamma^\mu \partial_\mu \psi + \frac{1}{2}|\partial\phi|^2 + \frac{1}{2}|\partial\sigma|^2 - \lambda \left(\vec{\phi}^2 + \sigma^2 - F_\pi^2\right)^2 - g\bar{\psi}(\sigma + i\vec{\tau}\cdot\vec{\phi}\gamma_5)\psi. \qquad (399)$$

Notice that there are no mass terms for the nucleons. Nevertheless since σ acquires a non-zero vev (equal to F_π), the Yukawa coupling implies an effective nucleon mass $m_N = gF_\pi$. This relation between nucleon mass, Yukawa coupling and pion decay constant has been experimentally verified. We will say more about this type of Yukawa coupling in the context of the abelian Higgs model.

• In reality, pions are not exactly massless. Their masses can be accommodated by via 'explicit' breaking of O(4), i.e., through the addition of a small O(4) breaking term, say proportional to σ in the potential.

11 Higgs mechanism

• Weak interactions are very feeble and very short ranged. Fermi's 4-fermion contact interaction model did a good job at tree level, but led to incurable divergences when quantum corrections were included, it is not renormalizable, G_F has negative mass dimension $10^{-5} m_p^{-2}$. The feeble and short-ranged nature of weak interactions can be explained if they are mediated by very heavy vector bosons ($M_{W^\pm}$ = 80 GeV). The propagator of a W-boson $1/(p^2 - M_W^2)$ is suppressed for momenta small compared to M_W. However, a theory of massive vector bosons is not renormalizable either. Vector boson masses violate gauge invariance. Without gauge symmetry, the quantum theory cannot be cured of infinities.

• The Higgs mechanism (due to Brout, Englert, Higgs, Kibble, Guralnik and Hagen in 1964) gives a way of making gauge bosons appear massive without spoiling the gauge symmetry. Essentially, minimal coupling of gauge bosons to scalar fields with non-zero vev can make the gauge bosons appear massive at low energies: the gauge symmetry is spontaneously broken. The gauge boson mass is proportional to the vev of the scalar field. The massive gauge boson can then mediate a short ranged force. The Higgs mechanism was used by Weinberg and Salam in 1967 to propose a gauge theoretic model for electroweak interactions, based on an earlier proposal of Glashow. However, it was unclear whether the model could be made free of divergences and used for calculating physical quantities. Gauge theories spontaneously broken by the Higgs mechanism were shown to be renormalizable by Veltman and 't Hooft by 1971-72. In the process, they gave a method (based on dimensional regularization and renormalization) to perform calculations and reliably predict physical quantities. The resulting theory is the electroweak standard model, which has been tested to great accuracy (2-loops in many cases). A not-so-technical account of these developments may be found in Frank Close's recent book, The Infinity Puzzle.

• Analogy with super-conductivity: Non-zero electric and magnetic fields can exist inside an insulator, they are transparent to these fields. The mobile electrons in conductors on the other hand, cancel out any external electric field, so conductors expel electric fields, though they are transparent to magnetic fields. Indeed, we use iron as a core for solenoid. Superconductors are special in that they expel both electric and magnetic fields (Meissner effect). The magnetic field only penetrates a short distance into the super conductor, the London penetration depth. It is as if the photon has become massive in a super conductor, it transmits a short range force. This phenomenon can be described by spontaneous breaking of the U(1) gauge symmetry of Maxwell theory. The Abelian Higgs model (AHM) does precisely this.

11.1 Abelian Higgs Model (AHM) for generating vector boson mass

• In the standard model, the weak gauge bosons 'get their masses' from the Higgs mechanism via the spontaneous breaking of the $SU(2) \times U(1)$ electroweak gauge symmetry. A simpler place to study the Higgs mechanism is the spontaneous breaking of a U(1) gauge symmetry.

• The Abelian Higgs model consists of a complex scalar subject to the so-called Higgs/Mexican hat potential, minimally coupled to a U(1) gauge field. If $D_\mu = \partial_\mu - iA_\mu$, then

$$\mathcal{L} = -\frac{1}{4e^2} F^{\mu\nu} F_{\mu\nu} + |D_\mu \phi|^2 - V(|\phi|). \tag{400}$$

The gauge-invariant quartic self-interaction potential $V(|\phi|) = -m^2|\phi|^2 + \frac{\lambda}{4}|\phi|^4$ is chosen such that for $m^2 > 0$, $\phi = 0$ is a local maximum while the points on the circle $|\phi|^2 = 2m^2/\lambda$ are a 1-parameter family of global minima. Up to an irrelevant additive constant, it is convenient to write

$$V(|\phi|) = \frac{\lambda}{4}(|\phi|^2 - v^2)^2 \quad \text{where} \quad v^2 = \frac{2m^2}{\lambda}. \tag{401}$$

The potential is clearly minimized when $|\phi| = v = \sqrt{\frac{2m^2}{\lambda}}$, and v is called the vacuum expectation (vev) value of the scalar field since in the ground state $\langle |\phi| \rangle = v$.

• If $m^2 < 0$, then the minimum of the potential is at $\phi = 0$, and this minimum energy configuration is invariant under the U(1) symmetry $\phi \to e^{i\theta}\phi$. So for $m^2 < 0$, the g.s. retains this symmetry of the Hamiltonian. However, for $m^2 > 0$, we have a degenerate family of ground states $\phi = ve^{i\theta}$, none of which is invariant under U(1) phase changes. So for $m^2 > 0$, the ground states do not posses the symmetry of the hamiltonian and the U(1) symmetry is said to be spontaneously broken by the non-zero vev v of the scalar field.

• In general, the symmetry group G (U(1) in the case) may not be completely broken, there may be a remnant, the unbroken subgroup H. It is defined as the isotropy group of the vacuum, i.e., the subgroup of G which leaves the vacuum invariant. In our case, the vacuum is any one of $\phi_\theta = ve^{i\theta}$. For $g \in U(1)$ to leave ϕ_θ invariant, we must have $g\phi_\theta = \phi_\theta$, this is possible only if $g = 1$. So in this case, the unbroken subgroup $H = \{1\}$ is trivial. We say that the U(1) gauge symmetry is completely broken by the non-zero vev of the scalar field. In the Weinberg-Salam model, $G = SU(2)_W \times U(1)_Y$ is broken to the subgroup $U(1)_{\rm EM}$ of electromagnetism.

• When a continuous global symmetry is broken (see Gell-Mann-Levy linear sigma model) we have a massless boson in the spectrum, the Nambu-Goldstone boson. The case of spontaneous gauge-symmetry breaking is more intricate. Around a symmetry-broken vacuum, the AHM describes a massive spin one particle (massive vector boson like photon in a superconductor, or Z boson) and a massive neutral scalar particle (like the Higgs particle).

• First let us count field degrees of freedom. If $m^2 < 0$ and the symmetry is unbroken, we have a massless U(1) gauge field (A_μ, a photon) which has two polarizations as well as a complex scalar field ϕ. In total these have 4 real field degrees of freedom (DoF) at each space-time point. The number of DoF should not change upon SSB. Indeed, a massive vector boson has three possible polarization states and a massive real scalar accounts for one field degree of freedom.

• To see these features around a symmetry broken vacuum, we write the Lagrangian in terms of new variables. Let $\phi = (v+h)e^{i\theta}$ where h is a real field. If $h = 0$ and θ is a constant, then this is a minimum energy state. h will describe fluctuations around the vacuum (a massive neutral scalar, like the Higgs particle) and we will see that θ can be eliminated from the Lagrangian. This has to be the case, since we expect the gauge field A_μ to describe three field dof (massive vector boson) and the real scalar h should be the 4th field degree of freedom. To begin with, the scalar potential is simply

$$V = \frac{\lambda}{4}(|\phi|^2 - v^2)^2 = \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda h^4}{4}. \tag{402}$$

The first term is a mass term for the scalar h field while the next two describe cubic and quartic self-interaction vertices of h. λ is called the Higgs self-coupling. $\theta(x)$ dropped out of V since it only depends on $|\phi|^2$ to satisfy gauge invariance. The square of the covariant derivative of ϕ, $|D_\mu\phi|^2$ can also be made independent of θ by a suitable gauge transformation. Indeed, if we view $\phi \to e^{i\theta}\phi$ as a gauge transformation, then the corresponding gauge transformation takes $A \to A + \partial\theta$. So this suggests we define a new gauge field $B_\mu = A_\mu - \partial_\mu\theta$. Then

$$\begin{aligned} \partial_\mu\phi &= \partial_\mu\left[(v+h)e^{i\theta}\right] = \left[\partial_\mu h + i(v+h)\partial_\mu\theta\right]e^{i\theta} \quad \text{and} \\ D_\mu\phi &= (\partial_\mu - i(B_\mu + \partial_\mu\theta))\phi = \left[\partial_\mu h - iB_\mu(v+h)\right]e^{i\theta} \end{aligned} \tag{403}$$

Thus θ drops out of the square of the covariant derivative

$$|D_\mu\phi|^2 = |\partial_\mu h - iB_\mu(v+h)|^2 = (\partial h)^2 + v^2B^2 + 2vhB^2 + B^2h^2. \tag{404}$$

$(\partial h)^2$ is the kinetic term for a real scalar, v^2B^2 is a mass term for the gauge field while the last two terms describe interactions between the massive gauge field and the scalar particle. Finally, the Maxwell Lagrangian $-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$ for the gauge field A_μ being gauge-invariant takes the same form in terms of B_μ

$$-\frac{1}{4e^2}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) = -\frac{1}{4e^2}(\partial_\mu B_\nu - \partial_\nu B_\mu)(\partial^\mu B^\nu - \partial^\nu B^\mu). \tag{405}$$

Adding the three terms, the Lagrangian in the new variables upon extracting the gauge coupling from the gauge field, ($B \to eB$) is

$$\mathcal{L} = -\frac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)(\partial^\mu B^\nu - \partial^\nu B^\mu) + e^2v^2B^2 + (\partial h)^2 - \lambda v^2h^2 - \lambda v h^3 - \frac{\lambda h^4}{4} + 2e^2vhB^2 + e^2B^2h^2. \tag{406}$$

This Lagrangian describes a massive vector field B_μ (massive photon, $m_B^2 = 2e^2v^2$, see below) coupled to a massive real scalar field h (Higgs field $m_h^2 = \lambda v^2$). Notice that the scalar does not have derivative interactions with the vector field: the scalar is uncharged (neutral) with respect to the $U(1)$ gauge symmetry. Moreover, in this form, the Lagrangian is not manifestly gauge-invariant, due to the vector boson mass term $e^2v^2B^\mu B_\mu$. However, this is an artifact of writing $\mathcal{L}$ in variables adapted to a vacuum state $\phi = ve^{i\theta}$ that is not invariant under the U(1) symmetry. The model is in fact gauge-invariant as it was in the original variables.

• There are 4 types of interaction vertices, $\lambda v h^3 + \frac{\lambda h^4}{4} + 2e^2vhB^2 + e^2B^2h^2$ which describe self-interactions of the Highs particle and interactions between the Higgs and the massive vector boson. They describe scattering and decay vertices (possibly virtual) such as $h \to hh$ (higgs radiating a virtual higgs), $hh \to hh$ (higgs-higgs scattering), $h \to BB$ and $hB \to hB$.

• The massive vector boson B^μ has three physical degrees of freedom at each location. One way to see this is to consider the above Lagrangian for B, ignoring interactions with h. This is the non-gauge-invariant Proca Lagrangian

$$\mathcal{L}_{\text{Proca}} = -\frac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)(\partial^\mu B^\nu - \partial^\nu B^\mu) + e^2v^2B^2. \tag{407}$$

The resulting equation of motion is the Proca equation $\partial_\mu F^{\mu\nu} + 2e^2v^2B^\nu = 0$ where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. Since $F^{\mu\nu}$ is antisymmetric, taking the divergence of the Proca equation gives us the consistency condition $\partial_\nu B^\nu = 0$[53]. This is one constraint on the 4 components of B^μ, leaving only three propagating degrees of freedom for a massive vector field. Each component of the Proca field B^μ satisfies the massive KG equation for a real scalar field

$$\partial_\mu \partial^\mu B^\nu + 2e^2v^2B^\nu = 0 \quad \text{or} \quad (\Box + m_B^2)B^\nu = 0. \tag{408}$$

From this we read off the mass of the photon $m_B^2 = 2e^2v^2$ resulting from the Higgs mechanism. This is the tree-level mass ignoring quantum effects of interactions, which can *and do* modify the mass slightly, as one finds in perturbation theory.

[53]If we had kept the interaction terms, we would get a more complicated consistency condition depending on the other fields, but it too would impose one constraint among the four components of B.

11.2 Dirac fermion masses from Abelian Higgs mechanism

• In the electroweak standard model, the quarks and charged leptons 'get their masses' by the Higgs mechanism via the spontaneous breaking of the $SU(2) \times U(1)$ electroweak gauge symmetry. The Abelian Higgs model is a simpler context in which to study how spin half Dirac fermions can get masses.

• Consider a Dirac field ψ Yukawa-coupled (g below is the Yukawa coupling constant) to a complex scalar $\phi = \phi_1 + i\phi_2$ subject to a Mexican hat potential

$$\mathcal{L} = |\partial\phi|^2 + \bar{\psi} i\gamma \cdot \partial\psi - \frac{\lambda}{4}(|\phi|^2 - v^2)^2 - g\bar{\psi}(\phi_1 + i\gamma_5\phi_2)\psi \tag{409}$$

Notice that the real part of ϕ couples to the scalar $\bar{\psi}\psi$ while the imaginary part couples to the pseudo-scalar $\bar{\psi}\gamma_5\psi$. ϕ_1 is assumed to be a scalar while ϕ_2 is taken to be a pseudoscalar, so that as a whole, $\mathcal{L}$ is parity even. This seemingly peculiar coupling is needed to ensure that $\mathcal{L}$ possesses a U(1) global symmetry that we mention below, the symmetry would be lost if both ϕ_1 and ϕ_2 coupled to the scalar $\bar{\psi}\psi$.

• The energy is minimized when $|\phi| = v$. So if ϕ_1 has a non-zero vacuum expectation value v, then the Yukawa interaction mimics a mass term for the Dirac field, $m_\psi = gv$.

• This Lagrangian has an interesting $U(1)$ global symmetry under the compensating phase changes of the complex scalar and Dirac spinor

$$\phi \to e^{i\theta}\phi, \quad \text{and} \quad \psi \to e^{-\frac{i}{2}\gamma_5\theta}\psi. \tag{410}$$

Use the fact that $\gamma_5^2 = I$ (so $e^{i\theta\gamma_5} = \cos\theta + i\sin\theta$) and that $\{\gamma_5, \gamma^\mu\} = 0$ to show that under this transformation,

$$\bar{\psi} \to \bar{\psi} e^{-\frac{i}{2}\gamma_5\theta} \quad \text{and} \quad (\phi_1 + i\gamma_5\phi_2) \to e^{i\theta\gamma_5}(\phi_1 + i\gamma_5\phi_2). \tag{411}$$

As a consequence, both $\bar{\psi}\partial\!\!\!/\psi$ and $\bar{\psi}(\phi_1 + i\gamma_5\phi_2)\psi$ are invariant under this global U(1) symmetry. The seemingly peculiar coupling of the real and imaginary parts of ϕ to the scalar and pseudoscalar bilinears in Dirac fields was essential for this to work.

• So far, we have a complex scalar in the Mexican hat potential coupled to a Dirac fermion with Yukawa interactions giving a mass to the fermion when the scalar has a non-zero vev. We may gauge the above global U(1) symmetry and arrive at a theory of spin zero, half and one fields where the Higgs mechanism gives masses to the fermion as well as gauge boson.

• Minimal coupling of the scalar and Dirac fields to a U(1) gauge field A_μ is effected via different covariant derivatives

$$D_\mu\phi = (\partial_\mu - iA_\mu)\phi \quad \text{and} \quad D_\mu\psi = (\partial_\mu + \frac{1}{2}i\gamma_5 A_\mu)\psi. \tag{412}$$

These ensure that $D_\mu\phi$ and $D_\mu\psi$ transform in the same way as ϕ, ψ. Indeed,

$$\psi \to g\psi \qquad \text{where} \;\; g(x) = e^{-\frac{i}{2}\gamma_5\theta}\psi \quad \text{and} \quad A_\mu \to A_\mu + \partial_\mu\theta$$

$$\Rightarrow \qquad D_\mu\psi \to \left(g\partial_\mu\psi - \frac{1}{2}i\gamma_5(\partial_\mu\theta)g + \frac{1}{2}i\gamma_5 A_\mu g + \frac{1}{2}i\gamma_5(\partial_\mu\theta)g\right)\psi = gD_\mu\psi. \quad (413)$$

[$g(x)$ in these equations is not to be confused with the Youkawa coupling constant g!] Thus we may write down a gauge invariant Lagrangian

$$\mathcal{L} = -\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma\cdot D - g(\phi_1 + i\gamma_5\phi_2))\psi + (D_\mu\phi)^*(D_\mu\phi) - \frac{\lambda}{4}(|\phi|^2 - v^2)^2. \quad (414)$$

By transforming to variables adapted to the vacua $|\phi| = v$:

$$\phi = (v+h)e^{i\theta}, \quad \psi = e^{-\frac{1}{2}i\gamma_5\theta}\chi, \quad B_\mu = A_\mu - \partial_\mu\theta. \quad (415)$$

it is possible to write the Lagrangian in terms of a massive neutral scalar Higgs field h, a massive fermion χ and a massive photon B_μ just as was done for the abelian Higgs model in the previous section. θ drops out of the Lagrangian as before, due to the gauge symmetry. Show that the masses are

$$m_B = \sqrt{2}ev, \quad m_h = \sqrt{\lambda}v \quad \text{and} \quad m_\chi = gv. \quad (416)$$

We notice that the fermion and vector boson masses are proportional to the vev v of the scalar field. It is the non-zero vev of the scalar field that is responsible for these particles having non-zero masses. Writing out the interaction terms in the Lagrangian one also finds a trilinear coupling of the fermions to the Higgs scalar $g\bar{\chi}h\chi$. Thus both the mass of the fermion as well as its Yukawa coupling are proportional to g. This means a heavier fermion couples more strongly to the Higgs than a light fermion. This feature continues to hold in the standard model, where there are several fermions - the charged leptons and quarks, all of which get their masses via the spontaneous breaking of the non-abelian EW gauge symmetry SU$(2)\times$ U(1) by the Higgs mechanism. The top quark couples much more strongly to the Higgs particle than the electron does. What is more, due to the γ_5 in $D_\mu\psi$, the massive vector boson B_μ couples to the axial vector current $\bar{\chi}\gamma^\mu\gamma_5\chi$ of the massive Dirac fermion rather than to the vector current.

11.3 Higgs mechanism in an SU(2) gauge theory

• We now generalize the abelian (U(1)) Higgs model to a gauge theory associated to a non-abelian group SU(2). This is a step closer to the SU$(1)\times$ U(1) gauge theory relevant to the electroweak standard model. Since there are three weak gauge bosons and $su(2)$ is a 3d Lie algebra, it is a natural example to consider. Upper case SU(2) refers to the group and lower case $su(2)$ to the Lie algebra.

• The non-abelian gauge fields of $su(2)$, A_μ transform in the adjoint representation of SU(2) (so they are 2×2 matrices) and are coupled to a complex doublet of scalar fields $\phi = \begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}$ that transform in the fundamental representation of SU(2). Since $su(2)$ is three dimensional, there are three gauge bosons. The covariant derivative is $D_\mu\phi = \partial_\mu\phi - ieW_\mu\phi$ where $W_\mu = W_\mu^a t^a$, $t^a = \frac{1}{2}\sigma^a$ and σ^a are the Pauli matrices which furnish a basis for the Lie algebra, with $\mathrm{tr}\ t^a t^b = \frac{1}{2}\delta^{ab}$. Then the Lagrangian is

$$\mathcal{L} = -\frac{1}{2}\,\mathrm{tr}\ F^{\mu\nu}F_{\mu\nu} + (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi). \quad (417)$$

If $V(\phi) = 0$, since there is no mass term for the gauge bosons, this would describe 10 propagating field degrees of freedom: three gauge bosons, each with two independent polarizations and four real scalar fields.

• A scalar potential that is a function of $\phi^\dagger \phi$ is automatically gauge invariant, and we take the Mexican hat potential

$$V(\phi) = \frac{\lambda}{4}(\phi^\dagger \phi - v^2)^2 \tag{418}$$

The scalar potential is minimized when $\phi^\dagger \phi = |\phi_1|^2 + |\phi_2|^2 = v^2$. The space of minima (vacua) is called the vacuum manifold. In this case it is a 3-sphere S^3 embedded in $\mathbb{R}^4$ since the complex doublet of scalar fields have four real components. If the vacuum value of the scalar field is non-zero (e.g. $\phi = \begin{pmatrix} 0 \\ v \end{pmatrix}$), then the SU(2) symmetry is spontaneously broken.

• $G =$ SU(2) is completely broken to $H = \{I\}$ since the isotropy subgroup of the vacuum (sub group that leaves the vacuum invariant) is trivial. For,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 \\ v \end{pmatrix} = \begin{pmatrix} bv \\ dv \end{pmatrix} = \begin{pmatrix} 0 \\ v \end{pmatrix} \quad \Rightarrow \quad b = 0, \; d = 1 \tag{419}$$

and unitarity and determinant one force $c = 0$ and $a = 1$.

• If SU(2) were a global symmetry then there would be three massless Nambu-Goldstone bosons corresponding to the three zero modes of the potential about any vacuum (3 independent directions we may move in S^3 without changing the potential). When the local gauge symmetry SU(2) is spontaneously broken, it is as if these three massless Goldstone bosons are eaten by the three vector bosons, making them massive. A massive spin one gauge field has three modes of polarization (three spin projections) unlike a massless spin one particle, which only has two helicities. It is as if the Goldstone bosons become the longitudinal polarization states of the vector bosons. We will see below that SU(2) is completely broken, all three vector bosons become equally massive and there is also a neutral scalar Higgs particle in the spectrum. Thus the number of propagating fields remains the same (10) after SSB: 3 massive spin one particles and a spin zero particle.

• To find the masses of the gauge bosons after SSB we work in the gauge $\phi = \begin{pmatrix} 0 \\ v+h \end{pmatrix}$ where v, h are real and expand out the square of the covariant derivative of ϕ keeping terms quadratic in fields. Denote $W^\pm = W^1 \pm i W^2$. Then,

$$\begin{aligned} D_\mu \phi &= \partial_\mu \phi - ie W_\mu \phi = \begin{pmatrix} 0 \\ \partial_\mu h \end{pmatrix} - \frac{ie}{2} \begin{pmatrix} W^3_\mu & W^-_\mu \\ W^+_\mu & -W^3_\mu \end{pmatrix} \begin{pmatrix} 0 \\ v+h \end{pmatrix} = \begin{pmatrix} -\frac{ie}{2} W^-_\mu (v+h) \\ \partial_\mu h + \frac{ie}{2} W^3_\mu (v+h) \end{pmatrix} \\ \text{and} \quad & (D_\mu \phi)^\dagger (D^\mu \phi) = (\partial h)^2 + \frac{e^2 v^2}{4} \left[(W^1_\mu)^2 + (W^2_\mu)^2 + (W^3_\mu)^2 \right] + \text{interactions}. \end{aligned} \tag{420}$$

The second term is the mass term for gauge bosons, all three are equally massive. To identify the numerical factors in the mass, one compares with the kinetic term for gauge fields from the YM Lagrangian

$$-\frac{1}{2} \operatorname{tr} F_{\mu\nu} F^{\mu\nu} = -\frac{1}{2} \operatorname{tr} t^a t^b \left[2\partial_\mu W^a_\nu \partial^\mu W^{\nu b} - 2\partial_\mu W^a_\nu \partial^\nu W^{\mu b} + \text{cubic and quartic interactions} \right]$$

$$= \frac{1}{2}\left[W^a_\nu \Box W^{\nu a} - W^a_\nu \partial^\nu \partial_\mu W^{\mu a} + \text{divergence and interaction terms}\right] \quad (421)$$

The equation of motion for W^a_μ is $\Box W^a_\mu + \frac{e^2v^2}{2} W^a_\mu + \ldots = 0$. So we have three spin one gauge bosons with equal masses $m_{W^a} = \frac{ev}{\sqrt{2}}$ for $a = 1, 2, 3$.

• The mass of the the Higgs field h is similarly obtained by expanding the potential. Note that $\phi^\dagger\phi = (v+h)^2$, so the Higgs Lagrangian becomes

$$(\partial h)^2 - V(\phi) = (\partial h)^2 - \frac{\lambda}{4}(\phi^\dagger\phi - v^2)^2 = (\partial h)^2 - \frac{\lambda}{4}(4v^2h^2 + 4vh^3 + h^4) = (\partial h)^2 - (\lambda v^2)h^2 + \text{interactions.} \quad (422)$$

We read off $m_H = v\sqrt{\lambda}$.

• This spontaneously broken $su(2)$ gauge theory is interesting but cannot be appropriate for the weak interactions since the W and Z do not have the same masses. Moreover, the strengths of the neutral and charged weak interactions aren't quite the same. This indicates $su(2)$ cannot be the gauge Lie algebra of the weak interactions. We now proceed to the idea that worked.

12 SU$(2)_W\times$ U$(1)_Y$ Electroweak gauge theory

Parity transforms a LH Dirac fermion into a RH one. Parity violation in the weak interactions is because the LH and RH components couple differently to the weak gauge bosons. Indeed, Marshak and Sudarshan, Feynman and Gell-Mann found that the experimental data favors the charged weak currents to have a V-A structure. As a consequence, only the LH projections of the quarks and leptons participate in the charge-changing weak interactions mediated by $W^\pm$, parity is maximally violated. $W^\pm$ must be very heavy, to explain the weakness of the weak force in beta decay. In 1973, more feeble 'neutral' weak interactions were discovered, mediated by the Z^0. The masses of the weak gauge bosons were measured in 1983, $M_W = 80, M_Z = 91$ GeV. Unlike the $W^\pm$, the Z couples to both LH and RH components of fermions, though not in the same way, so Z interactions also violate parity. The only renormalizable theory of massive vector bosons we know of, is a spontaneously broken (via the Higgs mechanism) gauge theory coupled to scalar fields. The gauge group would have to have dimension 3 to accommodate $W^\pm$ and Z. However, it cannot be SU(2) which leads to 3 gauge bosons of equal mass while $M_W \neq M_Z$. Moreover, the fact that $W^\pm$ do not couple to RH fermions while Z^0 couples to both LH and RH fermions, and that the strengths of the charged and neutral weak interactions are slightly different suggests we cannot have a simple gauge group: we need two coupling constants. Moreover, we should bear in mind that SSB via the Higgs mechanism can be partial, some gauge symmetries may be broken while others may survive. Now recall that in addition to these weak gauge bosons, we also have the massless photon of an unbroken U(1) gauge theory: electromagnetism. The photon couples in the same way to both LH and RH Dirac fermions and conserves parity. So we seek a 4d gauge lie algebra that can be broken to the u(1) of EM, leaving three massive weak gauge bosons.

12.1 Electroweak mixing, Higgs mechanism, $W^{\pm}, Z^0, \gamma$

• The 4d unitary Lie algebra $u(2) \cong su(2)_W \times u(1)_Y$ is an obvious choice, it is a product of weak isospin and weak hyper charge. It consists of 2×2 hermitian matrices. Any such hermitian matrix is a linear combination of the identity and the Pauli matrices. The identity generates $U(1)_Y$ while the Pauli matrices generate $SU(2)_W$. Since $u(2)$ is a product of a simple factor and a u(1), the gauge theory can have two independent coupling constants e_2 and e_1 respectively.

• As we will see, $su(2) \times u(1)$ can be broken by the Higgs mechanism to $u(1)$ of electromagnetism (which is not the same as $u(1)_Y$). The three gauge bosons $W_\mu^\pm, W_\mu^0$ of $su(2)_W$ couple only to LH components of quarks and leptons. $W^\pm$ become massive as a consequence of SSB via the Higgs mechanism. The weak hypercharge gauge boson Y_μ couples to both LH and RH components, but in different ways. There is a linear combination of W^0 and Y which couples in the same way to both LH and RH fermions, i.e. to the vector current constructed from Dirac spinors. It is identified with the photon field A_μ, it remains massless and corresponds to the unbroken u(1) of electric charge. The coefficients in the linear combination are determined by the gauge couplings e_1 and e_2. The orthogonal linear combination of W_μ^0 and Y_μ is the massive Z_μ^0. It mediates the weak neutral interactions which violate parity. In this section we focus on the electroweak gauge bosons and the scalar fields responsible for SSB. We will include the quarks and leptons subsequently.

• As in the case of SU(2) gauge theory, we consider a complex doublet of scalar fields $\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$ which transform in the fundamental representation of $\mathrm{U}(2)$ gauge group. The gauge-invariant scalar potential we consider is as before: $V(\phi) = \frac{\lambda}{4}(\phi^\dagger \phi - v^2)^2$.

• The u(2) gauge field is a hermitian 2×2 matrix at each space-time point. We choose as basis the identity and Pauli matrices, which span commuting $u(1)$ and $su(2)$ sub-algebras, $W_\mu = W_\mu^a \frac{\sigma_a}{2}$. We may write the covariant derivative of the scalar field as

$$D_\mu \phi = \partial_\mu \phi - i \left(\frac{1}{2} e_1 Y_\mu + e_2 \sum_{a=1}^{3} \frac{1}{2} \sigma_a W_\mu^a \right) \phi. \tag{423}$$

As before we use $W^\pm$ to denote the combinations $W_\mu^\pm = W_\mu^1 \pm i W_\mu^2$.

• The gauge field kinetic and self-interactions are given by the Yang-Mills Lagrangian

$$-\frac{1}{4} Y^{\mu\nu} Y_{\mu\nu} - \frac{1}{2} \operatorname{tr} W^{\mu\nu} W_{\mu\nu}. \tag{424}$$

The field strengths are

$$Y_{\mu\nu} = \partial_\mu Y_\nu - \partial_\nu Y_\mu \quad \text{and} \quad W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - i e_2 [W_\mu, W_\nu]. \tag{425}$$

The possibility of having two independent coupling constants can also be understood by thinking directly in terms of a $u(2)$ gauge field A_μ which is a hermitian matrix field with no constraint on its trace (unlike in the case of $su(2)$ it need not be traceless). Let $F_{\mu\nu} =$

$\partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]$. Now one can form two independent gauge- invariant and Lorentz invariant quantities that involve at most two derivatives, these are $\operatorname{tr} F_{\mu\nu}F^{\mu\nu}$ and $\operatorname{tr} F_{\mu\nu} \operatorname{tr} F^{\mu\nu}$. The latter was identically zero for $su(2)$ but is non-trivial here. The Lagrangian can then be an arbitrary linear combination of these two, giving two independent coupling constants.

• The minima of the scalar potential $\phi^\dagger\phi = v^2$ form a 3-sphere. We choose the vacuum as $\phi = \begin{pmatrix} 0 \\ v \end{pmatrix}$ where v is a real number. The isotropy subgroup of the vacuum is non-trivial in this case, in fact it is $\mathrm{U}(1)$ (good since we want an unbroken $\mathrm{U}(1)$ of electromagnetism):

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 \\ v \end{pmatrix} = \begin{pmatrix} bv \\ dv \end{pmatrix} = \begin{pmatrix} 0 \\ v \end{pmatrix} \quad \Rightarrow \quad b = 0, d = 1 \tag{426}$$

and unitarity implies $c = 0$ and $|a| = 1$, leaving an arbitrary phase $a \in U(1)$. Thus the residual symmetry group is $\mathrm{U}(1)$, we will see that the corresponding massless gauge boson is the photon field.

• The hypercharges of W_μ^a and Y_μ are all zero while the third component of weak isospin T_3 is ± 1 for $W^\pm$ and zero for W_μ^3 and Y_μ. We will define the electric charge as $Q = T_3 + Y/2$ (for reasons to be indicated below), so $W^\pm$ have electric charges ± 1 in units of the positron charge, while W^3 and Y are electrically neutral.

Minimal coupling to the $\mathrm{U}(2)$ scalar doublet is effected via the covariant derivative

$$D_\mu\phi = \partial_\mu\phi - i\left(\frac{1}{2}e_1 Y_\mu + e_2 \frac{\sigma_a}{2} W_\mu^a\right)\phi = \begin{pmatrix} -\frac{ie_2}{2}(v+h)W_\mu^- \\ \partial_\mu h - \frac{ie_1}{2}(v+h)Y + \frac{ie_2}{2}(v+h)W_\mu^3 \end{pmatrix}. \tag{427}$$

The Lagrangian is

$$\mathcal{L} = -\frac{1}{4}Y^{\mu\nu}Y_{\mu\nu} - \frac{1}{4}W^{a\mu\nu}W_{\mu\nu}^a + (D_\mu\phi)^\dagger(D_\mu\phi) - \frac{\lambda}{4}(\phi^\dagger\phi - v^2)^2. \tag{428}$$

The quadratic terms in gauge fields are normalized as for real scalar fields. Up to terms that are 4-divergences,

$$\mathcal{L} = \frac{1}{2}\left(Y_\mu \Box Y^\mu - Y_\nu \partial^\nu \partial_\mu Y^\mu + W_\mu^a \Box W^{\mu a} - W_\nu^a \partial^\nu \partial_\mu W^{\mu a}\right) + \ldots. \tag{429}$$

To identify the gauge and Higgs boson masses, we go to the gauge where $\phi = \begin{pmatrix} 0 \\ v+h \end{pmatrix}$ and expand the Lagrangian to quadratic order in the fluctuating fields W, Y, h. The square of the covariant derivative of the scalar fields is

$$\begin{aligned} (D_\mu\phi)^\dagger(D^\mu\phi) &= (\partial h)^2 + \frac{e_2^2(v+h)^2}{4}W_\mu^- W^{\mu +} + \frac{e_1^2(v+h)^2}{4}Y^2 + \frac{e_2^2(v+h)^2}{4}(W_\mu^3)^2 - \frac{e_1 e_2 (v+h)^2}{2}Y_\mu W^{3\mu} \\ &= (\partial h)^2 + \frac{e_1^2 v^2}{4}Y^2 + \frac{e_2^2 v^2}{4}\left((W_\mu^1)^2 + (W_\mu^2)^2 + (W_\mu^3)^2\right) - \frac{e_1 e_2 v^2}{2}W_\mu^3 Y_\mu + \text{interactions}. \end{aligned} \tag{430}$$

Putting $|D\phi|^2 = (\partial h)^2 + \frac{1}{2}M_{ab}X_\mu^a X^{b\mu}$ where $X_\mu = (Y_\mu, W_\mu^1, W_\mu^2, W_\mu^3)$ we identify a *symmetric* mass^2 matrix for gauge bosons

$$M_{ab} = \frac{v^2}{2}\begin{pmatrix} e_1^2 & 0 & 0 & -e_1 e_2 \\ 0 & e_2^2 & 0 & 0 \\ 0 & 0 & e_2^2 & 0 \\ -e_1 e_2 & 0 & 0 & e_2^2 \end{pmatrix}. \tag{431}$$

It is clear that W^1 and W^2 (or the charged weak bosons $W^{\pm}$) have definite and equal masses $m_W^2 = \frac{1}{2} e_2^2 v^2$. But Y and W^3 do not have definite masses as the mass^2 matrix is not diagonal in this basis. The neutral electroweak gauge bosons with definite masses (which we identify with the photon A_μ and Z^0 boson Z_μ) are the remaining eigenvectors of the mass^2 matrix. Since the matrix has zero determinant, the remaining eigenvalues are 0 and $\frac{1}{2} v^2 (e_1^2 + e_2^2)$ with corresponding eigenvectors

$$A_\mu = \frac{e_1 W_\mu^3 + e_2 Y_\mu}{\sqrt{e_1^2 + e_2^2}} = \sin\theta_W W_\mu^3 + \cos\theta Y_\mu \quad \text{and} \quad Z_\mu = \frac{e_2 W_\mu^3 - e_1 Y_\mu}{\sqrt{e_1^2 + e_2^2}} = \cos\theta W_\mu^3 - \sin\theta_W Y_\mu. \tag{432}$$

It is natural to identify A_μ with the photon field, since it is the only massless gauge boson in the model. Thus, the masses of the four gauge bosons are

$$m_{W^\pm} = \frac{1}{2} v^2 e_2^2, \quad m_Z = \frac{1}{2} v^2 (e_1^2 + e_2^2) \quad \text{and} \quad m_\gamma = 0. \tag{433}$$

We defined the weak mixing angle or Weinberg angle θ_W (originally introduced by Glashow)

$$\sin\theta_W = \frac{e_1}{\sqrt{e_1^2 + e_2^2}}, \quad \cos\theta_W = \frac{e_2}{\sqrt{e_1^2 + e_2^2}}, \quad \tan\theta_W = \frac{e_1}{e_2} \;\; \text{and} \;\; e_2 \sin\theta_W = e_1 \cos\theta_W. \tag{434}$$

θ_W measures the extent of mixing between weak hypercharge and weak isospin [if $\theta_W = 0$, then $A_\mu = Y_\mu$ and $Z_\mu = W_\mu^3$] or the relative strengths of the corresponding gauge couplings e_1 and e_2. θ_W is a 'running' parameter, it has been measured to be about 30° or $\sin^2\theta_W \approx 0.23$ at an energy scale corresponding to the mass of the Z boson, 91 GeV. The ratio of the masses of the W and Z bosons is

$$\frac{m_W}{m_Z} = \frac{v e_2}{v\sqrt{e_1^2 + e_2^2}} = \cos\theta_W. \tag{435}$$

• So far, the EM coupling and electric charge have not appeared in the Weinberg-Salam model. They need to be *defined* in such a way that the model agrees with QED when we consider the interactions of charged particles with photons. To relate the electromagnetic coupling e to the couplings e_1 and e_2 we write the covariant derivative in terms of A_μ and Z_μ (instead of Y_μ and W_μ^3) and compare with the usual electromagnetic coupling $\partial_\mu - ieQA_\mu$ to charged particles, where $e = \sqrt{4\pi\hbar c\alpha}$ is the electromagnetic coupling and Q is the charge in units of the positron charge. Here we work with the covariant derivative of the Higgs doublet. Being a complex scalar, it is charged, though the two components can have different charges, as we will see. Using

$$Y_\mu = \cos\theta_W A_\mu - \sin\theta_W Z_\mu \quad \text{and} \quad W_\mu^3 = \sin\theta_W A_\mu + \cos\theta_W Z_\mu, \tag{436}$$

the covariant derivative for the scalar doublet

$$\begin{aligned} D_\mu &= \partial_\mu - \frac{ie_2}{2}(\sigma_1 W_\mu^1 + \sigma_2 W_\mu^2) - \frac{ie_1}{2} Y_\mu - \frac{ie_2}{2}\sigma_3 W_\mu^3 \\ &= \partial_\mu - \frac{ie_2}{2}(\sigma_1 W_\mu^1 + \sigma_2 W_\mu^2) - \frac{i}{2}\left(e_1 I \cos\theta + e_2 \sigma_3 \sin\theta\right) A_\mu - \frac{i}{2}\left(e_2 \sigma_3 \cos\theta - e_1 I \sin\theta\right) Z_\mu. \end{aligned}$$

However, this must reduce to the usual electromagnetic coupling to the photon $\partial_\mu - iQeA_\mu$, where Q is the charge matrix for the particles under consideration, and e is the electromagnetic coupling. Comparing, we see that eQ must be a linear combination of $\frac{1}{2}\sigma_3 = T_3$ (corresponding to the 3rd component of weak isospin) and the identity $I/2$ (corresponding to weak hypercharge) and e should be linearly related to e_1 and e_2 through coefficients that are trigonometric functions of the weak mixing angle. e, e_1, e_2, θ_W are all 'running' parameters while Q is a constant matrix depending on the charged particles being considered. There is some freedom in how one arranges the factors. We will define the charge matrix as the linear combination $Q = \frac{\sigma_3}{2} + \frac{Y}{2}$, where Y is the hypercharge matrix and e is the electromagnetic coupling constant (some authors absorb the half into Y). So we must have

$$-ieQA_\mu = -ie\left(\frac{\sigma_3}{2} + \frac{Y}{2}\right)A_\mu = -\frac{i}{2}(e_2 \sin\theta\sigma_3 + e_1\cos\theta\, I)A_\mu \tag{437}$$

Now the hypercharge matrix of the scalar doublet is $Y = I$ (this is explained in the section on fermion masses) so we must have

$$e = e_2 \sin\theta_W = e_1 \cos\theta_W. \tag{438}$$

• We have seen that the Fermi coupling of low energy beta decay is proportional to the square of the weak coupling e_2 and inversely proportional to the square of M_W (from the W propagator)

$$G_F = \frac{e_2^2}{4\sqrt{2}m_W^2} = \frac{1}{2\sqrt{2}v^2} \quad \Rightarrow \quad v = \frac{1}{\sqrt{2\sqrt{2}G_F}} \approx 174.3\,\text{GeV}. \tag{439}$$

We used $m_W^2 = \frac{1}{2}v^2e_2^2$ from the GWS model and the known value of $G_F \approx 1.165 \times 10^{-5}$ GeV^{-2} to estimate the vev of the Higgs field. Sometimes $\sqrt{2}v = 246.4$ GeV is quoted. The weak mixing angle was determined in the mid 1970s through a measurement of the strength of weak neutral currents (relative to the weak charged currents), giving $\sin^2\theta_W \approx 0.23$. Thus it became possible to predict the masses of the W and Z bosons using known values for the Higgs vev $v \approx 174.3$ GeV, electromagnetic coupling $e = \sqrt{4\pi\alpha_{em}} \approx 0.31$ (we use the value $\alpha = 1/129$ valid at 91 GeV) and $\sin\theta_W \approx 0.48$, $\cos\theta \approx .88$:

$$m_W = \frac{1}{\sqrt{2}}ve_2 = \frac{ve}{\sqrt{2}\sin\theta_W} \approx 79.6\text{GeV} \quad \text{and} \quad m_Z = \frac{m_W}{\cos\theta_W} = \frac{ev}{\sqrt{2}\sin\theta\cos\theta} \approx 90.4\text{GeV}. \tag{440}$$

The W and Z were discovered at roughly the predicted masses in 1983 at CERN.

• The mass of the Higgs field h is similarly obtained by expanding the potential. Note that $\phi^\dagger\phi = (v+h)^2$, so the Higgs Lagrangian becomes

$$(\partial h)^2 - V(\phi) = (\partial h)^2 - \frac{\lambda}{4}(\phi^\dagger\phi - v^2)^2 = (\partial h)^2 - \frac{\lambda}{4}(4v^2h^2 + 4vh^3 + h^4) = (\partial h)^2 - (\lambda v^2)h^2 + \text{interactions}. \tag{441}$$

We read off $m_H = v\sqrt{\lambda}$. The spin zero Higgs particle with mass $m_H \approx 125$ GeV was discovered at CERN in 2012. This fixes the dimensionless Higgs self coupling $\lambda = (m_H/v)^2 \approx 0.5$. The fact that the measured value of λ is not large justifies a perturbative treatment of the scalar sector of the SM.

12.2 Minimal coupling to fermions in the SU$(2)\times$ U(1) model

Quarks and leptons are modeled as Dirac spinors in the standard model. To accommodate parity violation as encoded in the V-A theory of weak interactions, the left and right handed projections of the Dirac fermions (quarks and leptons) behave differently[54] under SU$(2)\times$ U(1). Let us first consider the first generation ((ν_e, e), and (u,d)), ignoring mixing among generations. The LH components of a Dirac spinor are those that live in the -1 eigenspace of γ_5 while the RH components lie in the $+1$ eigenspace of γ_5. So $\gamma_5 \nu_{eL} = -\nu_{eL}$ and $P_L \nu_{eL} = \frac{1}{2}(I - \gamma_5)\nu_{eL} = \nu_{eL}$ etc. The property of being right or left handed is a Lorentz-invariant concept, unlike positive or negative helicity which is well-defined only for massless particles.

• Parity violation and V-A implies that only the LH projections of quarks and leptons participate in the charge changing weak interactions. Since these $W^\pm$ interactions arise from the $SU(2)_W$ group, the LH projections of quarks and leptons must transform non-trivially under weak isospin, the simplest possibility is that they transform in the fundamental representation. On the other hand, the RH projections do not participate in the charge changing weak interactions, so they must be singlets under SU$(2)_W$.

• The LH components of the quarks and leptons each transform as a doublet in the fundamental representation of weak isospin. In the case of quarks, we have three such doublets, one for each color $\alpha = 1,2,3$:

$$\text{LH weak isospin doublets} \quad l_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad \text{and} \quad q_{\alpha L} = \begin{pmatrix} u_\alpha \\ d_\alpha \end{pmatrix}_L. \tag{442}$$

Weak isospin gauge transformations can rotate the LH electron neutrino into a LH electron for instance. The electron neutrino and up quarks (of all colors) have third component of weak isospin $T_3 = \frac{1}{2}$ while $T_3 = -\frac{1}{2}$ for the electron and down quarks. The weak hypercharges of these LH components are fixed by their electric charges and the relation $Q = T_3 + Y/2$. Thus $Y_{\nu_{eL}} = Y_{e_L} = -1, Y_{u_{\alpha L}} = Y_{d_{\alpha L}} = \frac{1}{3}$. Under a U$(1)_Y$ transformation, a field ψ of weak-hypercharge Y transforms as $\psi \to e^{iY\theta}\psi$.

• On the other hand, the RH components $\nu_R, e_R, u_{\alpha R}$ and $d_{\alpha R}$ are singlets under weak isospin (i.e. they transform in the trivial representation). In the standard model ν_R does not interact with any other particle and decouples from the theory. So $T_3 = 0$ for all the RH quarks and leptons. Thus their weak hypercharges are simply twice their electric charges $Y = 2Q$ or $Y_{\nu_{eR}} = 0, Y_{e_R} = -2, Y_{u_{\alpha R}} = \frac{4}{3}, Y_{d_{\alpha R}} = -\frac{2}{3}$.

• The weak ispspins and weak hypercharges of the second and third generations of quarks and leptons are assigned in the same way.

• It follows that the RH anti-leptons and RH anti-quarks also form weak isospin doublets: $\begin{pmatrix} e^+ \\ \nu_{e^+} \end{pmatrix}_L$ etc. Similarly, the LH anti-leptons and anti-quarks are singlets under weak isospin. The hypercharges of the anti-particles are the negatives of the Y-values of the corresponding particles.

[54]By contrast the left and RH components of the quarks transform in the same way under color SU(3).

• These weak isospins and hypercharges are important in determining the covariant derivatives of the fermion fields, and consequently their coupling to electromagnetic and weak gauge bosons. For instance, the covariant derivative of the LH neutrino-electron doublet $l = \begin{pmatrix} \nu_e \\ e \end{pmatrix}$ is

$$D_\mu l_a = \partial_\mu l_a - i\left(W_{\mu a}^{\ b} - Y_\mu \delta_a^b\right) l_b \tag{443}$$

Here $a = 1, 2$ labels the $T_3 = \frac{1}{2}, -\frac{1}{2}$ components of the weak isospin doublet, $l_1 = \nu_{eL}, l_2 = e_L$. Here $W_{\mu a}^{\ b}$ and Y_μ are the gauge fields corresponding to weak isospin and weak hypercharge (the coupling constants e_2 and e_1) have been absorbed into the respective gauge fields.

• The RH electron is a weak isospin singlet, it does not couple to W_μ, so its covariant derivative is

$$D_\mu e_R = \left[\partial_\mu - i(-2Y_\mu)\right] e_R. \tag{444}$$

• For the LH quarks $q_{\alpha a}$ $\alpha = 1, 2, 3$, is a color index while $a = 1, 2$ label the two weak isospin projections ($q_{\alpha 1} = u_{\alpha L}, q_{\alpha 2} = d_{\alpha L}$). If $A_{\mu \alpha}^{\ \beta}$ are the gluon fields, then

$$D_\mu q_{\alpha a} = \partial_\mu q_{\alpha a} - i\left[W_{\mu a}^{\ b}\delta_\alpha^\beta + \frac{1}{3}Y_\mu \delta_\alpha^\beta \delta_a^b + A_{\mu \alpha}^{\ \beta}\right] q_{\beta b} \tag{445}$$

• The RH up and down quarks do not couple to W_μ since they are weak isospin singlets. Thus

$$D_\mu u_{\alpha R} = \partial_\mu u_{\alpha R} - i\left[\frac{4}{3}Y_\mu \delta_\alpha^\beta + A_{\mu \alpha}^{\ \beta}\right] u_{\beta R} \quad \text{and} \quad D_\mu d_{\alpha R} = \partial_\mu d_{\alpha R} - i\left[-\frac{2}{3}Y_\mu \delta_\alpha^\beta + A_{\mu \alpha}^{\ \beta}\right] d_{\beta R}. \tag{446}$$

• The RH neutrinos transform trivially under the whole SM gauge group. So they do not couple to any gauge fields $D_\mu \nu_{eR} = \partial_\mu \nu_{eR}$. They do not interact with any other particles and have not been detected.

12.3 Fermion masses from Yukawa couplings sans mixing of generations

• Quarks and leptons are spin-half Dirac fermions in the standard model. Mass terms for Dirac spinors take the form $-m\bar{\psi}\psi = -m\bar{\psi}_R\psi_L - m\bar{\psi}_L\psi_R$. Such mass terms for quarks and leptons are not gauge-invariant since the LH components transform as weak isospin doublets while the RH components are weak isospin singlets. Moreover, such a mass term is also not invariant under $\mathrm{U}(1)_Y$ since the LH and RH fermions do not have the same weak hypercharges Y. (Note that hypercharge of $\bar{e}_R$ is 2 while that of e_L is -1. To be U(1) invariant, the sum of the hypercharges of the fields in a product should be zero.

On the other hand, Yukawa coupling of a Dirac spinor to a scalar field can lead to a mass term for the Dirac spinor if the scalar has a non-zero vev. To get a gauge invariant mass term this way, we exploit the fact that both the scalar field and the LH fermions are doublets under weak isospin. There are two distinct and interesting ways in which we may introduce gauge invariant Yukawa couplings. They are based on the observation that both the 'dot' and 'cross' products of a pair of SU(2) doublets are invariant under SU(2). Suppose

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \quad \text{and} \quad l = \begin{pmatrix} l_1 \\ l_2 \end{pmatrix} = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad \text{or} \quad \begin{pmatrix} u \\ d \end{pmatrix}_L \tag{447}$$

are the Higgs scalar and the LH lepton (or quark) $SU(2)_W$ doublets. Then the 'dot product' $\phi^\dagger l$ is obviously SU(2) invariant. Similarly, the 'cross' product

$$\phi \times l = \epsilon_{ab}\phi_a l_b = \phi_1 l_2 - \phi_2 l_1 = \phi^t \, (i\sigma_2) \, l \tag{448}$$

is also invariant under $\phi \to g\phi, l \to gl$. This is familiar to us as the spin singlet combination of two spin halves. But we may show it explicitly by using the general parametrization of an SU(2) matrix $g = \begin{pmatrix} z & w \\ -w^* & z^* \end{pmatrix}$ where z and w are complex numbers with $\det g = |z|^2 + |w|^2 = 1$. Then an SU(2) transformation takes the Higgs doublet to

$$\phi_1 \to z\phi_1 + w\phi_2 \quad \text{and} \quad \phi_2 \to -w^*\phi_1 + z^*\phi_2 \tag{449}$$

and similarly for the quark/lepton doublet l. It follows that the cross product

$$\begin{aligned} \phi_1 l_2 - \phi_2 l_1 &\to (z\phi_1 + w\phi_2)(-w^* l_1 + z^* l_2) - (-w^*\phi_1 + z^*\phi_2)(z l_1 + w l_2) \\ &= (|z|^2 + |w|^2)(\phi_1 l_2 - \phi_2 l_1) = \det g(\phi_1 l_2 - \phi_2 l_1) = (\phi_1 l_2 - \phi_2 l_1). \end{aligned} \tag{450}$$

is SU(2) invariant.

• The 'cross-product' type of Yukawa coupling generates masses for the up-type quarks $q = \begin{pmatrix} u \\ d \end{pmatrix}_L$.

$$g_u \bar{u}_R^\alpha \phi^t (i\sigma_2) \begin{pmatrix} u_\alpha \\ d_\alpha \end{pmatrix}_L + \text{ h.c.} = g_u \bar{u}_R^\alpha \epsilon_{ab} \phi_a q_b + \text{ h.c.} \tag{451}$$

As discussed, this is invariant under weak isospin gauge transformations. To be invariant under $\mathrm{U}(1)_Y$ we need to ensures that the weak hypercharges of $\bar{u}_R(Y = -4/3), \phi$ and $q_L(Y = 1/3)$ add up to zero. *This requirement fixes the hypercharge of the scalar field to be* 1 *as mentioned earlier.* Choosing 'unitary' gauge where $\phi = \begin{pmatrix} 0 \\ v + h \end{pmatrix}$ we see that the up quark mass is $g_u v$:

$$g_u \bar{u}_R^\alpha \begin{pmatrix} 0 & v + h \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} u_\alpha \\ d_\alpha \end{pmatrix} + \text{ h.c.} = -g_u v \bar{u}_R^\alpha u_{L\alpha} - g_u h \bar{u}_R^\alpha u_L + \text{h.c.} \tag{452}$$

The cross-product type of Yukawa coupling can also generate masses for the charged leptons (see below). The second term describes the trilinear vertex at which up quarks and up anti-quarks couple to the Higgs. Both the up quark mass and its Yukawa coupling to the Higgs are proportional to g_u. So a heavier Dirac fermion couples more strongly to the Higgs, as we noticed in the Abelian Higgs model. This is like more massive bodies feeling a stronger gravitational force. In particular, the top quark has the strongest coupling to the Higgs among all the quarks and leptons and provides a possible way of producing and studying the Higgs particle.

• Note: the Yukawa coupling g_u is a priori a complex number. However, it can be taken to be a non-negative real number (quark mass divided by Higgs vev) since its phase can be absorbed into a redefinition of the quark fields. More on this in the next section.

• The above 'cross-product' Yukawa couplings cannot lead to masses for down-type quarks (or neutrinos). The down quark gets its mass $g_d v$ using the 'other' gauge-invariant 'dot-product' Yukawa coupling to the Higgs. In the same gauge,

$$-g_d \bar{d}_R^{\alpha} \phi^{\dagger} \begin{pmatrix} u_\alpha \\ d_\alpha \end{pmatrix}_L + \text{h.c.} = -g_d \bar{d}_R^{\alpha} \begin{pmatrix} 0 & v+h \end{pmatrix} \begin{pmatrix} u_\alpha \\ d_\alpha \end{pmatrix}_L + \text{h.c.} = -g_d \bar{d}_R^{\alpha} (v+h) d_{\alpha L} + \text{h.c.} \quad (453)$$

• In the SM, the electron and other charged leptons (though not the neutrinos) receive their masses through the Yukawa coupling to the scalar doublet. To write down the Yukawa couplings and identify the masses (using the same gauge as above), it is convenient to define $l = \begin{pmatrix} e^+ \\ \nu_{e^+} \end{pmatrix}_R$ to be the RH (positron, anti-neutrino) weak isospin doublet (with $Y = 1$) rather than the LH neutrino, electron doublet (with $Y = -1$). If g_e is the electron (and positron) Yukawa coupling constant, then

$$g_e \bar{e}_L (\phi \times l_R) + h.c. = g_e \bar{e}_L \phi^t (i\sigma_2) \begin{pmatrix} e \\ \nu_e \end{pmatrix}_R + h.c. = g_e \bar{e}_L \epsilon_{ab} \phi_a l_b + h.c. \quad (454)$$

For the reasons given above, this is invariant under weak isospin gauge transformations. Check that it is also invariant under weak hypercharge if we assign $Y_\phi = 1$, which is consistent with what we found above. To find the resulting electron (and positron) mass (which is gauge invariant), we work in the gauge in which $\phi = \begin{pmatrix} 0 \\ v+h \end{pmatrix}$, where v is the vev of the scalar and h the Higgs scalar. Then the Yukawa coupling term is

$$g_e \bar{e}_L \begin{pmatrix} 0 & v+h \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} e \\ \nu_e \end{pmatrix}_R + h.c. = -g_e v \bar{e}_L e_R - g_e \bar{e}_L h e_R + h.c. \quad (455)$$

The electron mass is then $m_e = g_e v$. The second term describes the trilinear vertex at which electrons and positrons couple to the Higgs.

• The cross-product Yukawa couplings can be used to obtain neutrino masses, though it leaves unexplained why neutrinos are so much lighter than the other charged leptons. For that matter, we do not know why the top quark is so much heavier than the up quark either!

12.4 Quark mixing: CKM matrix

• The above Yukawa coupling of quarks and leptons to the scalar doublet can give masses to fermions in each generation. Let us focus on the quarks (same reasoning applies to leptons and can lead to masses for charged leptons and neutrinos). Generations are weak eigenstates, quarks that couple in a specific way to weak gauge bosons (e.g. the LH ones from each generation form a weak isospin doublet). So we might imagine obtaining masses for each of the two quarks in each generation. However, it is found that the weak eigenstates of quarks do not have definite masses. Mass eigenstates are those that propagate a definite ways while the weak eigenstates are those that interact with weak gauge bosons in definite ways, the two do not coincide. To accommodate this, we notice that each generation of quarks transforms in the same way under $\mathrm{U}(1)_Y$. So without losing gauge invariance, we can have a Yukawa coupling between a RH up-quark, Higgs doublet and a LH (c, s) doublet just as we can have a Yukawa coupling between a RH u-quark, Higgs doublet and a LH (u,d) doublet, generations can mix. So there aren't just 6 Yukawa couplings but a 3×3 complex matrix g of Yukawa couplings for up-type quarks and another 3×3 complex matrix $\tilde{g}$ for the down-type quarks. Let us use $i = 1, 2, 3$ to label generations and suppress color indices, then the Yukawa couplings are

$$\mathcal{L}_{\text{Yuk}} = -\tilde{g}_i^j \bar{d}_R^i \phi^\dagger \begin{pmatrix} u_j \\ d_j \end{pmatrix}_L + g_i^j \bar{u}_R^i \phi^t (i\sigma_2) \begin{pmatrix} u_j \\ d_j \end{pmatrix}_L + \text{h.c.} \tag{456}$$

A general complex matrix g cannot be diagonalized by a similarity transformation. But it can be diagonalized by two different unitary transformations, one acting on the right and one on the left! To see this, note that by polar decomposition, any complex matrix can be written as $g = P\Theta$ where P is a positive matrix (hermitian with non-negative eigenvalues) and Θ is a unitary matrix. Moreover, suppose U is a unitary matrix that diagonalizes P, $U^\dagger P U = D$. Now if we let $V = \Theta^\dagger U$ we find that $U^\dagger g V = D$. The entries of the diagonal matrix D are called the singular values of g, they are non-negative real numbers, the positive square-roots of the positive matrix $gg^\dagger$ ($g^\dagger g$ and $gg^\dagger$ have the same eigenvalues).

• So let us suppose we have unitary matrices A_L, A_R and $\tilde{A}_L, \tilde{A}_R$ that diagonalize g and $\tilde{g}$ respectively. The singular values of g are the masses of the up-type quarks in units of the Higgs vev v while the singular values of $\tilde{g}$ are the masses of the down-type quarks in units of v:

$$g = \frac{1}{v} A_R M A_L^\dagger \quad \text{and} \quad \tilde{g} = \frac{1}{v} \tilde{A}_R \tilde{M} \tilde{A}_L^\dagger \quad \text{or} \quad A_R^\dagger g A_L = \frac{M}{v} \quad \text{and} \quad \tilde{A}_R^\dagger \tilde{g} \tilde{A}_L = \frac{\tilde{M}}{v}. \tag{457}$$

Here $M = \text{diag}(m_u, m_c, m_t)$ and $\tilde{M} = \text{diag}(m_d, m_s, m_b)$ are the diagonal 3×3 matrices of u-type quark masses and d-type quark masses.

• In the 'unitary' gauge $\phi = (0, v + h)$, the Yukawa Lagrangian, omitting interactions with h becomes

$$\mathcal{L}_{\text{Yuk}} = -v \begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}_R g \begin{pmatrix} u \\ c \\ t \end{pmatrix}_L - v \begin{pmatrix} \bar{d} & \bar{s} & \bar{b} \end{pmatrix}_R \tilde{g} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L + \text{h.c.} + \text{interactions.} \tag{458}$$

• Now we define linear combinations of the weak eigenstates so as to diagonalize g and $\tilde{g}$. Let

$$\begin{pmatrix} u \\ c \\ t \end{pmatrix}_L = A_L \begin{pmatrix} u \\ c \\ t \end{pmatrix}_L^\prime, \quad \begin{pmatrix} u \\ c \\ t \end{pmatrix}_R = A_R \begin{pmatrix} u \\ c \\ t \end{pmatrix}_R^\prime, \quad \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L = \tilde{A}_L \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L^\prime, \quad \begin{pmatrix} d \\ s \\ b \end{pmatrix}_R = \tilde{A}_R \begin{pmatrix} d \\ s \\ b \end{pmatrix}_R^\prime. \tag{459}$$

In terms of these new 'primed' quark fields the Yukawa Lagrangian simplifies. Below we use u and d as short-hand notation for the three up-type and three down-type quarks:

$$\begin{aligned} \mathcal{L}_{\rm Yuk} &= -v\bar{u}'_R A_R^\dagger g A_L u'_L - v\bar{d}'_R \tilde{A}_R^\dagger \tilde{g}\tilde{A}_L d'_L + \text{h.c.} + \text{interactions} \\ &= -\bar{u}'_R M u'_L - \bar{d}'_R \tilde{M} d'_L - \bar{u}'_L M u'_R - \bar{d}'_L \tilde{M} d'_R + \text{interactions}. \end{aligned} \tag{460}$$

We see that the Yukawa couplings have lead to Dirac masses for the six quarks. The primed quark fields have definite masses are are called the mass eigenstates. They are linear combinations of the un-primed weak eigenstates.

• These mass terms only involve the 6 quark masses, which are the singular values of the complex Yukawa coupling matrices g and $\tilde{g}$. Do the remaining parameters in $g, \tilde{g}$ (or equivalently the parameters in the unitary matrices $A_L, A_R, \tilde{A}_L, \tilde{A}_R$) play any role? The passage to mass eigenstates for quark fields affects the interaction terms between quarks and gauge fields. We may write the EM, neutral weak and charged weak interactions in terms of mass eigenstates. It turns out that the EM and neutral weak interactions are unaffected (check in the same way as we do below for the charge changing weak interactions below). Something interesting happens in the charge-changing weak interactions, a combination of the A_L, $\tilde{A}_L$ matrices show up there.

• Recall that the charge weak interactions only affect the left handed quark and lepton doublets. Focusing on quarks, consider the LH doublets $q_{Li\alpha a}$ where $i = 1, 2, 3$ label 3 generations, $a = 1, 2$ label a basis in the fundamental rep of weak isospin and $\alpha = 1, 2, 3$ label a basis in the fundamental representation of color. The covariant derivative is

$$D_\mu q_{Li\alpha a} = \partial_\mu q_{Li\alpha a} - i\left[G_{\mu}{}_{\alpha}^{\beta}\delta_a^b + W_{\mu}{}_{a}^{b}\delta_\alpha^\beta + \frac{2}{3}Y_\mu\right] q_{Li\beta b}. \tag{461}$$

The Dirac Lagrangian is then

$$\mathcal{L}_D = \bar{q}_L^{i\alpha a} i\gamma^\mu D_\mu q_{Li\alpha a} \tag{462}$$

Let us focus on the *change changing* weak interactions that involve $W_\mu^\pm$ (we suppress color indices below)

$$\begin{aligned} \mathcal{L}_{cc} &= \frac{1}{2}\bar{q}_L^i \gamma^\mu \begin{pmatrix} 0 & W^- \\ W^+ & 0 \end{pmatrix} q_{Li} = \frac{1}{2}\begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}_L \gamma^\mu W_\mu^- \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L + \text{h.c.} \\ &= \frac{1}{2}\begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}'_L A_L^\dagger \gamma^\mu W_\mu^- \tilde{A}_L \begin{pmatrix} d \\ s \\ b \end{pmatrix}'_L + \text{h.c.} = \frac{1}{2}\begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}'_L \gamma^\mu W_\mu^- C \begin{pmatrix} d \\ s \\ b \end{pmatrix}'_L + \text{h.c} \end{aligned} \tag{463}$$

Here we defined the 3×3 unitary CKM matrix $C = A_L^\dagger \tilde{A}_L$. In general, the same unitary matrices do not diagonalize both g and $\tilde{g}$. C measures the difference. If $A_L = \tilde{A}_L$, then there is no mixing. We see that the charged current weak interactions can convert an up quark mass eigenstate into a linear combination of mass eigenstates of the down type quarks accompanied by the emission of a W^+. While cc weak interactions operate within generations of weak or flavor eigenstates, they do not operate within families of mass eigenstates. The Yukawa couplings enter the Lagrangian only through the quark masses and CKM matrix elements.

• The CKM matrix introduces several parameters into the standard model beyond the quark masses, let us count them. Suppose there are N_g generations. Being a unitary $N_g \times N_g$ matrix, C depends on N_g^2 parameters. However, there is an equivalence relation among CKM matrices, two distinct unitary matrices $A_L, \tilde{A}_L$ may correspond to the same Yukawa couplings. Indeed, A_L and A_R are not unique. Multiplying A_L and A_R by a diagonal unitary matrix diag$(e^{i\theta_1}, \cdots, e^{i\theta_{N_g}})$ on the right does not affect g. Similarly, multiplying $\tilde{A}_L$ and $\tilde{A}_R$ by a diagonal unitary matrix diag$(e^{i\tilde{\theta}_1}, \cdots, e^{i\tilde{\theta}_{N_g}})$ on the right does not affect $\tilde{g}$. So the CKM matrix $C_{ij} = (A_L^\dagger \tilde{A}_L)_{ij}$ is physically equivalent to $C_{ij} e^{i(\tilde{\theta}_j - \theta_i)}$. However, a constant addition to both the θ's and $\tilde{\theta}$'s does not affect C. We can use these phases to eliminate $2N_g - 1$ parameters from the N_g^2 parameters that C can depend on. Thus, the CKM matrix depends on $N_g^2 - 2N_g + 1 = (N_g - 1)^2$ free parameters. For two generations there is one one such parameter, which is the Cabibbo mixing angle and was introduced to incorporate the possibility that both the strange and down quark mass eigenstates could decay to an up quark mass eigenstate.

• For $N_g = 3$, there are 4 parameters in the CKM matrix. Three of these can be regarded as generalized Cabibbo mixing angles as they are present in a 3×3 (real) orthogonal matrix, which depends on $\frac{1}{2}3(3-1) = 3$ parameters. The fourth parameter is a (complex) phase. A complex phase in the CKM matrix makes the weak interaction Lagrangian not invariant under time-reversal (which involves complex conjugation). By the CPT invariance of the same Lagrangian this implies CP violation in the weak interactions. Indeed, Kobayashi and Maskawa in 1973 predicted a third generation of quarks in order to make it possible to incorporate CP violation in the weak interactions, which had been unexpectedly experimentally discovered in 1964 by Cronin and Fitch.

13 Summary of the Standard Model

• The standard model of particle physics (aside from the inclusion of the third generation of quark and leptons) reached its current form around 1973-1974 when it became generally accepted that QCD was the best candidate for a theory of the strong interactions, displaying the crucial property of asymptotic freedom (Gross, Wilczeck and Politzer, based on the work of ’t Hooft and Veltman on quantization of non-abelian gauge theories). The need for asymptotic freedom arose as a consequence of the results of the deep inelastic scattering experiments of the late 1960s which showed that the structure functions of the proton were scale invariant, at sufficiently small scales. This was explained by Bjorken and Feynman in terms of the parton model, the constituents of the proton had to be point-like nearly free particles. So the force that bound these partons had to be weak at short distances (asymptotically free) but strong at long distances (infrared slavery). QCD had this remarkable property. The discovery (Ting and Richter) of the J/ψ meson containing a charm quark convinced most that quarks were real. Partons were identified as quarks and gluons. The charm quark had been predicted by Glashow Iliopolous and Maiani (GIM mechanism which involves mixing among quarks of two generations).

• Following the discovery of parity violation in 1956 and the development of the V-A theory of weak interactions, Glashow proposed an $SU(2) \times U(1)$ gauge theory of massive weak gauge bosons in 1961, but it was not renormalizable and the mass terms for the gauge fields broke the local gauge symmetry. The Higgs mechanism for a gauge invariant way to generate vector boson masses was proposed in 1964 and by 1967-68 Weinberg and Salam combined the Higgs mechanism with Glashow's model to arrive at the Lagrangian of the electroweak part of the SM (for two families of leptons). However, this model had to wait for the work of 't Hooft and Veltman (1972) to be shown to be a viable quantum theory in a perturbation approximation. Kobayashi and Maskawa introduced three generations of quarks with mixing among them in 1973 to incorporate CP violation. While the Lagrangian of QCD is relatively simple since it is not a chiral gauge theory, the Lagrangian of the electro-weak theory is more intricate due to parity violation, Yukawa coupling to scalars and mixing among quark and lepton families. Mixing among two lepton families was introduced by Maki, Nakagawa and Sakata in 1962 based on Pontecorvo's 1957 prediction of neutrino oscillations. However, this was generally accepted and extended to three lepton families only in the late 1990s with experimental confirmation of neutrino oscillations coming in 2001.

• The standard model is a non-abelian Yang-Mills gauge theory with gauge Lie algebra $su(3) \times su(2) \times u(1)$, corresponding to color, weak isospin and weak-hypercharge symmetries. In addition to the 8 gluons and 4 electroweak gauge bosons, it also has a complex doublet of scalar fields and a set of fermion fields for the quarks and leptons. The gauge fields corresponding to color, weak isospin and weak hypercharge are G_μ, W_μ and Y_μ. For each $\mu = 0, 1, 2, 3$ and location x they are traceless hermitian 3×3 matrices, traceless hermitian 2×2 matrices and a real number respectively, each transforming in the adjoint representation of the respective gauge group while being a singlet under the other gauge groups. We will use $\alpha = 1, 2, 3$ to denote color indices, i.e. to enumerate a basis for the fundamental (triplet) representation of color SU(3). $a = 1, 2$ label a basis for the fundamental (doublet) representation of weak isospin.

• The kinetic terms and self-interactions of the gauge fields are simply described by the squares of the corresponding field strengths ($Y^{\mu\nu} = \partial^\mu Y^\nu - \partial^\nu Y^\mu$ and $W^{\mu\nu} = \partial^\mu W^\nu - \partial^\nu W^\mu - i[W^\mu, W^\nu]$ and $G^{\mu\nu} = \partial^\mu G^\nu - \partial^\nu G^\mu - i[G^\mu, G^\nu]$)

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{2(e_1^2/4)} Y^{\mu\nu} Y_{\mu\nu} - \frac{1}{2e_2^2} \text{ tr } W^{\mu\nu} W_{\mu\nu} - \frac{1}{2e_3^2} \text{ tr } G^{\mu\nu} G_{\mu\nu} \tag{464}$$

Here $e_{1,2,3}$ are the coupling constants of $u(1)_Y, su(2)_W$ and $su(3)_C$. We absorb the coupling constants into the corresponding gauge fields, so they appear as pre-factors in the Yang-Mills Lagrangian.

• The scalar and fermion fields are introduced by specifying the representation of the SM gauge group G according to which they transform. This allows us to write covariant derivatives for each of the spin zero and half matter fields, from which their coupling to gauge fields follows by the Lorentz minimal coupling prescription. The scalar fields are color singlets, transform as a doublet under weak isospin and have $Y = 1$ (in order that the Yukawa couplings to quarks and leptons be gauge invariant). So they belong to the $(1, 2, 1)$ representation of G. The covariant derivative of ϕ is

$$(D_\mu \phi)_a = \partial_\mu \phi - i[W_{\mu}{}_a^b + Y_\mu \delta_a^b] \phi_a. \tag{465}$$

If we had not absorbed the couplings into the gauge fields, then they would appear here as $(e_1/2) Y_\mu \delta_a^b$ and $(e_2/2) W_\mu^A (\sigma_A)_b^a$.

• The propagation, self-interactions and interactions of scalars with electroweak gauge fields is governed by the Lagrangian density

$$\mathcal{L}_{\text{scalar}} = (D_\mu \phi)^\dagger (D^\mu \phi) - \frac{\lambda}{4} \left(\phi^\dagger \phi - v^2\right)^2 . \tag{466}$$

Under a gauge transformation $g \in U(2)$, $\phi \to g\phi$ and it is clear that the scalar potential is invariant.

• Now we come to the quarks and leptons, which are Dirac spinor fields whose LH and RH components are in different representations of G and consequently have different covariant derivatives and couplings to the gauge and Higgs fields. We will use the Greek letters $\alpha, \beta = 1, 2, \ldots, N_c$ for color indices ($N_c = 3$). Latin letters from the beginning of the alphabet $a, b = 1, 2$ label a basis for the fundamental representation of weak isospin $\text{SU}(2)_W$. Latin indices from the middle of the alphabet $i, j = 1, 2, \ldots, N_g$ label the $N_g = 3$ generations of quarks and leptons, i.e., the weak eigenstates (rather than the mass eigenstates). The number of quark and lepton generations are the same: this is needed for the cancelation of anomalies in the quantum theory, it is also an experimental fact. The covariant derivatives do not mix generations, the Yukawa couplings mix generations. Let us recall the covariant derivatives.

• We denote the LH quarks by $q_{\alpha a i}$, $\gamma_5 q = -q$. they transform in $(\mathbf{3}, \mathbf{2}, \frac{1}{3})$, i.e., in the fundamental representations of color SU(3) and weak isospin and have hyper charge $Y = 1/3$. It follows that the covariant derivatives of the LH quarks are

$$D_\mu q_{\alpha a i} = \partial_\mu q_{\alpha a i} - i \left[W^b_{\mu a} \delta^\beta_\alpha + \frac{1}{3} Y_\mu \delta^b_a \delta^\beta_\alpha + G^\beta_{\mu \alpha} \right] q_{\beta b i} \tag{467}$$

• The RH up type quarks $u_{\alpha i}$ transform in $(3, 1, \frac{4}{3})$ and the RH down type quarks $d_{\alpha i}$ belong to $(3, 1, -\frac{2}{3})$. So

$$D_\mu u_{\alpha i} = \partial_\mu u_{\alpha i} - i \left(G^\beta_{\mu\alpha} + \frac{4}{3} Y_\mu \delta^\beta_\alpha \right) u_{\beta i}, \quad \text{and} \quad D_\mu d_{\alpha i} = \partial_\mu d_{\alpha i} - i \left(G^\beta_{\mu\alpha} - \frac{2}{3} Y_\mu \delta^\beta_\alpha \right) d_{\beta i} \tag{468}$$

• The LH leptons l_{ai} ($\gamma_5 l = -l$) belong to $(\mathbf{1}, \mathbf{2}, -1)$, i.e., are color singlets, fundamental under weak isospin and have weak hyper-charge -1. Their covariant derivatives are

$$D_\mu l_{ai} = \partial_\mu l_{ai} - i \left[W^b_{\mu a} - Y_\mu \delta^b_a \right] l_{bi} \tag{469}$$

• The RH charged leptons $\gamma_5 e_i = e_i$ are singlets under both color and weak isospin and have weak hypercharge $Y = -2$. They belong to $(\mathbf{1}, \mathbf{1}, -2)$, so

$$D_\mu e_i = \partial_\mu e_i - i(-2Y_\mu) e_i . \tag{470}$$

• The RH neutrinos $\gamma_5 \nu_i = \nu_i$ belong to $(\mathbf{1}, \mathbf{1}, 0)$, i.e., they transform trivially under the full SM gauge group. So they do not interact with any of the SM gauge fields: $D_\mu \nu_i = \partial_\mu \nu_i$.

• Using these covariant derivatives, we can write down the Dirac Lagrangian for the massless quarks and leptons

$$\mathcal{L}_{\text{Dirac}} = \bar{q}^{\alpha a i}(i\gamma\cdot D)q_{\alpha a i}+\bar{u}^{\alpha i}(i\gamma\cdot D)u_{\alpha i}+\bar{d}^{\alpha i}(i\gamma\cdot D)d_{\alpha i}+\bar{l}^{a i}(i\gamma\cdot D)l_{a i}+\bar{e}^{i}(i\gamma\cdot D)e_{i}+\bar{\nu}^{i}(i\gamma\cdot D)\nu_{i}. \tag{471}$$

• The Yang-Mills, Higgs and Dirac Lagrangians treat each generation of massless quarks and leptons separately. The Yukawa couplings of fermions give them masses and can mix generations. Recall that the up type quarks and charged leptons have cross-product-type of Yukawa couplings ($\bar{u}\phi \times q$) while the down type quarks (and possibly the neutrinos) have dot-product-type of Yukawa couplings ($\bar{d}\phi^{\dagger}q$). The Yukawa coupling terms are

$$\mathcal{L}_{\text{Yukawa}} = g_i^j \bar{u}^{\alpha i}\phi_a \epsilon^{ab} q_{\alpha b j} + f_i^j \bar{e}^i \phi_a \epsilon^{ab} l_{bj} + \tilde{g}_i^j \bar{d}^{\alpha i} {\phi^{\dagger}}^a q_{\alpha a j} + \tilde{f}_i^j \bar{\nu}^i {\phi^{\dagger}}^a l_{aj}. \tag{472}$$

Here g and $\tilde{g}$ are the up-type and down-type complex Yukawa coupling matrices while $f, \tilde{f}$ are the charged lepton and neutrino Yukawa coupling matrices. Let M and m be the diagonal mass matrices of up-type-quarks and charged leptons (masses M_i and m_i are the singular values of g, f in units of the Higgs vev v), and similarly $\tilde{M}_i$ and $\tilde{m}_i$ are masses of the down-type quarks and neutrinos respectively. Then

$$g = \frac{1}{v} A_R M A_L^{\dagger} \quad \text{and} \quad \tilde{g} = \frac{1}{v} \tilde{A}_R \tilde{M} \tilde{A}_L^{\dagger} \quad \text{and} \quad f = \frac{1}{v} B_R m B_L^{\dagger} \quad \text{and} \quad \tilde{f} = \frac{1}{v} \tilde{B}_R \tilde{M} \tilde{B}_L^{\dagger}. \tag{473}$$

Here the A's and B's are unitary matrices relating the mass eigenstates to the weak eigenstates as explained in the last section. The 3×3 unitary CKM and PMNS matrices $C = A_L^{\dagger} A_L$ and $P = B_L^{\dagger} B_L$ encode mixing among mass eigenstates of down-type quarks and neutrinos when they participate in charge changing weak interactions (the CKM matrix is usually denoted V).

• Here $M, m, \tilde{M}, \tilde{m}$ are diagonal matrices with entries M_i, m_i, $\tilde{M}_i$, $\tilde{m}_i$ respectively. In the original SM, the last term was absent as neutrinos were taken to be massless $\tilde{m} \equiv 0$.

• Let us count the number of free input parameters in the SM that are to be determined from experiment. In the gauge sector, there are three dimensionless gauge couplings e_1, e_2, e_3 (e_3 is determined in terms of Λ_{QCD}). In the scalar sector we have the dimensionless scalar (Higgs) self-coupling λ and the scalar vev $\sqrt{2}v \approx 246$ GeV. All of these have been measured. The rest of the parameters are the Yukawa couplings, or equivalently, the masses of the quarks and leptons and the dimensionless mixing angles in the unitary CKM and PMNS matrices C and P. Suppose there are N_g generations of quarks and leptons. Then there are $4N_g$ quark and lepton masses and $(N_g - 1)^2$ mixing angles and phases each in C and P, after accounting for the freedom to redefine phases of fermion fields. Thus the SM with three generations has $5+4N_g+2(N_g-1)^2 = 25$ free parameters. The masses of the charged leptons and heavy quarks are known reasonably accurately while the light quark masses $m_{u,d,s}$ are somewhat imprecisely known. The CKM matrix elements have all been measured, three angles (beginning with the Cabibbo angle from weak decays of charged Kaons followed by two more mixing angles from weak decays of heavy mesons containing c and b quarks) and a CP violating phase. The three neutrino mixing angles have also been measured from neutrino oscillations (the atmospheric, solar and reactor angles). The neutrino masses and CP violating phase in the PMNS matrix remain to be measured.

• The vacuum expectation value of the scalar field v is the *only dimensional parameter* in the electroweak part of the standard model: quark and lepton masses are multiples of v as determined by the dimensionless Yukawa couplings. It is a mystery why some of the Yukawa couplings are so small.

• The number of colors N_c is an integer and experiments favor $N_c = 3$, we did not include it with the continuous free parameters above. Measurements of the Z^0 decay width favor $N_g = 3$ generations.

• Note that we do not count the speed of light or Planck's constant among the parameters of the SM. $\hbar$ and c may be regarded as parameters of the frameworks of quantum theory and relativity and do not pertain to the specific dynamical model of forces specified by the SM. We may work in units where they are equal to one. $\hbar$ and c are used to set the scale for two out of the three dimensions of mass, length and time.

• In addition to the 25 continuous free parameters of the SM needed to explain strong, weak and electromagnetic phenomena, we need two more parameters to use general relativity to predict classical gravitational phenomena: Newton's gravitational constant and the cosmological constant.

• Though most of the Lagrangian of the SM is now confirmed, we are still quite far from understanding its consequences, especially in the low-energy behavior of the strong interactions and the high-energy behavior of electroweak interactions. Much progress has been made, primarily in weak coupling perturbation theory but many things (like the masses and structure of hadrons) remain beyond the reach of perturbation theory. It is a little bit like being able to integrate Newton's equations for the motion of planets for short times, but not knowing the medium and long term behavior (shapes of orbits and their stability).